

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГБОУ ВО «Поволжский государственный технологический университет»



# ГрафиКон 2025

Материалы 35-й Международной конференции  
по компьютерной графике и машинному зрению

Научное текстовое локальное электронное издание

---

Поволжский государственный технологический университет  
Йошкар-Ола, 30 сентября – 2 октября 2025 года  
<https://graphicon.srcript.ru/>

УДК 004.92  
ББК 32.973.26-018.2  
Г 78

**Редакционная коллегия:**

*А. А. Роженцов*, доктор технических наук, профессор,  
первый проректор ПГТУ (*отв. редактор*);  
*М. А. Берберова*, кандидат технических наук, доцент кафедры  
промышленной информатики РТУ МИРЭА, директор АНО «НИЦФТИ»;  
*А. А. Кислицын*, кандидат технических наук, доцент кафедры  
радиотехники и связи ПГТУ (*отв. за выпуск*)

Г 78     **ГрафиКон 2025** : материалы 35-й Международной конференции по компьютерной графике и машинному зрению (Россия, Йошкар-Ола, 30 сентября – 2 октября 2025 г.) / Поволжский государственный технологический университет; редкол.: А. А. Роженцов (отв. ред.) [и др.]. – Электронные данные. – Йошкар-Ола: Поволжский государственный технологический университет, 2025. – 1095 с. – URL:  
ISBN 978-5-8158-2474-4

Представлены материалы 35-й Международной конференции «ГрафиКон 2025», проходившей на базе Поволжского государственного технологического университета.

В сборник вошли доклады участников конференции, посвященные методам и технологиям компьютерного анализа изображений, визуальной и когнитивной аналитики, 3D-реконструкции, визуальной навигации и человеко-машинного взаимодействия, виртуальной и дополненной реальности, распознавания образов и др.

Издание адресовано сотрудникам научно-исследовательских и образовательных организаций, специалистам предприятий ИТ-индустрии, аспирантам, студентам.

---

*Электронное научное издание*

**ГрафиКон 2025**

Материалы 35-й Международной конференции по компьютерной графике и машинному зрению (Россия, Йошкар-Ола, 30 сентября – 2 октября 2025 года)

*Ответственность за содержание материалов несут авторы*

Оригинал-макет издания подготовлен редакционно-издательским отделом ПГТУ

Редакторы *Л. С. Емельянова, П. Г. Павловская*

Компьютерная верстка *Е. А. Головина, Т. В. Отмахова, С. Н. Эштыкова*

Переводы на английский язык предоставлены авторами

Для дизайна использованы материалы из открытых интернет-источников

Сборник разработан с помощью программного обеспечения Microsoft Office Word, Adobe Acrobat Pro.

Подписано к использованию 01.12.2025. Объем издания 98,7 Мб.

Поволжский государственный технологический университет. 424000 Йошкар-Ола, пл. Ленина, 3

---

ISBN 978-5-8158-2474-4

© Поволжский государственный  
технологический университет, 2025



## Организационный комитет конференции ГрафиКон 2025

### Руководство организационного комитета

#### *Председатель*

**Петухов Игорь Валерьевич,**

ректор Поволжского государственного  
технологического университета,  
доктор технических наук, профессор



#### *Сопредседатель*

**Роженцов Алексей Аркадьевич,**

первый проректор Поволжского государственного  
технологического университета,  
доктор технических наук, профессор



#### *Заместитель председателя*

**Берберова Мария Александровна,**

доцент кафедры промышленной информатики РТУ МИРЭА,  
директор АНО «НИЦФТИ», кандидат технических наук



## Члены оргкомитета

### **Бабаян Павел Варганович,**

к.т.н., доцент, заведующий кафедрой автоматизации и информационных технологий в управлении РГРТУ им. В.Ф. Уткина, Рязань, Россия

### **Бондарев Александр Евгеньевич,**

к.ф.-м.н., старший научный сотрудник ИПМ им. М.В. Келдыша РАН, доцент кафедры высшей математики РТУ МИРЭА, Москва, Россия

### **Будак Владимир Павлович,**

д.т.н., профессор, профессор кафедры светотехники НИУ МЭИ, главный редактор журнала «Светотехника / Light & Engineering», Москва, Россия

### **Волобой Алексей Геннадьевич,**

д.ф.-м.н., доцент, ведущий научный сотрудник ИПМ им. М.В. Келдыша РАН, Москва, Россия

### **Галактионов Владимир Александрович,**

д.ф.-м.н., профессор, главный научный сотрудник ИПМ им. М.В. Келдыша РАН, Москва, Россия

### **Грязин Владимир Альбертович,**

к.т.н., доцент, заместитель директора по научной работе института механики и машиностроения Поволжского государственного технологического университета, Йошкар-Ола, Россия

### **Дебелов Виктор Алексеевич,**

д.т.н., доцент, Новосибирск, Россия

### **Ерёмченко Евгений Николаевич,**

научный сотрудник МГУ им. М.В. Ломоносова, Москва, Россия

### **Карпов Леонид Евгеньевич,**

д.т.н., главный научный сотрудник ИСП им. В.П. Иванникова РАН, ответственный секретарь редколлегии журнала «Программирование», доцент кафедры системного программирования ВМК МГУ им. М.В. Ломоносова, Москва, Россия

### **Конопацкий Евгений Викторович,**

д.т.н., доцент, директор Института информационных технологий ННГАСУ, Нижний Новгород, Россия

### **Крылов Андрей Серджевич,**

д.ф.-м.н., профессор, заведующий лабораторией математических методов обработки изображений факультета ВМК МГУ им. М.В. Ломоносова, Москва, Россия

### **Лагунова Марина Викторовна,**

д. пед. н., профессор, заместитель заведующего кафедрой инженерной графики и информационного моделирования ННГАСУ, Нижний Новгород, Россия

### **Ласточкин Денис Михайлович,**

к.т.н., доцент, заместитель директора по научной работе института механики и машиностроения Поволжского государственного технологического университета, Йошкар-Ола, Россия

### **Панчук Константин Леонидович,**

д.т.н., доцент, профессор кафедры инженерной геометрии и САПР Омского государственного технического университета, Омск, Россия

### **Подвесовский Александр Георгиевич,**

к.т.н., доцент, профессор кафедры информатики и программного обеспечения БГТУ, Брянск, Россия

### **Ротков Сергей Игоревич,**

д.т.н., профессор, профессор кафедры инженерной графики и информационного моделирования ННГАСУ, Нижний Новгород, Россия

### **Толок Алексей Вячеславович,**

д.т.н., профессор, главный научный сотрудник ИПУ РАН им. В.А. Трапезникова, заведующий кафедрой инженерной графики МГТУ «СТАНКИН», Москва, Россия

### **Турлапов Вадим Евгеньевич,**

д.т.н., доцент, профессор кафедры высокопроизводительных вычислений и истемного программирования ННГУ им. Н.И. Лобачевского, Нижний Новгород, Россия

### **Филинских Александр Дмитриевич,**

к.т.н., доцент, заведующий кафедрой графических информационных систем НГТУ им. Р.Е.Алексеева, Нижний Новгород, Россия

### **Хранилов Валерий Павлович,**

д.т.н., профессор, заместитель директора Учебно-научного института радиоэлектроники и информационных технологий по научной работе НГТУ им. Р.Е. Алексеева, Нижний Новгород, Россия

## Программный комитет конференции ГрафиКон 2025



*Председатель*

**Галактионов Владимир Александрович,**

доктор физико-математических наук, профессор,  
главный научный сотрудник Института прикладной  
математики им. М.В. Келдыша РАН

### Члены программного комитета

**Алексеев Алексей Кириллович**

ОАО «РКК «Энергия», Москва

**Бабаян Павел Варганович**

РГРТУ им. В.Ф. Уткина, Рязань

*(председатель секции)*

**Барладян Борис Хаимович**

ИПМ им. М.В. Келдыша РАН, Москва

**Бондарев Александр Евгеньевич**

ИПМ им. М.В. Келдыша РАН, РТУ

МИРЭА, Москва *(председатель секции)*

**Будак Владимир Павлович**

НИУ МЭИ, Москва

**Васильев Евгений Павлович**

ННГУ им. Н.И. Лобачевского,

Нижний Новгород

**Ватолин Дмитрий Сергеевич**

МГУ им. М.В. Ломоносова, Москва

**Волобой Алексей Геннадьевич**

ИПМ им. М.В. Келдыша РАН, Москва

*(председатель секции)*

**Дубнишев Юрий Николаевич**

Институт теплофизики СО РАН,

Новосибирск

**Гетманская Александра Александровна**

ННГУ им. Лобачевского,

Нижний Новгород

**Грязин Владимир Альбертович**

ПГТУ, Йошкар-Ола

*(председатель секции)*

**Дебелов Виктор Алексеевич**

ИВМиМГ СО РАН, Новосибирск

**Ерёмченко Евгений Николаевич**

МГУ им. М.В. Ломоносова, Москва

*(председатель секции)*

**Ершов Сергей Валентинович**

ИПМ им. М.В. Келдыша РАН, Москва

**Жданов Андрей Дмитриевич**

Университет ИТМО, Санкт-Петербург

**Жданов Дмитрий Дмитриевич**

Университет ИТМО, Санкт-Петербург

**Захарова Алёна Александровна**

ИПУ РАН, Москва

**Знаменская Ирина Александровна**

МГУ, Москва

**Исаев Руслан Александрович**

БГТУ, Брянск

**Конопацкий Евгений Викторович**

ННГАСУ, Нижний Новгород

*(председатель секции)*

**Коростелев Дмитрий Александрович**

БГТУ, Брянск

**Костров Борис Васильевич**

РГРТУ им. В.Ф. Уткина, Рязань

**Крылов Андрей Серджевич**

МГУ им. М.В. Ломоносова, Москва

*(председатель секции)*

**Курачий Михаил Иванович**

ТУСУР, Томск

**Лагунова Марина Викторовна**

ННГАСУ, Нижний Новгород

*(председатель секции)*

**Ласточкин Денис Михайлович**

ПГТУ, Йошкар-Ола

*(председатель секции)*

**Ложников Павел Сергеевич**

ОмГТУ, Омск

**Макарова Елена Андреевна**

ООО «К-Скай», Петрозаводск

**Михайлюк Михаил Васильевич**

НИИСИ РАН, Москва

**Небаба Степан Геннадьевич**

ТПУ, Томск

**Пилюгин Виктор Васильевич**

МИФИ, Москва

**Подвесовский Александр Георгиевич**

БГТУ, Брянск

**Поройков Антон Юрьевич**

НИУ МЭИ, Москва

Потёмин Игорь Станиславович,

Университет ИТМО, Санкт-Петербург

**Приоров Андрей Леонидович**

ЯрГУ им. П.Г. Демидова, Ярославль

**Пчелинцев Яков Антонович**

МГУ им. М.В. Ломоносова, Москва

**Ротков Сергей Игоревич**

ННГАСУ, Нижний Новгород

*(председатель секции)*

**Саблина Виктория Александровна**

РГРТУ им. В.Ф. Уткина, Рязань

**Семенов Виталий Адольфович**

ИСП им. В.П. Иванникова РАН, Москва

**Скорнякова Надежда Михайловна**

НИУ МЭИ, Москва

**Сорокин Дмитрий Васильевич**

МГУ им. М.В. Ломоносова, Москва

**Толок Алексей Вячеславович**

ИПУ РАН, МГТУ «СТАНКИН», Москва

*(председатель секции)*

**Трубаков Андрей Олегович**

БГТУ, Брянск

**Турлапов Вадим Евгеньевич**

ННГУ им. Н.И. Лобачевского,

Нижний Новгород *(председатель секции)*

**Филинских Александр Дмитриевич**

НГТУ им. Р.Е. Алексеева,

Нижний Новгород *(председатель секции)*

**Фролов Владимир Александрович**

МГУ им. М.В. Ломоносова, Москва

**Хвостиков Александр Владимирович**

МГУ им. М.В. Ломоносова, Москва

**Холопов Иван Сергеевич**

ООО «Квантрон Групп», Рязань

**Хранилов Валерий Павлович**

НГТУ им. Р.Е. Алексеева,

Нижний Новгород *(председатель секции)*

**Чуприна Светлана Игоревна**

ПГНИУ, Пермь

**Шкляр Алексей Викторович**

ТПУ, Томск

**Якименко Игорь Владимирович**

филиал НИУ МЭИ, Смоленск

## ОРГАНИЗАТОРЫ



АНО Научное общество  
«ГрафиКон»



**ПОВОЛЖСКИЙ**  
ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ  
УНИВЕРСИТЕТ

Поволжский государственный  
технологический университет



Омский государственный  
технический университет



Институт  
прикладной математики  
им. М.В. Келдыша РАН



Нижегородский государственный  
архитектурно-строительный  
университет



АНО «Научно-исследовательский центр  
физико-технической информатики»



Нижегородский государственный технический  
университет им. Р.Е. Алексеева



Рязанский государственный  
радиотехнический университет  
им. В.Ф. Уткина



Брянский государственный  
технический университет



МИРЭА – Российский  
технологический университет



## Предисловие

С 30 сентября по 2 октября 2025 года г. Йошкар-Ола, столица Республики Марий Эл, впервые приняла эстафету организации и проведения 35-й Международной конференции «ГрафиКон». Поволжский государственный технологический университет (Волгатех) стал центром притяжения научной элиты в области компьютерной графики, компьютерного зрения и визуализации данных.

«ГрафиКон» – одна из наиболее известных и авторитетных конференций, проводимых ежегодно в России с 1991 года. Это многоплановая научно-дискуссионная площадка, на которой отечественные и зарубежные ученые докладывают и обсуждают результаты своих исследований в областях науки, связанных с компьютерной графикой, компьютерным зрением, визуализацией данных, а также их приложениями в различных сферах науки, практики и образования. Конференция традиционно способствует развитию научных связей, объединяя усилия университетов, научных организаций и промышленных предприятий по разработке и внедрению инновационных проектов.

Кроме участия авторитетных ученых, известных в нашей стране и за рубежом, важной традицией является максимальное вовлечение магистрантов, аспирантов и молодых исследователей, для которых «ГрафиКон» служит стартовой площадкой или очередным этапом в развитии их научных идей.

На этот раз ведущей организацией для проведения столь значимого научного форума стал Поволжский государственный технологический университет. Конференция «ГрафиКон 2025» собрала более ста докладчиков из ведущих организаций России, подтвердив статус Йошкар-Олы как одного из центров развития IT-технологий. Университет располагает развитой инновационной структурой, способствующей производству новых высокотехнологичных продуктов и передовых научных исследований. Волгатех обладает мощным научным потенциалом. Исследования ученых охватывают широкий круг областей информационных, телекоммуникационных, радиоэлектронных и энергосберегающих систем и технологий, энергетики, машиностроения, строительства и биотехнологий, рационального природопользования.

Конференция стала площадкой для открытой дискуссии о насущных проблемах отрасли. Ключевые темы отражали растущую потребность в квалифицированных кадрах в области визуализации данных, необходимость ускорения интеграции научных разработок в промышленность, а также этические вопросы и актуальные проблемы безопасности в применении компьютерного зрения и искусственного интеллекта. Синергетический эффект от взаимодействия участников привел к рождению новых идей и созданию востребованных технологий.

От имени организационного и программного комитетов конференции «ГрафиКон2025» выражаю признательность её участникам за плодотворную работу, яркие доклады и конструктивные дискуссии, а также благодарю всех, кто предоставил материалы для публикации в данном сборнике. Желаю всем новых творческих успехов и дальнейших достижений в развитии компьютерной графики и смежных областей!

*А. А. Рожнецов, доктор технических наук, профессор, первый проректор  
Поволжского государственного технологического университета,  
сопредседатель организационного комитета конференции ГрафиКон 2025*





# ***ПЛЕНАРНЫЕ ДОКЛАДЫ***

---

## **PLENARY REPORTS**

**GraphiCon 2025**  
30 сентября – 2 октября 2025  
Йошкар-Ола, Россия

УДК 681.5

DOI: 10.25686/978-5-8158-2474-4-2025-10-12

## Графические технологии расширенного взаимодействия с сервисными роботами

В. Е. Пряничников<sup>1-3</sup>, В. В. Ястребов<sup>1,2</sup>, С. В. Кувшинов<sup>3</sup>, К. В. Харин<sup>3</sup>

<sup>1</sup>Институт прикладной математики им. М.В.Келдыша Российской академии наук, Москва, Россия

<sup>2</sup>Московский государственный технологический университет «СТАНКИН», Москва, Россия

<sup>3</sup>МИНОТ РГТУ Институт перспективных медиатехнологий и интеллектуальной роботоники, Москва, Россия

**Аннотация.** В работе рассматривается концептуальная структура, названная «онтологией сотрудничества», исследующая взаимосвязи и особенности совместной работы человека и машин, в особенности в контексте совместного использования роботов и технологий искусственного интеллекта. Уделяется внимание интуитивному и безопасному взаимодействию, формированию «доверия» между человеком и машиной, а также минимизации рисков, связанных с ошибками или непредсказуемым поведением. Дается обзор иммерсивных технологий и мультимодального построения интерфейсов, обеспечивающих более адекватное взаимодействие в системе «человек-робот» на основе использования психологических, профессиональных, гуманитарных особенностей человека, в том числе для подмены или комплексирования информации от различных органов чувств человека, что соответствует явлению так называемой «синестезии». Предложен и протестирован вид общения человека и робота посредством синтеза контурных рисунков эмоций, шаржей, отображаемых с помощью лазера в векторной графике на окружающих поверхностях общего для человека и робота операционного пространства либо передаваемых дистанционно через типовые интерфейсные системы. Акцент сделан на образной передаче информации.

**Ключевые слова:** взаимодействие человека и робота (HRI – human-robot interaction), онтология сотрудничества, мультимодальный интерфейс, иммерсивные технологии, синестезия, контурные рисунки, лазерная векторная графика, образная передача информации, дополненная реальность, виртуальная реальность, смешанная реальность.

## Graphical technologies for enhanced interaction with service robots

V. E. Pryanichnikov<sup>1-3</sup>, V. V. Iastrebov<sup>1,2</sup>, S. V. Kuvshinov<sup>3</sup>, K. V. Kharin<sup>3</sup>

<sup>1</sup> Keldysh Institute of Applied Mathematics (KIAM) Russian Academy of Sciences, Moscow, Russia

<sup>2</sup> Moscow State University of Technology "STANKIN" (MSTU STANKIN), Moscow, Russia

<sup>3</sup> INET RSUH Institute of Advanced Media Technologies and Intelligent Robotics, Moscow, Russia

**Abstract.** This paper considers a conceptual structure called the "ontology of cooperation", which explores the interrelationships and features of human and machine collaboration, especially in the context of the joint use of robots and artificial intelligence technologies. Attention is paid to intuitive and safe interaction, trust between a human and a machine, as well as minimizing the risks associated with errors or unpredictable behavior. An overview is given of immersive technologies and multimodal interface construction that ensure more adequate interaction in the human-robot system based on the use of psychological, professional, and humanitarian human characteristics, including for substituting or integrating information from various human sensory organs, which corresponds to the phenomenon of the so-called "synesthesia." A type of human-robot communication is proposed and tested by means of contour drawings in the form of emotions, cartoons displayed using a laser in vector graphics on the surrounding surfaces of a space common to humans and robots, or transmitted remotely through the standard interface systems. The emphasis is on the imaginative transfer of information.

**Keywords:** human-robot interaction (HRI), ontology of cooperation, multimodal interface, immersive technologies, synesthesia, contour drawings, laser vector graphics, imaginative information transfer, AR-augmented reality, VR-virtual reality, MR-mixed reality.

### Введение

Сотрудничество человека и робота может быть представлено в виде концептуальной структуры, которая исследует взаимосвязи и особенности совместной работы человека и машин, особенно в контексте использования роботов и технологий искусственного интеллекта. Такую концептуальную структуру можно назвать «онтологией сотрудничества», которая рассматривает, как люди и роботы могут эффективно и «синергетически» работать вместе, учитывая особенности графических интерфейсов, психологические, социальные, этические и технические факторы (некоторые публикации из данной области приведены в работах [1-9]). Одним из ключевых элементов в этой онтологии является понимание ролей каждого из участников сотрудничества.



### Субъективные факторы проектирования интерфейсов человек-робот

Люди и роботы имеют различающиеся сильные стороны и недостатки: люди могут проявлять творчество, эмоциональный интеллект и способность к принятию решений в условиях неопределенности, в то время как роботы, системы ИИ и машинного обучения могут обеспечивать высокую производительность, точность и способность обрабатывать большие объемы данных.

Онтология сотрудничества рассматривает то, как эти сильные стороны могут дополнять друг друга, чтобы достичь общей цели, будь то в производственной среде, в сфере обслуживания или в других областях, в том числе с учётом динамики взаимодействия. Это предполагает необходимость проектирования систем, обеспечивающих интуитивное и безопасное взаимодействие, доверие между человеком и машиной, а также минимизацию рисков, связанных с ошибками или непредсказуемым поведением. В этом контексте вопросы этики и ответственности также становятся значимыми, поскольку они касаются того, кто отвечает за действия робота (разработчик, эксплуатант, пациент).

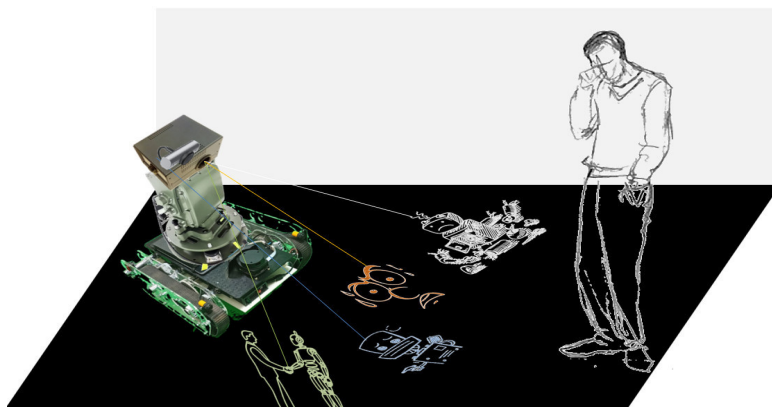
Онтология сотрудничества включает в себя изучение контекстных факторов, таких как культурные различия, специфика отрасли и операционной среды. Например, в сфере, где высока необходимость в гуманитарном подходе и социальном взаимодействии, роль человека будет более значительной, тогда как в производственных областях роботы могут занять более активную позицию.

Таким образом, онтология сотрудничества человека и робота предполагает многофакторный подход к пониманию динамики совместной работы в быстро развивающемся технологическом ландшафте (*rapidly evolving technological landscape*). Она помогает находить баланс между интеграцией инновационных технологий и сохранением человеческих ценностей и отношений, что является актуальным вопросом для современного общества, в том числе при использовании технологий виртуальной реальности.

### Разработки коллаборативных сервисных роботов

В докладе приведён краткий обзор иммерсивных технологий и мультимодального построения интерфейсов, обеспечивающих более адекватное взаимодействие в системе «человек-робот» на основе использования психологических, профессиональных, гуманитарных особенностей человека, в том числе для подмены или комплексирования информации от различных органов чувств человека (явление так называемой синестезии).

На основании этого обзора был выполнен ряд работ, приведших к созданию соответствующих программно-аппаратных графических средств и интерфейсов для коллаборативных роботов, взаимодействующих с человеком в общей операционной среде. Разработанные сервисные мобильные роботы (см. рисунок) и сенсорные системы различного назначения были сведены в роботариум, который стал основой для проведения научно-исследовательских работ и для подготовки специалистов в области робототехники.



Вычерчивание роботом эмоциональной и образной информации с помощью RGB-лазера

### Заключение

В представленном обзоре рассмотрены и проанализированы разнообразные способы общения и сотрудничества человека и робота. Уделено внимание физиологическим и психологическим аспектам этого общения и сотрудничества.

На основании проведённого анализа предложен и протестирован вид общения человека и робота посредством контурных рисунков типа эмоций, шаржей, отображаемых с помощью лазера в векторной графике на окружающих поверхностях общего для человека и робота пространства либо передаваемых дистанционно через типовые интерфейсные системы. Акцент был сделан на образной передаче информации. Она воспринимается правым полушарием и может мгновенно быть воспринята всеми органами чувств, включая обоняние, тактильные ощущения. В этом смысле можно говорить об искусственной синестезии, с помощью которой робот создаёт у человека ощущения восприятия одновременно всеми органами чувств.

Векторная лазерная графика предложена также с целью подмены роботом (который не является человекоподобным) своего имиджа, имитирующего человеческое общение. Соображения, представленные в этом обзоре, легли в основу наших разработок коллаборативных сервисных роботов, лазерных и сенсорных систем, используемых как в исследовательских целях, так и для ведения совместной образовательной деятельности в рамках проекта «Интеллектуальная робототроника».

### **Источник финансирования**

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (тема 07, FFMN-2025-0022).

### **Список литературы**

1. Пряничников В. Е., Ястребов В. В. Способ свёртки изображений для последующей идентификации объектов сервисными роботами // Запись и воспроизведение объёмных изображений в кинематографе, науке, образовании, медиа и в других областях: XVI ежегодная научно-практическая конференция: материалы и доклады. Часть II / под общ. ред. О. Н. Раева. – Москва: КУНА, 2024. С. 128-135. ISBN 978-5-98547-148-9. URL: [https://www.rsuh.ru/education/minot/conf\\_3d-2024.php](https://www.rsuh.ru/education/minot/conf_3d-2024.php)
2. Давыдов О. И., Пряничников В. Е. Архитектура системы управления мобильного сервисного робота // Информационно-измерительные и управляющие системы. 2015. Т. 13, № 7. С. 41-50.
3. Davydov D. V., Eprikov S. R., Kirsanov K. B., Pryanichnikov V. E. (2017). Service Robots Integrating Software and Remote Reprogramming, Proceedings of the 28th DAAAM International Symposium, pp. 1234-1240, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-11-2, ISSN 1726-9679, Vienna, Austria. DOI: 10.2507/28th.daaam.proceedings.172.
4. Smith J.R., & Johnson M.K. (2023). Immersive Technologies for Human-Robot Interaction: Current Trends and Future Directions. Robotics and Autonomous Systems, 159, 103-121. <https://doi.org/10.1016/j.robot.2023.103121>
5. Lee S., Kim H., & Park T. (2022). Augmented Reality Interfaces for Collaborative Robotics. IEEE Transactions on Human-Machine Systems, 52(4), 678-690.
6. Chen X., Wang Y., & Zhang Q. (2021). Virtual Reality in Robot Teleoperation: A Systematic Review. International Journal of Social Robotics, 13(5), 1023-1045.
7. Hernandez M.G., & Williams B. (2020). The Psychology of Immersive HRI: Building Trust Through Virtual Environments. ACM Transactions on Human-Robot Interaction, 9(3), 1-24.
8. Anderson P.L., & Taylor R.S. (2019). Mixed Reality for Industrial Human-Robot Collaboration. Springer Series on Automation and Control Engineering. Springer.

УДК 621.39

DOI: 10.25686/978-5-8158-2474-4-2025-13-21

## Метод обнаружения принимаемых мод сигнала на изображении ионограммы радиозондирования ионосферы с использованием машинного обучения

В. А. Иванов, Н. А. Конкин, В. В. Овчинников, М. И. Рябова

Поволжский государственный технологический университет, Йошкар-Ола, Россия

**Аннотация.** В статье представлен комплексный метод автоматической обработки изображений ионограмм, полученных в результате наклонного зондирования ионосферы, с целью выявления полезных сигналов на фоне флуктуационных и сосредоточенных помех. Предлагаемый метод сочетает классические подходы к фильтрации (обеляющее и медианное эквалайзирование) с методами машинного обучения на основе кластерного анализа. Основным компонентом является алгоритм машинного обучения на основе кластеризации k-means, применяемый к изображению ионограммы, преобразованному в формат цветового пространства RGB, для автоматического выделения мод сигнала. Предварительная фильтрация обеспечивает стандартизацию распределения интенсивности сигнала и снижение влияния шумов, что позволяет повысить точность сегментации. В серии натурных экспериментов, проведённых на радиотрассе Кипр – Йошкар-Ола, показано, что оптимальное число кластеров (11) обеспечивает наилучшее разделение компонент ионограммы по уровню сигнал/шум. Разработанный метод может быть использован в составе когнитивных радиосистем для автоматического анализа и выбора рабочих частот радиоканалов в условиях динамично изменяющейся среды.

**Ключевые слова:** ионограмма, зашумленное изображение, фильтрация помех, кластеризация, k-means, машинное обучение, автоматическое обнаружение сигнала, КВ-связь, когнитивные системы.

## Echo Signal detection on ionograms of ionospheric sounding with the use of machine learning methods

V. A. Ivanov, N. A. Konkin, V. V. Ovchinnikov, M. I. Ryabova

Volga State University of Technology, Yoshkar-Ola, Russia

**Abstract.** This paper presents a comprehensive method for the automatic processing of ionogram images obtained through oblique ionospheric sounding, aimed at identifying echo signals amid both fluctuating and narrowband interference. The proposed approach combines classical filtering techniques—specifically whitening and median equalization—with machine learning methods based on cluster analysis. The core component of the method is the k-means clustering algorithm, which is applied to the ionogram image converted into the RGB color space to automatically extract signal modes. The preliminary filtering stage standardizes the signal intensity distribution and reduces noise influence, thereby improving segmentation accuracy. A series of field experiments conducted along the Cyprus–Yoshkar-Ola radio path demonstrated that using 11 clusters yields the most effective separation of ionogram components based on their signal-to-noise ratio. The developed method can be integrated into cognitive radio systems for automatic analysis and selection of operational frequency bands in highly dynamic propagation environments.

**Keywords:** ionogram, noisy image, interference filtering, clustering, k-means, machine learning, automatic signal detection, HF communication, cognitive systems.

### Введение

Методы машинного обучения для обработки изображений нашли широкое применение в различных областях техники [1]. Важным направлением в научных исследованиях является обработка изображений с целью дальнейшего получения информации об изучаемом объекте [2]. В частности, к этому направлению относится и задача извлечения информации о параметрах многомерного ионосферного канала ВЧ-связи из изображений ионограмм, получаемых в результате зондирования ионосферы. Ионограммы, полученные при помощи вертикального и наклонного зондирования ионосферы, являются ключевыми источниками информации о многомерной структуре среды распространения. Однако их практическое использование осложняется наличием шумов, вызванных как природными, так и антропогенными источниками, а также высокой изменчивостью ионосферных условий. При этом полезные сигналы и сигналы помех являются стохастическими процессами. В условиях когнитивных радиосистем и автономных радиосенсоров возникает необходимость автоматического анализа ионограмм для выделения частотных каналов с наилучшим отношением сигнал/шум. В этой задаче на первом этапе требуется получить изображение ионограммы, очищенное от помех. В таком случае оставшееся изображение может применяться для дальнейшего анализа при помощи машинного обучения с целью выделения полезного сигнала. Такая очистка позволит

определять в автоматическом режиме необходимые для пользователей параметры канала ВЧ-связи, а также осуществлять автоматизированную обработку большого количества ионограмм.

Вопросам очистки ионограмм посвящено достаточно много работ различных авторов [3, 4]. В настоящем исследовании предлагается комплексный метод, сочетающий классические фильтрационные подходы (обеляющее и медианное эквалайзирование) с машинным обучением, в частности алгоритмом кластеризации k-means [5], для автоматического выделения полезного сигнала на ионограммах. В отличие от нейросетевых решений [6], требующих длительного обучения и больших объёмов размеченных данных, используемый подход обеспечивает быструю и интерпретируемую сегментацию, адаптируемую к конкретным условиям помех. Новым будет являться разработка комплексного метода с применением алгоритма машинного обучения, который можно реализовать в автоматическом режиме. Он предназначен для определения по результатам интеллектуальной обработки ионограмм частотной зависимости отношения сигнал-шум для всего множества каналов ВЧ-диапазона и является актуальным для выбора оптимального из них.

### Постановка задачи

В КВ-диапазоне имеют место изменчивые флуктуационные и сосредоточенные помехи. Последние обусловлены работой посторонних систем КВ-связи. Полосы частот сосредоточенных помех обычно составляют 3 кГц или кратны этой величине. Время стационарности вариаций в основном не меньше 10 с.

Спектр комплексной амплитуды принимаемой смеси сигнала и помех можно записать в виде следующей суммы:

$$Y_R(j\omega) = U_R(j\omega) + \sum_{k=1}^{k_0} N_k(j\omega) + N_0(j\omega),$$

где  $U_R(j\omega)$  – спектр зондирующего сигнала;  $N_k(j\omega)$  – спектр отдельной сосредоточенной помехи;  $N_0(j\omega)$  – спектр флуктуационной помехи.

Спектр мощности сигнала разностной частоты  $\Omega$  можно представить в виде:

$$S_\tau^2(\Omega = \tau \cdot \dot{\omega}) = S^2(\Omega = \tau \cdot \dot{\omega}) + \sum_{k=1}^{k_0} N_k^2(\Omega) + N_0^2(\Omega).$$

Видно, что шум присутствует как на задержках принимаемого сигнала, так и на остальных задержках из области априорной неопределенности. Решение задачи автоматического выделения в матрице–ионограмме  $S_\tau^2(\Omega = \tau \cdot \dot{\omega})$  отсчетов зондирующего сигнала требует применения эффективных методов фильтрации отсчетов помех и машинного обучения для обнаружения сигнала на фоне оставшегося при фильтрации шума. Для фильтрации помех необходимо последовательно применять ряд адаптивных алгоритмов, таких как обеляющее и медианное эквалайзирование. Их рассмотрение для решения поставленной задачи представлено в работах [4, 7, 8].

Целью настоящей работы является разработка и экспериментальная проверка метода автоматического обнаружения мод принимаемого сигнала на ионограммах наклонного зондирования ионосферы в условиях флуктуационных и сосредоточенных помех. Метод должен обеспечивать: подавление помех и шумов за счёт предварительной адаптивной обработки ВЧ-сигнала до его сжатия; нормализацию распределения амплитудных отсчётов для повышения устойчивости алгоритма обработки; автоматическую сегментацию ионограммы с применением алгоритма машинного обучения для выделения кластеров с различным отношением сигнал/шум.

Задача формализуется как поэтапная обработка с предварительным обеляющим эквалайзированием широкополосного принимаемого сигнала на этапе его спектрального представления и последующая обработка матрицы амплитудных отсчётов с целью выделения кластеров, характеризующихся заданными уровнями отношения сигнал/шум.

### Теория

Для исследования и анализа методов фильтрации помех и получения подготовленных изображений ионограмм для обучения алгоритма выделения полезного сигнала нами будут рассматриваться вариационные ряды для величины

$$G(m, n) = 10 \lg \frac{S_r^2(m, n)}{Me(S_r^2)_n} = 10 \lg \hat{S}_r^2(m, n), \quad (1)$$

где  $m$  – номер строки матрицы-ионограммы, соответствующий задержке;  $n$  – номер столбца этой матрицы, соответствующий средней частоте парциального канала;  $Me(S_r^2)_n$  – медиана отсчетов  $S_r^2(m, n)$  для столбца.

Формула (1) иницирует вторую матрицу того же размера, но с элементами  $G(m, n)$ . Наши исследования показывают, что действие этих алгоритмов целесообразно проанализировать на основе характеристик вариационных рядов, для которых вариантами являются как отсчеты  $S_r^2(m, n)$ , так и  $G(m, n)$ . Для обеих функций частоты (номер варианты в ряду, отнесенный к объему выборки) определяют вероятность того, что варианты с меньшими значениями не превысят его, а разность между единицей и этой частотой – вероятность того, что имеются варианты с большими чем в данном отсчете значениями. Поэтому вариационный ряд можно рассматривать как функцию, обратную функции распределения вероятности вариантов.

Обеление осуществляется до операции сжатия сигнала в приёмнике. На начальном шаге находится текущий спектр мощности принимаемого сигнала  $Y_R(\omega)$  на заданных с некоторым шагом временных выборок, определяющим разрешение в спектре по частоте. Далее спектр мощности усредняется по выборкам, полученным за время его стационарности, равное 10 с [7]. Усредненный спектр амплитуд  $\sqrt{\langle Y_R^2(\omega) \rangle}$  используется для эквалайзирования последующих далее спектров, пока за время стационарности набирается следующая для усреднения выборка. Адаптация заключается в том, что в данном алгоритме учитывается изменение во времени усреднённого спектра обеляемых помех. Имеющая место изменчивость помех приводит к тому, что алгоритм все же пропускает на выход небольшое число сосредоточенных помех с уровнем до 10–15 дБ [7, 8, 9]. На рисунке 1 представлен алгоритм действия обеляющего эквалайзирования.

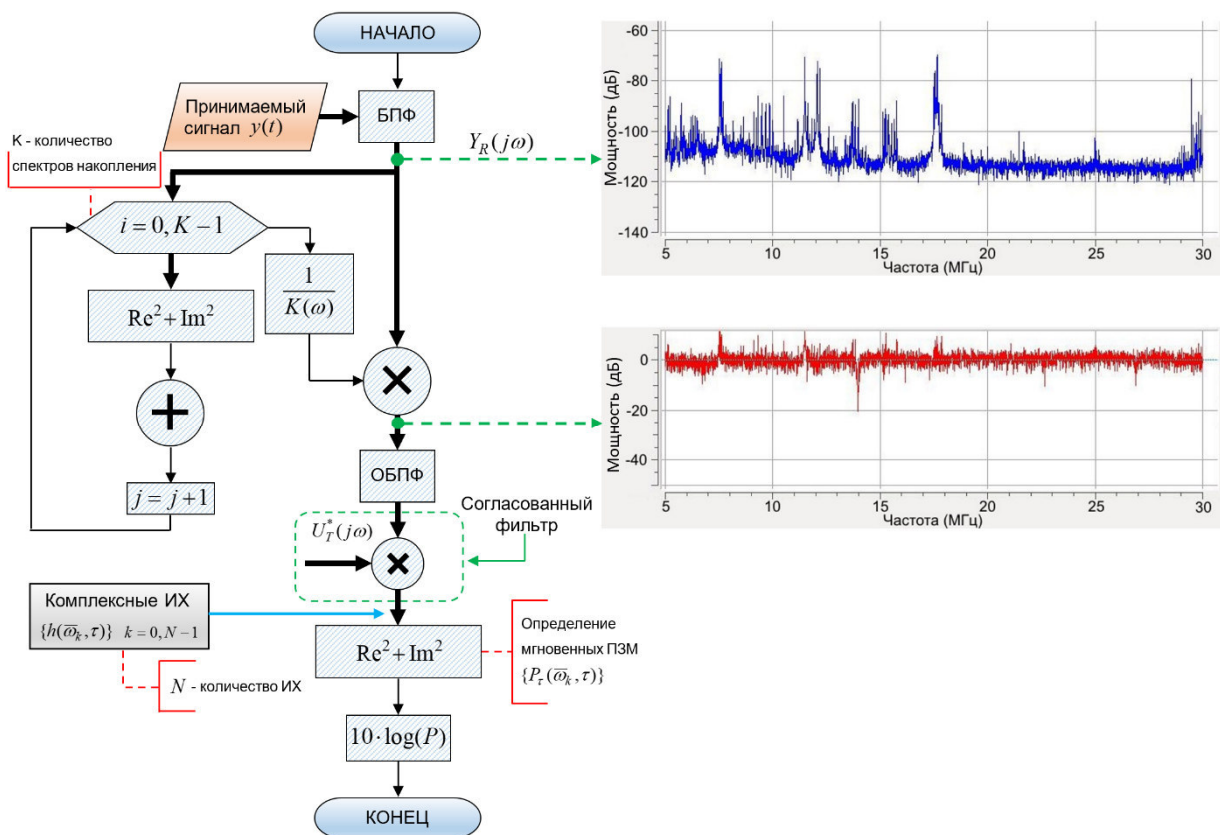


Рис. 1. Алгоритм действия обеляющего эквалайзирования

На рисунке 2 в качестве примера действия алгоритма обеления приведены вариационные ряды для столбцов матрицы с элементами  $S_{\tau}^2(m, n)$  до обеления и после него.

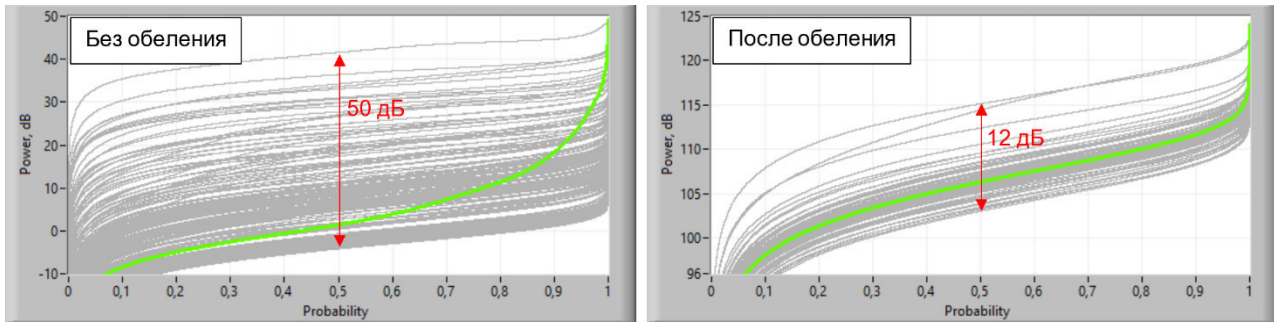


Рис. 2. Вариационные ряды полного множества вектор-столбцов матрицы-ионограммы:  
а) без обеления; б) после обеления

В задаче получения ионограммы варианты, принадлежащие сигналу, находятся в малой левой окрестности вероятности, равной 1. Поэтому подавляющее большинство отсчетов варианты соответствует только шуму. Обеление приводит к существенному уменьшению разброса значений варианты шума, который до фильтрации достигал значений 50 дБ, а после нее уменьшился до 12 дБ.

Различия в вариационных рядах связаны с отличием для шумов параметров их функции распределения вероятностей. Обратим внимание на то, что обратная функция к вариационному ряду является по сути законом распределения вероятности. Известно, что для однопараметрического закона путем деления варианты на некоторый параметр можно стандартизировать функцию распределения. Очевидно, что в нашем случае в качестве такого параметра следует выбрать медиану  $Me(S_{\tau}^2)_n$  для вектор-столбца. Вариационный ряд для столбца  $(S_{\tau}^2)_n$  позволяет оценить медиану, которая является характеристикой шума, так как его отсчеты преобладают в элементах столбца. Деление каждого столбца матрицы на его медиану называют медианным эквалайзиром. На рисунке 3 представлен алгоритм действия медианного эквалайзера, выполняющегося после процедуры обеления.

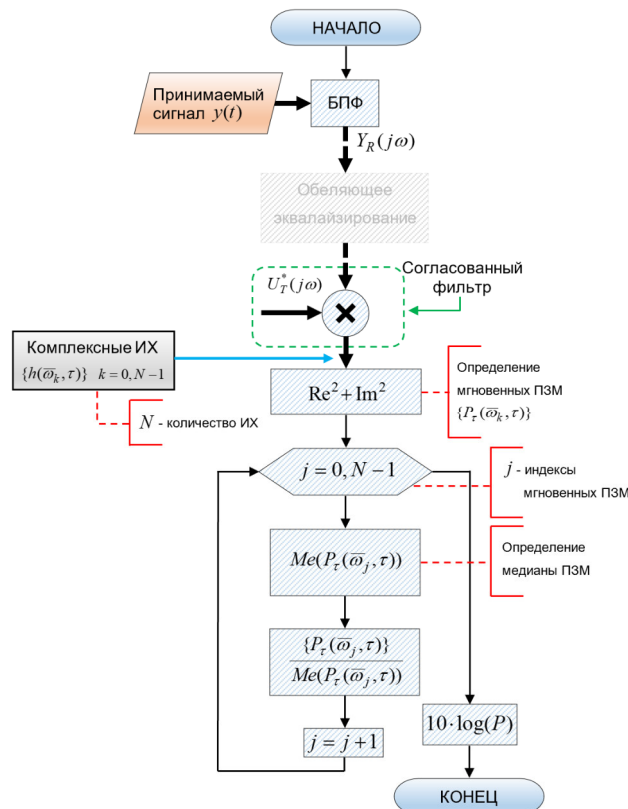


Рис. 3. Алгоритм действия медианного эквалайзера



Рисунок 4 иллюстрирует действие медианного эквалайзирования в случаях без предварительного обеления и с обелением.

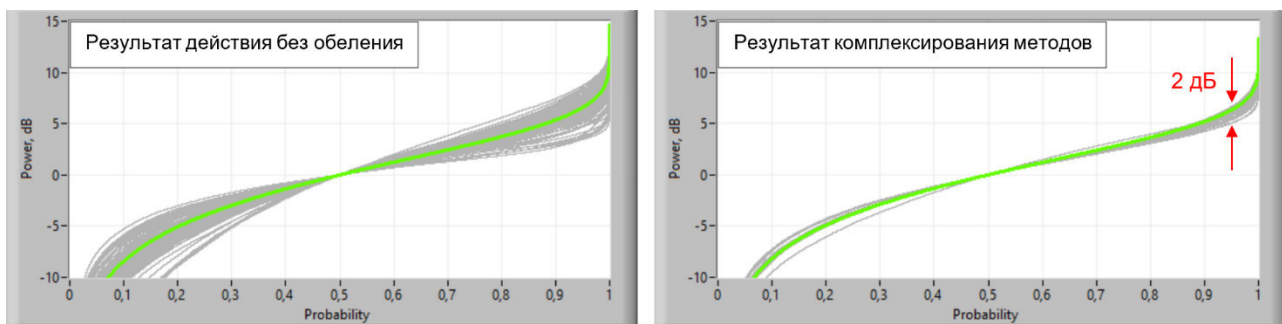


Рис. 4. Вариационные ряды полного множества вектор-столбцов матрицы-ионограммы: без обеления, но с медианным эквалайзированием (слева); с обелением и медианным эквалайзированием (справа)

Как и следовало ожидать, медианный эквалайзер, особенно при обеляющем эквалайзировании, практически стандартизирует функцию распределения, так как разброс варианты составляет всего 2 дБ (см. рис. 4). В результате такого эквалайзирования середина каждой функции  $G(m, n)$  смещается в нуль, в том числе и для вариационного ряда из элементов всей матрицы. Вариационные ряды для такого множества отсчетов изображены на рисунке 4 зеленым цветом. Обратим внимание на то, что в случае применения обеления и медианного эквалайзирования вариационный ряд для всей матрицы практически совпадает с рядом для столбца с самыми высокими значениями варианты. Этот эффект позволяет применять полученные изображения ионограмм для обучения алгоритма обнаружения полезного сигнала.

На рисунке 5 представлены ионограммы до обработки и после обработки комплексным алгоритмом [10]. Вертикальные прямые являются следствием воздействия на соответствующие каналы мощных сосредоточенных помех, а принимаемые с сигналом флуктуационные помехи создают зашумленность изображения. Диапазон частот, где присутствует отраженный сигнал, определяет полосу прозрачности линии КВ-связи (полосу допустимых рабочих частот). Ее верхней границей является максимально наблюдаемая частота (МНЧ).

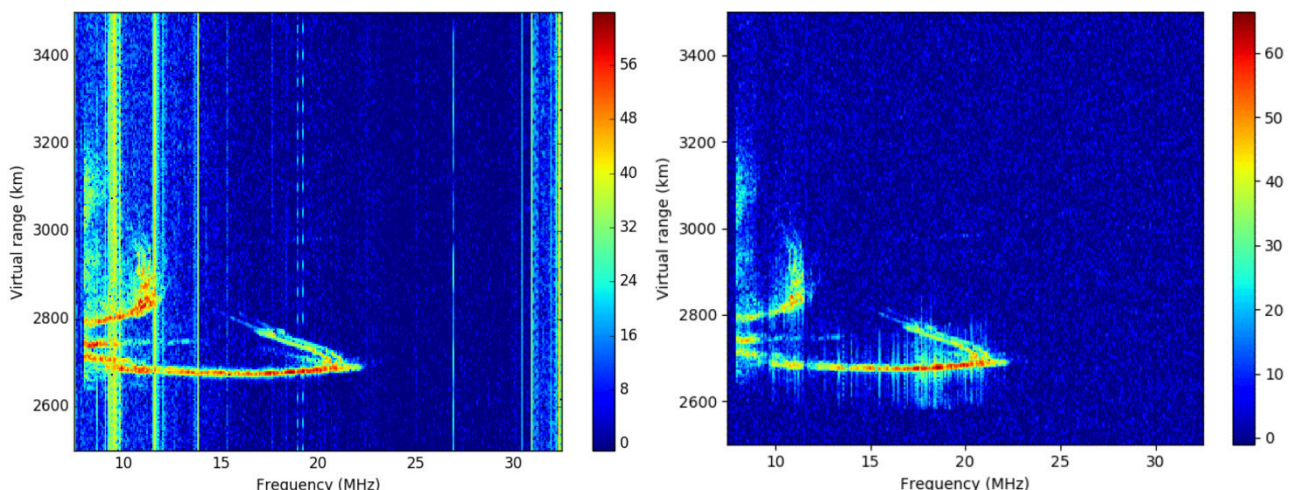


Рис. 5. Ионограммы до обработки (слева) и после обработки (справа)

Полученное после обработки изображение ионограммы содержит только отсчёты полезного сигнала и фонового шума, близкого к белому. Данные изображения применялись далее для обучения алгоритма автоматического обнаружения принимаемых мод сигнала на основе кластерного анализа данных.

В настоящей работе кластерный анализ является частью разведочного и используется для выделения полезного сигнала при анализе ионограмм наклонного зондирования ионосферы. Ключевой задачей является автоматическое определение групп точек, значение сигнал/шум которых удовлетворяет заданному порогу. Для решения данной задачи применен алгоритм машинного обучения k-means [5]

как наиболее удобный для работы с изображениями. Суть алгоритма заключается в грубом распределении центров кластеров исследуемого набора данных  $(x_i, y_i)$ , после чего центр  $(\bar{x}_l, \bar{y}_l)$   $l$ -го кластера вычисляется как

$$\bar{x}_l = \frac{1}{q_l} \sum_{i \in \text{кластер } l} x_i, \bar{y}_l = \frac{1}{q_l} \sum_{i \in \text{кластер } l} y_i,$$

где  $q$  – суммарный размер массива ионограммы.

Дальнейшее разделение исходной выборки на подвыборки по принципу минимизации суммы квадратов расстояний от каждой точки подвыборки до центра кластера. Сумма квадратов внутри кластера определяется следующим образом:

$$NN_l = \sum_{i \in \text{кластер } l} (x_i - \bar{x}_l)^2 + (y_i - \bar{y}_l)^2.$$

На следующем этапе итеративно выполняется коррекция центров каждого кластера за счет минимизации дисперсии до достижения заданного порога по точности или по количеству итераций по следующей формуле:

$$\arg \min_N \sum_{i=1}^l NN_i$$

где  $\arg \min$  – аргумент минимизации;  $N$  – набор всех кластеров.

Ионограммы, полученные после предварительной фильтрации шумов, перед применением алгоритма k-means преобразуются в трехмерный массив средствами библиотеки OpenCV (Open Source Computer Vision Library) [11], доступной для использования на языке программирования Python. Полученный массив содержит количество строк, равное высоте изображения (ионограммы), количество столбцов, равное ширине изображения, а каждая ячейка по высоте и ширине содержит значения уровня сигнал/шум, представленного в виде трех значений кодировки цветности по стандарту RGB. Полученный трехмерный массив является входным аргументом для алгоритма k-means. В исследовании использовалась функция «kmeans» библиотеки OpenCV, в параметрах настройки которой задаются требуемое количество кластеров, количество итераций до остановки алгоритма, количество итераций, когда группировка выполняется заново с использованием разных начальных центров, и тип функции выбора начальной точки кластера.

### Результаты экспериментов

Верификация предложенного метода в натурных экспериментах на среднеширотной радиотрассе о. Кипр – г. Йошкар-Ола. В экспериментах анализировались ионограммы с применением различного количества кластеров. В алгоритме машинного обучения диапазон значений числа выбираемых кластеров на изображении изменялся от 5 до 11. В результате экспериментов было установлено, что увеличение количества кластеров модели k-means позволяет более точно выделить границы мод ионограммы с низким отношением сигнал/шум. На рисунке 6 представлены пример оригинального изображения ионограммы после предварительной обработки и результат выделения двух полученных кластеров из 11.

На рисунке 6 (снизу) показан принцип выделения кластеров с максимальными значениями сигнал/шум на изображении ионограммы с помощью представленного в предыдущем пункте алгоритма машинного обучения. Красным цветом отображены варианты сигнала, которые содержат максимальные значения отношения сигнал/шум, оранжевым – зона с меньшим уровнем полезного сигнала. В результате обработки удалено 8 из 11 кластеров. Варианты с низким отношением сигнал/шум (менее 25 дБ) были удалены алгоритмом с изображения.

На рисунке 7 представлена ионограмма, очищенная с использованием идентичных 11 кластеров, но с сохранением уровня более высокого диапазона сигнал/шум. В таком случае алгоритмом детектируется сигнал с низким отношением сигнал/шум (2-5 дБ).



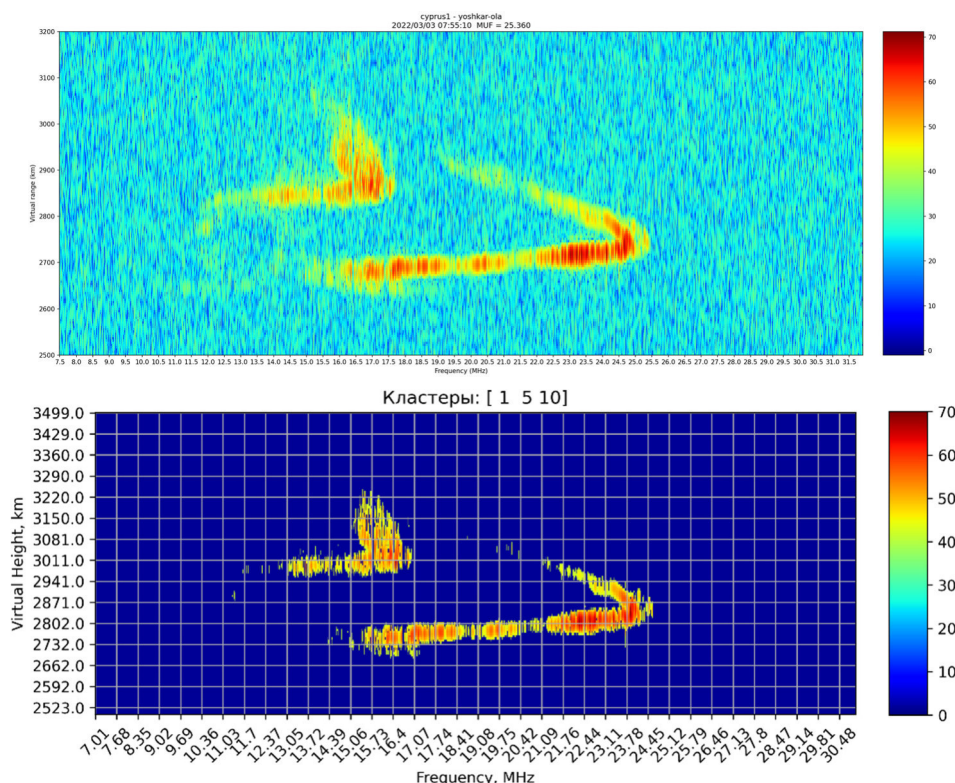


Рис. 6. Ионogramмы наклонного зондирования ионосферы: оригинальная ионogramма (сверху) и ионogramма (снизу) с выделенными кластерами максимального уровня сигнал/шум, очищенная с использованием 11 кластеров

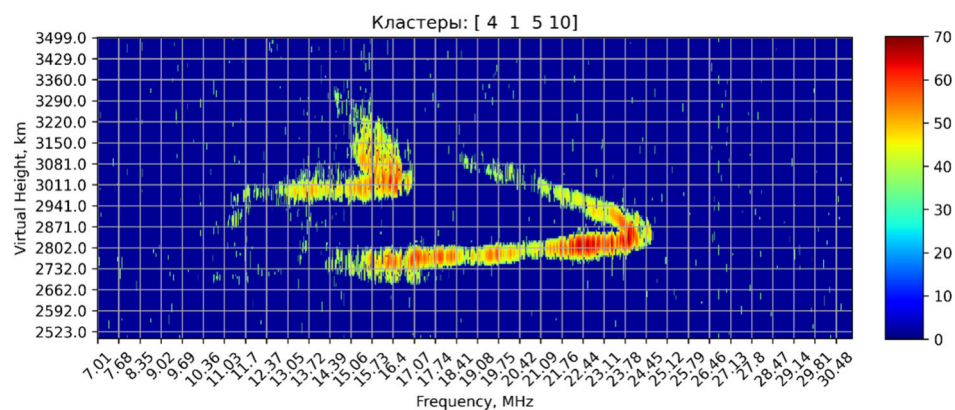


Рис. 7. Ионogramмы наклонного зондирования ионосферы, очищенная с использованием 11 кластеров (7 из 11 отфильтрованы), с сохранением сигналов с низким отношением сигнал/шум дБ

### Обсуждение результатов

Проведённый анализ влияния входных параметров алгоритма k-means на результат кластеризации показал, что ключевым параметром для настройки является используемое количество кластеров. При выборе количества кластеров от 8 до 11 полезная область ионogramмы определяется минимум тремя кластерами, центральным (см. рис. 6), где сосредоточены максимальные значения отношения сигнал/шум и побочными со средним и минимальным уровнем. При количестве кластеров до 7 алгоритм детектирует значительно большее количество сигнала с минимальным отношением сигнал/шум 2-5 дБ за один кластер, но в таком случае отсутствует возможность коррекции побочных составляющих сигнала за счёт добавления или удаления дополнительных кластеров. Полученное после обработки изображение позволяет осуществлять в автоматическом режиме дальнейшие вычисления требуемых параметров коротковолновой линии связи, таких как максимально применимая частота, наименьшая применимая частота, оптимальная радиочастота.

Многочисленные эксперименты, проведённые на трассе о. Кипр – г. Йошкар-Ола, показали, что при использовании количества кластеров от 5 до 7 целесообразно выделять только один кластер, а от 8 до 11 – по три кластера. Обнаружение вариант с минимальными потерями полезного сигнала достигается при использовании 11 кластеров. Для автоматизации вычисления номера кластера, который несет полезное значение сигнал/шум, вычисляется средний уровень отношения сигнал/шум [12] по точкам кластера с последующим выбором двух максимальных.

### Заключение

При зондировании многомерного КВ-радиоканала и ионосферы возникает задача автоматической обработки изображений ионограмм с целью выделения полезного сигнала над помехами. Ее сложность заключается в необходимости учета изменчивости помеховой обстановки за время не менее 10 с, изменчивости среды распространения за время, не меньшее 5 мин. Для решения задачи в работе предложен комплексный метод обнаружения принимаемых мод сигнала на изображении ионограммы с применением алгоритмов обеляющего эквалайзирования ВЧ-сигнала, медианного эквалайзирования сигнала после его сжатия и машинного обучения на основе кластерного анализа изображений ионограмм.

Экспериментально показано, что предварительная обработка алгоритмами обеляющего и медианного эквалайзирования стандартизирует функции распределения вариант отсчётов сигнал+шум, уменьшая разброс значений от 50 до 2 дБ. В натурных экспериментах исследовалось применение различного числа кластеров изображения ионограммы для автоматического обнаружения сигнала. Многочисленные эксперименты, проведённые на трассе о. Кипр – г. Йошкар-Ола, показали, что при использовании количества кластеров от 5 до 7 целесообразно выделять только один кластер, а от 8 до 11 – по три кластера. Обнаружение вариант с минимальными потерями полезного сигнала достигается при использовании 11 кластеров.

Предложенный метод может быть внедрён в состав когнитивных радиосенсоров и систем адаптивной радиосвязи. В перспективе работа может быть расширена с применением сверточных нейросетей или гибридных подходов с предварительной кластеризацией.

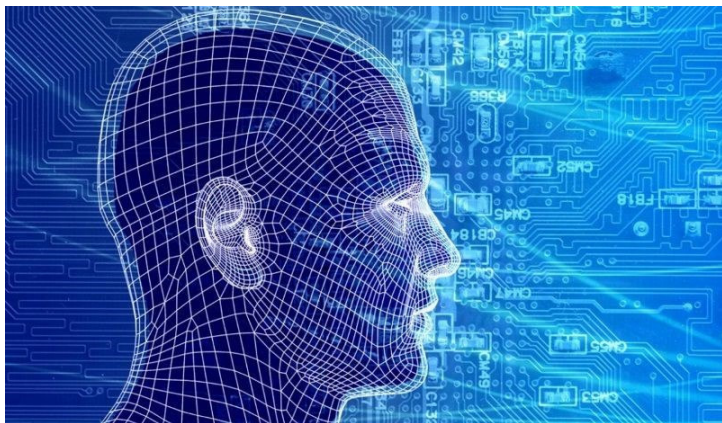
### Источник финансирования

Работа выполнена при финансовой поддержке Российского научного фонда, грант (Теория, методы и средства сенсорной диагностики доступности каналов когнитивной КВ радиосвязи с учетом частотно-временной дисперсии, канальных помех и с применением технологий программно-конфигурируемых радиосистем и анализом больших данных, №24-29-00442).

### Список литературы

1. A. Bruce, P.C. Bruce, and P. Gedeck, "Practical Statistics for Data Scientists," O'Reilly Media, Inc., 2020 (ISBN 10: 9781491952962).
2. H. Singh, "Practical Machine Learning and Image Processing: For Facial Recognition, Object Detection, and Pattern Recognition Using Python," Apress, February 2019 (ISBN 10: 9781484241493).
3. C. De La Jara, and C. Olivares, "Ionospheric echo detection in digital ionograms using convolutional neural networks," *Radio Science*, vol. 56, no. 8, pp. 1-15, Aug. 2021, doi: 10.1029/2020RS007258.
4. V.A. Ivanov, D.V. Ivanov, N.V. Ryabova, M.I. Ryabova, A.A. Chernov, A.A. Elsukov, "Computer filtering of images of ionograms of HF ionospheric radio communication paths and an algorithm for determining signal-to-noise ratio," *Journal of Applied Engineering Science*, 16(1), pp. 116–124, 2018. doi: 10.5937/jaes16-11790.
5. J. Grus, "K-means and hierarchical clustering with Python," O'Reilly Media, Inc., August 2016 (ISBN 10: 9781491966174).
6. P. Benchawattananon, A. Siritatiwat, P. Supnithi, et al. "Automatic classification of spread-F types in ionogram images using support vector machine and convolutional neural network", *Earth Planets Space* 76, 56, 2024.
7. V.A. Ivanov, D.V. Ivanov, N.V. Ryabova, M.I. Ryabova, A.A. Chernov, and A.A. Elsukov, "Algorithms and software of whitening of noise in the reception of broadband chirp signals in the HF communication channel," *Journal of Applied Engineering Science*, 13(4), pp. 245–250, 2015. doi: 10.5937/jaes13-9342.
8. D.V. Ivanov, V.A. Ivanov, A.A. Elsukov, N.V. Ryabova, V.V. Ovchinnikov, and M.I. Ryabova "Algorithms of adaptive narrow-band interferences suppression and automatic signal detection for swept-frequency sounding of multiple HF radio channels," *URSI AP-RASC 2019*, New Delhi, India, 09 - 15 March 2019.

- 
9. H.M. Finn and R.S. Johnson, "Adaptive detection mode with threshold control as a function of spatially sampled clutter level estimates," *RCA Rev.*, vol. 29, no. 3, pp. 414–464.
  10. D. Ivanov, V. Ivanov, N. Ryabova, N. Konkin, and V. Ovchinnikov, "Method for Detecting Modes of Received Echo Signal on Image of Ionogram of Radio Sounding with the Use of Machine Learning," 2022 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), St. Petersburg, Russian Federation, 2022, pp. 1-5, doi: 10.1109/WECONF55058.2022.9803515.
  11. OpenCV. (2022, April 13). Open Source Computer Vision Library: OpenCV modules. <https://docs.opencv.org/4.x/>
  12. D.V. Ivanov, V.A. Ivanov, R.R. Belgibaev, and A.R. Lashevsky, "Adaptive Algorithms for Filtering Ionograms and Measuring Channel Parameters to Estimate its Availability," 2019 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), 2019, pp. 1-4, doi: 10.1109/SYNCHROINFO.2019.8813917.



Секция 1

***РЕАЛИСТИЧНАЯ КОМПЬЮТЕРНАЯ  
ГРАФИКА И ВЫЧИСЛИТЕЛЬНАЯ  
ОПТИКА***

---

**REALISTIC COMPUTER GRAPHICS  
AND COMPUTATIONAL OPTICS**



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия

## Сравнительный анализ выполнения нейронных моделей для графики

Н. А. Милин<sup>1,2</sup>, Р. О. Родионов<sup>1,2</sup>, В. А. Галактионов<sup>2</sup>, В. А. Фролов<sup>1,2</sup>

<sup>1</sup>Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем  
МГУ имени М.В. Ломоносова, Москва, Россия

<sup>2</sup>Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В данной работе представлено исследование производительности различных подходов к реализации выполнения нейронных моделей в задачах компьютерной графики, связанных с представлением поверхностей и материалов трёхмерных объектов. Рассматриваются четыре различных метода, среди них как универсальные, поддерживаемые на большинстве устройств, так и использующие специальные аппаратные возможности. Экспериментальное исследование проведено на двух задачах: задание поверхности в виде нейронной функции дистанции и задание двухлучевой функции отражательной способности в виде нейронной модели. Результаты демонстрируют существенные различия в производительности между методами. Выводы, сделанные в рамках исследования, позволяют нам представить практические рекомендации для разработчиков, стремящихся интегрировать нейронные методы в приложения компьютерной графики.

**Ключевые слова:** компьютерная графика, рендеринг реального времени, нейронные сети, функции дистанции со знаком.

## A comparative study of the performance of neural models for computer graphics

N. A. Milin<sup>1,2</sup>, R. O. Rodionov<sup>1,2</sup>, V. A. Galaktionov<sup>2</sup>, V. A. Frolov<sup>1,2</sup>

<sup>1</sup>IAI Moscow State University, Moscow, Russia

<sup>2</sup>Keldysh Institute of Applied Mathematics, Moscow, Russia

**Abstract.** This paper presents a performance study of different approaches to the implementation of neural models in computer graphics applications related to the representation of surfaces and materials of three-dimensional objects. Four different methods are considered, among them both universal, supported on most devices, and those that use specialized hardware capabilities. Experimental studies are performed on two tasks: representing a surface as a neural distance function and representing a bidirectional reflectance distribution function as a neural model. The results show significant differences in performance between the methods. The findings of the study allow us to provide practical recommendations for developers seeking to integrate neural methods into computer graphics applications.

**Keywords:** computer graphics, real-time rendering, neural networks, signed distance functions.

### Введение

Современная компьютерная графика переживает значительные изменения, связанные с внедрением технологий искусственного интеллекта (ИИ) в существующие алгоритмы рендеринга. Нейронные подходы к представлению геометрии и материалов открывают новые возможности для создания высококачественных визуальных эффектов, такие как моделирование многослойных материалов [9] и геометрии [11]. Кроме того, нейронные модели используются в дифференцируемом рендеринге [12, 13] и могут являться составной частью инструментов генеративного ИИ [15].

Однако интеграция нейронных методов в графические приложения реального времени сталкивается с существенными вызовами производительности [9]. Выполнение нейронных сетей на GPU требует тщательной оптимизации алгоритмов и эффективного использования архитектурных особенностей современных графических процессоров. Мотивацией данной работы является нахождение оптимальных методов выполнения нейронных моделей на графических процессорах.

### Обзор существующих технологий

В современных задачах компьютерной графики применяются разнообразные архитектуры нейронных сетей, каждая из которых адаптирована под специфические требования графических вычислений. Многослойные перцептроны (MLP) широко используются в задачах трассировки лучей и расчета освещения благодаря своей способности аппроксимировать сложные функции излучения и отражения [1]. Сверточные нейронные сети (CNN) находят применение в задачах шумоподавления и повышения разрешения изображений, где пространственная корреляция данных играет ключевую

роль [2]. Рекуррентные архитектуры, включая LSTM и GRU, применяются для временной стабилизации в алгоритмах глобального освещения [3]. Архитектуры на основе трансформеров адаптируются для задач синтеза изображений, где важна способность к обработке последовательностей различной длины [4, 7].

Значительная часть исследовательских работ в области компьютерной графики опирается на библиотеки глубокого обучения общего назначения, такие как PyTorch, TensorFlow. Это обусловлено простотой их использования и обширным набором инструментов разработки. Работы [3, 6, 7, 11] используют PyTorch, однако его применение в контексте графики реального времени сопряжено с существенными накладными расходами, связанными с интерпретацией Python-кода и невозможностью низкоуровневой ручной оптимизации. Решения, основанные на технологии CUDA, используемые в работах [1, 5, 8], отличаются высокой степенью оптимизации за счет адаптации к аппаратным возможностям устройств NVIDIA, однако существенным недостатком является привязка к конкретной экосистеме, что ограничивает портируемость решения на другие платформы и создает зависимость от проприетарных технологий.

Отдельно стоит отметить, что традиционные библиотеки глубокого обучения оптимизированы для согласованного выполнения нейронной сети, где входные данные собраны в один набор (батч) и высокая пропускная способность достигается за счет параллельной обработки большого количества образцов. Однако в графическом конвейере может потребоваться выполнить нейронную модель только для части пикселей или выполнить несколько различных моделей с разными данными на входе. Как отмечено в работе [9], инструменты для интеграции нейронных сетей с потенциально расходящимся исполнением в языки программирования шейдеров, такие как GLSL или HLSL, практически отсутствуют. Помимо этого, для повышения производительности предпочтительно использовать аппаратное ускорение матричного умножения, доступное в устройствах NVIDIA, AMD и Intel. Далее обсудим как решать эти проблемы и сравним несколько перспективных методов выполнения нейронных сетей в задачах компьютерной графики.

### **Описание исследуемых методов**

#### *Метод 1: Компилируемые веса*

Очевидный способ – сохранить все веса нейронной сети в самом шейдере. Так они будут доступны компилятору, что позволит поместить их в кэш и провести дополнительные оптимизации, такие как разворачивание циклов и удаление избыточных инструкций. Данный метод требует создания отдельного шейдера для каждой конкретной нейронной сети, что усложняет его использование, если требуется выполнить произвольное число моделей в течение рендеринга одной сцены.

#### *Метод 2: Веса в буфере*

При данном подходе параметры нейронной сети загружаются динамически в глобальную память GPU и доступ к ним осуществляется через индексирование внутри шейдера. Это позволяет использовать код одного шейдера для моделей различных размеров, передавая данные об архитектуре сети как константы. Существенным недостатком является то, что каждый отдельный поток многократно загружает параметры модели из видеопамати. Улучшением могло бы служить использование быстрой разделяемой памяти путем предварительной загрузки небольшого числа параметров в нее и выполнение блочного матричного умножения, но данная возможность доступна только в вычислительных (compute) шейдерах.

#### *Метод 3: cuDNN*

Чтобы ускорить стандартные для нейронных сетей операции, можно воспользоваться библиотекой cuDNN [10], которая помимо эффективного выполнения матричного умножения позволяет совмещать несколько последовательных операций в один граф и тем самым оптимизировать обращения к памяти. Для этого потребуется создать входные данные в отдельном ядре, сохранить их в глобальную память GPU, затем вызвать функцию исполнения построенного cuDNN графа операций и обработать полученный выход в следующем ядре. Таким образом можно сократить время прямого прохода сети, но потребуется использовать дополнительную память и вызывать несколько ядер, что увеличивает общие накладные расходы. Недостатком является то, что данный метод доступен только на устройствах NVIDIA и плохо интегрируется в графический конвейер.

#### Метод 4: Аппаратное ускорение с использованием Cooperative Vectors

С появлением этого расширения для современных графических API<sup>123</sup> стала возможной реализация прямого прохода нейросети, совместимая с существующей моделью программирования графических шейдеров. Его цель – предоставить высокоуровневый инструмент для описания матричного умножения в коде шейдера, в котором компилятор отвечает за объединение входных векторов с разных потоков в матрицу для эффективного перемножения с использованием тензорных ядер. Расширение не требует, чтобы данные в разных потоках были одинаковые, что позволяет обрабатывать несколько нейронных моделей в одном шейдере. Повышение производительности ожидается за счет неявной загрузки параметров модели по необходимости и выполнения умножения матриц и активаций слоев с аппаратным ускорением. Данный метод требует поддержки расширения в графическом драйвере.

Эти методы выбраны для исследования по следующим причинам. Методы 1, 2 могут быть реализованы при помощи стандартных операций графических API, поэтому применимы на большинстве платформ, включая мобильные. При этом первый теоретически способен показать лучшую производительность, если компилятору удастся расположить параметры модели в более быстрой памяти, а второй не зависит от самих параметров, предоставляет возможность их динамической загрузки. Третий использует алгоритмы умножения матриц с наилучшей оптимизацией для устройств NVIDIA, четвертый легко встраивается в существующий конвейер рендеринга простой заменой аналитических вычислений на оценку нейронной модели и предоставляет возможность выполнения нейронных сетей во всех типах шейдеров, в том числе в конвейере трассировки лучей. В методах 3, 4 также поддерживается объединение нескольких последовательных операций, таких как умножение матрицы входных данных на матрицу весов, прибавление матрицы сдвига, применение функции активации, что позволяет избежать промежуточных преобразований данных для их непрерывной обработки на тензорных ядрах.

#### Описание экспериментов

В ходе экспериментов была проверена работа каждого из методов для разных размеров нейронных моделей. Методы 1, 2 и 4 были реализованы с использованием Vulkan API, в методе 4 дополнительно включено расширение VK\_NV\_cooperative\_vector. Метод 3 был реализован на платформе CUDA по причине хорошей совместимости с cuDNN. В качестве языка для написания графических шейдеров использовался язык Slang с открытым исходным кодом [16]. Для измерения производительности в методах 1, 2 и 4 применялись временные метки Vulkan, в методе 3 – системный таймер Windows 10. Все эксперименты проводились на видеокарте NVIDIA RTX 3080 10GB с версией драйвера 572.42 в разрешении 1920x1080, параметры моделей были представлены в числах с плавающей точкой половинной точности (float16). Для оценки и сравнения методов выполнения нейронных моделей были выбраны следующие задачи:

##### 1. Рендеринг нейронных SDF.

Функция расстояния со знаком в трёхмерном метрическом пространстве – отображение  $R^3 \rightarrow R$ , показывающее расстояние от данной точки до границы некоторого подмножества пространства. Знак функции определяет, находится ли точка внутри множества. Такие функции являются универсальным способом представления геометрии, способным приближать поверхности, заданные в ином виде (мэш из треугольников, воксели и другие). Традиционно SDF задаются аналитически при помощи технологии конструктивной блочной геометрии, где поверхность строится из примитивов и операций над ними. Хотя данным способом возможно задать довольно сложные поверхности, его трудно использовать для моделирования произвольных объектов. В то же время нейронные модели способны аппроксимировать сложный сигнал, такой как функция расстояния, и могут являться непрерывным и дифференцируемым представлением [11], что позволяет использовать их для задач обратного рендеринга [12, 13].

<sup>1</sup> <https://developer.nvidia.com/blog/neural-rendering-in-nvidia-optix-using-cooperative-vectors>

<sup>2</sup> [https://registry.khronos.org/vulkan/specs/latest/man/html/VK\\_NV\\_cooperative\\_vector.html](https://registry.khronos.org/vulkan/specs/latest/man/html/VK_NV_cooperative_vector.html)

<sup>3</sup> <https://devblogs.microsoft.com/directx/cooperative-vector>



В сравнении для рендеринга SDF использовался стандартный итерационный алгоритм Sphere Tracing [17] с максимальным числом шагов, равным 100. Нейронные SDF были обучены методом, предложенным в работе [11] на модели Armadillo [14].

## 2. Моделирование материалов с помощью нейронных BRDF.

Функции двунаправленного отражения света (Bidirectional Reflectance Distribution Function, BRDF) являются одними из главных компонентов физически корректного рендеринга. Нейронные BRDF позволяют аппроксимировать сложные и анизотропные материалы, которые трудно описать аналитическими моделями [9], и потенциально обеспечивают более точное воспроизведение реальных материалов на основе измерений [18]. В эксперименте использовались нейронные BRDF с числом параметров от 275 до 9179 и функцией активации RELU. Также проведено сравнение времени их выполнения с классической аналитической моделью Кука-Торренса.

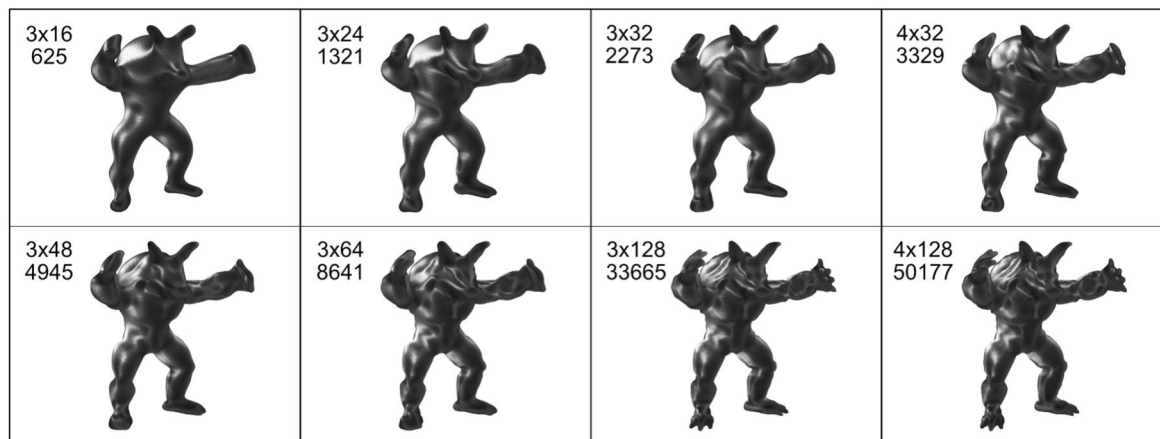


Рисунок 1. Нейронная SDF, построенная по модели Armadillo при разных размерах сети (в обозначении: количество скрытых слоев, размер скрытого слоя, на следующей строке – общее число параметров)

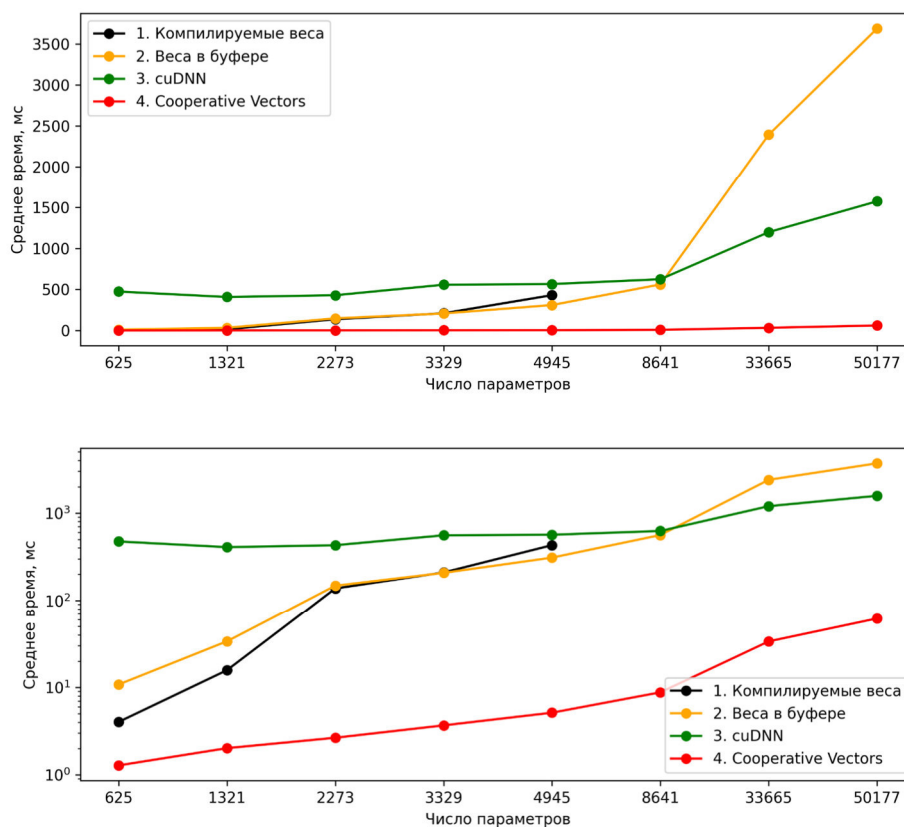


Рисунок 2. Время выполнения нейронной SDF при разном числе параметров модели (в линейной и логарифмической шкалах)



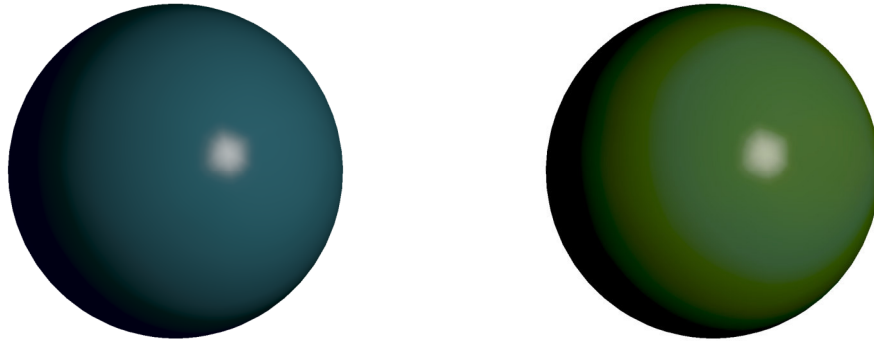


Рисунок 3. Изображения сферы, полученные с использованием нейронных BRDF

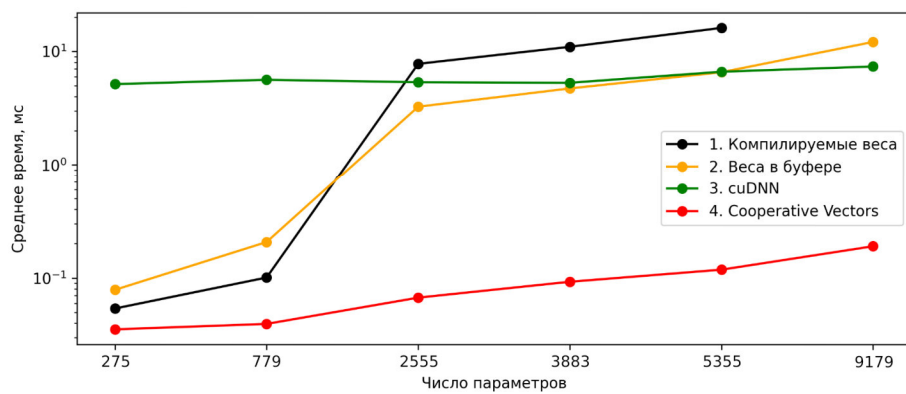


Рисунок 4. Время выполнения нейронной BRDF при разном числе параметров модели (в логарифмической шкале)

Таблица 1. Время выполнения нейронной SDF

Метод реализации	Количество параметров модели							
	3x16 625	3x24 1321	3x32 2273	4x32 3329	3x48 4945	3x64 8641	3x128 33665	4x128 50177
	Среднее время, мс							
1. Компилируемые веса	4.018	15.748	137.843	209.946	430.141	E	E	E
2. Веса в буфере	10.797	33.752	148.149	207.818	309.995	561.047	2393.544	3690.585
3. cuDNN	474.025	407.932	429.275	557.170	565.323	623.787	1197.341	1572.553
4. Cooperative Vectors	1d.277	2.014	2.650	3.673	5.131	8.787	33.703	61.728

Таблица 2. Время выполнения нейронной BRDF

Метод реализации	Количество параметров модели					
	275	779	2555	3883	5355	9179
	Среднее время, мс					
1. Компилируемые веса	0.054	0.101	7.780	10.984	16.111	E
2. Веса в буфере	0.079	0.207	3.263	4.727	6.560	12.119
3. cuDNN	5.157	5.629	5.357	5.296	6.630	7.371
4. Cooperative Vectors	0.035	0.040	0.067	0.093	0.118	0.190
Классическая Cook-Torrance	0.027					

### Анализ результатов

В таблицах 1-2 приведены результаты замеров времени выполнения нейронных SDF и BRDF методами 1-4 при разном числе параметров модели. Эксперименты показали, что различия во времени выполнения для одной и той же модели могут достигать трех порядков. Так, метод с компилируемыми весами демонстрирует относительно хорошую производительность при малом числе параметров (примерно до 2000), однако затем время выполнения резко возрастает, после чего возникают ошибки создания конвейера в драйвере (обозначены как Е в таблицах) и его применение становится невозможным при дальнейшем увеличении размера модели. Метод 2 показывает линейный рост времени выполнения при увеличении числа параметров. Для малых моделей он уступает в производительности, но сохраняет применимость при любом размере модели, хотя и с существенными расходами на доступ к памяти. Метод 3, основанный на библиотеке cuDNN, несмотря на ожидаемую оптимизацию, демонстрирует высокие и практически постоянные накладные расходы, связанные с переключением между ядрами и синхронизацией, что делает его малоэффективным для моделей с небольшим числом параметров, особенно в задаче рендеринга SDF, где приходится многократно запускать разные ядра. При моделировании BRDF, однако, он обходит методы 1-2 уже при меньших размерах нейронной сети. Метод 4 (Cooperative Vectors) показывает лучшую производительность и стабильно работает при всех рассмотренных размерах моделей.

В целом для задач рендеринга SDF в реальном времени подходят только модели небольшого размера, так как в традиционно используемом алгоритме Sphere Tracing требуется многократная оценка функции дистанции. Из рассмотренных методов для этой задачи наиболее применимы метод 4 и методы 1, 2 но только для наиболее малых моделей. Будущие исследования могут включать разработку специальной архитектуры нейронных SDF, для которых нахождение хотя бы приближенного решения  $Sdf(x) = 0$  будет возможно без выполнения модели десятки раз.

Задачи моделирования BRDF менее требовательны к вычислениям, и все методы показывают приемлемое время выполнения, а для небольших моделей метод 4 достигает производительности, сопоставимой с классической реализацией аналитической модели Кука-Торренса, что делает возможным прямую замену аналитических моделей нейронными.

Таким образом, можно сделать следующие обобщения. Методом 1 можно выполнять только небольшие модели (до примерно 2000 параметров на видеокарте NVIDIA RTX 3080), но он позволяет получить высокую производительность и не ограничен конкретной платформой. Метод 2 подходит, если нужно переносимое решение, работающее с любым числом параметров, и задача не требует большого числа оценок модели. Метод 3 не рекомендуется для графических приложений реального времени из-за высоких накладных расходов и сложной интеграции с графическими API, несмотря на его эффективность в традиционных задачах глубокого обучения. Метод 4 предпочтительнее, если целевым оборудованием являются современные графические ускорители, имеющие поддержку Cooperative Vectors.

### Заключение

В рамках исследования был проведен сравнительный анализ различных методов выполнения нейронных моделей в задачах рендеринга геометрии и материалов в контексте графики реального времени. Эксперименты включали реализацию и измерение производительности различных подходов к выполнению нейронных SDF и BRDF. Полученные данные позволяют сделать ряд обоснованных выводов о применимости, эффективности и ограничениях каждого метода. Результаты подтверждают, что эффективное выполнение нейронных моделей в задачах графики реального времени возможно, однако требует тщательного выбора стратегии реализации с учетом специфики задачи и целевой платформы. Предметом будущих исследований может служить разработка новых методов, а также развитие предложенных в данной работе.

### Источник финансирования

Исследование выполнено за счет гранта Российского научного фонда № 25-11-00054, <https://rscf.ru/project/25-11-00054/>

**Список литературы**

1. Müller, T., et al. “Instant Neural Graphics Primitives with a Multiresolution Hash Encoding.” ACM Transactions on Graphics, 2022.
2. Chaitanya, C. R. A., et al. “Interactive Reconstruction of Monte Carlo Image Sequences using a Recurrent Denoising Autoencoder.” ACM Transactions on Graphics, 2017.
3. Munkberg, J., et al. “Neural Denoising with Layer Embeddings.” Computer Graphics Forum, 2020.
4. Kulháněk, J., et al. “ViewFormer: NeRF-free Neural Rendering from Few Images Using Transformers” European Conference on Computer Vision, 2022.
5. Meister, D., et al. “N-BVH: Neural Bounding Volume Hierarchies for Real-Time Ray Tracing.” SIGGRAPH Conference Papers, 2024
6. Mildenhall, B., et al. “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.” European Conference on Computer Vision, 2020.
7. “RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination.” ACM SIGGRAPH 2025 Conference Papers, 2025
8. Müller, T., et al. “Instant neural graphics primitives with a multiresolution hash encoding.” ACM Transactions on Graphics, 2022.
9. Zeltner, T., et al. “Real-Time Neural Appearance Models.” ACM Transactions on Graphics, 2024
10. Chetlur, S., et al. “cuDNN: Efficient Primitives for Deep Learning.” ArXiv abs/1410.0759, 2014
11. Sitzmann, V., et al. “Implicit Neural Representations with Periodic Activation Functions.” Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020
12. Zhang, D., et al. “RISE-SDF: A Relightable Information-Shared Signed Distance Field for Glossy Object Inverse Rendering.” International Conference on 3D Vision, 2025
13. Zhang, K., et al. “IRON: Inverse Rendering by Optimizing Neural SDFs and Materials from Photometric Images.” IEEE Conference on Computer Vision and Pattern Recognition, 2022
14. The Stanford 3D Scanning Repository. URL: <http://graphics.stanford.edu/data/3Dscanrep>
15. Li, Z., et al. “Sparc3D: Sparse Representation and Construction for High-Resolution 3D Shapes Modeling.” arXiv preprint arXiv:2505.14521, 2025
16. Bangaru, S. P., et al. “SLANG.D: Fast, Modular and Differentiable Shader Programming.” ACM Transactions on Graphics, 2023
17. Hart, J. “Sphere Tracing: A Geometric Method for the Antialiased Ray Tracing of Implicit Surfaces.” The Visual Computer, 1995
18. Sztrajman, A., et al. “Neural BRDF Representation and Importance Sampling.” Computer Graphics Forum, 2021

## Метод спектрального рендеринга с использованием рядов Фурье

Р. О. Родионов<sup>1,2</sup>, Е. В. Приходько<sup>1</sup>, В. А. Фролов<sup>1,2</sup>, А. Г. Волобой<sup>2</sup>

<sup>1</sup>Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ имени М.В. Ломоносова, Москва, Россия

<sup>2</sup>Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В данной статье рассматриваются задача спектрального рендеринга и ключевые проблемы, связанные с хранением и обработкой спектральных данных. Предлагается подход, основанный на использовании ограниченного числа коэффициентов ряда Фурье. Метод ориентирован на быстрый предварительный просмотр и интерактивное использование. Также решается проблема представления материалов в Фурье-базисе, в том числе с использованием нейронных сетей для аппроксимации спектральных распределений. Рассмотрена эффективность такого подхода для уменьшения дисперсии в задаче спектрального рендеринга.

**Ключевые слова:** спектральный рендеринг, ряд Фурье, нейронные модели материалов.

## Spectral rendering method using Fourier series

R. O. Rodionov<sup>1,2</sup>, E. V. Prikhodko<sup>1</sup>, V. A. Frolov<sup>1,2</sup>, A. G. Voloboy<sup>2</sup>

<sup>1</sup>IAI Moscow State University, Moscow, Russia

<sup>2</sup>Keldysh Institute of Applied Mathematics, Moscow, Russia

**Abstract.** This paper addresses the problem of spectral rendering and key issues related to storing and processing spectral data. An approach based on using a limited number of Fourier series coefficients is proposed. The method is focused on fast preview and interactive use. The problem of representing materials in the Fourier basis is also solved, including using neural networks to approximate spectral distributions. The effectiveness of this approach for reducing variance in spectral rendering is considered.

**Keywords:** spectral rendering, Fourier series, neural material models.

### Введение

Одним из наиболее перспективных направлений физически корректного рендеринга является спектральный рендеринг: вместо привычного RGB-цвета в сценах применяется вещественная функция длины волны, описывающая спектральную плотность излучения света (или отражательной способности материалов). Спектр обычно представлен в рендер-системах и системах оптического моделирования в виде набора пар длина волны – значение спектра. Этот способ используется, например, в рендер-системе PBRT [1]. Данным способом могут быть представлены значения спектра для материалов, источников освещения, текстур. Кроме того, отдельные лучи или фотоны, переносящие спектр, как правило, сохраняют не весь спектр, а лишь некоторую выборку из него. При этом типичное число длин волн в таких выборках, как правило, является небольшим: 4, 8, 16, 32 [2]. Такое представление неоптимально как по памяти, так и по производительности.

К сожалению, даже простая выборка из такого спектра на практике приводит к необходимости реализации бинарного поиска в массиве, из-за чего некоторые рендер-системы [1, 3] для спектров материалов и источников заранее отображают спектр в массив с некоторым константным шагом, например в 1 нм.

Обычно это приводит к заметному (20–30%) увеличению скорости работы рендер-систем. Однако и метод с бинарным поиском, и тем более метод спектра с константным шагом на практике мало применимы для задания спектральных текстур, т.к. исходные 2D-текстуры превращаются в 3D-текстуры и начинают потреблять на 1–2 порядка больше памяти, чем обычные текстуры. На практике нам бы хотелось иметь хранилище значений 3D-функции, которое было бы относительно компактно и достаточно эффективно для обращения.

Поскольку большинство функций спектра, используемых на практике, являются непрерывными, их можно представить с помощью преобразования Фурье. Получаемые коэффициенты могут быть далее восстановлены в первоначальную функцию с помощью, например, рядов Фурье.

Современные исследования способов хранения спектра уже обращались к преобразованию Фурье как к компактному и эффективному представлению спектрального распределения световых источников или отражательных способностей материалов. На практике для хранения достаточно гладких распределений без сильной потери точности достаточно небольшого числа коэффициентов. Однако ни одна из рассмотренных работ не предлагает использование представления Фурье непосредственно в процессе рендеринга, что могло бы потенциально снизить затраты памяти (в т.ч. оперативной) при рендеринге и хранении сцен, снизить эффект цветового шума за счет использования полной информации о спектре вместо подвыборки и ускорить время получения изображения.

Целью данной статьи является исследование возможностей применения спектра в пространстве коэффициентов Фурье в системах рендеринга на основе трассировки путей.

### Обзор существующих методов

Наиболее распространённым способом представления спектра в задачах спектрального рендеринга является дискретное хранение его значений на фиксированном наборе длин волн. Как правило, спектр представляется таблицей значений, равномерно распределённых по спектральному диапазону. И хотя в некоторых физических моделях материалов, таких как идеальные диэлектрики или интерференционные покрытия, отражённый или преломлённый луч может переносить вклад лишь одной длины волны [4], в большинстве случаев переносится сразу весь световой спектр. В таких случаях дискретное представление становится избыточным и неэффективным, что мотивирует поиск более компактных способов хранения спектра.

Работа [5] предлагает использовать небольшой вектор коэффициентов полинома для представления спектра отражения. Сумма таких векторов будет соответствовать сумме их спектров. Однако данный метод предназначен для получения синтетического спектра на основе известного цвета и возможности его использования для представления известного спектра не исследовались.

Также функцию спектрального распределения можно представить с помощью коэффициентов Фурье. Существует несколько работ, предлагающих данный подход. Авторы статьи [6] для восстановления функции предлагают использовать метод MESE (спектральная оценка максимальной энтропии), а также его модификацию – Bounded MESE для представления спектра отражения. Показано, что, применяя MESE, непрерывную функцию спектра можно с достаточной точностью представить всего шестью коэффициентами Фурье. Метод позволяет избежать артефактов и искажений, возникающих при прямой реконструкции усечённого ряда Фурье (рис. 1).

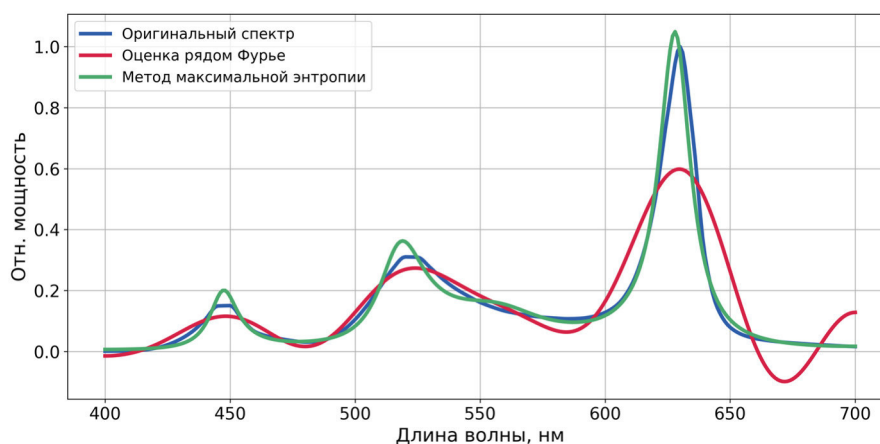


Рисунок 1. Реконструкция спектра излучения через усечённый ряд Фурье и с помощью метода максимальной энтропии

Связанный с таким подходом метод хранения спектральных текстур был представлен в статье [7]. Здесь показано, что спектральные текстуры, спектры которых конвертированы в коэффициенты Фурье, достаточно хорошо поддаются сжатию алгоритмом JPEG. Авторы данной работы не обращаются к методу MESE [6], для того чтобы сохранить точные значения спектра в точках выборки, и используют ряды Фурье для реконструкции спектра.

Работа [8] предлагает гибридную систему рендеринга, представляющую низкочастотные компоненты спектра в виде усеченного ряда Фурье, а высокочастотные с помощью традиционного способа представления спектра. При рендеринге спектр, представленный в виде коэффициентов Фурье, конвертируется в выборку значений спектра. Поскольку операция вычисления усеченного ряда Фурье выполняется за линейное время, такой подход дает возможность сохранить линейную сложность операций со спектром, позволяя передавать сразу весь спектр.

### Предложенный метод

#### 1. Косинусный ряд Фурье для представления спектра

Рассмотрим функцию сигнала  $f(\phi)$  с периодом  $2\pi$ ,  $f(\phi) \geq 0$ . Далее представим эту функцию в виде  $m + 1$  коэффициентов Фурье, используя набор базисных функций:

$$c_j(\phi) = \frac{1}{2\pi} (\exp(-ij\phi)) \in \mathbb{C}, j = 0 \dots m.$$

Полученные коэффициенты будут иметь следующий вид:

$$f_j = \int_{-\pi}^{\pi} f(\phi) c_j(\phi) d\phi \in \mathbb{C}, j = 0 \dots m.$$

Далее рассмотрим функцию спектрального распределения  $g(\lambda)$  на видимом интервале  $\Lambda = [\lambda_0, \lambda_1]$ . Представим ее в виде четного сигнала  $f(\phi)$ : с помощью замены переменной

$$\phi(\lambda) = \pi \frac{\lambda - \lambda_0}{\lambda_1 - \lambda_0} - \pi$$

отобразим ее на отрезок  $[-\pi, 0]$  и отзеркалим относительно нуля.

Поскольку функция  $f(\phi)$  чётная, получим следующее представление сигнала в виде набора коэффициентов:

$$f_j = \frac{1}{\pi} \int_{-\pi}^0 f(\phi) \cos(j\phi) d\phi \in \mathbb{R}, j = 0 \dots m.$$

Таким образом, получим возможность извлекать приближение исходного сигнала в виде тригонометрического ряда Фурье, представляя высокочастотные компоненты нулевыми значениями:

$$f(\phi) \cong f_0 + 2 \sum_{j=1}^m f_j \cos(j\phi).$$

#### 2. Фурье базис для спектрального рендеринга

Одна из актуальных проблем спектрального рендеринга – цветной шум, возникающий из-за того, что для оценки спектра используется разреженная выборка на видимом диапазоне длин волн. Особенно сильно этот эффект виден при наличии в сцене источников или материалов с резкими пиками в спектральном распределении, прозрачных объектов с показателем преломления, зависимым от длины волн проходящего света. Использование большой выборки может помочь решить первую проблему, но приводит к повышенным вычислительным затратам.

Пусть рассчитываемый световой луч сталкивается с поверхностью материала или источника света, описываемой некоторой двулучевой функцией отражательной способности (ДФОС, англ. Bidirectional reflectance distribution function – BRDF). Энергия входного луча и распределение отражательной способности для выбранного направления отражённого луча описываются своими коэффициентами Фурье:

$$f(\phi) = \sum_{i=0}^m a_i \cos(i\phi), g(\phi) = \sum_{j=0}^m b_j \cos(j\phi).$$

Тогда энергия отражённого луча будет иметь распределение

$$f(\phi)g(\phi) = \sum_{i=0}^m \sum_{j=0}^m a_i b_j \cos(i\phi) \cos(j\phi).$$

Далее раскрываем произведение косинусов

$$\cos(i\phi) \cos(j\phi) = \frac{1}{2} [\cos((i+j)\phi) + \cos((i-j)\phi)],$$

где  $\cos(-k\phi) = \cos(k\phi)$ .

Поскольку наше представление хранит фиксированное число коэффициентов Фурье, приходится отсекал высокочастотные компоненты, что для множества пересечений может привести к накоплению ошибки и сильным искажениям спектра.

Полученное распределение в итоге описывается простой сверткой исходных распределений, представленных в виде усеченного ряда Фурье:

$$f(\phi)g(\phi) = \sum_{i=0}^m c_i \cos(i\phi),$$

$$c_i = \frac{1}{2} \sum_{p+q=i} a_p b_q + \frac{1}{2} \sum_{|p-q|=i} a_p b_q.$$

Итоговая формула позволяет посчитать коэффициенты Фурье для произведения двух сигналов.

Полученные в результате рендеринга коэффициенты Фурье-спектра необходимо преобразовать в цвет. Это можно сделать несколькими способами. Наиболее простым способом конвертации является вычисление спектра с помощью ряда Фурье. Полученный спектр далее переводится в координаты цвета цветовой модели XYZ:

$$X = \int_{\Lambda} s(\lambda) \bar{x}(\lambda) d\lambda, Y = \int_{\Lambda} s(\lambda) \bar{y}(\lambda) d\lambda, Z = \int_{\Lambda} s(\lambda) \bar{z}(\lambda) d\lambda.$$

Заметим, что поскольку перевод спектра в цвет в программах как правило [3] производится интегрированием на заранее определенных длинах волн, возможно предварительно подсчитать значения  $\cos(i\phi)$  для ряда Фурье и использовать эти значения при конвертации.

Другим способом преобразования является свертка с предварительно подсчитанными коэффициентами Фурье для функций  $\bar{x}$ ,  $\bar{y}$ ,  $\bar{z}$ . Полученные коэффициенты будут соответствовать произведению спектра с функциями чувствительности. Далее, поскольку для коэффициента Фурье  $a_0$  функции  $f(\phi)$ , соответствующей спектру  $g(\lambda)$  верно

$$a_0 = \int_{-\pi}^0 f(\phi) d\phi = \frac{1}{\lambda_1 - \lambda_0} \int_{\lambda_0}^{\lambda_1} g(\lambda) d\lambda,$$

то значения  $X$ ,  $Y$ ,  $Z$  будут равны соответствующим нулевым коэффициентам Фурье, умноженным на  $(\lambda_1 - \lambda_0)$ .

Также вместо рядов Фурье для преобразования коэффициентов в спектр можно использовать метод MESE.

Отметим, что при обычном рендеринге на выборках длин волн преобразование спектра в цвет производится на каждой итерации (что позволяет избежать необходимости хранения большого фреймбуфера). В случае коэффициентов Фурье данная проблема выражена гораздо меньше, так как число коэффициентов Фурье может быть значительно меньше, чем количество длин волн, что позволяет хранить фреймбуфер коэффициентов и производить вычисление цвета уже после трассировки путей.

### 3. Фурье-базис для уменьшения дисперсии

Помимо прямого расчёта спектрального освещения с помощью рядов Фурье, можно рассмотреть использование данного метода для уменьшения дисперсии в классическом методе спектрального рендеринга через выборку длин волн. Поскольку метод, использующий Фурье, даёт смещённую оценку распределения на всём диапазоне спектра, коррелирующую с истинным спектром, естественным шагом является использование техники контрольных переменных (control variates) [9].

Допустим, нам требуется оценить матожидание функции расчёта цвета пикселя на изображении  $\ell = \mathbb{E}[H(X)]$ , где  $X$  – выборка длин волн на видимом спектре,  $Y = H(X)$  – оценка цвета пикселя по данной выборке в представлении CIE XYZ.

Оригинальный метод Монте Карло оценивает цвет следующим образом:

$$\widehat{Y}_n = \frac{H(X_1) + \dots + H(X_n)}{n}.$$

Дисперсия такой величины оценивается как  $\text{Var}[\widehat{Y}_n] = \frac{1}{n} \text{Var}[H(X)]$ .

Предположим, что мы хотим уменьшить дисперсию для одного луча в трассировке пути, используя малую выборку длин волн. Обозначим  $C = F(X)$  – оценка цвета пикселя по выборке  $X$  длин волн из Фурье-спектра,  $\mu = \mathbb{E}[C]$ . Оценка методом контрольной переменной:

$$\widehat{\ell} = \frac{1}{n} \sum_{i=1}^n (Y_i - \alpha(C_i - \mu)) = \widehat{Y}_n - \alpha(\widehat{C}_n - \mu).$$

Имея оптимально подобранный параметр  $\alpha^* = \frac{\text{Cov}(Y, C)}{\text{Var}[C]}$ , можно было бы ожидать снижения дисперсии:

$$\text{Var}[\widehat{\ell}] = (1 - \rho^2) \text{Var}[\widehat{Y}_n] < \text{Var}[\widehat{Y}_n],$$

где  $\rho$  – корреляция между  $Y$  и  $C$  [10]. Однако оптимальное значение коэффициента  $\alpha^*$  невозможно найти заранее, поэтому зачастую прибегают к аппроксимациям, оценивающим параметр некоторой небольшой стартовой выборкой или на каждом отдельном батче, что, в теории, создаёт небольшое смещение относительно оптимального решения.

#### 4. Вычисление и хранение спектральных материалов

Главная особенность спектрального рендеринга – возможность визуализации явлений, являющихся прямым следствием спектральной природы света. Только такой подход позволяет корректно показать взаимодействие световых источников и материалов со сложными спектральными распределениями, явления интерференции, дифракции, дисперсии и поляризации света. Явным ограничением рассматриваемого метода является отсутствие зависимости направления распространения лучей от длины волны, поскольку этот метод переносит энергию для всего видимого диапазона. Поддержка такого типа явлений требует особого подхода, зачастую приводящего к сильному зашумлению итогового изображения. Данный метод, напротив, способен дать относительно быстрое приближение для визуализации спектральной сцены с минимумом затрат по памяти и практически полным отсутствием цветного шума ценой игнорирования вышеупомянутого явления.

Возможность распространения всего спектра целиком ведёт и к другому ограничению – сложность расчёта произвольных BRDF. Отдельные примеры не вызывают сложностей. Так, диффузные материалы, описываемые моделью Ламберта, сводятся к вычислению одного скалярного умножения с домножением получившегося значения на спектр альбеда. Требуется лишь хранить отдельный спектр или спектральную текстуру. Для универсальных моделей, таких как модель Кука–Торренса, а также для диэлектриков, проводников и множества других материалов, вызывает сложность вычисление значения функции Френеля. Одним из вариантов решения проблемы является использование аппроксимаций, таких как модель Шлика [11] или более точная модель на основе декомпозиции спектра отражения [12].

Другой вариант – использование предрассчитанных таблиц отражательной и пропускательной способности. Таблицы имеют разрешение  $N \cdot L$ , где  $N$  – размер выборки длин волн;  $L$  – число выбранных углов. Для избежания алиасинга и цветовых искажений нет необходимости хранить большое число значений, однако это число можно сократить до  $M \cdot L$ , предполагая спектр достаточно гладким и хорошо представимым дискретным преобразованием Фурье, что обычно верно для спектров отражения и пропускания. Описанный подход особенно полезен для сложных материалов, таких как однослойные и многослойные тонкие плёнки.

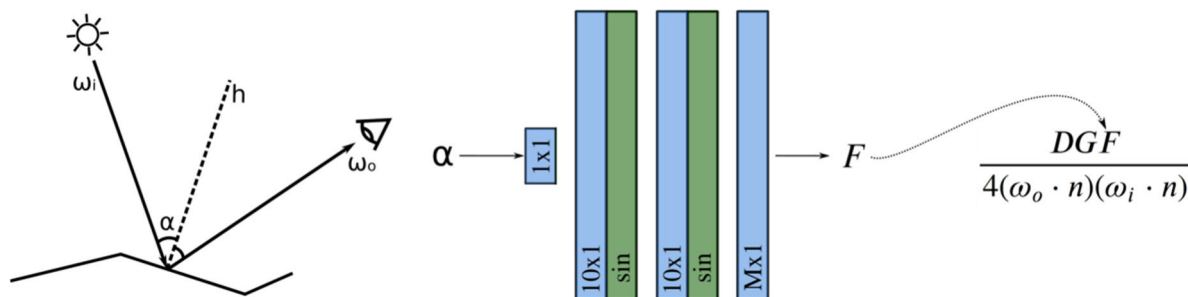


Рисунок 2. Вычисление компоненты Френеля для модели Кука–Торренса с помощью нейронной сети.

Пайплайн поддерживает задание BRDF или его компонент в виде нейронной сети. Пример вычисления компоненты Френеля в модели Кука–Торренса показан на рисунке 2. Аналогично работам [13, 14], нейронные функции BRDF представляются в виде многослойного перцептрона с малым числом внутренних слоёв и нейронов. Число выходных нейронов совпадает с числом используемых коэффициентов Фурье. Для обучения используется эталонная функция BRDF и случайные входные параметры. На рисунке 3 изображены оригинальная, реконструированная из Фурье-базиса (10 коэффициентов на каждый угол наблюдения) и реконструированная из предсказанного нейронной сетью Фурье-базиса таблица отражательной способности для диэлектрической тонкой плёнки. Для гладких распределений, таких как показанное в примере, не требуется большого числа параметров. Используемая на рисунке 2 нейронная сеть использует два внутренних слоя по 10 нейронов, периодическую функцию активации (синусоида) и 10 коэффициентов Фурье. Обучение занимает около 2 минут на RTX 4070.

Применение компактных нейросетей позволяет избежать хранения таблиц с высоким разрешением и сократить потребление памяти для реконструкции гладких сигналов. В отличие от похожей работы по нейронным спектральным BRDF [14], одному проходу нейросети соответствует весь спектр, а не значение на одной длине волны, что помогает сократить время рендеринга. Использование



нейронных сетей даёт и другую возможность – задание параметрических моделей (BRDF с зависимостью от параметров материала в конкретной точке). Входные параметры задаются в текстурах (заданных художником или полученных в ходе обучения нейросети, как в работе [13]), и подаются в нейронную сеть во время расчёта материала.

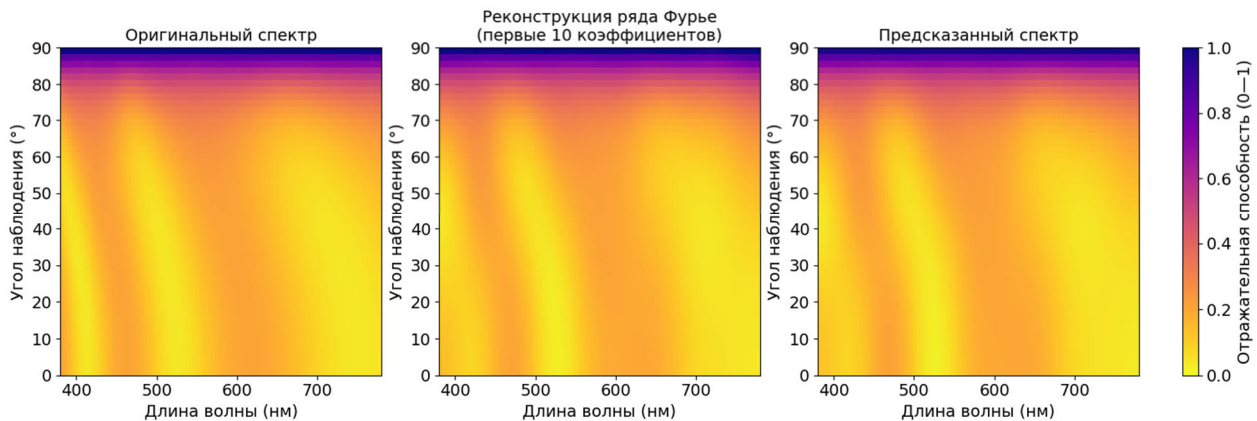


Рисунок 3. Таблица отражательной способности:

- 1) хранящаяся в явном виде; 2) реконструированная из предрассчитанного ряда Фурье;
- 3) реконструированная из предсказанного нейронной сетью ряда Фурье

На рисунке 4 показана сцена, содержащая описанные выше материалы. Для визуализации использовался метод на основе коэффициентов Фурье. Все материалы, кроме диффузных, используют предрассчитанные таблицы отражательной и пропускательной способности (для прозрачных материалов) либо представлены в весах нейронных сетей.

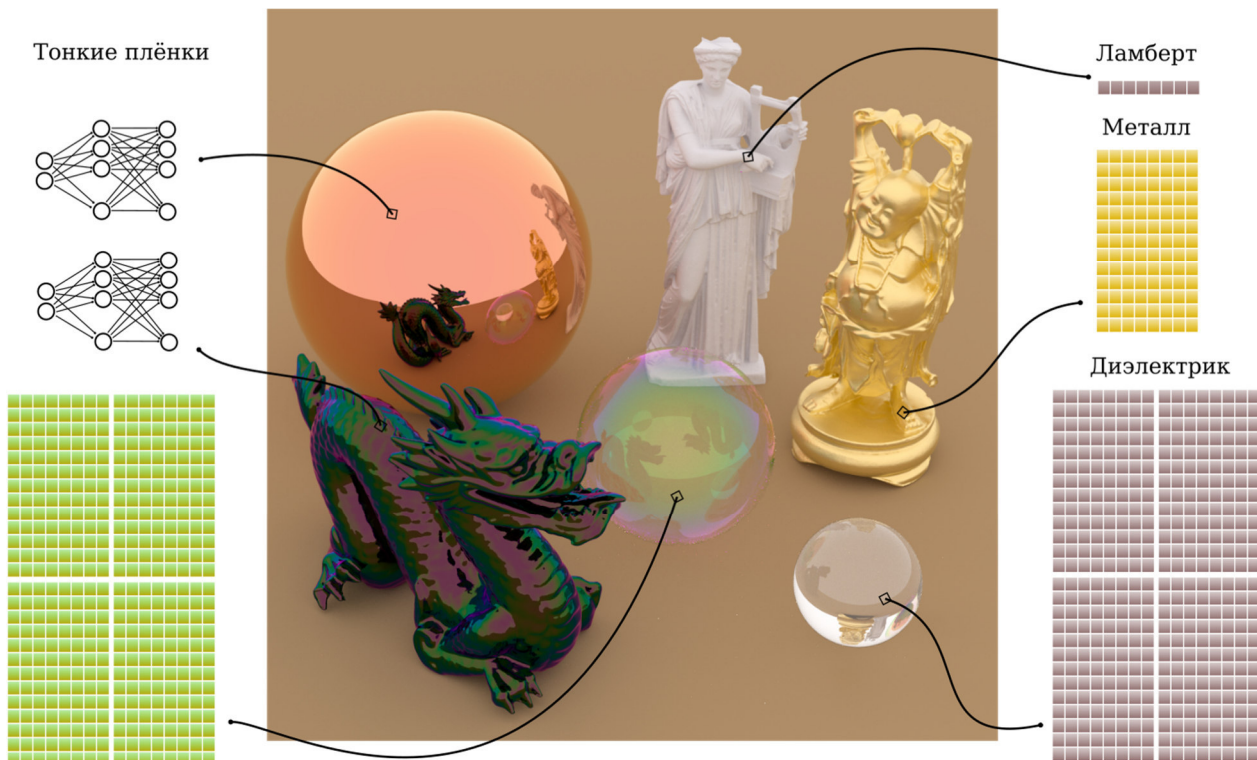


Рисунок 4. Поддерживаемые материалы и их представления

## Эксперименты

### 1. Реализация предложенного метода в системе спектрального рендеринга

На текущем этапе работы метод спектрального рендеринга с использованием метода максимальной энтропии и предложенный метод, использующий усечённые ряды Фурье, реализованы в рендер системе Hydra Core 3 [3].

Для работы с рядами Фурье были реализованы способы конвертации итоговых коэффициентов в цвет: с помощью рядов Фурье, с помощью таблицы поиска и с помощью свертки с функциями цветовой чувствительности. Рассматривались два случая: преобразование в цвет в конце рендеринга (коэффициенты Фурье хранились в отдельном буфере) и на каждой итерации.

Были проведены сравнения данных методов при рендеринге сцен (рис. 5-7). В качестве метрики качества использовалась метрика CIE  $\Delta E$ . Рендеринг производился с разрешением 512x512 и размером выборки для пикселя (числом итераций), равным 64. Результаты сравнения можно увидеть в таблице:

**Результаты сравнения различных способов рендеринга**

	Ряды Фурье	Таблица поиска	Нулевой коэфф.	MESE	Spec32 <sup>1</sup>
CIE $\Delta E$	11.5498	11.5498	11.5577	11.5578	11.6477
PSNR	17.30	17.30	17.29	17.29	17.39
Время рендеринга (спектральный буфер), с	7.8	7.5	7.4	9.8	—
Время рендеринга (RGB буфер), с	31.1	11.1	7.5	162.3	8.3

CPU: AMD Ryzen 7 9700x @ 3.80 GHz, 32 GB RAM. GPU не использовался.

Анализ результатов показывает, что метод преобразования на основе нулевого коэффициента Фурье оказался наиболее быстрым в обоих случаях. При этом он имеет малую ошибку по сравнению с использованием коэффициентов Фурье. Также из таблицы видно, что использование методов на основе коэффициентов Фурье позволяет достичь меньшей ошибки по сравнению с обычным спектральным рендерингом с размером выборки 32.

На рисунке 7 можно заметить, что средняя цветовая ошибка метода на основе рядов Фурье меньше чем отличие для спектра с размером выборки 32. Также видно, что метрика PSNR для обоих способов рендеринга остается приблизительно одинаковым (что можно объяснить тем, что пространственный шум гораздо сильнее влияет на значение метрики, чем цветовой).

Далее были проведены эксперименты с применением спектра с явно выраженными пиками (лазер). Спектральное распределение источника можно увидеть на рисунке 8. Результаты представлены на рисунках 9, 10.

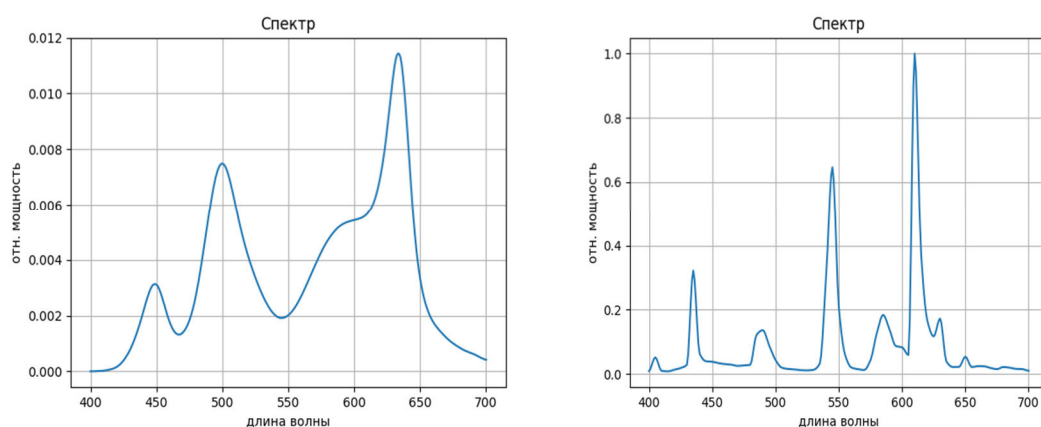


Рис. 5. Графики спектрального распределения ламп для сцены 1

<sup>1</sup> Время рендеринга с отдельным спектральным буфером не замерялось.

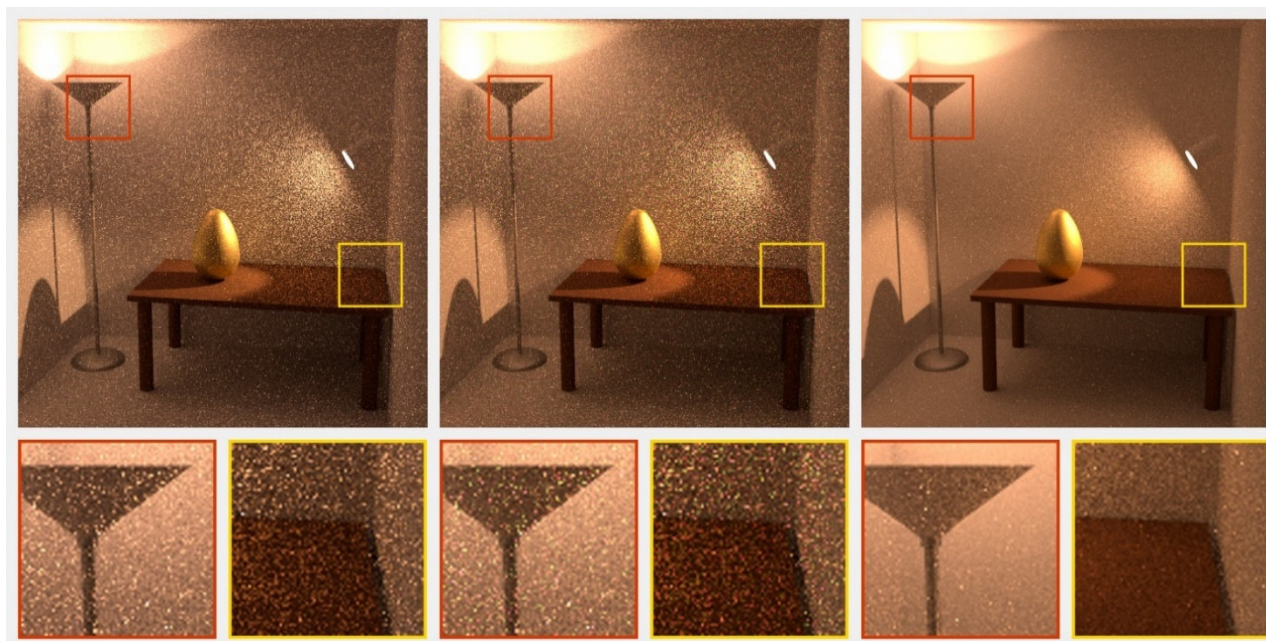


Рисунок 6. Примеры сцены 1 (размер выборки на пиксель – 256): предложенный метод, обычный спектральный рендеринг с размером выборки 32, изображение с большим размером выборки

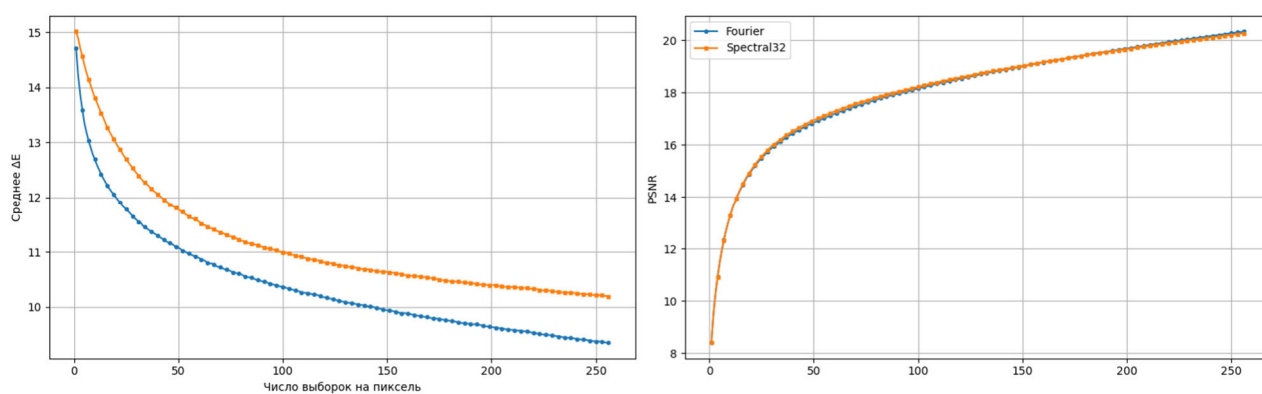


Рисунок 7. Графики средних значений CIE  $\Delta E$  и PSNR для рендеринга в зависимости от размера выборки на пиксель (сцена 1)

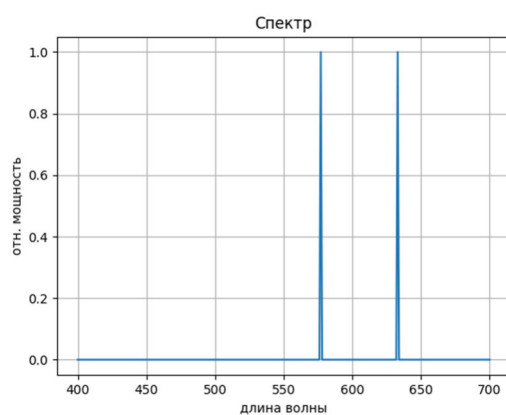


Рисунок 8. График спектрального распределения источника света для сцены 2



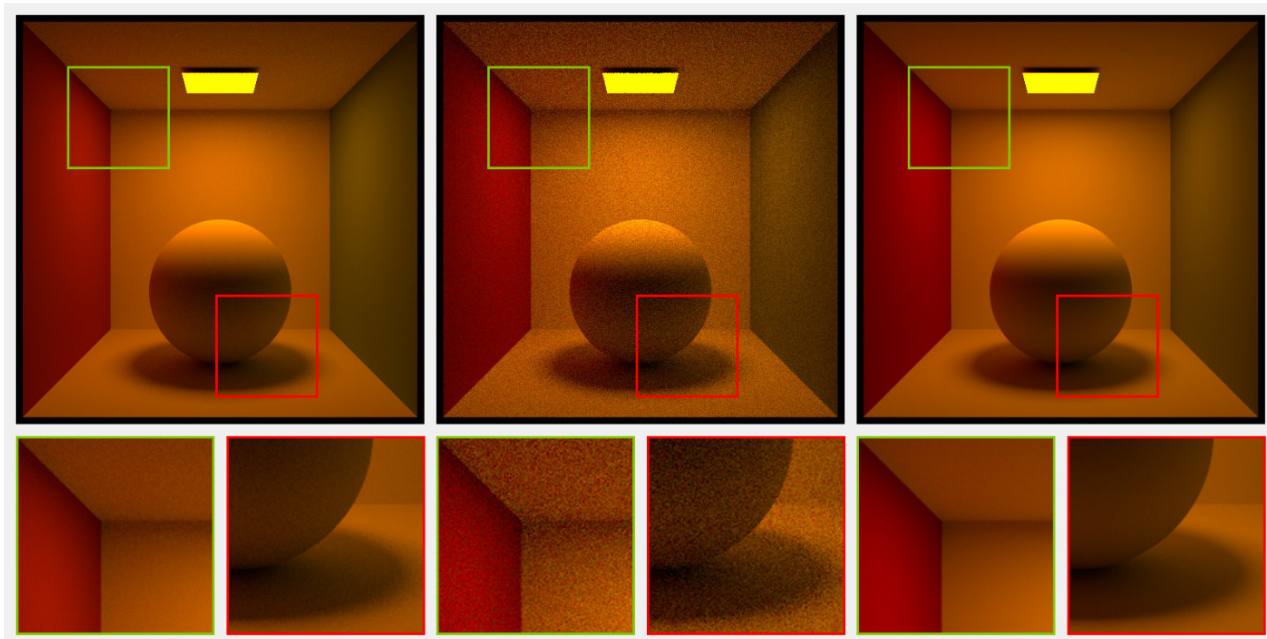


Рисунок 9. Примеры сцены 2 (размер выборки на пиксель – 256): предложенный метод, обычный спектральный рендеринг с размером выборки 32, эталонное изображение

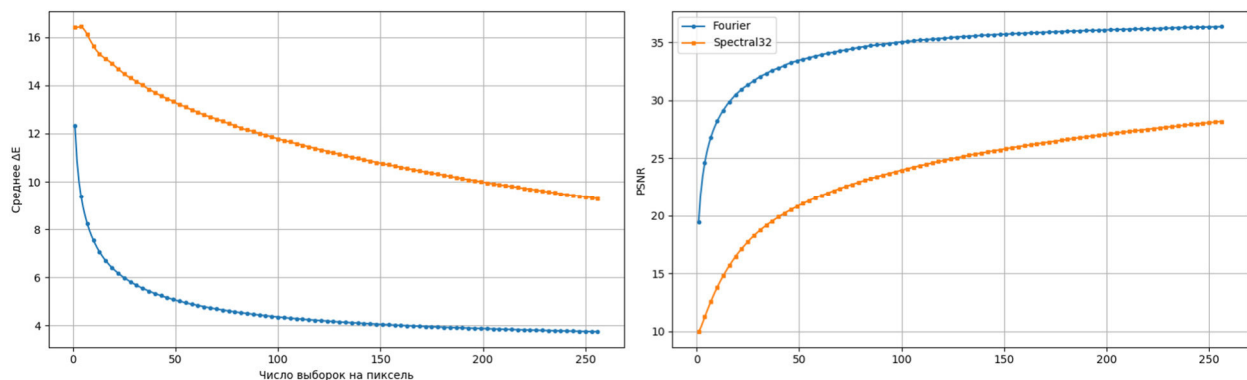


Рисунок 10. Графики средних значений CIE  $\Delta E$  и PSNR для рендеринга в зависимости от размера выборки на пиксель (сцена 2)

По результатам рендеринга сцены 2 можно заметить, что предложенный метод позволил значительно снизить цветовой шум.

## 2. Метод уменьшения дисперсии

Для тестирования метода Фурье при использовании его в качестве контрольной переменной проведён эксперимент: прямое освещение цветовой калибровочной мишени рассчитывается для различных источников, для каждого метода результаты замеряются и сравниваются с эталоном.

Используемая калибровочная мишень (ColorChecker® Classic 2002 GretagMacBeth, рис. 11) имеет 24 цвета. Взаимодействие источника с поверхностью рассчитывается следующими методами:

- 1) равномерная выборка длин волн на спектре;
- 2) стратифицированная выборка (равномерная выборка на фиксированных диапазонах);
- 3) выборка на расстояниях с фиксированным шагом относительно случайно выбранной длины волны (Hero Wavelength Spectral Sampling [15]);
- 4) предложенный метод на основе Фурье-базиса;
- 5) Фурье-базис в качестве контрольной переменной для методов 1–3.

Для оценки отклонения полученного цвета от эталонного выбрана метрика  $\Delta E$ .

На рисунках 12 и 13 показаны замеры для двух источников. Методы, основанные на выборке длин волн, используют 12 коэффициентов для переноса излучения. Объединённый метод с контрольной переменной использует 8 длин волн и 4 коэффициента Фурье.



Рисунок 11. Цветовая калибровочная мишень, освещённая двумя различными источниками: D6500 и фосфорный светодиод

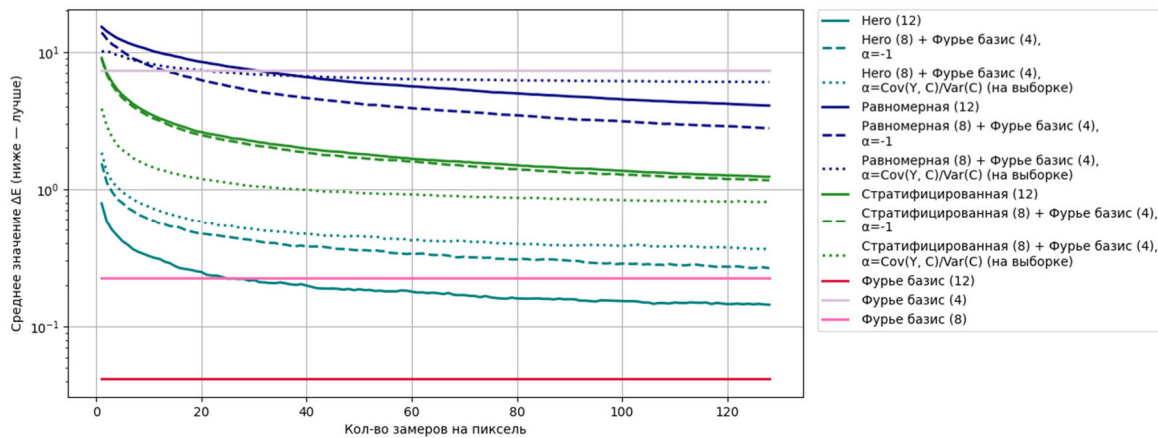


Рисунок 12. Среднее значение  $\Delta E$  для прямого освещения с источником D6500

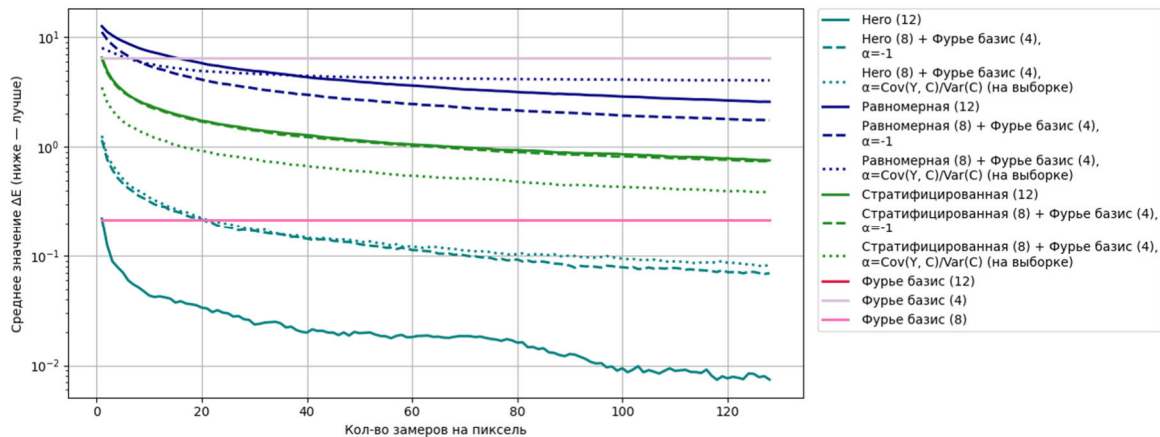


Рисунок 13. Среднее значение  $\Delta E$  для прямого освещения с фосфорным светодиодом

### Обсуждение результатов

Результаты сравнения рассматриваемых вариантов рендеринга с помощью коэффициентов Фурье, представленные выше в таблице, демонстрируют, что предложенный метод позволяет достичь качества рендеринга, сравнимого с традиционным подходом. На сценах с источниками освещения, имеющими спектральное распределение с заметными пиками, метод позволяет получить выигрыш по метрике  $\Delta E$ .

Также из результатов следует, что наиболее выгодным способом преобразования итогового спектра в цвет является использование нулевого коэффициента свертки спектра с кривыми цветовой чувствительности. Важным фактом является то, что метод MESE хоть и зарекомендовал себя в качестве способа хранения спектра [6], все же не смог дать существенного выигрыша в комбинации с нашим методом и существенно увеличил время рендеринга.

Предложенный метод показал значительное улучшение результатов для источников с явно выраженными пиками (как например, лазеров). Это может быть связано с тем, что значения спектра

в пиках (имеющие наибольший вклад в результат рендеринга) могут не попадать в выборку спектра. Такая проблема не возникает при представлении спектра рядами Фурье, так как в данном случае для трассировки путей используется весь спектр, а не его подвыборка.

Поскольку предложенный метод позволяет достичь малого цветового шума даже на небольшом размере выборки на пиксель, из результатов следует возможность использования данного метода для быстрой генерации изображений предварительного просмотра сцен.

Как показали эксперименты, рассмотренный метод уменьшения дисперсии, использующий контрольные переменные, способен дать ощутимый прирост качества по метрике  $\Delta E$ , но только при условии равномерной или стратифицированной выборки длин волн. Использование более продвинутых методов выборки приводит к отсутствию выигрыша от предложенной идеи.

### Заключение

В данной работе был предложен метод спектрального рендеринга, основанный на коэффициентах усеченного ряда Фурье. Представлено исследование различных стратегий преобразования коэффициентов во время рендеринга. Показано достижение меньшего уровня цветового шума на полученном изображении по сравнению с традиционными методами при сопоставимом уровне качества и времени рендеринга.

Рассмотренный метод целесообразно применять для быстрого предпросмотра или интерактивного просмотра сцен, особенно при низком разрешении, когда цветовой шум наиболее заметен и существенно влияет на восприятие. Данный подход потенциально применим и в других приложениях, которые будут подробно рассмотрены в последующих исследованиях.

### Источник финансирования

Исследование выполнено за счет гранта Российского научного фонда № 25-11-00054, <https://rscf.ru/project/25-11-00054/>.

### Список литературы

1. Система реалистичного рендеринга PBRT 4. URL: <https://github.com/mmp/pbrt-v4> (Дата обращения: 12.07.2025).
2. Pharr M., Jakob W., Humphreys G. Physically Based Rendering: from theory to implementation. Fourth ed. Section 4.6.5. URL: [https://www.pbr-book.org/4ed/Radiometry,\\_Spectra,\\_and\\_Color/Color#ChoosingtheNumberofWavelengthSamples](https://www.pbr-book.org/4ed/Radiometry,_Spectra,_and_Color/Color#ChoosingtheNumberofWavelengthSamples) (дата обращения: 12.07.2025).
3. Система реалистичного рендеринга HydraCore3 // URL: <https://github.com/Ray-Tracing-Systems/HydraCore3> (Дата обращения: 12.07.2025).
4. Evans G. F., Mccool M. D. Stratified wavelength clusters for efficient spectral Monte Carlo rendering // Graphics Interface '99. Morgan Kaufmann. 1999. P. 42–49.
5. Jakob W., Hanika J. A low-dimensional function space for efficient spectral upsampling. // Computer Graphics Forum. 2019. Vol. 38, no. 2. P. 147–155.
6. Using moments to represent bounded signals for spectral rendering / C. Peters [and others] // ACM Transactions on Graphics. 2019. Vol. 38, no. 4. P. 1–14.
7. Fichet A., Peters C., Compression of Spectral Images using Spectral JPEG XL // Journal of Computer Graphics Techniques (JCGT). 2025. Vol. 14, no. 1. P. 49–69.
8. A spectrally based framework for realistic image synthesis / Y. Sun [and others] // Visual Comp. 2001. Vol. 17. P. 429–444.
9. Lemieux C. Control Variates // Wiley StatsRef: Statistics Reference Online. 2017. P. 1–8.
10. Botev Z., Ridder A. Variance reduction // Wiley StatsRef: Statistics Reference Online. 2017. P. 1–6.
11. Schlick C. An inexpensive BRDF model for physically-based rendering // Computer graphics forum. 1994. Vol. 13. P. 233–246.
12. Belcour L, Bati M, Barla P. Bringing an accurate fresnel to real-time rendering: a preintegrable decomposition // SIGGRAPH 2020: special interest group on computer graphics and interactive techniques conference talks. ACM. 2020. P. 1–2.
13. Real-time Neural Appearance Models / T. Zeltner [and others] // ACM Transactions on Graphics. 2024. Vol. 43, no. 3. P. 1–17.
14. Chitnis S., Sole A., Chandran S. Spectral bidirectional reflectance distribution function simplification // Journal of Imaging. 2025. Vol. 11, no. 1. 18 p.
15. Hero wavelength spectral sampling / A. Wilkie [and others] // Computer Graphics Forum. 2014. Vol. 33, no. 4. P. 123–131.

## Трассировка лучей нейронных геометрических моделей

А. А. Николаев<sup>2</sup>, Р. О. Родионов<sup>1</sup>, В. А. Фролов<sup>1,2</sup>

<sup>1</sup>Институт искусственного интеллекта МГУ, Москва, Россия

<sup>2</sup>Институт прикладной математики им.М.В.Келдыша РАН, Москва, Россия

**Аннотация.** В работе предлагается метод неявного представления поверхностей на основе нейронных направленных полей расстояний. Такие поля для начала и направления луча содержат расстояние до его пересечения поверхностью. Данный метод позволяет компактно хранить поверхности за счёт грубого приближения BVH-деревом и последующего уточнения нейронной сетью, усиленной пространственным кодированием хэш-таблиц с вектор-признаками. В сравнении с существующими методами предложенный метод потребляет в 3 раза меньше памяти при том же уровне точности представления поверхностей.

**Ключевые слова:** нейронный рендеринг, нейронная геометрия, трассировка лучей, сжатие геометрии, поля расстояний

## Ray Tracing with Neural Geometry

A. A. Nikolaev<sup>2</sup>, R. O. Rodionov<sup>1</sup>, V. A. Frolov<sup>1,2</sup>

<sup>1</sup>IAI Moscow State University, Moscow, Russia

<sup>2</sup>Keldysh Institute of Applied Mathematics, Moscow, Russia

**Abstract.** We introduce an implicit surface representation method built on neural directed distance fields. Given a ray defined by origin and direction, these fields output the distance to the intersection point. We compress surfaces by combining a coarse BVH approximation with a neural network refined through spatial hash-table encoding of vector features. Compared with existing approaches, our method achieves the same surface representation accuracy while using a third of a storage.

**Keywords:** neural rendering, neural geometry, ray tracing, geometry compression, distance fields

### Введение

Полигональные сетки (меш) широко используются для представления объектов в компьютерной графике, однако плохо предназначены для хранения сложных моделей – попытки изобразить плавные поверхности и высокочастотные детали на объекте приводят к значительному увеличению размера представления.

Альтернативным представлением является представление нейронными сетями. Распространение нейронных моделей и улучшение их аппаратной и программной поддержки значительно расширило класс функций, которые можно эффективно приближать такими моделями, включив в него кодирование трехмерных сцен. Так или иначе, все алгоритмы синтеза изображений (трассировка лучей, поля расстояний со знаком SDF [1]) являются функциями, которые можно приближать нейронными сетями. Объект переводится в набор весов, а необходимая для рендера информация вычисляется прямым проходом нейронной сети.

Класс функций нейронных сетей достаточно широк [2]. Правильно подобранная архитектура и процедура обучения позволяют получить функции, соответствующие поверхностям, которые трудно представить полигональной сеткой. В то же время такой подход позволяет довольно быстро интегрировать нейронные представления в существующие рендер-системы – достаточно заменить запросы к свойствам трехмерного объекта на запросы к нейронной сети.

Основные характеристики нейронных представлений – точность, с которой восстанавливается исходная поверхность, скорость синтеза изображений и компактность. Оптимизация по всем трем значениям определяет фундаментальную проблему представления трехмерных объектов. В данной работе область ограничивается методами, подходящими для рендера в реальном времени, а фокусом является оптимальность по двум остальным свойствам.

### Обзор существующих решений

#### 1. Нейронные представления

Методы DeepSDF [3] и DeepLS [4] приближают функцию расстояния со знаком (SDF [1]). Авторы выделяют широкий набор сцен и обучают векторные представления, уникальные для каждой сцены,



и универсальный декодировщик. Такое представление имеет меньший размер, чем исходная полигональная сетка, однако для синтеза изображений требуется многошаговый алгоритм Sphere Tracing [5]. Работа [6] в дальнейшем сместила фокус с обучения латентного вектора под каждую сцену с общим декодировщиком к обучению отдельной нейронной сети для каждого объекта.

## *2. Пространственное кодирование*

В NeRF [7] и других методах прямым проходом координаты точки в пространстве дополнительно кодируются через подстановку их в периодические функции с разными частотами.

ReLU Fields [8] предлагает определить в пространстве воксельную сетку и поместить в ее узлы обучаемые F-мерные векторы. Тогда кодирование точки в пространстве вычисляется как линейная интерполяция векторов в ближайших узлах сетки. Такой метод позволяет улучшить точность представления, но имеет количество обучаемых весов, растущее пропорционально количеству вокселей в сетке.

В работе [9] предлагается кодирование HashGrid: определяется таблица из T обучаемых F-мерных вектор-признаков и хэш-функция, осуществляющая (неинъективное) отображение узлов сетки в векторы. Так количество параметров перестает зависеть от разрешения сетки, однако возникающие коллизии ухудшают результат.

VQAD [10] является модификацией HashGrid, в котором отображение из узлов в таблицу становится обучаемым. В каждый узел сетки помещается индекс в таблицу признаков, который настраивается во время обучения. Недостатком метода является большое потребление памяти во время обучения, так как для непрерывной оптимизации каждый индекс требуется заменить на T весов для каждой строки таблицы.

Наконец, в работе [11] обобщается понятие функции отображения вокселей в таблицу вектор-признаков и предлагается альтернатива такой функции в VQAD [10], которая требует меньшее количество вычислений.

Появление описанных пространственных кодирований во многом снизило необходимость подбора функций активации и других конфигураций, как предлагалось в [8, 12, 13].

## *3. Направленные поля расстояний*

Поля расстояний со знаком для данной точки пространства хранят расстояние до ближайшей точки поверхности. Для получения пересечения вдоль заданного направления приходится использовать алгоритм Sphere Tracing [5], требующий многократного обращения к нейронной сети. Альтернативным представлением являются направленные поля расстояний (Directed Distance Fields, DDF) [14], которые за одну итерацию позволяют получить расстояние до пересечения вдоль направления. Такой подход имеет свои недостатки, например, в отличие от SDF, DDF представляет разрывную функцию, что усложняет его приближение нейронными сетями.

В работе [14] приводится строгое определение такого поля, а также выявляются некоторые математические ограничения на функции, являющиеся направленными полями расстояний. В исследовании [15] рассматривается более эффективное представление луча для DDF, содержащее меньше степеней свободы. В работе [16] предлагается определять луч через две точки на сфере, содержащей сцену, а также модифицировать процедуру обучения для получения лучшей консистентности.

## *4. Применение ускоряющих структур*

Для задачи построения нейронного представления по заданной полигональной сетке имеет смысл использовать ускоряющие структуры – методы классической компьютерной графики, предназначенные для более быстрого проведения запросов на пересечении лучей.

В работе [17] представлена обучающаяся нейронная модель, предсказывающая наличие пересечения для вторичных лучей. Модель затем интегрируется в классическую схему трассировки лучей, ускоряя ее. Алгоритм NGLoD [18], в свою очередь, строит вокруг заданного объекта октодерево, помещая в его узлы вектор-признаки. Такая схема позволяет перед запросом к нейронной SDF пропустить заведомо пустое пространство с помощью октодерева. При этом глубину обхода дерева можно менять после обучения, определяя разные уровни детализации.

NBVH [19] в сравнении с NGLoD [18, 19] предлагает использовать вместо октодерева дерево BVH [20], а вектор-признаки хранить по принципу HashGrid [9]. SDF заменено на направленное поле расстояний, которое также хранит информацию о нормалях и других параметрах поверхности. Для генерации

обучающей выборки (множества лучей) лучи генерируются случайно в увеличенном корневом узле дерева. Для каждого луча проводится обход дерева до первой точки пересечения.

При пересечении луча с узлом BVH внутри узла на луче семплируются  $k$  точек. Каждая точка кодируется HashGrid-вектором, векторы всех точек конкатенируются и формируют вход в нейронную сеть.

При хорошей степени сжатия NBVH [19] ограничен в точности представления геометрии. Изменение конфигурации для улучшения точности приводит к увеличению количества параметров нейронной сети и к уменьшению степени сжатия.

Работа [21] показывает, что применение нейросетей позволяет смягчить проблему дивергенции лучей при параллельной реализации трассировки лучей и получить ускорение. Алгоритм Neural Prefiltering [22] делит сцену на воксели и кодирует в весах нейронной сети результаты предрасчета трассировки лучей внутри каждого вокселя. Разные размеры вокселей определяют уровни детализации.

LSNIF [23] предлагает на нижних уровнях BVH определять воксельную сетку и обучать нейронную модель под пересечение внутри вокселей. Кроме того, узлы сетки вокселей и сетки HashGrid [9] выровнены. Начало и конец отрезка луча всегда попадают на грань вокселей HashGrid, что ускоряет интерполяцию значений вектор-признаков.

Во всех упомянутых методах нейронная сеть представляет собой многослойный перцептрон (MLP).

### Предложенный метод

В работе предлагается метод, основанный на NBVH [19], улучшающий его сжимающую способность и гибкость в выборе параметров. Предложены модификации следующих компонентов: пространственного кодирования, нейросетевой архитектуры, генерации обучающей выборки.

#### 1. Пространственное кодирование

Предложенный метод кодирования – HashVQAD – вдохновлен [11], сохраняет преимущества VQAD [10] и требует меньший объем памяти как для обучения, так и для хранения после фиксации весов.

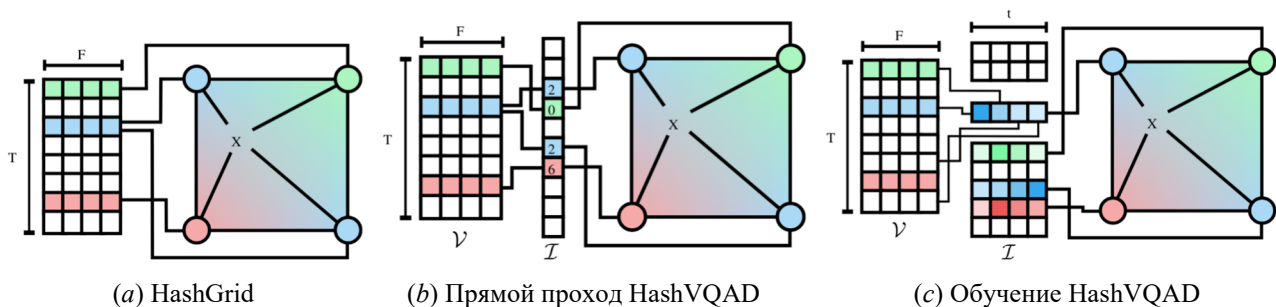


Рисунок 1. (a) HashGrid [9] определяет таблицу из  $T$  обучаемых вектор-признаков и фиксированное отображение узлов в строки таблицы хэш-функцией. (b) HashVQAD определяет промежуточную таблицу обучаемых индексов, указывающих на строки таблицы признаков. (c) Во время обучения в каждой ячейке таблицы индексов подбираются  $t < T$  весов. Индексы  $[0, t)$  отображаются в  $[0, T)$  с помощью хэш-функции

Пусть воксельная сетка имеет размерность  $N \times N \times N$ .  $F$ -мерные вектор-признаки хранятся в таблице  $V$  размера  $T$  аналогично [9]. При этом задается вторая таблица  $I$  размера  $H$ , каждая строка которой хранит единственное число  $[0, T - 1]$  – индекс в таблицу признаков.

Каждому вокселю ставится в соответствие элемент таблицы  $I$  с помощью хэш-функции, аналогичной [9]. В свою очередь, каждый элемент  $i$  таблицы  $I$  указывает на некоторую строку  $V$ . Так для каждого вокселя определяется вектор-признак (рис. 1 (b)).

Элементы  $I$  обучаемы. Чтобы избежать решения задачи дискретной оптимизации в целых числах, можно на время обучения каждый элемент  $I$  заменить на  $T$  вещественных весов. Вектор-признак тогда считается как взвешенная сумма строк  $V$ . Для уменьшения потребления памяти предлагается настраивать вес не для каждого из  $T$  векторов, а для  $t < T$ .

В таблице 1 приведены примеры значений гиперпараметров.

Таблица 1. Конфигурации позиционного кодирования ( $F = 16$ ,  $T = 2^8$ ,  $N = 7$ )

Метод	Хранение	Обучение	Параметры
VQAD	1 МБ	1 ГБ	-
HashVQAD	64 КБ	512 КБ	$H = 2^{15}$ , $t = 16$

После обучения элементы  $I$  переводятся в целые числа операцией  $\text{argmax}$ . Вообще говоря, выходные значения кодирования до и после применения такой операции отличаются. Поэтому после основного этапа обучения проводятся еще несколько итераций оптимизации с уменьшенной температурой  $\text{softmax}$  ( $x$  – вектор весов,  $\tau$  – температура):

$$\text{softmax}(x, \tau)_i = \frac{\exp(x_i/\tau)}{\sum_j \exp(x_j/\tau)}.$$

При уменьшении температуры распределение по строкам  $V$  приближается к вырожденному, потому переход на использование лишь одного индекса с наибольшим весом вносит минимальные изменения в выход модели. После завершения обучения индексы  $[0, t]$  переводятся хэш-функцией в значения  $[0, T]$ .

Вычисление HashVQAD по количеству операций сравнимо с вычислением HashGrid. Такой подход не требует построения воксельного дерева, как в [10], в котором при  $N = 7$  количество узлов достигает  $2^{20}$ . Индекс в таблице признаков хранится не в каждом узле дерева, а в отдельной таблице, размер которой можно регулировать.

## 2. Нейросетевая архитектура

На отрезке, в котором ищется точка пересечения, семплируются дополнительные внутренние точки. Как правило, увеличение количества точек  $k$  приводит к лучшему качеству. При количестве точек  $\geq 8$  размерность первого слоя становится слишком большой и может составлять больше половины всех параметров MLP. К тому же ширина второго слоя не меняется, что приводит к потере значительной части дополнительной информации, предоставляемой отсемплированными точками.

В этой работе предлагается рассмотреть альтернативную архитектуру нейронной сети – KMLP.

Метод прямого прохода организован следующим образом (рис. 2):

- выделяется один MLP;
- среди  $k$  точек выделяются  $k - 1$  пар соседних точек. Представления пространственного кодирования для каждой пары точек конкатенируются и по отдельности подаются на вход MLP;
- MLP для каждой пары точек вычисляет данные о пересечении на локальном отрезке между двумя точками. Далее данные для всех пар объединяются для получения данных о пересечении на всем отрезке.

В таком подходе количество параметров нейронной сети не зависит от количества семплируемых точек, а MLP не ограничивает точность метода. Кроме того, решается проблема разрывов значений поля DDF, если при разрыве происходит смена отрезка.

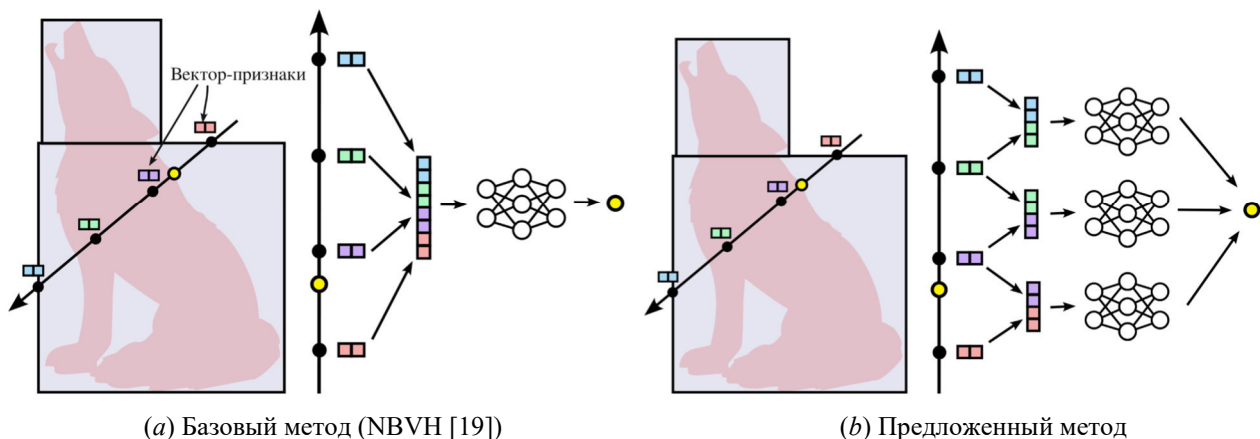


Рисунок 2. (a) NBVH [19] конкатенирует векторы всех точек. При увеличенном количестве точек точность метода ограничена одним MLP, а первый слой принимает на вход неестественно большой вектор. (b) Несколько MLP с идентичными весами позволяют увеличивать количество точек и получать улучшение точности

### 3. Генерация обучающей выборки

Общепринятые алгоритмы построения BVH-дерева для каждого узла определяют множество примитивов, в нем содержащихся. При разбиении узла это множество примитивов делится на два непересекающихся множества. Форма ограничивающих объемов (выровненные по осям координат параллелепипеды, AABB) и частое отсутствие разделяющей гиперплоскости для двух подмножеств приводит к тому, что ограничивающие объемы двух новых узлов могут пересекаться.

Эта особенность делает возможной ситуацию, когда при обходе дерева точка пересечения луча с поверхностью находится внутри AABB-узла, но примитив, содержащий эту точку, принадлежит другому узлу (рис. 3). Это создает неконсистентную обучающую выборку – корректную с точки зрения обхода дерева, но некорректную с точки зрения реального нахождения пересечения в узле.

В LSNIIF [23] такой проблемы не возникает, так как нейронная сеть вычисляет пересечения внутри сетки вокселей, а не произвольных ограничивающих объемов.

В данной работе предлагается генерировать обучающую выборку в два этапа. Сначала для каждого луча находятся все точки пересечения путем полного обхода дерева. Второй обход каждому листу ставит в соответствие ближайшую точку, которая лежит внутри AABB этого листа (если такая существует). Во время второго обхода принадлежность примитива точки листу не имеет значения, и в очередном узле точка учитывается, даже если ее примитив не принадлежит узлу (рис. 3).

## Результаты

### 1. Набор данных и параметры обучения

Тестирование проводилось на нескольких классических моделях, а также на моделях из датасета Thingi10k [24]. Было отобрано 20 моделей разных размеров (рис. 4). Отобранные модели содержат достаточно разных участков поверхностей, чтобы на них можно было провести уверенное сравнение.

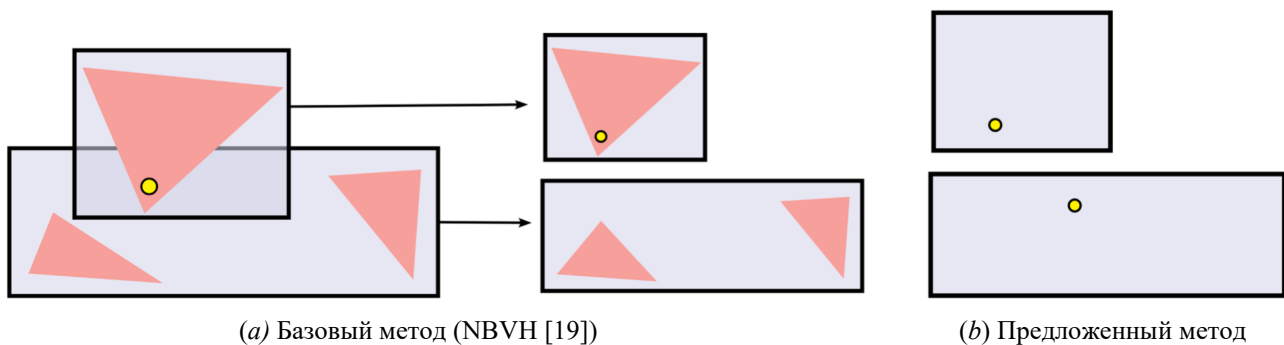


Рис. 3. (a) Точка пересечения учтена только в одном из узлов. (b) Точка пересечения учтена оба раза вне зависимости от примитива, которому она принадлежит

Использовалась следующая конфигурация обучения и валидации:

- оптимизатор Adam [25];
- Learning Rate 0.001, остальные параметры оптимизатора по умолчанию;
- хранение весов в float32, Mixed Precision [26] не применяется;
- размер батча 400000 лучей, длительность обучения – 20 эпох по 800 батчей;
- размер изображения  $1000 \times 1000$ .

Каждый узел BVH-дерева содержит шесть чисел float32 (границы ограничивающего параллелепипеда) и индекс int32 левого сына. Левый и правый сыновья всегда идут подряд. С выравниванием до 32 байт каждый узел содержит 32 байта. Во всех экспериментах строилось 4096 узлов дерева, то есть все дерево занимает 128 КБ.

MLP содержит 4 внутренних слоя по 64 нейрона. На отрезке семплируется 8 точек. С учетом первого слоя базовая архитектура MLP занимает 200 КБ, архитектура KMLP – 100 КБ.

В экспериментах для HashGrid и HashVQAD зафиксированы количество уровней  $L = 4$ , разрешения нижнего и верхнего уровней  $N_{\min} = 16$ ,  $N_{\max} = 128$ , размерность вектор-признаков  $F = 16$ . Для HashVQAD размер таблицы индексов  $H$  связан с размером таблицы признаков  $T$  соотношением  $H = T \cdot 2^7$ .

Для каждого индекса обучается  $t = 16$  весов. Таким образом, единственный регулируемый параметр – размер таблицы признаков  $T$ .

Наблюдения и результаты других работ [9, 19, 10] показывают, что регулировки  $T$  достаточно, чтобы рассматривать кодировщики разного размера.

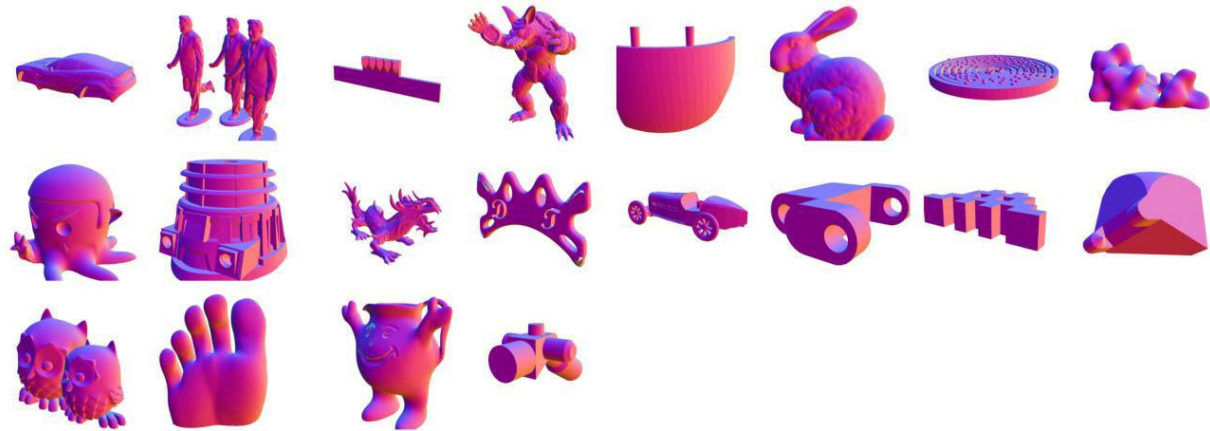


Рисунок 4. Примеры моделей, на которых проводилось тестирование

Качество нейронного представления определяется среднеквадратичной ошибкой между оригинальным изображением и изображением, полученным из этого представления. Синтез изображения проводится следующим образом: в сцене фиксируется точечный белый источник света, цвет пикселя определяется как косинус между нормалью в точке пересечения и направлением к источнику света. Такая формулировка учитывает предсказанные значения и для нормалей, и для расстояний до пересечения.

После завершения обучения синтезировалось 100 изображений с разных положений камеры, результаты усреднялись. В результатах экспериментов приведены значения MSE, усредненные по отобранным моделям.

## 2. Результаты

**HashVQAD.** В таблице 2 приведены результаты тестирования. Можно утверждать, что предложенное кодирование показывает аналогичные результаты при вдвое меньшем числе параметров.

Таблица 2. Результаты тестирования кодирования HashVQAD.

Тип кодирования	$T$	Размер кодирования	Полный размер	MSE
HashGrid	$2^9$	128 КБ	456 КБ	0.00255
HashVQAD	$2^8$	128 КБ	456 КБ	0.00230
HashGrid	$2^{10}$	256 КБ	584 КБ	0.00232
HashVQAD	$2^{10}$	256 КБ	584 КБ	0.00201
HashGrid	$2^{11}$	512 КБ	840 КБ	0.00197
HashVQAD	$2^{10}$	512 КБ	840 КБ	0.00188

Кодирование HashVQAD стабильно показывает лучшие результаты при том же количестве параметров.

**KMLP.** Результаты тестирования приведены в таблице 3. Разделение точек на соседние пары позволяет одновременно уменьшить количество параметров нейронной сети и получить прирост в качестве предсказаний.

**Генерация обучающей выборки (лучей).** В таблице 4 приведены результаты тестирования, показывающие улучшение точности при использовании новой стратегии генерации лучей.

**Комбинация улучшений.** В таблице 5 приведены результаты при последовательном применении всех предложенных улучшений. Размеры кодировщика подобраны так, чтобы качество предсказания оказывалось примерно одинаковым. Можно наблюдать, что качество, аналогичное базовому методу, достигается при втрое меньшем количестве параметров.



Таблица 3. Результаты тестирования архитектуры KMLP

Архитектура	T	Полный размер	MSE
MLP	512	456 КБ	0.00255
KMLP	512	356 КБ	0.00241
MLP	1024	584 КБ	0.00232
KMLP	1024	484 КБ	0.00219
MLP	2048	840 КБ	0.00197
KMLP	2048	740 КБ	0.00186

Таблица 4. Результаты тестирования предложенного метода генерации лучей

Метод генерации	T	Полный размер	MSE
Базовый	512	456 КБ	0.00255
Предложенный	512	456 КБ	0.00233
Базовый	1024	584 КБ	0.00232
Предложенный	1024	584 КБ	0.00210
Базовый	2048	840 КБ	0.00197
Предложенный	2048	840 КБ	0.00181

Таблица 5. Результаты тестирования всех предложенных методов: I: базовый метод + KMLP; II: I + улучшенная генерация лучей; III: II + HashVQAD

Метод	T	Полный размер	MSE
Базовый	2560	968 КБ	0.00188
I	1536	612 КБ	0.00198
II	1024	484 КБ	0.00190
III	512	356 КБ	0.00192

На рисунках 5 и 6 показаны примеры синтезированных изображений из сжатых представлений. Некоторые модели можно сжать в  $\approx 40$  раз и получать изображения, не отличимые с расстояния от оригинальных.

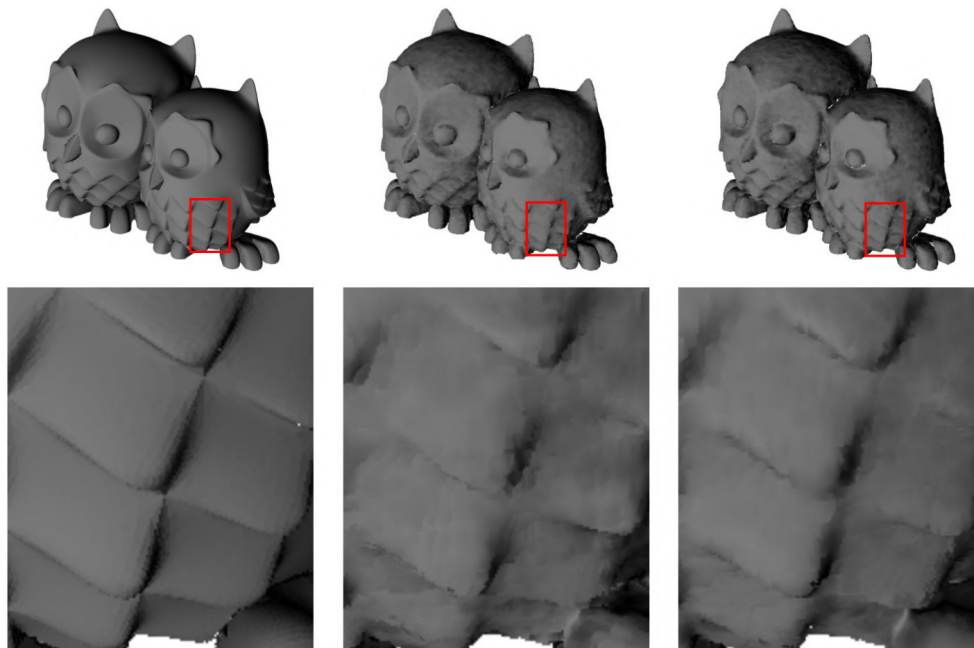


Рисунок 5. Синтез изображений из нейронных представлений. Слева направо: эталон (27 МБ), NBVH [19] (968 КБ, 0.00228 MSE), предложенный метод (868 КБ, 0.00176 MSE)

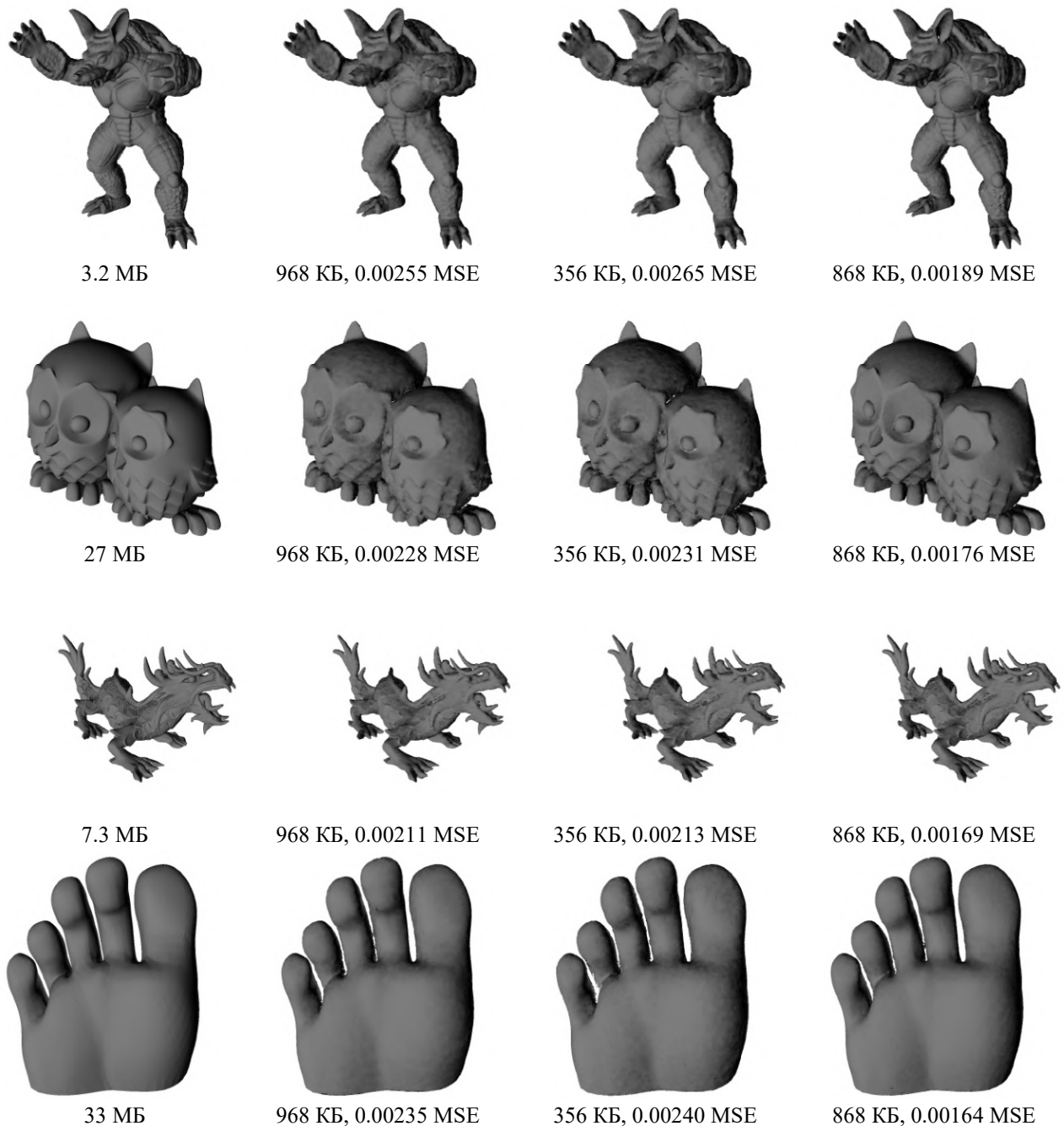


Рисунок 6. Изображения, по которым считается MSE. Фиксируется один точечный источник белого света, цвет пикселя определяется как косинус между нормалью и направлением к свету в точке пересечения (отрицательные значения заменяются нулем). Слева направо: оригинальное изображение, базовый метод, предложенный метод (356 КБ), предложенный метод (356 КБ).

На рисунке 7 показано сравнение базового метода и предложенного с визуализацией нормалей.

Приближение по центру показывает поверхность из большого количества мелких повторяющихся деталей, каждая из которых в полигональной сетке представляется отдельными треугольниками. В то же время нейронное представление позволяет сжимать поверхности такого вида.

Приближение снизу слева показывает сегмент, на котором проявляется преимущество KMLP: центральные отверстия неглубоки и попадают в один узел BVH. Разбиение отрезка на несколько меньших позволяет более точно приблизить разрывы, связанные с отверстиями.

Приближение снизу справа показывает недостаток, которым обладают оба метода: недостаточно качественное предсказание точек пересечения, которые находятся не в первых узлах при обходе BVH. Улучшенная генерация лучей в предложенном методе слегка улучшает качество в подобных участках.



Приведенные результаты показывают, что предложенный метод проявляет лучшую способность к сжатию, чем базовый метод. Предложенный метод позволяет получить то же качество изображений, что и базовый метод, используя в  $\approx 3$  раза меньше параметров. При этом преимущества предложенного метода можно проинтерпретировать, рассмотрев особо сложные участки поверхностей моделей.

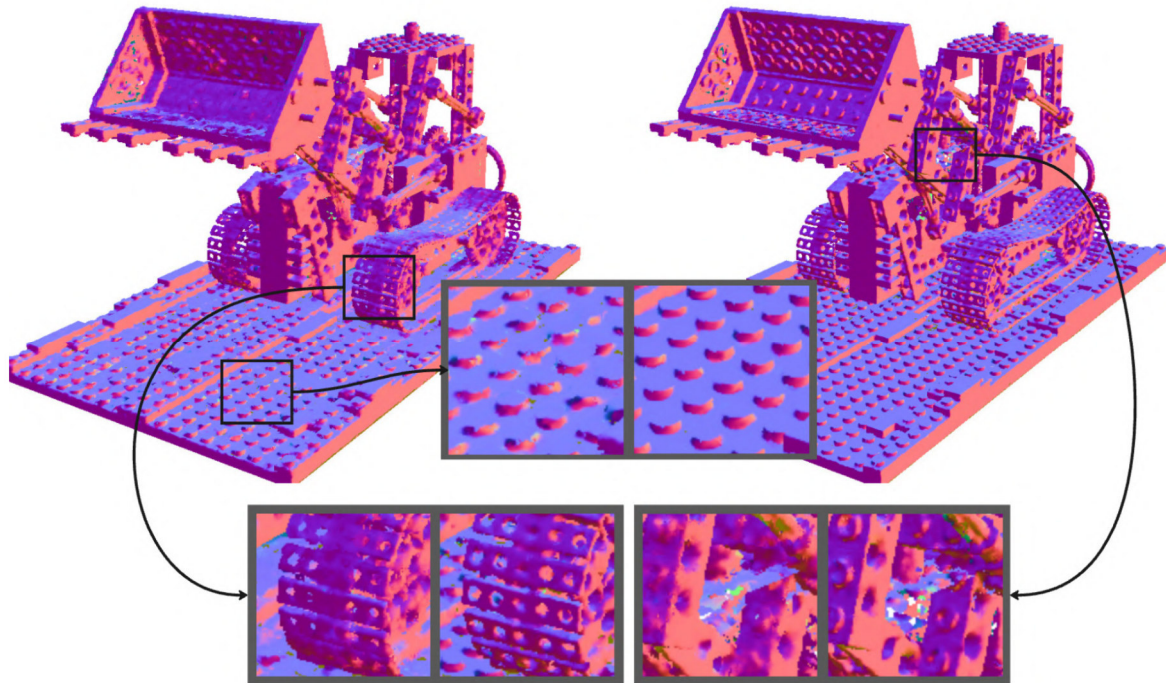


Рисунок 7. Базовый метод (слева) и предложенный (справа). Оба нейронных представления занимают 5 МБ, оригинальная модель занимает 82 МБ

### Заключение

В работе описан и протестирован метод сжимающего нейронного представления, которое повышает точность базового метода NBVN [19]. Показано, как проявляют себя предложенные модификации в применении к разным поверхностям.

Улучшение точности достигается ценой относительно небольшого увеличения вычислительной сложности алгоритма: внедрение предложенных методов генерации обучающей выборки и пространственного кодирования не меняет количество операций, потому KMLP – единственный компонент, с увеличенным временем работы в сравнении с NBVN. В таблице 6 указано количество сложений и умножений, необходимых для прямого прохода MLP и KMLP.

Таблица 6. Количество операций для KMLP и MLP.

Используются четыре скрытых слоя по 64 нейрона, выходной слой имеет 5 нейронов. Размерность вектор-признаков каждой из  $k$  точек  $F = 16$ .

Метод	$k$	Количество операций
MLP	4	16965
KMLP	4	44751
MLP	8	21061
KMLP	8	104419

Описанный метод имеет ограничения, унаследованные от NBVN [19]. Так, после изменения поверхности нейронное представление потребует обучения заново. Кроме того, предложенные модификации пространственного кодирования и архитектуры нейронной сети используют больше обращений к памяти, что усложнит эффективную реализацию. Решению описанных проблем будут посвящены будущие исследования.

Отдельной задачей для будущих работ является эффективная реализация на графическом процессоре с учетом особенности метода – совмещения обхода дерева и запуска нейронной сети для каждого листа. Наивная реализация этапов обучения и применения метода на фреймворке PyTorch [27] позволяет обучить представление на рисунке 7 примерно за минуту (10 эпох) и отрендерить изображение из нейронного представления за  $\approx 220$  мс. Для экспериментов использовалась видеокарта RTX 4080 SUPER.

### Список литературы

1. Curless Brian. A volumetric method for building complex models from range images // Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '96. New York, NY, USA: Association for Computing Machinery, 1996. P. 303–312. <https://doi.org/10.1145/237170.237269>.
2. Augustine Midhun T. A Survey on Universal Approximation Theorems. 2024.
3. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation / Jeong Joon Park, Peter Florence, Julian Straub, et al. // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019. P. 165–174. <http://dx.doi.org/10.1109/CVPR.2019.00025>.
4. Deep Local Shapes: Learning Local SDF Priors for Detailed 3D Reconstruction / Rohan Chabra, Jan E. Lenssen, Eddy Ilg, et al. // Computer Vision – ECCV 2020. Springer International Publishing, 2020. P. 608–625. [http://dx.doi.org/10.1007/978-3-030-58526-6\\_36](http://dx.doi.org/10.1007/978-3-030-58526-6_36).
5. Hart John. Sphere Tracing: A Geometric Method for the Antialiased Ray Tracing of Implicit Surfaces // *The Visual Computer*. 1995. 06. Vol. 12.
6. Davies Thomas. On the Effectiveness of Weight-Encoded Neural Implicit 3D Shapes. 2020.
7. NeRF: representing scenes as neural radiance fields for view synthesis / Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, et al. // *Communications of the ACM*. 2021. Vol. 65, no. 1. P. 99–106. <http://dx.doi.org/10.1145/3503250>.
8. ReLU Fields: The Little Non-linearity That Could / Animesh Karnewar, Tobias Ritschel, Oliver Wang, Niloy Mitra // Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings. SIGGRAPH '22. ACM, 2022. <http://dx.doi.org/10.1145/3528233.3530707>.
9. Instant neural graphics primitives with a multiresolution hash encoding / Thomas Muller, Alex Evans, Christoph Schied, Alexander Keller // *ACM Transactions on Graphics*. 2022. Vol. 41, no. 4. P. 1–15. <http://dx.doi.org/10.1145/3528223.3530127>.
10. Variable Bitrate Neural Fields / Towaki Takikawa, Alex Evans, Jonathan Tremblay, et al. // Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings. SIGGRAPH '22. ACM, 2022. P. 1–9. <http://dx.doi.org/10.1145/3528233.3530727>.
11. Takikawa Towaki. Compact Neural Graphics Primitives with Learned Hash Probing. 2023. <https://arxiv.org/abs/2312.17241>.
12. Tancik Matthew. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. 2020.
13. Sitzmann Vincent. Implicit Neural Representations with Periodic Activation Functions. 2020
14. Representing 3D Shapes with Probabilistic Directed Distance Fields / Tristan Aumentado-Armstrong, Stavros Tsogkas, Sven Dickinson, Allan Jepson // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022. P. 19321–19332. <http://dx.doi.org/10.1109/CVPR52688.2022.01874>.
15. Feng Brandon Yushan. PRIF: Primary Ray-based Implicit Function. 2022.
16. Liu Zhuoman. RayDF: Neural Ray-surface Distance Fields with Multi-view Consistency. 2023.
17. Fujieda Shin. Neural Intersection Function. 2023.
18. Neural Geometric Level of Detail: Real-time Rendering with Implicit 3D Shapes / Towaki Takikawa, Joey Litalien, Kangxue Yin, et al. // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021. P. 11353–11362. <http://dx.doi.org/10.1109/CVPR46437.2021.01120>.
19. N-BVH: Neural ray queries with bounding volume hierarchies / Philippe Weier, Alexander Rath, Elie Michel, et al. // Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers '24. SIGGRAPH '24. ACM, 2024. P. 1–11. <http://dx.doi.org/10.1145/3641519.3657464>.
20. A Survey on Bounding Volume Hierarchies for Ray Tracing / Daniel Meister, Shinji Ogaki, Carsten Benthin, et al. // *Computer Graphics Forum*. 2021. Vol. 40, no. 2. Pp. 683–712. <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.142662>.
21. Data Parallel Ray Tracing of Massive Scenes based on Neural Proxy / Shunkang Xu, Xiang Xu, Yanning Xu, Lu Wang // Pacific Graphics Conference Papers and Posters / Ed. by Renjie Chen, Tobias Ritschel, Emily Whiting. The Eurographics Association, 2024.
22. Neural Prefiltering for Correlation-Aware Levels of Detail / Philippe Weier, Tobias Zirr, Anton Kaplanyan, et al. // *ACM Trans. Graph*. 2023. Vol. 42, no. 4. – 16 pp. <https://doi.org/10.1145/3592443>.

23. Fujieda Shin. LSNIIF: Locally-Subdivided Neural Intersection Function / Shin Fujieda, Chih-Chen Kao, Takahiro Harada // *Proceedings of the ACM on Computer Graphics and Interactive Techniques*. 2025. Vol. 8, no. 1. P. 1–18. <http://dx.doi.org/10.1145/3728295>.
24. Zhou Qingnan. Thingi10K: A Dataset of 10,000 3D-Printing Models / Qingnan Zhou, Alec Jacobson // *arXiv preprint arXiv:1605.04797*. 2016.
25. Kingma Diederik P. Adam: A Method for Stochastic Optimization. 2017. <https://arxiv.org/abs/1412.6980>.
26. Micikevicius Paulius. Mixed Precision Training. 2017.
27. Paszke Adam. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019. <https://arxiv.org/abs/1912.01703>.
28. Plenoxels: Radiance Fields without Neural Networks / Sara Fridovich-Keil, Alex Yu, Matthew Tancik, et al. // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022. P. 5491–5500. <http://dx.doi.org/10.1109/CVPR52688.2022.00542>.
29. 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions / Andy Zeng, Shuran Song, Matthias Niessner, et al. // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017. P. 199–208. <http://dx.doi.org/10.1109/CVPR.2017.29>.
30. Wu Zhirong. 3D ShapeNets: A Deep Representation for Volumetric Shapes. 2014.
31. Perlin K., Hoffert E. M. Hypertexture // *SIGGRAPH Comput. Graph.* 1989. Vol. 23, no. 3. P. 253–262. <https://doi.org/10.1145/74334.74359>.
32. Perlin K., Hoffert E. M. Hypertexture // *Proceedings of the 16th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '89*. New York, NY, USA: Association for Computing Machinery, 1989. P. 253–262. <https://doi.org/10.1145/74333.74359>.
33. Hu Edward J. LoRA: Low-Rank Adaptation of Large Language Models. 2021. <https://arxiv.org/abs/2106.09685>.
34. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields / Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, et al. // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2021. <http://dx.doi.org/10.1109/ICCV48922.2021.00580>.
35. Adaptive Shells for Efficient Neural Radiance Field Rendering / Zian Wang, Tianchang Shen, Merlin Nimier-David, et al. // *ACM Transactions on Graphics*. 2023. Vol. 42, no. 6. P. 1–15. <http://dx.doi.org/10.1145/3618390>.
36. 3D Gaussian Splatting for Real-Time Radiance Field Rendering / Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, George Drettakis // *ACM Transactions on Graphics*. 2023. Vol. 42, no. 4. P. 1–14. <http://dx.doi.org/10.1145/3592433>.
37. Guedon Antoine. Gaussian Frosting: Editable Complex Radiance Fields with Real-Time Rendering. 2024.
38. KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs / Christian Reiser, Songyou Peng, Yiyi Liao, Andreas Geiger // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2021. <http://dx.doi.org/10.1109/ICCV48922.2021.01407>.
39. Yariv Lior. BakedSDF: Meshing Neural SDFs for Real-Time View Synthesis. 2023.
40. Baking Neural Radiance Fields for Real-Time View Synthesis / Peter Hedman, Pratul P. Srinivasan, Ben Mildenhall, et al. // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2021. <http://dx.doi.org/10.1109/ICCV48922.2021.00582>.
41. Knodt Julian. Neural Ray-Tracing: Learning Surfaces and Reflectance for Relighting and View Synthesis. 2021.
42. Kim Doyub, Lee Minjae, Museth Ken. NeuralVDB: High-resolution Sparse Volume Representation using Hierarchical Neural Networks // *ACM Transactions on Graphics*. 2024. Vol. 43, no. 2. P. 1–21. <http://dx.doi.org/10.1145/3641817>.
43. Wang Peng. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. 2021.
44. Occupancy Networks: Learning 3D Reconstruction in Function Space / Lars Mescheder, Michael Oechsle, Michael Niemeyer, et al. // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019. <http://dx.doi.org/10.1109/CVPR.2019.00459>.
45. Chibane Julian. Neural Unsigned Distance Fields for Implicit Function Learning. 2020.
46. Sitzmann Vincent. Light Field Networks: Neural Scene Representations with Single-Evaluation Rendering. 2021.
47. Li Zhong. NeuLF: Efficient Novel View Synthesis with Neural 4D Light Field. 2021.
48. Kulkarni Nilesh. What's Behind the Couch? Directed Ray Distance Functions (DRDF) for 3D Scene Reconstruction. 2021.
49. Feng Brandon Yushan, Varshney Amitabh. SIGNET: Efficient Neural Representation for Light Fields // *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021. October. Pp. 14224–14233.
50. Scene Representation Transformer: Geometry-Free Novel View Synthesis Through Set-Latent Scene Representations / Mehdi S.M. Sajjadi, Henning Meyer, Etienne Pot, et al. // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022. P. 6219–6228. <http://dx.doi.org/10.1109/CVPR52688.2022.00613>.
51. Wang Peng. Progressively-connected Light Field Network for Efficient View Synthesis. 2022.

- 
52. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis / Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, et al. // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021. P. 7491–7500. <http://dx.doi.org/10.1109/CVPR46437.2021.00741>.
  53. Moenne-Loccoz Nicolas. 3D Gaussian Ray Tracing: Fast Tracing of Particle Scenes. 2024. <https://arxiv.org/abs/2407.07090>.
  54. Epic Games Inc. Nanite. 2020.
  55. Zhang Kai. NeRF++: Analyzing and Improving Neural Radiance Fields. 2020. <https://arxiv.org/abs/2010.07492>.
  56. Deng Kangle. Depth-supervised NeRF: Fewer Views and Faster Training for Free. 2024. <https://arxiv.org/abs/2107.02791>.

## Приближенный рендеринг В-сплайновых поверхностей с триммингом в реальном времени

М. С. Кунц<sup>1</sup>, А. Р. Гарифуллин<sup>2</sup>, И.О. Коротаев<sup>1</sup>, В.А. Фролов<sup>1,2</sup>

<sup>1</sup>Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ имени М.В. Ломоносова, Москва, Россия

<sup>2</sup>Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В данной работе предлагаются два улучшения существующего метода рендеринга поверхностей на основе трассировки лучей. Первое улучшение нацелено на борьбу с артефактами рендеринга, возникающими из-за того, что в существующих подходах на основе метода Ньютона не гарантирована сходимость к ближайшему пересечению. Предложен метод стохастического выбора начального приближения, позволяющий получить корректное изображение сцены, состоящей из NURBS поверхностей, в течение нескольких кадров за счет сохранения правильного начального приближения. Второе улучшение нацелено на ускорение процесса тримминга (или обрезки) поверхности после того, как пересечение уже найдено. Предложенный метод на основе двумерных функций расстояния ускоряет рендеринг в среднем от 1.5 до 2 раз ценой незначительного понижения точности тримминга и, кроме того, позволяет практически без накладных расходов отображать ребра объекта.

**Ключевые слова:** трассировка лучей, NURBS, функции расстояния со знаком.

## Approximate real-time rendering of trimmed B-spline surfaces

M. S. Kunts<sup>1</sup>, A.R. Garifullin<sup>2</sup>, I.O. Korotaev<sup>1</sup>, V.A. Frolov<sup>1,2</sup>

<sup>1</sup>IAI Moscow State University, Moscow, Russia

<sup>2</sup>Keldysh Institute of Applied Mathematics, Moscow, Russia

**Abstract.** This paper proposes two enhancements to an existing ray tracing-based surface rendering method. The first enhancement addresses rendering artifacts that arise because existing Newton-based approaches do not guarantee convergence to the closest intersection. A stochastic initial guess selection method is proposed, which enables the generation of a correct image for a scene composed of NURBS surfaces within a few frames by preserving a valid initial approximation. The second enhancement aims to accelerate the surface trimming (or clipping) process after the intersection has been found. The proposed method, based on two-dimensional distance functions, accelerates rendering by an average factor of 1.5 to 2 at the cost of a minor reduction in trimming accuracy. Furthermore, it allows for the visualization of object edges with virtually no computational overhead.

**Keywords:** ray tracing, NURBS, signed distance functions.

### Введение

NURBS (Non Uniform Rational B-Spline) поверхности являются стандартом геометрического моделирования в современных САПР-системах благодаря своей гибкости и точности представления как элементарных, так и сложных форм. Они широко применяются в промышленном дизайне, авиа- и машиностроении, а также в других областях, требующих работы с параметрическими поверхностями с высокой точностью.

Однако рендеринг таких поверхностей является сложной и не до конца решенной в настоящий момент проблемой. Фундаментальная сложность этой задачи заключается в оптимизации трех ключевых характеристик алгоритма рендеринга: точность, скорость, потребляемая память. На практике приходится соблюдать баланс между этими характеристиками, подходящий для целевой задачи и вычислительного оборудования.

В данной работе мы предлагаем улучшение существующего метода рендеринга В-сплайновых поверхностей на основе трассировки лучей и позиционируем разработанные алгоритмы для приложений предварительного просмотра инженерных моделей, состоящих из В-сплайновых поверхностей, на ограниченном по ресурсам и возможностям графическом оборудовании (GPU), а также для рендеринга на центральном процессоре (CPU).

### Обзор существующих работ

#### 1. Растеризация

Растеризация триангулированных поверхностей в настоящий момент является основным методом рендеринга В-сплайновых моделей. Например, этот подход использует библиотека OpenCASCADE [1].

Данный подход имеет принципиальные ограничения, связанные с необходимостью преобразования параметрических поверхностей в полигональные сетки. Это сказывается на точности получаемого изображения и на потребляемой памяти. Кроме того, растеризация плохо масштабируется и показывает низкую производительность на сложных сценах, когда размер треугольника сопоставим с размером пикселя или меньше его.

Чтобы обойти указанные ограничения, был предложен метод динамической тесселяции ETER [2], адаптирующий плотность триангуляции на основе расстояния до камеры и локальной кривизны поверхности. Это программное окружение реализует алгоритмы, завязанные на такие аппаратные возможности GPU, как консервативная растеризация, тензорные ядра и меш-шейдеры. Реализация метода динамической тесселяции из работы [2] в теории возможна и на графических процессорах, не имеющих полной поддержки указанной аппаратной функциональности, но эффективность такого решения находится под вопросом.

## 2. Трассировка лучей

Численный метод пересечения лучей с усеченными рациональными поверхностями Безье был предложен в работе [3]. В этом подходе задача нахождения пересечений луча с поверхностью сводится к задаче построения выпуклой оболочки поверхности в специальном пространстве, которая используется для оценки диапазонов параметров, гарантированно содержащих искомые точки. Хотя метод обеспечивает высокую точность и гарантирует нахождение всех точек пересечения, он может давать сбой в некоторых особых случаях. Кроме того, метод обладает высокой вычислительной сложностью, поскольку требует выделения локальных участков поверхности и построения их выпуклых оболочек. Также он потребляет память, зависящую от параметров поверхности, что затрудняет его эффективную реализацию на видеокартах.

Для ускорения обработки сложных моделей с В-сплайновыми поверхностями был разработан подход, основанный на использовании BVH (иерархии ограничивающих объемов) в сочетании с методом Ньютона-Рафсона для поиска пересечений лучей с поверхностью [4]. Ключевая особенность метода – использование эвристики, учитывающей кривизну поверхности для выбора оптимального уровня подразделения, минимизирующего вероятность расхождения метода. Однако пересечение с В-сплайновыми поверхностями без преобразований имеет высокую вычислительную сложность, что может повлиять на производительность.

Последующие исследования были направлены на оптимизацию алгоритмов пересечения. В работе [5] предложены усовершенствования метода отсечения Безье. Комбинация отсечения Безье с методом Ньютона-Рафсона позволила увеличить скорость сходимости на 30-50 %, а введение адаптивных порогов ошибок устранило проблемы с бесконечными циклами.

Современные подходы к рендерингу усеченных В-сплайновых поверхностей демонстрируют высокую точность и качество сглаживания [6]. Ключевой особенностью является использование kd-деревьев для эффективной организации обрезающих кривых, сгруппированных в «монотонные наборы»: последовательность связанных кривых одинаковой монотонности. Применение метода бисекции обеспечивает точное определение областей усечения, а ускоряющие структуры позволяют достичь высокой производительности.

Дальнейшая оптимизация была достигнута в работе [7], которая была направлена на минимизацию дорогостоящего поиска пересечения с кривыми и на расширение случаев быстрой классификации, не требующей сложных вычислений. Сохраняя точность рендеринга, данный подход в настоящее время является наиболее производительным решением для трассировки лучей на GPU, демонстрируя скорость обработки до нескольких сотен миллионов лучей в секунду для среднесложных сцен.

Работа [8] использует в качестве начального приближения пересечение с полигональной сеткой, которое может быть эффективно посчитано, используя существующие библиотеки и средства аппаратного ускорения трассировки лучей на CPU и на GPU. В этой работе использовалось пересечение с единичной поверхностью линзы, и не было реализовано усечение (тримминг). Мы считаем это перспективным направлением исследований, так как неочевидна его применимость к рендерингу объектов, состоящих из нескольких NURBS поверхностей, а также поверхностей с триммингом.

### Предложенный метод

Наша работа является развитием метода, предложенного в [7]. Этот метод состоит из трех основных этапов: (1) обход BVH-дерева; (2) поиск пересечения луча и Безье-поверхности методом Ньютона–Рафсона; (3) усечение (тримминг) поверхности с использованием двумерного kd-дерева. Предложенный метод модифицирует два последних шага. В действительности этот подход требует предобработку, так как алгоритмы вычисления точек и производных для В-сплайновых поверхностей  $S(u, v)$  обладают высокой вычислительной сложностью и требуют значительных объемов памяти [9], что делает их неэффективными для реализации на GPU. Поэтому перед построением BVH-дерева работа [7], как и наша работа, использует преобразование В-сплайновых поверхностей в поверхности Безье с помощью алгоритма вставки узлов [10].

#### 1. Метод стохастического выбора начального приближения

Метод Ньютона–Рафсона, несмотря на обеспечиваемую квадратичную скорость сходимости, критически зависит от выбора начального приближения и в общем случае не гарантирует сходимость даже при достаточно сильном разбиении геометрии BVH-деревом (хотя в этом случае вероятность появления артефактов уменьшается ценой повышения потребления памяти) [7]. Основная идея нашего метода заключается в том, чтобы варьировать начальное приближение на разных кадрах. Если хотя бы в одном кадре было получено ближайшее пересечение, дальнейшие вариации будут отброшены (рис. 1).

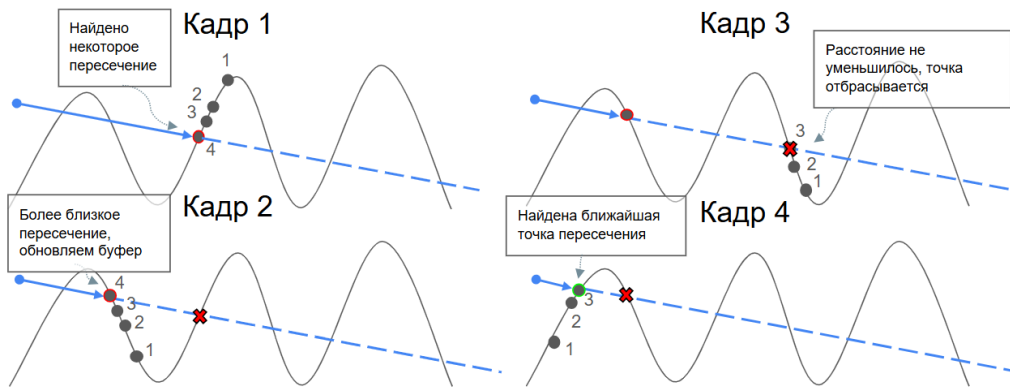


Рисунок 1. Иллюстрация идеи работы предложенного метода для нескольких кадров

Для формализации подхода введем унифицированное параметрическое пространство. Область определения каждого подучастка поверхности отобразим в единичный квадрат  $[0, 1]^2$ . Такая нормализация позволяет единообразно задавать начальные приближения для всех участков поверхности. Например, чтобы для всех участков начальные приближения были в их параметрических центрах, можно использовать глобально заданные в унифицированном пространстве значения  $(\hat{u}, \hat{v}) = (\frac{1}{2}, \frac{1}{2})$ . В таком случае для каждого участка начальное приближение вычисляется следующим образом:

$$(u_0, v_0) = \left( \frac{u_{min} + u_{max}}{2}, \frac{v_{min} + v_{max}}{2} \right). \quad (1)$$

В общем случае для глобально заданного значения  $(\hat{u}, \hat{v}) \in [0, 1]^2$  начальное приближение конкретного участка можно получить линейной интерполяцией граничных значений параметров:

$$(u_0, v_0) = (u_{min} + \hat{u} \cdot (u_{max} - u_{min}), v_{min} + \hat{v} \cdot (v_{max} - v_{min})). \quad (2)$$

Для первого кадра естественным выбором является использование параметрических центров: значения  $(\hat{u}_0, \hat{v}_0) = (\frac{1}{2}, \frac{1}{2})$ . Таким образом формируется базовое покрытие изображения. Для каждого следующего кадра начальное приближение выбирается случайно, используя равномерное распределение:

$$(\hat{u}_i, \hat{v}_i) \sim U([0, 1] \times [0, 1]), i \geq 1. \quad (3)$$



Остается задействовать информацию с предыдущих кадров. Для этого используется буфер глубины, в котором записаны значения расстояния до точки пересечения в каждом пикселе изображения. Для каждого пикселя в текущем кадре цвет изменяется только в том случае, если найденная точка пересечения ближе к камере, чем предыдущий результат. В результате чего в каждом пикселе будет рассматриваться ближайшая найденная на текущий момент времени точка пересечения. Таким образом, предложенный метод при неизменной позиции камеры с течением времени позволяет получить корректное изображение (рис. 2), даже если трассировка лучей не всегда гарантирует корректный выбор начального приближения.

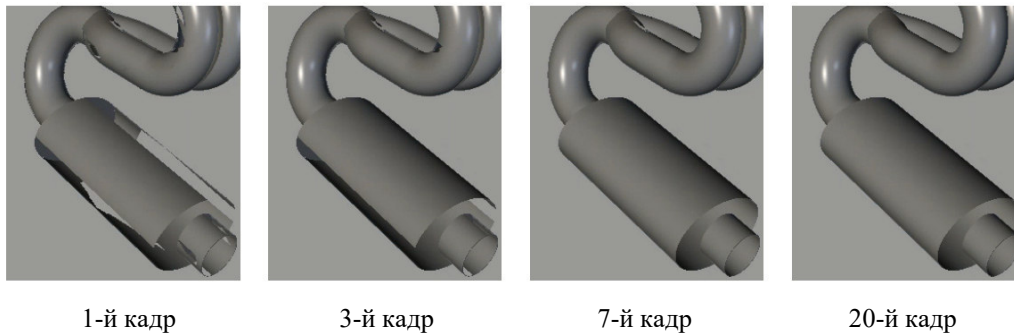


Рисунок 2. Демонстрация работы предложенного метода с первого по 20-й кадр

## 2. Метод усечения на основе двумерных SDF-текстур

SDF (англ. Signed Distance Field) представляет собой скалярное поле  $\Phi(x): \mathbb{R}^n \rightarrow \mathbb{R}$ , значение которого в точке по модулю равно расстоянию до границы объекта и положительно или отрицательно, если точка находится снаружи или внутри объекта соответственно. В компьютерной графике наиболее часто используются двумерные ( $n = 2$ ) и трехмерные ( $n = 3$ ) поля расстояний. Для практического применения поле расстояния часто дискретизируют и сохраняют в виде текстуры. Такой метод используется в [11] для рендеринга текста. В предложенном методе SDF-текстуры используются в рамках усечения поверхности, где SDF может служить универсальным представлением для описания сложных границ обрезки.

Генерация текстур обрезки на основе полей расстояний со знаком (SDF) начинается с создания высокоточной бинарной маски, которая служит основой для последующих вычислений. Затем вычисляется расстояние до ближайшего пиксела с другим значением маски (1, если текущий пиксел 0, или 0, если текущий пиксел 1). На финальном этапе обработки выполняется квантование расстояний в 8-битное представление с последующей битовой упаковкой в 32-битной текстуре с целью сокращения затрат памяти. Использование предложенного метода вносит незначительные изменения в алгоритм рендеринга: вместо запроса к kd-дереву над кривыми обрезки производится выборка из SDF-текстуры с использованием билинейной интерполяции (листинг 1). Данная техника позволяет хранить информацию об усеченной области в сжатом виде без существенной потери качества (рис. 3).

```

1 def isTrimmed(surface, uv):
2     trimKdTree ←
        surface.trimKdTree
3     return trimKdTree.query(uv)
4
```

```

1 def isTrimmed(surface, uv):
2     trimSDF ← surface.trimSDF
3     distance ← trimSDF.sampleBilinear(uv)
4     return distance < 0
5
```

Листинг 1: Сопоставление алгоритмов тримминга в базовом (слева, [7]) и предложенном (справа) методах. В предложенном методе вычислительно сложная операция поиска в kd-дереве заменяется на одну текстурную выборку с билинейной интерполяцией значений.

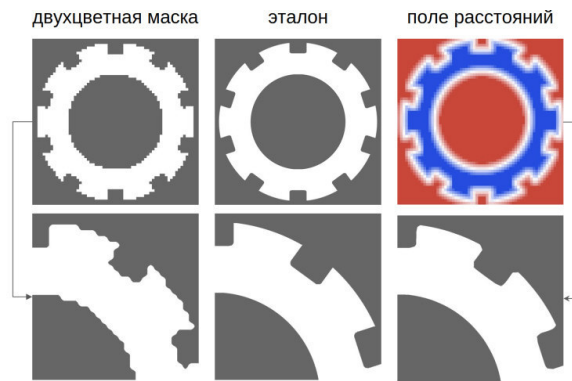


Рисунок 3. Сравнение представления трафарета через битовую маску и SDF-текстуру (64x64, 8 бит на пиксель) при одинаковом разрешении текстуры

### 2.1. Оптимальное распределение памяти для SDF-текстур

Наиболее важным параметром, определяющим качество усечения, является разрешение SDF-текстур. В силу ограниченности ресурсов необходимо распределить имеющуюся память под текстуры так, чтобы визуальное качество генерируемого изображения было оптимальным. Предложенный метод основывается на предположении, что большие поверхности будут занимать большую часть экранного пространства. Распределение памяти осуществляется по формуле

$$M_i = \left\lfloor M_{total} \times \frac{A_i}{\sum_{k=1}^n A_k} \right\rfloor, \quad (4)$$

где  $i = \overline{1 \dots n}$  – номер поверхности;

$M_{total}$  – общее количество выделенной памяти для SDF-текстур в байтах;

$A_k$  – оценка площади k-й поверхности;

$M_i$  – количество выделенной памяти для SDF-текстуры i-й поверхности в байтах.

Рассматриваются квадратные SDF-текстуры  $H_i \times H_i$ , размер  $H_i$  которых вычисляется по формуле

$$H_i = \lfloor \sqrt{M_i} \rfloor. \quad (5)$$

### 2.2. Визуализация ребер и границ

Использование SDF-текстур открывает возможность отображения ребер и границ объектов без существенных накладных расходов (рис. 4). Поскольку расстояние до границы уже записано в SDF-текстуре, для определения принадлежности точки к границе используется условие

$$p \in \delta\Omega \approx \Phi\Omega(p) < \epsilon, \quad (6)$$

где  $\epsilon$  – пороговое значение, определяющее толщину линии.

Для плавного перехода можно использовать интерполяцию:

$$color_{final} = \alpha \cdot color_{edge} + (1 - \alpha) \cdot color_{surface}, \quad (7)$$

где  $\alpha$  – функция от  $\Phi(p)$ , обеспечивающая гладкий переход.

Примеры использования SDF-текстур для реализации данных режимов отображения приведены на рисунке 4.

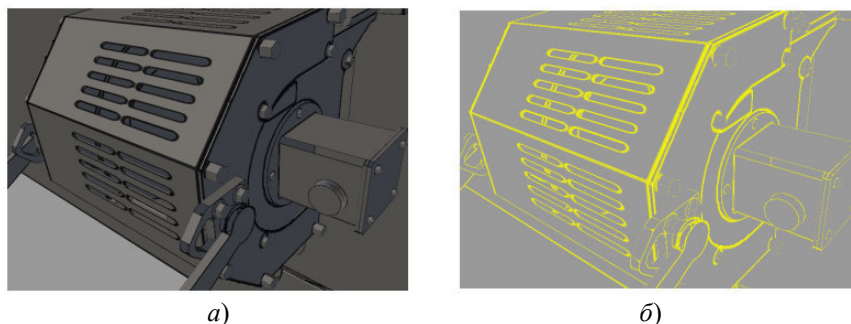


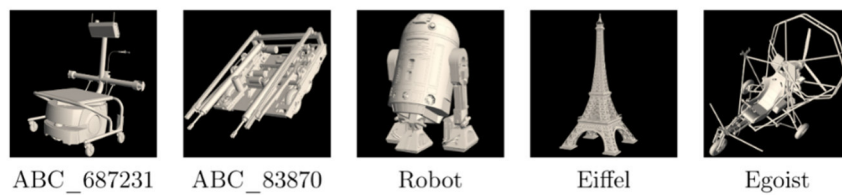
Рисунок 4. Демонстрация визуализации ребер с помощью предложенного метода:  
а – отображение с границами; б – только видимые границы

## Результаты и сравнение

Для оценки предложенных методов были выбраны две модели из открытого набора данных ABC [12] и три модели из работы [7], которые также находились в открытом доступе. Основные параметры используемых моделей представлены в таблице 1. В-сплайновые кривые и поверхности соответствуют исходным данным модели, а поверхности и монотонные кривые Безье – полученным результатам после преобразований.

Таблица 1. Характеристики тестовых моделей

Модель	NURBS Поверхности	NURBS Кривые	Поверхности Безье	Кривые Безье	Потребление памяти (МБ)
ABC_687231	27306	139526	99721	460252	116.3
ABC_83870	92458	498024	238374	787783	210.0
Robot	4292	21702	5786	30862	27.8
Eiffel	14774	132926	16626	157914	38.9
Egoist	22363	113509	102233	328927	126.1



### 1. Метод выбора начального приближения

Для метода стохастического выбора начального приближения численный эксперимент показал улучшение точности со временем (табл. 2, рис. 2). Кроме того, он показал, что в среднем 10 кадров достаточно, чтобы убрать артефакты, существенно ухудшающие качество изображения (средний PSNR на 10-м кадре  $>35$  для всех моделей). Таким образом, в случае визуализации реального времени (более 25 кадров в секунду) данный процесс будет занимать доли секунды.

### 2. Метод тримминга поверхностей при помощи SDF-текстур

В качестве эталона рассматриваются изображения, получаемые использованием базового метода на основе kd-деревьев с высоким уровнем разбиения поверхностей на участки малой кривизны. Мы исследовали, как меняется точность предложенного метода при различном разрешении SDF-текстур (табл. 3).

По полученным в ходе экспериментов данным можно сказать, что SDF-текстуры демонстрируют качественный прирост производительности по сравнению с kd-деревом при сохранении приемлемого качества визуализации (PSNR  $>30$  на всех дистанциях) даже при двукратном уменьшении объема используемой памяти. Рисунок 6 демонстрирует визуальное сопоставление.

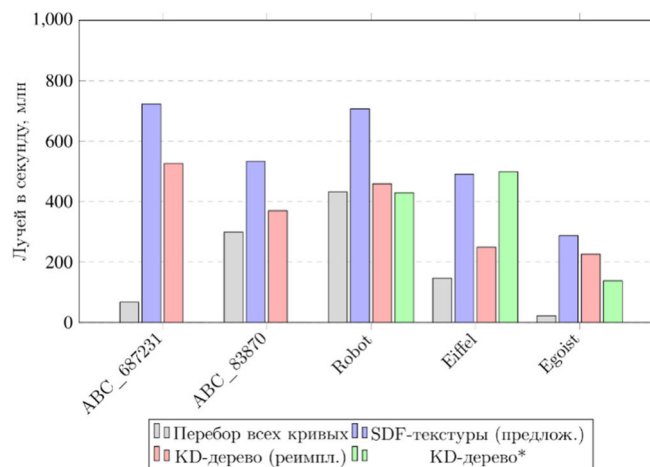


Рисунок 5. Сравнение производительности базового и предложенного методов.

Графический ускоритель аналогичен тому, что использовался в статье [7].

\*Значения, продемонстрированные в статье [7]

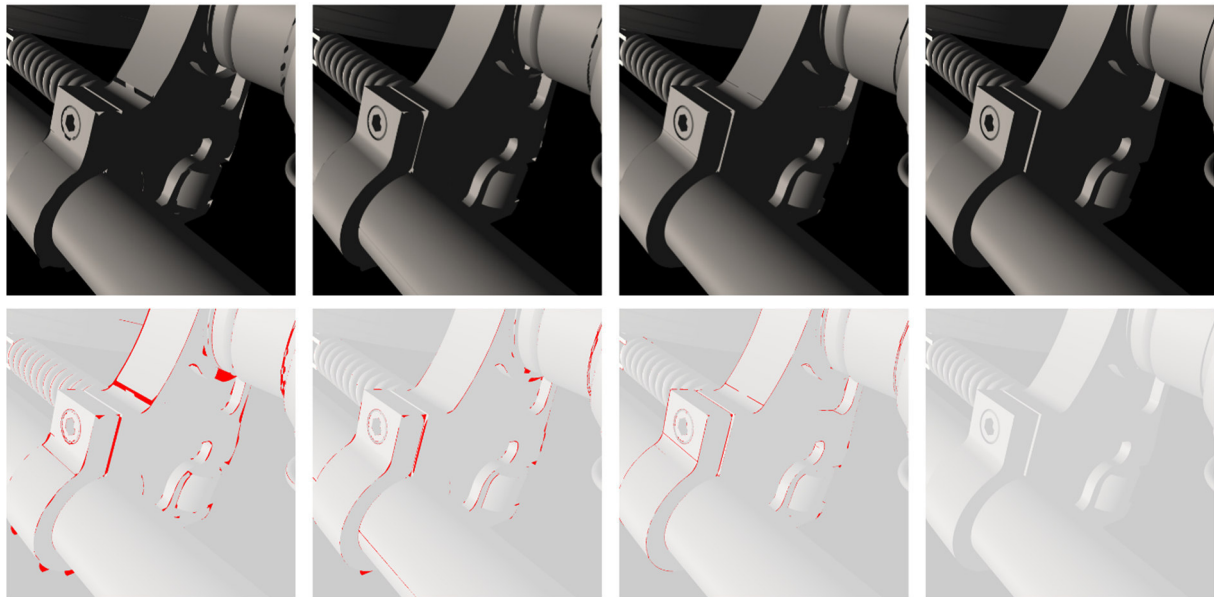


Рисунок 6. Разница изображений, генерируемых при помощи SDF-текстур, с эталоном при различном потреблении памяти.

Таблица 2. Качество изображений на 10-м кадре с применением алгоритма подавления артефактов

Модель	кадр 1	кадр 10
ABC_687231	20.3	40.2
ABC_83870	28.7	39.9
Robot	22.5	38.4
Eiffel	25.9	35.7
Egoist	21.1	45.5

Таблица 3. Точность визуализации при помощи SDF-текстур по метрике PSNR при различных затратах памяти на данные для тримминга (относительно базового метода)

Модель	50%	100%	200%	400%
ABC_687231	46.4	46.9	49.2	51.4
ABC_83870	32.2	36.5	38.1	41.7
Robot	34.2	39.4	40.7	44.8
Eiffel	32.0	32.6	37.0	37.2
Egoist	43.5	45.5	48.2	51.0

### Заключение

В работе предложены два метода улучшения рендеринга В-сплайновых поверхностей. Метод стохастического выбора начального приближения эффективно устраняет артефакты алгоритма Ньютона в течение нескольких кадров при статичной камере. Метод на основе SDF-текстур демонстрирует значительное ускорение рендеринга (в 1.5-2 раза) при регулируемых уровне качества и потреблении памяти, а также дополнительно предоставляет возможность рендеринга рёбер объектов. Оба метода показывают хорошую эффективность при работе на системах с ограниченными ресурсами.

Таким образом, предложенные методы позволяют повысить точность и скорость визуализации сложных NURBS-моделей на ограниченных по ресурсам системах, что делает их перспективными для применения в системах интерактивного предпросмотра и CAD-приложениях.

Для визуализации в условиях изменяющейся камеры или движения моделей перспективным направлением представляется использование репроекции данных с предыдущих кадров для инициализации поиска пересечений в новых условиях наблюдения. Альтернативный подход может заключаться в комбинации попиксельного стохастического выбора начальных приближений с алгоритмами шумоподавления. Реализация указанных улучшений позволит расширить область применения предложенных методов.

### Список литературы

1. "Opencascade3d", 2025. Available: <https://dev.opencascade.org/>. [Accessed 14 5 2025].
2. R. Xiong, Y. Lu, C. Chen, J. Zhu, Y. Zeng and L. Liu, "Eter: Elastic tessellation for real-time pixel-accurate rendering of large-scale nurbs models", *ACM Transactions on Graphics*, No. 2023, pp. 1-13.
3. T. Nishita, T. W. Sederberg and M. Kakimoto, "Ray tracing trimmed rational surface patches", *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, p. 337–345, 1990.

4. W. Martin, E. Cohen, R. Fish and P. Shirley, "Practical ray tracing of trimmed nurbs surfaces", *Journal of Graphics Tools*, p. 27–52, 2000.
5. A. Efremov, V. Havran and H.-P. Seidel, "Robust and numerically stable bezier clipping method for ray tracing nurbs surfaces", *Proceedings of the 21st Spring Conference on Computer Graphics*.
6. A. Schollmeyer and B. Froehlich, "Efficient and anti-aliased trimming for rendering large nurbs models", *IEEE Transactions on Visualization and Computer Graphics*, p. 1489–1498, 2019.
7. J. Sloup and V. Havran, "Optimizing ray tracing of trimmed nurbs surfaces on the GPU", *Computer Graphics Forum*, p. 161–172, 2021.
8. D. Zhdanov, I. Potemin and A. Zhdanov, "Using embree technologies for ray tracing in optical systems with free-form surfaces", *Proceedings of the International Conference on Computer Graphics and Vision "Graphicon"*, vol. 33, p. 97–107, 2023.
9. K. Jankauskas, "Time-efficient nurbs curve evaluation algorithms", *Information Technologies 2010: proceedings of the 16th international conference on Information and Software Technologies*, 2010.
10. L. Piegl and B. Tiller, *The NURBS Book*, 2 ed., Springer, 1997.
11. C. Green, "Improved alpha-tested magnification for vector textures and special effects," *ACM SIGGRAPH 2007 courses*, 2007.
12. S. Koch, A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin and D. Panozzo, "ABC: A big cad model dataset for geometric deep learning", *The IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

## Метод вычисления полиномиального разложения для высокоточной трассировки лучей в оптических системах

С. В. Ершов<sup>1</sup>, В. В. Санжаров<sup>2</sup>, В. А. Фролов<sup>1,2</sup>, В. А. Галактионов<sup>1</sup>

<sup>1</sup>Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

<sup>2</sup>Институт перспективных исследований проблем искусственного интеллекта  
и интеллектуальных систем МГУ им. М.В. Ломоносова, Москва, Россия

**Аннотация.** В работе рассматривается проблема точности суммирования полиномов, используемых в оптике для моделирования асферических поверхностей. При увеличении порядка полиномиальных членов, а также при больших значениях коэффициентов могут возникать ошибки округления, связанные с представлением чисел с плавающей точкой. Накопление таких ошибок, в свою очередь, может приводить к сложностям при использовании численных методов, таких как, например, метод Ньютона в задаче трассировки лучей. Предлагается метод, использующий разложение в ряд Тейлора в заранее выбранных точках моделируемой поверхности. Разработанный метод позволяет значительно уменьшить ошибки округления при вычислении суммы полиномиальных членов и тем самым повысить точность расчетов точки пересечения луча и асферической поверхности.

**Ключевые слова:** оптика, трассировка лучей, асферические поверхности

## Polynomial expansion computation method for high-precision ray tracing in optical systems

S. V. Ershov<sup>1</sup>, V. V. Sanzharov<sup>2</sup>, V.A. Frolov<sup>1,2</sup>, V.A. Galaktionov<sup>1</sup>

<sup>1</sup>Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

<sup>2</sup>IAI Moscow State University, Moscow, Russia

**Abstract.** In this paper we consider the problem of accuracy of summation of polynomials used in optics for modeling aspherical surfaces. When the order of polynomial terms increases, as well as at large values of coefficients, rounding errors associated with the representation of floating-point numbers may occur. The accumulation of such errors, in turn, can lead to difficulties in using numerical methods such as Newton's method, for example, in case of ray tracing of such surfaces. We propose a method that uses Taylor series expansion at pre-selected points of the modeled surface. The developed method allows us to significantly reduce round-off errors when calculating the sum of polynomial terms and thereby improve the accuracy of calculations of the intersection point of the ray and the aspherical surface.

**Keywords:** optics, ray tracing, aspheric surfaces

### Введение

Один из распространенных способов представления асферических поверхностей в оптических системах представляет собой функцию, содержащую сумму полиномов [Sch19]. Например,

$$f(r^2) = \frac{cr^2}{1 + \sqrt{1 - (1 + \kappa)c^2r^2}} + \sum_{m=1}^N a_m \left(\frac{r}{R}\right)^{2m} \quad (1)$$

содержит полиномиальную часть

$$p(\rho^2) = \sum_{m=1}^N a_m \cdot (\rho^2)^m, \quad (2)$$

где  $\rho \equiv \frac{r}{R}$ .

Максимальная степень полинома  $2N$  обычно никак не ограничена, и программное обеспечение для проектирования оптических систем позволяет задавать её произвольным образом. Однако увеличение степени может вызывать ошибки вычислений, имеющие осциллирующий характер и сходные с так называемым феноменом Рунге – при интерполяции функции полиномом некоторой степени полином начинает расходиться вне заданного интервала, при этом расхождение все больше с увеличением степени полинома.

Добавление членов высокого порядка может вызвать осцилляцию в вычисленных значениях, которая не пропадает при добавлении новых членов [Rayces06]. Эти ошибки суммирования во многом возникают из-за вычислений с конечной точностью и плохо обусловленной природы мономиальных базисов. Это ограничивает точность при прямом вычислении полиномов.



Значения коэффициентов  $a_m$  могут иметь высокий порядок, при этом значение всей суммы может быть невелико. На рисунке 1 представлен пример поверхности (первая поверхность слева) с коэффициентами  $a_m$  в таблице 1.

Таблица 1. Пример значений коэффициентов полинома

$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$
-74.856..	863.472..	-7975.92..	55309.02..	-284929.33..	1085886.42..	-3062225.10..
$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$	$a_{14}$	$a_{15}$
6386243.47..	-9791775.99..	10874632.94..	-8496288.36..	4422713.64..	-1375665.94..	193280.54..

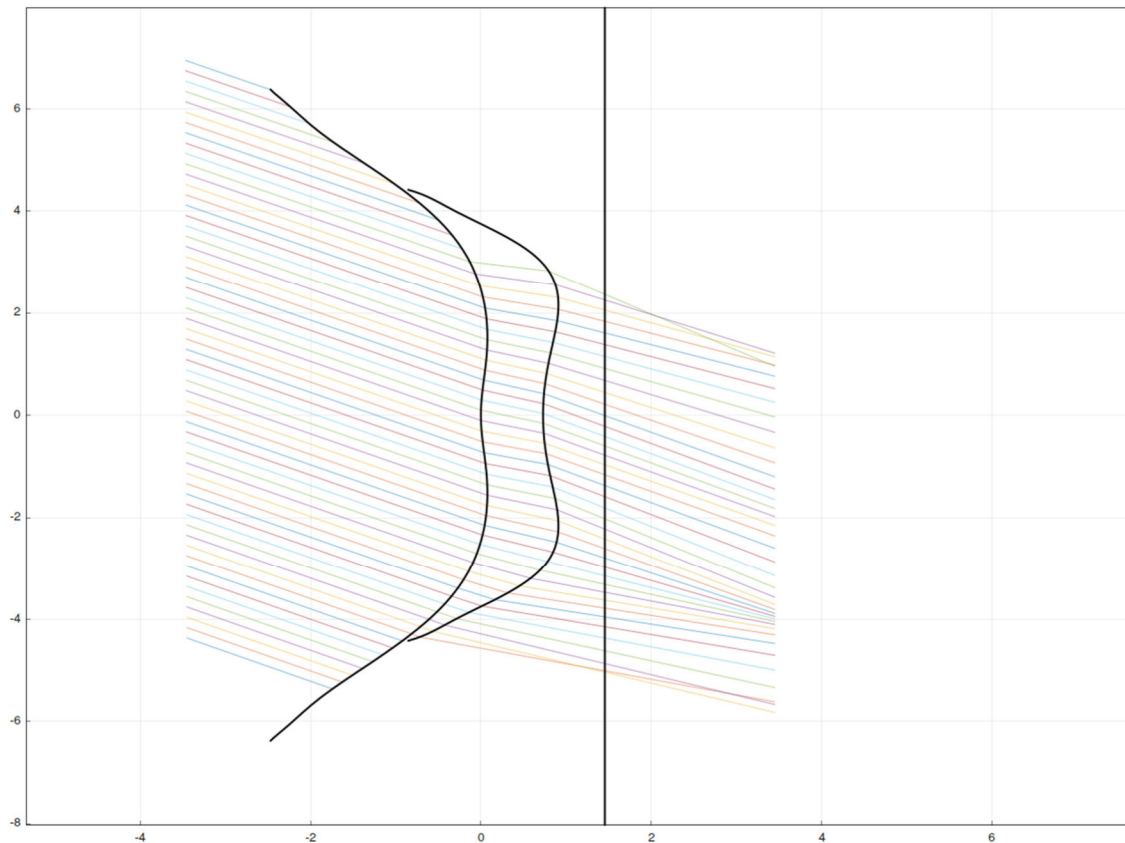


Рисунок 1. Пример линзы с асферической поверхностью с большими значениями коэффициентов полинома (таблица 1)

Вычисления с плавающей точкой двойной точности (т.е. с использованием типа double в языке C) дают точность порядка  $10^{-15}$ . То есть для (2) можно записать

$$p_{CPU}(\rho^2) = \sum_{m=1}^N (1 + \epsilon_m) a_m \cdot (\rho^2)^m = p(\rho^2) + \sum_{m=1}^N \epsilon_m a_m \cdot (\rho^2)^m,$$

где значение  $\epsilon_m \sim 10^{-15}$ . При этом значение  $\epsilon_m$  может быстро и непредсказуемо изменяться для разных  $\rho$  and  $m$ . То есть значения  $\epsilon_m$  можно рассматривать как условно «случайные», которые при этом умножаются на большие значения коэффициентов. Например, для  $m = 9$  (табл. 1)  $a_m$  имеет порядок  $10^7$ , что приводит к тому, что отклонение вычисленного значения суммы от истинного  $p_{CPU} - p$  (рис. 2) будет иметь порядок  $10^7 \cdot 10^{-15} = 10^{-8}$ . С увеличением значения  $\rho$  это отклонение будет увеличиваться, т.е. ошибка вычислений ближе к краям оптической поверхности.

Вычисление пересечения луча с поверхностью, заданной выражением (1), производится с помощью метода Ньютона. Ошибка суммирования приводит к тому, что итерационный процесс при достижении точности решения порядка ошибки суммирования начинает осциллировать около корня. В зависимости от значений коэффициентов  $a_m$  для моделируемой оптической поверхности и от требуемой точности вычисления пересечений это может привести к тому, что решение не будет найдено.

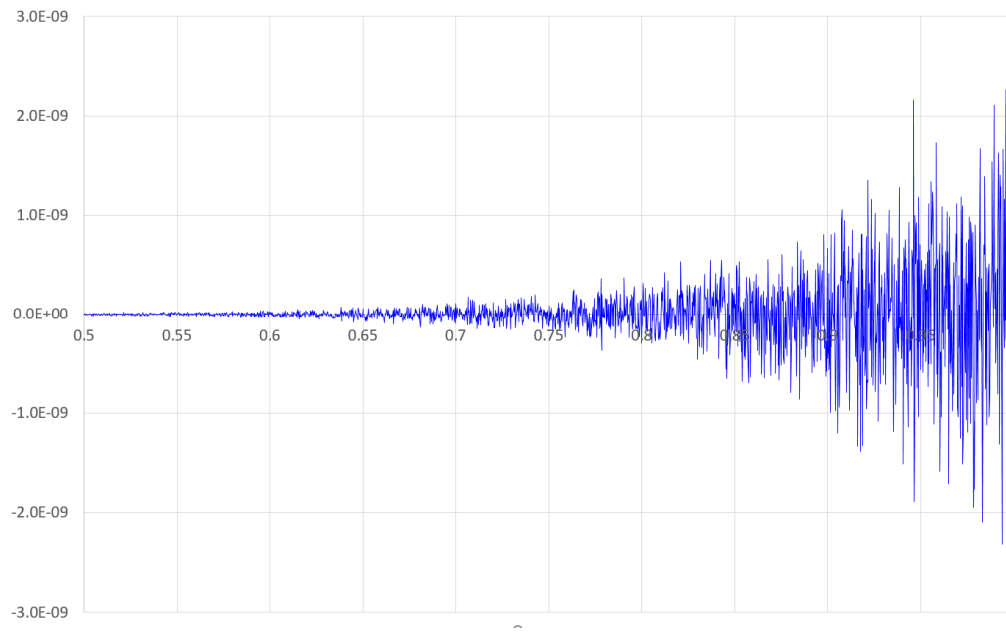


Рисунок 2. Разница между значением суммы (2), вычисленным с двойной точностью (тип double) и с высокой точностью (порядка 250 знаков после запятой) с помощью библиотеки qdpr [QDP]. По вертикальной оси показана ошибка суммирования, по горизонтальной оси отложено значение  $\rho^2$

Рассмотрим существующие подходы к проблеме точности вычислений в трассировке лучей в оптических системах.

#### *Существующие решения*

Один из подходов состоит в том, чтобы изменить полиномиальный базис или схему вычислений. В работе [Forbes10] предложен подход с использованием рекуррентных формул для ортогональных полиномов, позволяющий выполнять стабильные вычисления для высоких степеней ортогональных полиномов. Такой прием позволяет значительно уменьшить проблемы, возникающие при наивном суммировании. За счет этого становится возможным использование «обычной» арифметики с одинарной и двойной точностью, имеющей аппаратную поддержку во многих вычислительных системах, для вычислений с большими степенями полиномиальных представлений оптических поверхностей. Недостатком данного подхода является то, что он применим только для ортогональных полиномов, а также увеличение сложности программной реализации по сравнению с обычным суммированием.

Также следует упомянуть об общем решении для проблем с точностью при суммировании на основе использования алгоритмов компенсационного суммирования [Kahan65], которые позволяют уменьшить накопление погрешности при суммировании чисел с плавающей запятой. Но компенсационное суммирование не поможет при суммировании слагаемых значительно различающихся порядков, что вполне возможно для оптических систем. Также такие алгоритмы обычно увеличивают количество арифметических операций, в частности [Kahan65] требует в 4 раза большего числа вычислений. В работе [Bakas24] предлагается использовать разложения в ряд Тейлора в комбинации с типами данных с высокой точностью (авторы рассматривают 2000-битные типы) для устранения вычислительных ошибок, в том числе подобных феномену Рунге. Однако задача трассировки лучей в оптических системах может требовать вычисления миллионов пересечений лучей с поверхностью. Выполнение таких расчетов для 2000-битных типов потребует очень больших вычислительных ресурсов.

Радикальным решением является переход к другому представлению поверхностей, в частности использование сплайнов и кусочной интерполяции. Например, в работе [Morita10] предлагается использовать сетку из квадратичных интерполянтов с вектором нормали, так называемых «патчей», для моделирования асферической поверхности по частям. Метод позволяет получить высокую точность и скорость трассировки лучей. Проблемой как этого подхода, так и других подобных решений, являются сложность построения сетки, трудности при моделировании сложных форм и

производстве полученных линз. Кроме того, альтернативные представления поверхности также требуют вычислений с плавающей точкой двойной точности. В работе [Kimura19] используется схожий подход на основе патчей, для которого приводится сравнение вычислений с плавающей точкой одинарной и двойной точности в оптических расчетах на графическом процессоре (GPU). Эксперименты авторов показали, что только вычисления с двойной точностью позволяют достичь достаточно малого уровня ошибок. Время трассировки при этом возросло более чем в два раза.

В работе [Wu23] предложена модель ошибок округления на каждом шаге трассировки лучей в оптических системах (пересечение с поверхностью и вычисление нового направления луча) на основе IEEE-754 арифметики. Из этой модели авторы доказывают верхние границы («границы ошибок») для трассировки лучей.

Основываясь на этом анализе, в следующей работе [Wu24] авторы предлагают приемы компенсации этих ошибок:

- аппроксимация каждой поверхности «виртуальной касательно плоскостью» при вычислении пересечения;
- ре-проекция лучей в новую систему координат после каждого пересечения;
- вычисление корней для квадратных уравнений через формулу из метода Мюллера;
- нормализация вектора направления после каждого шага;
- вычисления в локальных координатах относительно точки пересечения.

Эти приемы позволяют «сбросить» накопление ошибок на каждом пересечении. На ряде экспериментов в [Wu24] демонстрируется более высокая точность вычислений, чем в коммерческих программных продуктах оптического моделирования. Основным недостатком этой работы является то, что приведенные авторами математические выкладки ориентированы на сферические и конические поверхности. Расширение данного подхода на асферические поверхности и поверхности свободной формы потребует построение другой, ещё более сложной модели. Кроме того, предложенные в [Wu24] способы компенсации ошибок добавляют дополнительные вычисления на каждом шаге трассировки лучей, что может сказаться на производительности. Авторы никак не анализируют затраты на дополнительные вычисления.

В применении трассировки лучей в других областях также исследовалась проблема точности вычислений с плавающей точкой. В работе [Dammertz06] предлагает поиск пересечения в локальных координатах объекта в комбинации с подразбиением ограничивающего параллелепипеда объекта для поиска трехмерного интервала, в котором гарантированно содержится пересечение. Поскольку найденный интервал представлен в локальных координатах объекта, уменьшаются ошибки вычислений с плавающей точкой. В [PBRT3] предлагается использовать специальный тип данных EFloat, который явно хранит интервалы для накопленных ошибок вычислений.

Таким образом, существующие подходы ориентированы на ортогональные полиномы [Forbes10], патчи [Morita10, Kimura19], треугольные сетки [PBRT3], отличные от рассматриваемого представления в нашей работе на основе выражения (1). Работы [Wu23, Wu24] рассматривают более простые поверхности и для применения к асферам требуют повторения математического вывода. Подходы на основе компенсационного суммирования [Kahan65] не работают, если слагаемые имеют сильно отличающийся порядок. Поэтому представляет интерес решение, которое бы позволило бороться с ошибками вычислений с плавающей точкой для асферических поверхностей, заданных полиномами, подобными (1).

## Предлагаемое решение

### 1. Основная идея

Описанная проблема с суммированием полиномиальной части уравнения (1) возникает при независимом вычислении суммы для близких значений аргумента. Предположим, что после некоторого числа итераций метода Ньютона была достигнута точность порядка  $10^{-5}$ . Мы можем заменить вычисление полинома на вычисление его разложения в ряд Тейлора в текущем значении точки  $\rho^2$  на всех последующих итерациях. В этом случае слагаемые суммы будут достаточно малыми и не будут вызывать осциллирующего поведения. Ясно, что при чрезмерном упрощении, например линеаризации, точность будет недостаточно высокой, поэтому лучше использовать больше членов ряда Тейлора.

Для представленного на рисунке 1 и в таблице 1 примера задано 14 коэффициентов полинома, начиная с 4 степени. Выполним его разложение в ряд Тейлора также с 14 членами, используя биномиальную формулу для вычисления новых коэффициентов:

$$p = \sum_{k=1}^{14} b_k \cdot (\rho^2 - \rho_0^2)^k, \quad (3)$$

$$b_k = \sum_{m=k}^{14} \frac{m!}{k!(m-k)!} (\rho_0^2)^{m-k} a_m, \quad (4)$$

где  $\rho_0^2$  – новая «базовая точка» или центр разложения.

При вычислении исходной суммы (2) около этой базовой точки, т.е. для  $\rho^2$ , достаточно близких к  $\rho_0^2$ , коэффициенты в формуле (3) будут быстро уменьшаться, т.е. по факту мы используем только несколько первых членов суммы. Это позволяет устранить осцилляции, вызываемые ошибками округления.

Важно отметить, что полином для близких значений аргумента  $\rho^2$  должен вычисляться для тех же самых значений  $b_k$ , так как в формулу (4) входят исходные потенциально большие значения  $a_m$  и поэтому  $b_k = b_k(\rho_0^2)$  также может осциллировать.

Алгоритм для вычисления коэффициентов  $b_k$  достаточно прост и отражает формулу (4).

---

**Алгоритм 1.** *Rebase*( $N, a, \rho_0$ ), вычисление коэффициентов  $b_k$

---

**Вход:**

$N$  – число коэффициентов,  $a$  – массив исходных коэффициентов полинома,  $\rho_0$  – центр разложения

---

**Выход:**

$b$  – массив новых коэффициентов

---

```

1   $k \leftarrow 0$ 
2  while  $k < N$  do
3       $sum \leftarrow a[k], prod \leftarrow 1, m \leftarrow 1$ 
4      while  $m < N - k$  do
5           $prod \leftarrow prod * \left( \rho_0^2 * \frac{(m+k)}{m} \right)$ 
10          $sum \leftarrow sum + a[1+k] * prod$ 
           $m \leftarrow m + 1$ 
11     end
12      $b[k] = sum$ 
13      $k \leftarrow k + 1$ 
19 end
20 return  $b$ 
```

---

Основная идея предлагаемого решения состоит в том, чтобы заменить непосредственное вычисление суммы полиномиальных коэффициентов (2) на вычисление разложения (3). То есть, например, при решении задачи трассировки лучей с помощью метода Ньютона, для вычисления значения функции поверхности, будет использоваться разложение (3) с заранее вычисленными по алгоритму 1 коэффициентами  $b_k$  для «базовой точки»  $\rho_0^2$ , ближайшей к текущему аргументу  $\rho^2$ .

## 2. Множество базовых точек

Как было сказано, необходимо вычислять сумму (3) в окрестности базовой точки  $\rho_0^2$ . Рассматриваемая поверхность, описываемая уравнением (1), обладает радиальной симметрией относительно оптической оси, поэтому удобно заранее вычислить коэффициенты  $b_k$  для некоторого предопределенного множества базовых точек, например:

$$\rho_0^2 = \{0, 0.1, 0.2, \dots, 0.9, 1.0\}$$

Тогда в процессе трассировки лучей, когда нам необходимо вычислить значение полинома в  $\rho^2$ , мы сначала находим ближайшее значение  $\rho_0^2$ . Затем берем соответствующие заранее рассчитанные значения коэффициентов  $b_k$  и подставляем в выражение (3). При небольших изменениях  $\rho^2$  на последующих итерациях метода Ньютона ближайшее значение  $\rho_0^2$  останется неизменным. Смена

коэффициентов  $b_k$  может потребоваться только в случае, если  $\rho^2$  перейдет среднее значение между двумя базовыми точками. Но если коэффициенты были рассчитаны с использованием арифметики с большой точностью (достаточно «четверной» точности, т.е. 128-битного типа данных, например, на основе библиотеки [QDP]), то это не приведет к каким-либо проблемам.

На рисунке 3 показана разница между вычислением исходного полинома (2) с использованием арифметики большой точности и его разложения (3) для коэффициентов  $b_k$ , также вычисленных с использованием арифметики большой точности.

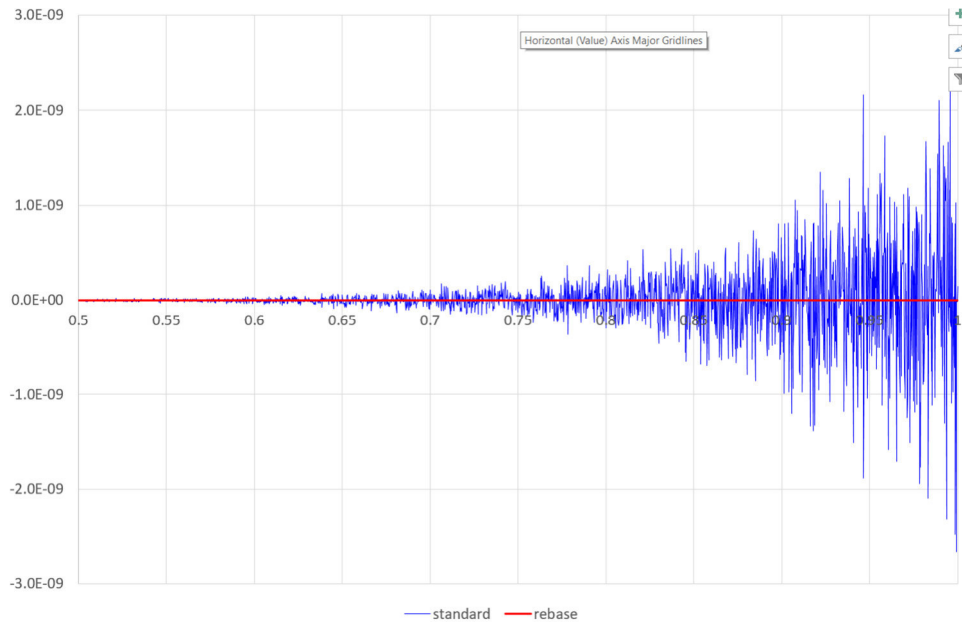


Рисунок 3. Разница между значениями исходного полинома (2), вычисленного со 128-битной арифметикой, и его разложения (3) с коэффициентами, также вычисленными со 128-битной арифметикой. Разница показана красной линией, синим показано значение исходного полинома (2), вычисленного с обычной двойной точностью

Если же коэффициенты  $b_k$  вычислить с обычной 64-битной арифметикой (т.е. с использованием типа double), то можно заметить скачки для разницы, которые происходят при значениях  $\rho_0^2$ , располагающихся между предрасчитанными значениями, т.е. для  $\rho_0^2 = 0.05, 0.15, 0.25$  и т.д. (рис. 4).

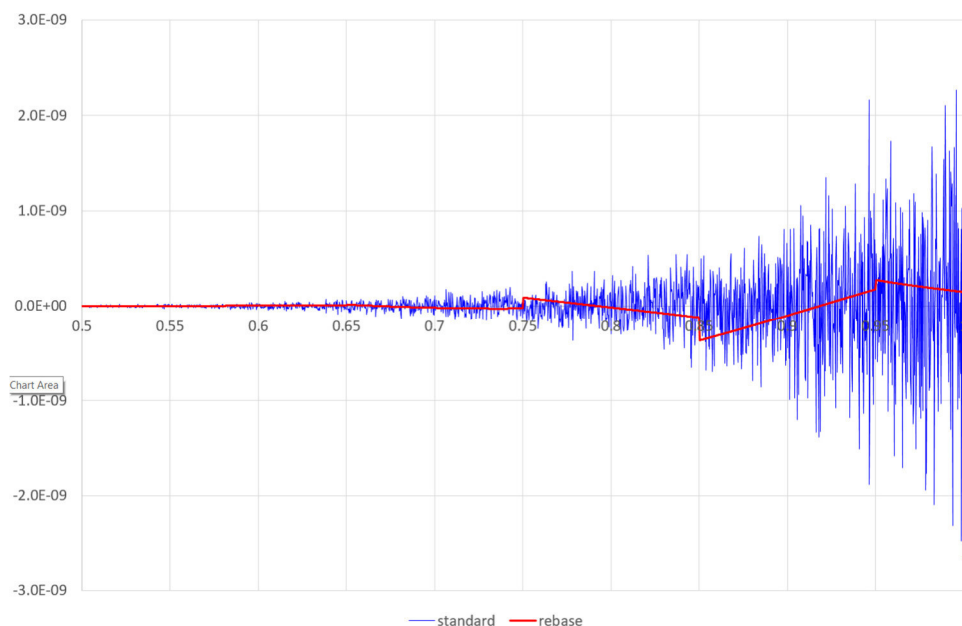


Рисунок 4. Разница между значениями исходного полинома (2), вычисленного со 128-битной арифметикой, и его разложения (3) с коэффициентами, также вычисленными с 64-битной арифметикой. Разница показана красной линией, синим показано значение исходного полинома (2), вычисленного с обычной двойной точностью

В данном примере линия разницы (рис. 4) состоит из сегментов прямых. Мы можем вычислить величину ошибки на границах этих сегментов и выполнить вычитание линейной функции, соответствующей тому или иному сегменту. Но в этом случае потребуется вычислить эталонное значение (2) с высокой точностью, хоть и в небольшом числе точек.

Другой вариант состоит в том, чтобы вычислить (3) для двух значений базовой точки  $\rho_0^2$  на концах интервала, в который попадает текущее значение  $\rho^2$  и выполнить линейную интерполяцию между этими значениями, чтобы сделать функцию ошибки гладкой (рис. 5)

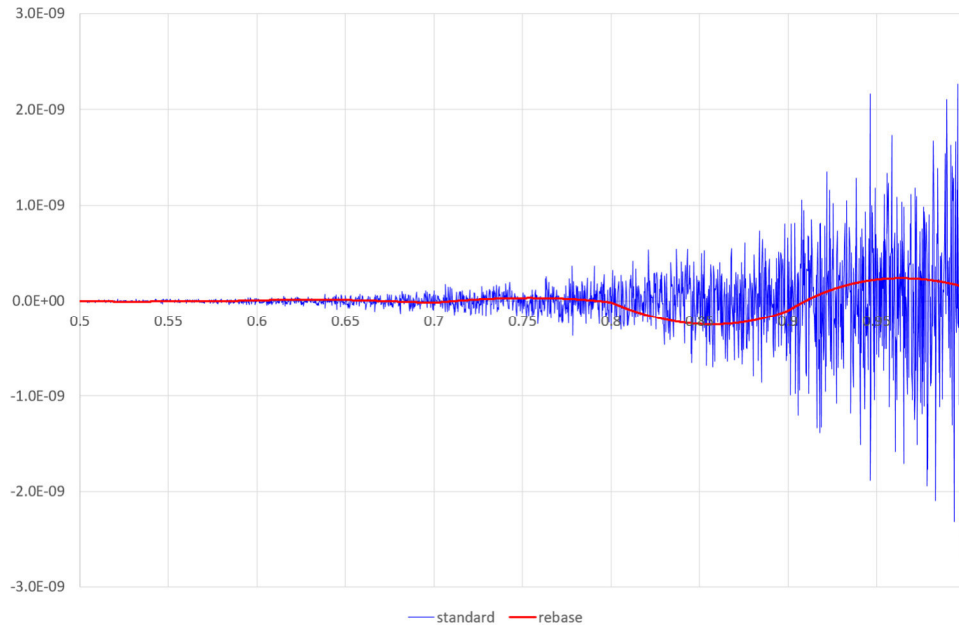


Рисунок 5. Разница между значениями исходного полинома (2), вычисленного со 128-битной арифметикой, и его разложения (3) с коэффициентами, также вычисленными с 64-битной арифметикой с интерполяцией между значениями базовых точек. Разница показана красной линией, синим показано значение исходного полинома (2), вычисленного с обычной двойной точностью

### 3. Предлагаемый метод

Основу предлагаемого метода составляют выражения (3), (4) и предварительный расчет коэффициентов для множества базовых точек. В экспериментах мы использовали значения для 11 точек  $\rho_0^2 = 0, 0,1, 0,2, \dots, 0,9, 1,0, 1,1$ . Принципиально метод может работать и для другого числа точек. Увеличение числа точек не приносит значимого улучшения результатов.

Таким образом, предлагаемый метод включает в себя два этапа:

1. Предварительный расчет при загрузке или обновлении поверхности:

а) задать множество базовых точек  $\rho_0^2$ ;

б) вычислить значения коэффициентов  $b_k$  для каждой базовой точки из множества с помощью алгоритма 1;

2. Основной этап при поиске пересечения или других вычислений, требующих значений функции поверхности:

а) найти ближайшее значение базовой точки  $\rho_0^2$  для требуемого значения  $\rho^2$ ;

б) используя значения коэффициентов  $b_k$  для найденной базовой точки, вычислить значение полинома в базовой точке с помощью формулы (3).

На рисунке 6 для поверхности, задаваемой коэффициентами из таблицы 1, показаны её сечение (синяя кривая), базовые точки  $\rho_0^2$  (синие маркеры), в которых выполнен предрасчет коэффициентов  $b_k$ . Для пучка лучей, идущего под углом в  $30^\circ$  к оптической оси, показаны найденные точки пересечения (красные треугольники).



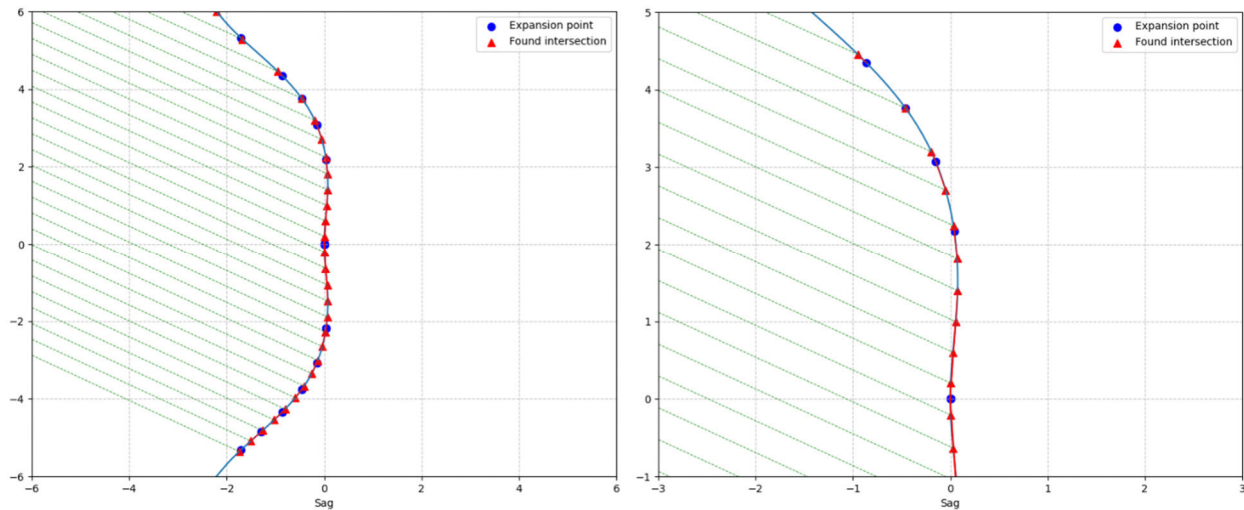


Рисунок 6. Визуализация поверхности, базовых точек разложения  $\rho_0^2$  (синие маркеры) и найденных пересечений (красные треугольники). Справа показан увеличенный участок; найденные пересечения и точки разложения, которые использовались для их расчета, соединены красными линиями

### Результаты экспериментов

Мы протестировали предложенный метод в задаче поиска пересечения с асферическими поверхностями вида (1). Для этого в методе Ньютона для вычисления значения функции поверхности мы использовали разложение (3) согласно описанию в предыдущем разделе.

В первом эксперименте были проведены расчеты для поверхности с большими значениями коэффициентов (рис. 1, табл. 1). Выполнялась трассировка пучка из  $\sim 1$  миллиона лучей (1024 на 1024), в качестве целевого значения точности пересечения  $acc_{target}$  было задано  $10^{-9}$  для суммирования полиномов в арифметике с двойной точностью (double) и  $10^{-6}$  с «обычной» плавающей точкой (float). Значение точности пересечения оценивалось как разница между точкой на луче  $\bar{p} = \bar{o} + \bar{d}t$  для найденного значения  $t$  и значением функции поверхности (1) для этой же точки  $f(p.x * p.x + p.y * p.y)$ .

Для лучей, у которых точка пересечения попадала внутрь апертуры, а точность пересечения  $acc$  лежала в интервале  $acc_{target} < acc < 10^{-5}$  для арифметики double и  $acc_{target} < acc < 10^{-4}$  для float, считалось, что пересечение найдено с низкой точностью. Для всех лучей, для которых были найдены пересечения, оценивалась средняя точность найденного пересечения. Также среднее число итераций метода Ньютона до достижения условия останова – достигнута требуемая точность или максимальное число итераций (равное 10). Результаты экспериментов приведены в таблицах 2 и 3. Замеры времени вычислений производились на CPU Ryzen 9 5950X.

Таблица 2. Замеры для пучка  $0^\circ$ , 1024 на 1024 луча

	Исходный полином		Предложенное разложение, 11 точек $\rho_0^2$		Предложенное разложение, 5 точек $\rho_0^2$		Предложенное разложение, 21 точка $\rho_0^2$	
	double	float	double	float	double	float	double	float
Число лучей с низкой точностью	4169	124752	0	0	0	782	0	0
Среднее значение ошибки для лучей с низкой точностью	1.26e-09	2.16e-05	-	-	-	1.05e-06	-	-
Среднее значение ошибки для всех лучей	5.24e-11	1.15e-05	2.62e-17	1.29e-07	2.96e-17	1.67e-07	2.56e-17	1.23e-07
Среднее число итераций метода Ньютона	2.91	4.96	2.87	2.61	2.87	2.61	2.87	2.59
Время вычислений, мс	59	64	59	55	57	55	59	56

Таблица 3. Замеры для пучка 30°, 1024 на 1024 луча

	Исходный полином		Предложенное разложение, 11 точек $\rho_0^2$		Предложенное разложение, 5 точек $\rho_0^2$		Предложенное разложение, 21 точка $\rho_0^2$	
	double	float	double	float	double	float	double	float
Число лучей с низкой точностью	3086	129831	0	12	0	210	0	0
Среднее значение ошибки для лучей с низкой точностью	1.27e-09	2.28e-05	-	1.02e-06	-	1.05e-06	-	-
Среднее значение ошибки для всех лучей	4.29e-11	1.14e-05	3.17e-16	1.18e-07	3.26e-16	1.56e-07	3.21e-16	1.19e-07
Среднее число итераций метода Ньютона	3.88	6.45	3.81	3.65	3.83	3.65	3.82	3.69
Время вычислений, мс	60	70	61	61	61	61	65	62

Из результатов видно, что предложенный метод позволяет достигнуть требуемой точности для всех лучей, для которых найдено пересечение. За счет этого среднее значение точности по всем лучам для двойной точности (double) при использовании предложенного метода оказывается значительно лучше, чем для исходного представления полинома, – на 6 порядков для пучка, параллельного оптической оси (табл. 2), и на 5 порядков для пучка под углом в 30° (табл. 3). За счет использования предложенного метода также быстрее достигается целевая точность и поэтому уменьшается среднее число итераций метода Ньютона. Время вычислений с помощью предложенного незначительно метода отличается от суммирования исходного полинома.

При суммировании полинома в float-арифметике предложенный метод также позволяет достичь целевой точности  $10^{-6}$  в отличие от суммирования исходного представления полинома (2). Средняя точность по всем лучам в этом случае в предложенном методе оказывается лучше на два порядка. Время вычислений для предложенного метода в float-арифметике несколько снижается за счет уменьшения числа итераций метода Ньютона в ~1.7-1.9 раз.

В float-арифметике заметно влияние числа базовых точек  $\rho_0^2$ . При пяти точках ещё остаются лучи низкой точности, при 11 точках такие лучи в небольшом количестве (12 лучей, около 0.1 %) остаются для пучка 30°, а при 21 точке нет ни одного луча низкой точности. С увеличением числа точек средняя ошибка по всем лучам также несколько падает, но остается в пределах одного порядка.

Для поверхностей, у которых значения коэффициентов полинома не такие большие, как в таблице 1, и находятся в пределах  $\sim 10^4$ , влияние предложенного метода не настолько выражено, но все же присутствует. Были проведены замеры на 24 различных поверхностях с такими коэффициентами и оценена средняя ошибка для всех лучей, для которых были найдены пересечения. Результаты приведены в таблице 4.

Таблица 4. Замеры для набора поверхностей с коэффициентами  $< 1e4$  для пучка 30°, 1024 на 1024 луча

	Исходный полином		Предложенное разложение	
	double	float	double	float
Среднее значение ошибки для всех лучей	4.4e-16	3.96e-08	4.2e-16	1.65e-08

### Заключение

Предложенный метод разложения полиномов позволяет повысить точность расчетов при моделировании оптических систем, содержащих асферические поверхности вида (1). Для отдельных случаев, когда значения коэффициентов полиномов имеют большие величины, это позволяет избежать непредсказуемых изменений значения суммы, подобных феномену Рунге, при небольших изменениях

точки поверхности, в которой производится расчет. При этом точность расчетов в задаче поиска пересечения луча с такими поверхностями методом Ньютона возрастает на 5-6 порядков в сравнении с обычным суммированием полинома в double-арифметике и на 2 порядка в float-арифметике. Для производства оптических поверхностей такие значения точности (в double-арифметике) избыточны, но за счет увеличения точности расчетов также снижается и количество итераций метода Ньютона, необходимое для достижения целевых значений точности. Кроме того, повышение точности расчетов открывает возможности для разработки алгоритмов в смешенной точности (mixed precision), когда, например, все итерации метода Ньютона, кроме последней, вычисляются в float, а последняя – в double-арифметике.

Использование предложенного метода сохраняет скорость расчетов и требует лишь небольшого объема дополнительной памяти для хранения коэффициентов разложения в 11 точках (для радиально-симметричных поверхностей). Для double-арифметики потенциально можно использовать и меньше точек (до 5). Таким образом, предложенный метод может быть использован в системах оптического моделирования, использующих представление поверхностей с помощью неортогональных полиномов, подобных (1), для повышения точности и стабильности расчетов.

### Список литературы

1. [Sch19] Schuhmann, Rainer. "Description of aspheric surfaces". *Advanced Optical Technologies* 8.3-4 (2019): 267-278.
2. [Rayces06] Rayces, Juan L. and Xuemin Cheng. "Numerical integration of an aspheric surface profile". *International Optical Design Conference 2006*. Vol. 6342. SPIE, 2006.
3. [QDP] Quad-double precision library. URL: <https://github.com/lhprojects/qdpp>
4. [Forbes10] Forbes, G. W. "Robust and fast computation for the polynomials of optics". *Optics express* 18.13 (2010): 13851-13862.
5. [Kahan65] Kahan, W. "Further remarks on reducing truncation errors". *Assoc. Comput. Mach* 8 (1965): 40.
6. [Bakas24] Bakas, Nikolaos. "Taylor Polynomials in a High Arithmetic Precision as Universal Approximators". *Computation* 12.3 (2024): 53.
7. [Kimura19] Hullin, Matthias B., Johannes Hanika and Wolfgang Heidrich. "Polynomial Optics: A construction kit for efficient ray-tracing of lens systems". *Computer Graphics Forum*. Vol. 31, no. 4. Oxford, UK: Blackwell Publishing Ltd, 2012.
8. [Wu23] Chang-Mao, Wu, et al. "High precision ray tracing method for space camera in optical design". *ACTA PHYSICA SINICA* 72.8 (2023).
9. [Wu24] Wu, Changmao, et al. "Mathematical modelling for high precision ray tracing in optical design". *Applied Mathematical Modelling* 128 (2024): 103-122.
10. [Dammertz06] Dammertz, Holger and Alexander Keller. "Improving ray tracing precision by object space intersection computation". *2006 IEEE Symposium on Interactive Ray Tracing*. IEEE, 2006.
11. [PBRT3] Pharr, Matt, Wenzel Jakob and Greg Humphreys. "Physically Based Rendering: From Theory to Implementation". 3d edition, Morgan Kaufmann, 2016.

## Выбор начального приближения в трассировке лучей оптических систем с асферическими поверхностями

В. В. Санжаров<sup>1</sup>, С. В. Ершов<sup>2</sup>, В. А. Фролов<sup>1,2</sup>, А. Г. Волобой<sup>2</sup>

<sup>1</sup>Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ им. М.В. Ломоносова, Москва, Россия

<sup>2</sup>Институт прикладной математики им. М.В.Келдыша РАН, Москва, Россия

**Аннотация.** В работе рассматривается проблема поиска пересечения луча с асферическими поверхностями в оптических системах. В качестве асферических поверхностей рассматривались поверхности, заданные полиномами высокого порядка. Для поиска пересечения луча с такими поверхностями обычно используется метод Ньютона, который чувствителен к начальному приближению. Предложен метод вычисления начального приближения, позволяющий повысить устойчивость расчетов пересечения луча с поверхностью. Разработанный метод позволяет с высоким уровнем точности находить пересечения с поверхностями сложной формы, в том числе для лучей, падающих под большим углом к оптической оси. В сравнении с протестированными существующими решениями предложенный метод позволяет вычислить пересечения для значительно большего числа лучей.

**Ключевые слова:** трассировка лучей, оптика, асферические поверхности

## Initial guess in ray tracing optical systems with aspheric surface

V. V. Sanzharov<sup>1</sup>, S. V. Ershov<sup>2</sup>, V. A. Frolov<sup>1,2</sup>, A. G. Voloboy<sup>2</sup>

<sup>1</sup>IAI Moscow State University, Moscow, Russia

<sup>2</sup>Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

**Аннотация.** The paper considers the problem of finding the intersection of a ray with aspherical surfaces in optical systems. Surfaces defined by high-order polynomials were considered as aspherical surfaces. The intersection of a ray with such surfaces is usually computed using Newton's method, which is sensitive to the initial approximation. A method for calculating the initial approximation has been proposed, which makes it possible to increase the stability of calculations of the intersection of the ray with the surface. The developed method makes it possible to find intersections with surfaces of complex shapes with a high precision, including rays incident at a large angle to the optical axis. In comparison with existing solutions, proposed method makes it possible to calculate intersections for a much larger number of rays.

**Ключевые слова:** ray tracing, optics, aspheric surface

### Введение

Точное пересечение луча с поверхностью имеет решающее значение при трассировке лучей в оптических системах. Для несферических преломляющих поверхностей (асферических линз или оптики свободной формы/сплайна) пересечение с лучом находится путем решения нелинейного уравнения по методу Ньютона. Однако для сходимости метода Ньютона требуется хорошее начальное предположение.

Выбор подходящих начальных предположений для метода Ньютона имеет решающее значение в приложениях трассировки лучей по следующим причинам:

- надежность сходимости: плохие начальные приближения могут привести к тому, что метод не будет сходиться или будет сходиться к неправильным точкам пересечения;
- эффективность вычислений: лучшие начальные приближения обычно требуют меньшего числа итераций для достижения желаемой точности.

В приложении к оптическим системам существует несколько подходов к вычислению начального приближения для метода Ньютона. Рассмотрим существующие решения.

#### 1. Методы на основе аналитической аппроксимации

Многие методы линеаризуют или упрощают поверхность локально, чтобы получить первое приближение. Самый простой подход – пересечение с плоскостью. В этом случае поверхность аппроксимируется плоскостью, перпендикулярной оптической оси. Это может быть касательная плоскость к оптической поверхности или же плоскость, проходящая через её центр. Производится

подстановка уравнения луча в уравнение плоскости и ищется решение. Это дает оценку  $t$ , которую можно вычислить легко и эффективно. На практике такая аппроксимация широко используется в трассировке лучей в оптических системах, в том числе в современных работах, таких как [Nie23, Wang22, Yang24].

Однако использование начального приближения, полученного таким путем, не гарантирует сходимость метода Ньютона. Для поверхностей со сложной формой и/или для лучей, идущих под большим углом, аппроксимации плоскостью может быть недостаточно (рис. 1).

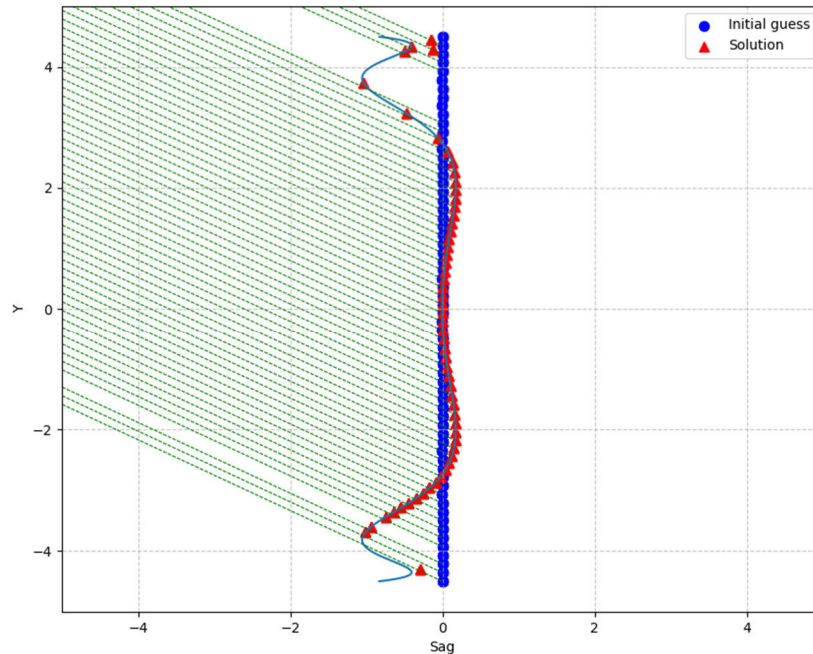


Рисунок 1. Пересечение пучка лучей с асферической поверхностью. Синими точками показаны начальные приближения для метода Ньютона, красными треугольниками – найденные решения

Если поверхность имеет перегиб или большую кривизну, то пересечение с касательной плоскостью может лежать за пределами физической апертуры, что приводит к тому, что метод Ньютона сходится к неверному решению.

Другой схожий подход – пересечение с "базовой" сферой. Многие асферические поверхности определяются как базовая коника (например, сфера или параболоид) плюс члены высших порядков. Можно найти пересечения луча с базовой сферой с заданным радиусом кривизны аналитически и использовать полученное значение  $t$  в качестве начального приближения. Такой подход может давать точку, расположенную ближе к реальной поверхности, чем в случае с аппроксимацией плоскостью. Пересечение со сферой тоже является вычислительно простым способом, так как требуется лишь решить квадратное уравнение. Несмотря на потенциально лучшее приближение для умеренно асферических поверхностей, данный подход имеет те же проблемы для сложных и сильно асферических поверхностей.

В работе [Joo16] авторы оценивают начальное приближение, шаг вдоль луча и интервал поиска пересечения, используя две аналитические поверхности, ограничивающие оптическую поверхность с обеих сторон. В качестве ограничивающих поверхностей предлагается использовать сферы или плоскости. При расчете пересечения лучей с поверхностями линз авторы отдают предпочтение методу бисекции с начальным интервалом поиска, заданным ограничивающими поверхностями. Основным недостатком метода бисекции является то, что для достижения высокого уровня точности, необходимого в оптических расчетах, ему требуется в несколько раз больше итераций, чем для метода Ньютона.

В целом, аналитические аппроксимации плоскостью или сферой просто реализовать. Они требуют минимальных вычислительных затрат, но им не хватает надежности для сложных форм – метод Ньютона может расходиться или сходиться к неправильному пересечению (например, за пределами апертуры). Как правило, такие аппроксимации работают, если поверхность достаточно близка к плоской или сферической в интересующей области.

## 2. Методы на основе подразбиений

Подходы из данной группы основаны на дискретизации оптической поверхности с целью поиска начального приближения для метода Ньютона. Общей идеей является аппроксимация поверхности сеткой из плоских участков, треугольников или точек. При поиске пересечения луча с поверхностью, заданной сплайнами или NURBS, часто строится грубая управляющая сетка или треугольная сетка. Лучи пересекаются с этой сеткой, чтобы найти приблизительную точку пересечения. Например, может использоваться равномерная сетка точек на поверхности, расстояние до пересечения с которой используется в качестве начального для метода Ньютона. Схожим образом можно выполнить разбиение для поверхности в параметрической области. Тогда каждый участок («патч») предоставляет потенциальный интервал, содержащий пересечение. Такие методы гарантируют, что начальное предположение лежит вблизи действительного корня при условии, что сетка достаточно мелкая. Эти методы хорошо справляются со сложными формами и множеством пересечений и применимы к поверхностям свободной формы и симметричным поверхностям. Однако для этого требуется плотная сетка с высоким уровнем подразбиения, что приводит к значительным вычислительным затратам.

В работе [Koning23] триангуляция для поиска пересечений используется для поверхностей свободной формы, моделируемых В-сплайнами. Авторы используют свойства В-сплайнов для ускорения процесса поиска пересечения лучей и треугольников. Однако эффективность предложенного подхода не исследовалась, и оптимизация предложенного алгоритма пересечения В-сплайнов оставлена для будущих работ.

Для повышения производительности поиска пересечения с представлениями поверхности в виде полигональной сетки могут также использоваться ускоряющие структуры (такие как BVH-дерево). В [Zhdanov22] такой подход предлагается для асферических поверхностей, а в [Zhdanov24] применение расширено на поверхности свободной формы и предложены дополнительные оптимизации расчетов. Пересечение с триангулированными сетками сначала дает хорошее начальное предположение, которое ускоряет сходимость, а также позволяет справиться с проблемными лучами, такими как касательные (или почти касательные) к поверхности. Однако в таких задачах, как проектирование оптических систем, когда целевая поверхность постоянно меняется в процессе оптимизации, использование такого подхода может вызвать трудности, поскольку триангулированную геометрию и ускоряющую структуру также потребуется постоянно обновлять.

Авторы работы [Morita10] предлагают аппроксимировать асферическую поверхность квадратичными интерполянтами (патчами). Пересечение между лучом и патчем может быть найдено аналитически. В работе показано, что найденное пересечение и нормаль в нем быстрее приближаются к истинным значениям с увеличением числа патчей, чем при использовании треугольников для аппроксимации поверхности. Авторы утверждают, что сходимость к пределу точности плавающей точки двойной точности достигается при использовании  $10^7$  патчей, в то время как для этого потребовалось бы  $10^{16}$  треугольников. Что, однако, все равно требует достаточно большого количества вычислений. Поэтому в целом потенциальные проблемы при таком подходе аналогичны тем, что возникают при использовании треугольников.

Методы на основе подразбиений могут работать с различными представлениями поверхностей, включая поверхности свободной формы, и дают начальное предположение, близкое к истинному пересечению, если сетка достаточно плотная. Однако это приводит к увеличению как вычислительных затрат, так и затрат памяти (для хранения сетки). А для задач проектирования оптических систем, когда целевая поверхность меняется в процессе оптимизации, необходимо также обновлять и новое представление поверхности (а также ускоряющую структуру, если она используется), что может быть неэффективным.

## 3. Методы на основе когерентных лучей и опорных точек

При трассировке лучей в оптических системах часто необходимо рассчитывать лучи, имеющие близкие направления. Методы на основе когерентности лучей используют этот факт для повторного использования результатов: можно использовать точку пересечения для соседних лучей (или предыдущего луча) в качестве начального предположения для текущего луча [Joy86]. В плотно дискретизированных пучках (например, последовательные лучи в пространстве изображения)



соседние пересечения оказываются близки. Этот подход очень дешев поскольку повторно использует уже рассчитанный результат и может значительно ускорить поиск пересечения. Но, например, если некоторый луч пропустит поверхность, на которую попали его соседи, метод не будет работать. Поэтому наивное повторное использование предыдущего пересечения иногда приводит к сходимости к неправильному корню. Кроме того, нужно каким-то образом получить предыдущее пересечение.

Более надежным вариантом является метод на основе опорных точек. Сначала задается набор опорных точек на поверхности, который может быть получен, например, варьированием координат по достаточно грубой сетке. Затем для каждого входящего луча находится опорная точка, ближайшая к нему по евклидову расстоянию. Наконец, эта точка проецируется перпендикулярно на луч и используется как начальное приближение для расстояния пересечения [Ren25]. По сути, такой подход выпускает луч вблизи известной точки поверхности. Это гарантирует, что начальное приближение находится вблизи поверхности, и улучшает сходимость для сильно асферических поверхностей или поверхностей свободной формы, позволяя избежать сходимости к неверным корням, которые могут возникать для методов, использующих аппроксимацию плоскостью. Недостатком являются дополнительные затраты на хранение и поиск опорных точек, а также вычисление ближайших расстояний. Кроме того, в задачах проектирования оптических систем, опорные точки требуется вычислять заново при итеративном изменении поверхности.

#### 4. Методы в пространстве параметров

Существуют специальные методы параметризации сплайновых поверхностей, позволяющие ослабить требования к начальному приближению. Например, в работе [Tang23] авторы заменяют параметры поверхности  $(u,v)$  преобразованной парой  $(\xi,\eta)$ , которая принадлежит интервалу  $[0,1]$ , даже если оценки для  $(u,v)$  выходят за его пределы. Таким образом, метод Ньютона будет сходиться к нужному решению даже для начальных значений  $(u_0,v_0)$ , выбранных из достаточно грубой сетки.

Другой параметрический подход – двумерное пересечение. Вместо решения уравнения луча можно решить эквивалентную систему в  $(u,v)$  путем пересечения луча с параметрической поверхностью, спроецированной на плоскость. Это часто приводит к двум нелинейным уравнениям в  $(u,v)$  и снова требует использования метода Ньютона. В таких случаях хорошие начальные приближения для  $(u,v)$  можно получить, проецируя направление луча на область параметров поверхности или используя среднюю точку ограничивающего параллелепипеда [Martin00].

В целом, данная группа методов специфична для сплайновых и NURBS-поверхностей. Они также усложняют процедуру поиска пересечений, поскольку необходимо реализовать преобразование и обратное преобразование.

Таким образом, существуют различные подходы, которые можно разделить на несколько классов, рассмотренных выше. Наиболее распространенными для асферических поверхностей, заданных полиномами, являются методы на основе аналитической аппроксимации плоскостями или сферами. Они просты в реализации и быстро работают, но в сложных случаях не дают достаточно хорошего начального приближения для обеспечения сходимости. Методы на основе подразбиений позволяют найти близкое к решению начальное приближение, но требуют вычислительных затрат и затрат памяти на хранение новых представлений, что также ограничивает их применимость. В методах на основе опорных точек также есть затраты на вычисление и хранение вспомогательных данных, хоть и меньшие.

Интерес представляет разработка метода, который позволял бы получать начальное приближение, позволяющее надежно находить пересечения в сложных случаях, но не требовал бы хранения и вычисления дополнительных структур данных. При этом желательным свойством нового метода также является адаптивность. Для распространенных простых случаев поиска пересечения он должен требовать меньше вычислений, чем для сложных случаев (лучи, идущие почти по касательной к поверхности и т.п.).

**Предлагаемое решение****1. Постановка задачи**

Рассматривается задача пересечения луча, заданного как

$$o + vt,$$

где  $o = (x_0, y_0, z_0)$  – координаты начальной точки;

$t$  – величина расстояния;

$v = (v_x, v_y, v_z)$  – вектор направления с асферической поверхностью.

В качестве примера будем рассматривать представление поверхности в виде полинома с четными степенями:

$$f(r) = \frac{cr^2}{1 + \sqrt{1 - (1+k)c^2r^2}} + \sum_{i=2}^n \alpha_i \rho^{2i},$$

$$\rho = \frac{r}{r_{norm}}$$

где  $f(r)$  – функция расстояния от оптической оси до точки с координатами  $(x, y)$ :  $r = \sqrt{x^2 + y^2}$ ;

$c$  – кривизна;

$k$  – коника;

$\alpha_i$  – коэффициенты полинома;

$r_{norm}$  – параметр нормализации радиуса.

Для дальнейших рассуждений детали представления поверхности не будут иметь значения, мы будем оперировать только значениями функции поверхности  $f(r)$  и её производной  $f'(r)$ . Таким образом, предложенное решение применимо и для других видов функции  $f(r)$ .

Запишем выражения для точки на луче  $p_{ray}(t)$  при некотором конкретном значении  $t$  и соответствующему ей значению функции  $z_{surf}(t)$ :

$$p_{ray}(t) = o + vt = (x_{ray}(t), y_{ray}(t), z_{ray}(t)),$$

$$z_{surf}(t) = f\left(\sqrt{x_{ray}^2(t) + y_{ray}^2(t)}\right).$$

При пересечении луча с поверхностью  $z_{ray}(t) \equiv z_{surf}(t)$ .

Пересечение луча с поверхностью приводит к уравнению относительно  $t$ :

$$f(r_0^2 + at^2 + bt) - (z_0 + v_z t) = 0, \quad (1)$$

$$r_0^2 \equiv x_0^2 + y_0^2,$$

$$a \equiv 1 - v_z^2,$$

$$b \equiv 2(x_0 v_x + y_0 v_y).$$

Тогда итерация метода Ньютона будет иметь вид

$$t_{k+1} = t_k - \frac{f(r_0^2 + at_k^2 + bt_k) - (v_z t_k + z_0)}{f'(r_0^2 + at_k^2 + bt_k)(2at_k + b) - v_z}.$$

Целью данной работы является разработка метода для вычисления начального приближения  $t_{k=0}$ , достаточно близкого к искомому  $t$ , чтобы обеспечить сходимости итераций Ньютона к искомому решению.

**2. Предложенный метод**

Часто оценка снизу для расстояния до пересечения предполагает, что поверхность достаточно гладкая, и исходит из оценок Липшица [Ban25], Харнака [Gillispie24]. Такой подход работает для определения поверхности через неявную функцию, поэтому он не может быть применен к явному определению поверхности в решаемой задаче «как есть». Предложенный подход использует базовые идеи из этих работ, адаптируя их к случаю гладкой и вращательно-симметричной поверхности. Все корни уравнения (1) являются точками пересечения. Для удобства будем считать, что начальная точка равна  $t=0$ . Поделим интервал  $r^2$  на некоторые точки  $R_i^2$ , на каждом отрезке  $[R_i^2, R_{i+1}^2]$  вычисляем и запоминаем границы  $f'$  в этом интервале:

$$f'_{min,i} \leq f'(r^2) \leq f'_{max,i} \text{ для } r^2 \in [R_i^2, R_{i+1}^2].$$

Теперь найдем интервал, содержащий  $r_0^2$ . Его индекс обозначим  $i$ , то есть  $r_0^2 \in [R_i^2, R_{i+1}^2]$ . Тогда

$$f(r_0^2) + (r^2 - r_0^2)f'_{\min,i} \leq f(r^2) \leq f(r_0^2) + (r^2 - r_0^2)f'_{\max,i} \quad (2)$$

для  $r^2 \in [R_i^2, R_{i+1}^2]$ . Подставив в уравнение (1), мы увидим, что корень  $t$  лежит в интервале положительных корней двух квадратных уравнений

$$\begin{aligned} f(r_0^2) + (at^2 + bt)f'_{\min,i} - (z_0 + v_z t) &= 0, \\ f(r_0^2) + (at^2 + bt)f'_{\max,i} - (z_0 + v_z t) &= 0. \end{aligned} \quad (3)$$

На рисунке 2 показан график (синим цветом) изменения разности между координатой  $z_{ray}$  точки на луче и координатой соответствующей точки на поверхности  $z_{surf}$  от значения искомого параметра расстояния  $t$ .

Однако неравенство (2) справедливо только для  $r^2 \in [R_i^2, R_{i+1}^2]$ , на рисунке 2 оно не выполняется, при  $r^2 \in r_0^2 + at^2 + bt = R_{i+1}^2$ . Соответствующая область (криволинейный треугольник) заштрихована желтым. Корень для второго уравнения из (3) находится вне этой области, поэтому его нельзя использовать для предсказания корня  $t$ . Другими словами, мы должны использовать уравнение (3) следующим образом:

1. Вычислить граничное значение  $t_c$  для допустимой области, т.е. так, что  $r_0^2 + at_c^2 + bt_c$  покидает  $[R_i^2, R_{i+1}^2]$ .

2. Для каждого из двух квадратных уравнений (3) вычислить их корни и взять наименьший положительный. Отсутствие положительных корней говорит о том, что парабола не пересекает ось. В этом случае мы принимаем в качестве корня бесконечность. Таким образом, имеем два значения (по одному для каждого из уравнений), которые обозначим через  $t_1$  (меньшее) и  $t_2$  (большее).

а. Если ни одно из этих значений не лежит в допустимой области, тогда нижняя граница искомого пересечения  $t = t_c$ , и мы сдвигаем начальную точку луча в  $t_c$ .

б. Если только одно из значений лежит в допустимой области (очевидно это  $t_1$ ), тогда нижняя граница искомого пересечения  $t = t_1$ , и мы сдвигаем начальную точку луча в  $t_1$ .

с. Если оба значения лежат в допустимой области, то искомое пересечение лежит между ними, т.е.  $t_1 \leq t \leq t_2$ , и мы сдвигаем начальную точку луча в  $t_2$ .

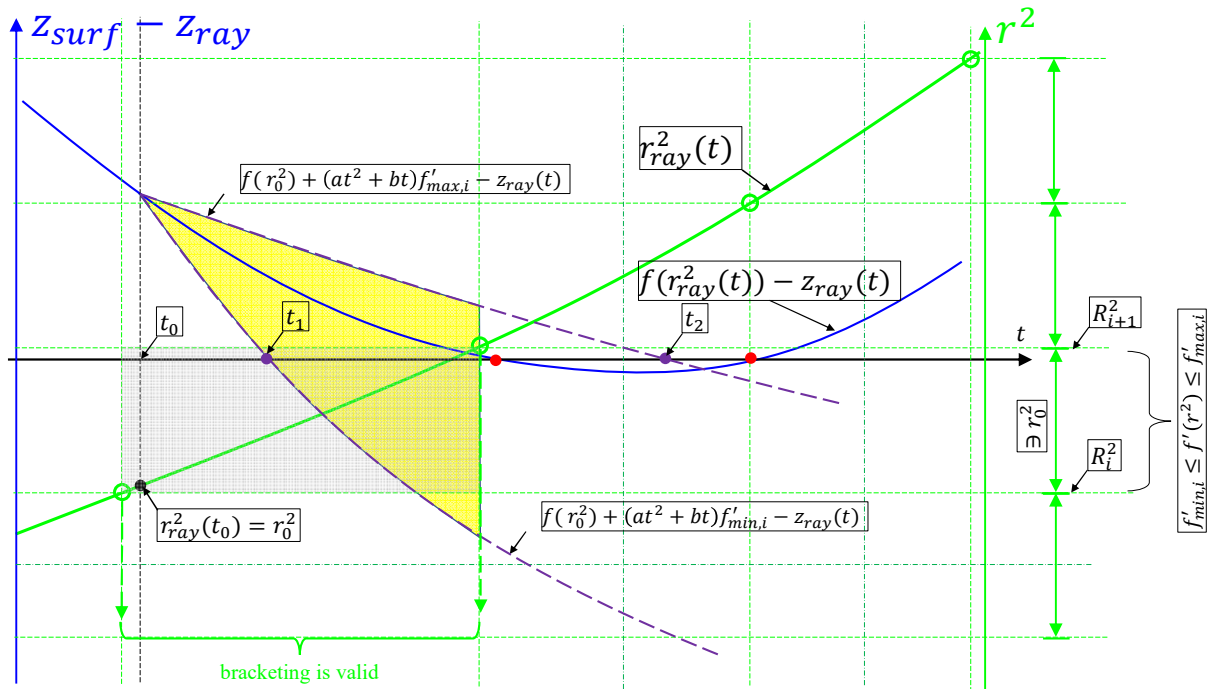


Рисунок 2. Границы корней уравнения (1), вычисленные из двух квадратных уравнений относительно  $\min f'$  и  $\max f'$ .

Красными точками отмечены корни уравнения (1), а фиолетовыми точками – положительные корни уравнения (3), по горизонтальной оси – параметр расстояния  $t$ , по левой вертикальной оси – разность между значением координаты  $z$  для точки на луче  $z_{ray}$  и для поверхности  $z_{surf}$ , по правой – значение  $r^2$

Это очень похоже на один шаг марширования по лучу. Как и там, после шага мы перемещаем начало луча в новую точку и повторяем процедуру.

Частота подразбиения области  $r^2$  на интервалы  $[R_i^2, R_{i+1}^2]$ , в которых мы берем границы  $f'$ , играет двойную роль:

- с одной стороны, чем чаще подразбиение, тем ближе (2), то есть стороны желтого треугольника на рисунке 2 проходят ближе к искомому  $f(r_0^2 + at^2 + bt)$ , и интервал для пересечения более плотный;
- с другой стороны, при более частом подразбиении допустимая область становится меньше, поэтому может случиться так, что  $r_0^2 + at^2 + bt$  покинет ее до того, как одна из парабол в (3) пересечет ось. В этом случае границы  $f'$  вообще не используются, и все, что мы можем сделать, это переместить луч на границу допустимой области  $t_c$ . Такое перемещение может быть слишком мало.

На рисунке 3 показана ситуация для двух уровней подразбиения (сегменты для второго уровня в два раза меньше).

Для второго уровня минимальный положительный корень (3), показанный фиолетовым крестиком на рисунке 3, больше, чем таковой для более грубого уровня (фиолетовая точка). Соответственно, мы могли бы подвинуть начальную точку луча дальше, если бы не ограничение, – этот корень для второго уровня лежит вне допустимой области (оранжевый треугольник). Таким образом, второй уровень подразбиения позволит нам подвинуть начало луча только до границы допустимой области, что все равно дальше, чем первый уровень. То есть в данном примере более мелкое подразбиение дает лучший результат, но при других параметрах более грубое подразбиение может быть лучше. На практике мы использовали 4-го уровня подразбиения. Для каждого из уровней вычисляется допустимый шаг по лучу и выбирается наибольший.

Луч от объекта в оптической системе может стартовать с некоторой произвольно удаленной точки и выполнять движение по лучу по описанному алгоритму в этом случае будет очень неэффективно. Поэтому необходимо выбрать начальное значение, с которого следует запускать алгоритм. Для этого предлагается использовать простой и быстрый метод на основе ограничивающих плоскостей в качестве предварительного этапа для предложенного метода (рис. 4).

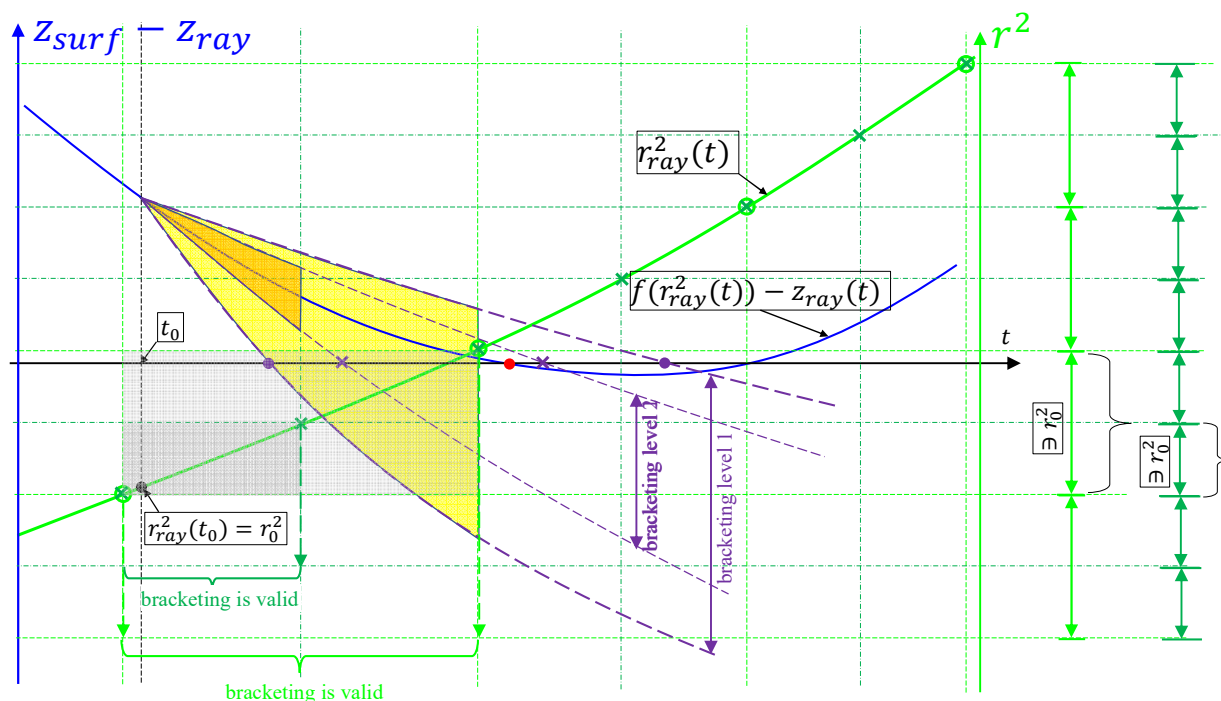


Рисунок 3. Два уровня подразбиения. Две параболы (3) для первого уровня показаны фиолетовыми толстыми пунктирными линиями, а для второго уровня – фиолетовыми тонкими пунктирными линиями. Допустимая область для второго уровня – оранжевый криволинейный треугольник. По горизонтальной оси – параметр расстояния  $t$ , по левой вертикальной оси – разница между значением координаты  $z$  для точки на луче  $z_{ray}$  и поверхности  $z_{surf}$ , по правой – значение  $r^2$

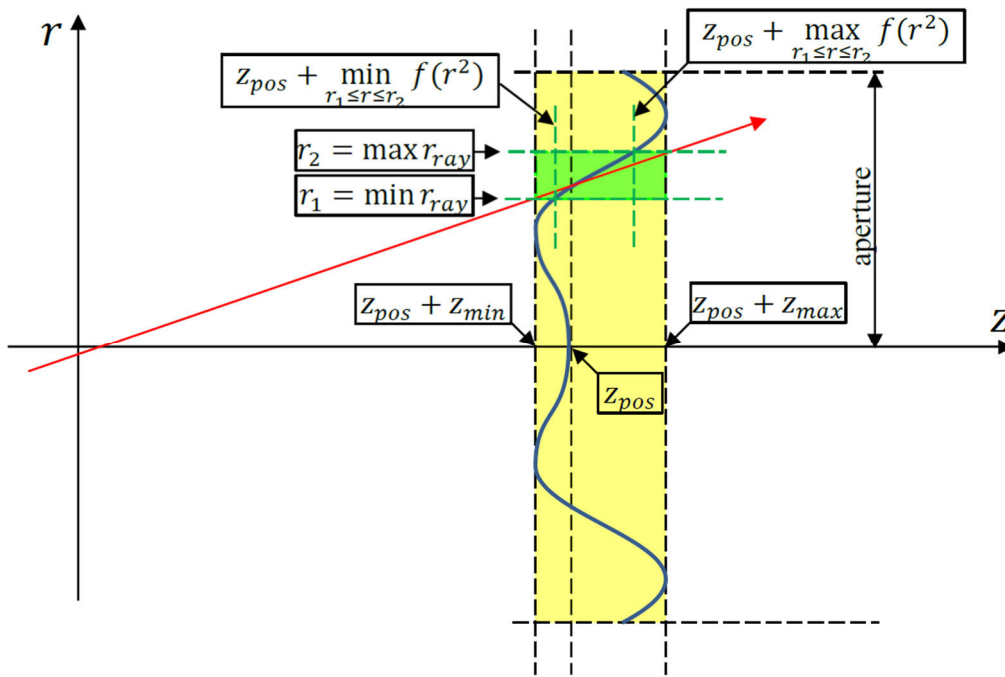


Рисунок 4. Получение начальной точки для основного алгоритма. Луч (красная стрелка) направлен на поверхность линзы (синяя кривая) издалека. Ограничивающие плоскости показаны вертикальными пунктирными линиями вместе с центральной плоскостью. Зеленая область показывает интервал, в котором лежит пересечение

В области, ограничивающей поверхность (обозначена желтым на рис. 4), значение  $r_{ray}(t)$  лежит в  $[r_1, r_2]$ . Для  $r_1 \leq r \leq r_2$  значения функции поверхности лежат в интервале, меньшем чем  $[z_{min}, z_{max}]$ . Этот интервал отмечен на рисунке 4 зелеными вертикальными пунктирными линиями. Основываясь на том, что поверхность радиально симметрична и меняется достаточно плавно в интервале допустимых значений радиуса, значения  $z_{min}$  и  $z_{max}$  могут быть вычислены как минимум и максимум значений функции поверхности  $f$  в  $N$  точках, равномерно расставленных от 0 до максимального значения радиуса  $r_{max}$ . В проведенных экспериментах использовалось  $N = 100$  точек, а для вычисленных значений был также добавлен небольшой «зазор» в 0.1 % от величины найденного интервала:

$$\delta = 0.001 * (z_{max} - z_{min}),$$

$$z_{min} - \delta, z_{max} + \delta.$$

Алгоритм определения начальной точки следующий:

- 1) вычислить пересечения луча  $t_1, t_2$  с ограничивающими плоскостями;
- 2) вычислить интервал  $[r_1 = \min(r_{ray}(t_1), r_{ray}(t_2)), r_2 = \max(r_{ray}(t_1), r_{ray}(t_2))]$ ;
- 3) вычислить интервал  $[z_1 = z_{pos} + f(r_1), z_2 = z_{pos} + f(r_2)]$ ;
- 4) вычислить пересечения луча  $t'_1, t'_2$  с плоскостями  $z = z_1$  и  $z = z_2$ .

Если угол между лучом и осью небольшой, то  $t'_1 \approx t'_2$  и точка на луче для  $t'_1$  лежит достаточно близко к поверхности. В общем случае можно ещё приблизиться к поверхности, повторив описанный алгоритм, взяв на первом шаге значения  $t'_1, t'_2$ . Для больших углов луча с осью Oz итеративный процесс может начать расходиться, поэтому необходимо остановиться, если на очередной итерации интервал значений  $t$  увеличивается.

### Экспериментальная оценка

Чтобы протестировать предложенный метод поиска начального приближения, были сгенерированы лучи с известными точными пересечениями. Для генерации лучей использовался следующий алгоритм:

1. Установить количество точек на поверхности выборки, количество значений для двух углов  $\varphi$  и  $\theta$  и расстояний от поверхности  $t$ .

- а. Точки равномерно распределены по половине поперечного сечения исследуемой поверхности (так как она симметрична) и не выходят за пределы допустимого радиуса или апертуры.
  - б. Угол  $\varphi$  равномерно выбирается от 0 до заданного максимального значения (например, 60 градусов).
  - в. Угол  $\theta$  выбирается равномерно от 0 до  $\pi$ .
  - г. Расстояние выбирается равномерно от небольшого расстояния до поверхности (например,  $1e-3$ ) до максимально допустимого  $r * 3$ .
2. Для каждой возможной комбинации точек, углов и расстояний мы вычисляем  $ray\_origin$  и  $ray\_dir$  следующим образом:
- а)  $dir1 = (\cos\varphi * \sin\theta, \sin\varphi * \sin\theta, \cos\theta)$ ;
  - б)  $dir2 = (-\cos\varphi * \sin\theta, -\sin\varphi * \sin\theta, \cos\theta)$ ;
  - в)  $ray_{dir} = concat(dir1, dir2)$ ;
  - г)  $ray_{origin} = sampled_{pt} - sampled_{dist} * ray_{dir}$ , где  $sampled_{pt}$  – выбранная точка на поверхности,  $sampled_{dist}$  – выбранное расстояние от поверхности.
3. Ищем пересечение этих лучей с поверхностью:
- а. Если для некоторого луча пересечение не найдено, считаем поиск пересечения неуспешным и увеличиваем счетчик ненайденных пересечений.
  - б. Если пересечение найдено, но отличается от истинного значения, то это может означать, что на луче между его началом  $ray_{origin}$  и точкой поверхности, от которой был построен луч, располагаются другие пересечения (рис. 5). Поэтому мы передвигаем начальную точку луча в найденную точку пересечения и повторяем поиск. Этот процесс повторяется до ста раз.
  - в. Если мы нашли пересечение, соответствующее истинной точке на поверхности, относительно которой строился луч, то считаем поиск успешным.

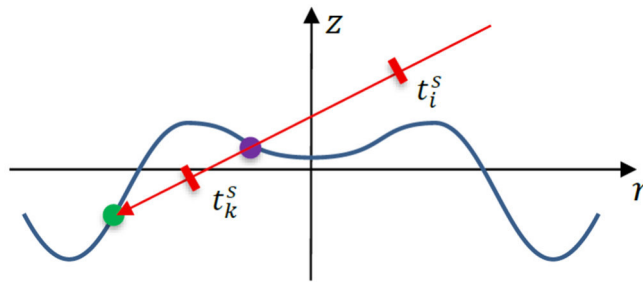


Рисунок 5. Ситуация нескольких пересечений с поверхностью (синяя линия) для сгенерированного тестового луча (красная линия). Как  $t_i^s$  и  $t_k^s$  обозначены разные выбранные расстояния и, соответственно, разные начальные точки луча. Для  $t_k^s$  будет сразу найдена точка пересечения (зеленая точка), из которой был построен луч и которая рассматривается тестом как истинная. Для  $t_i^s$  сначала будет найдено другое ближайшее к ней пересечение (фиолетовая точка), поэтому при тестировании будет произведена повторная трассировка луча с тем же направлением из этой точки

Был проведен эксперимент со следующими параметрами описанного алгоритма генерации лучей: 101 точка на поверхности, 21 значение угла  $\varphi$  от 0 до 60 градусов, 21 значение угла  $\theta$  от 0 до  $\pi$ , 21 значение расстояний до поверхности от  $1e-3$  до  $r * 3$ . Такие значения параметров дают набор из 1870722 лучей. Набор лучей был сгенерирован для каждой из тестируемых поверхностей. Использовался набор из 978 асферических поверхностей, для которых была выполнена трассировка сгенерированных наборов лучей с тремя разными способами выбора начального приближения:

- на основе ограничивающих плоскостей (широко используется в существующих работах (см. подраздел «Методы на основе аналитической аппроксимации» во введении);
- на основе опорных точек на поверхности [Ren25];
- предложенный метод.

На рисунке ниже представлены примеры объективов, поверхности из которых использовались для тестирования (рис. 6).



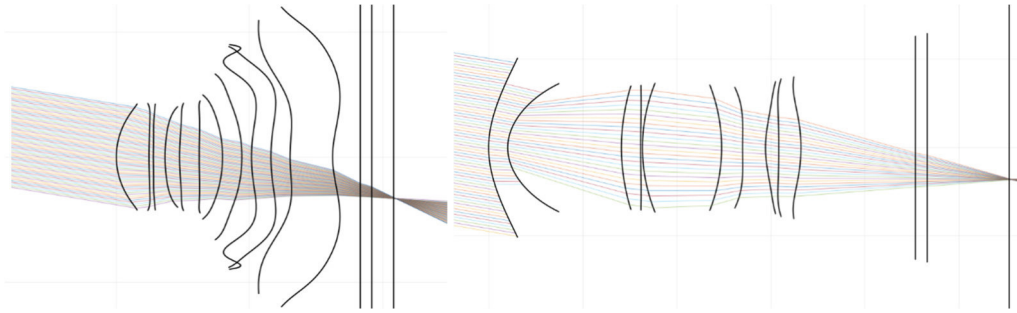


Рисунок 6. Примеры объективов с асферическими поверхностями, на которых проводилось тестирование

Для метода опорных точек [Ren25] использовалась реализация авторов, которая была перенесена с языка Python и библиотеки pytorch на C++. В качестве параметров в данном методе выступает разрешение сетки опорных точек на поверхности. Использовалось значение, выбранное авторами значением по умолчанию – 31 по  $x$  и  $y$  с удалением точек, лежащих вне апертуры.

Результаты эксперимента приведены в таблице 1. Как видно из них, предложенный метод позволяет уменьшить количество ненайденных пересечений на порядок по сравнению с двумя другими рассмотренными методами. При этом максимальное количество ненайденных лучей также остается небольшим, порядка  $\sim 0.51\%$ . В тоже время, метод на основе ограничивающих плоскостей и метод на основе опорных точек в отдельных случаях могут терять существенное количество пересечений – больше 5%, что для рассмотренного набора лучей составляет более 100 тыс. лучей.

Таблица 1. Тестирование поиска пересечения

Метод	Средний процент ненайденных лучей	Среднее количество ненайденных лучей	Наибольший процент ненайденных лучей	Наибольшее количество ненайденных лучей
Ограничивающие плоскости	0.25%	4676	6.12%	114576
Опорные точки [Ren25]	0.17%	3164	5.44%	101808
Предложенный метод	0.013%	237	0.51%	9576

Конечно, предложенный тест является синтетическим и демонстрирует значительно более сложную задачу поиска пересечений, нежели в типичном случае последовательной трассировки лучей через систему линз. Но и в обычных сценариях последовательной трассировки лучей через оптическую систему могут встречаться случаи, когда у более простых подходов возникают сложности с поиском пересечения. Ниже приведен пример одной из поверхностей сложной формы с визуализацией начальных приближений для каждого из методов и найденного решения (рис.7, рис. 8). Видно, что приближение на основе плоскости в этом случае очень грубое и приводит к тому, что для некоторых лучей решение не найдено. Метод на основе опорных точек уже лучи не теряет, но находит одно неверное решение.

Предложенный метод находит все пересечения, при этом все начальные приближения также располагаются близко к истинным решениям.

Были проведены отдельные эксперименты и для последовательной трассировки лучей. В этом случае отличия между разными методами выбора начальных приближений обычно сравнительно небольшие и проявляются при больших углах наклона.

Также были проведены замеры производительности. Для этого выполнялась трассировка пучка из 262144 лучей (512 на 512) для того же набора поверхностей. Замеры выполнялись на CPU Ryzen 9 5950X. Результаты представлены в таблице 2.

Предложенный метод ожидаемо работает в несколько раз медленнее, чем простой метод на основе ограничивающих плоскостей. При этом время выполнения может достаточно сильно варьироваться в зависимости от поверхности, что связано с тем, что у метода нет фиксированного количества итераций. Для простых случаев, таких как пучок лучей под углом  $0^\circ$  к оптической оси, предложенный

метод работает в только в 2-3 раза медленнее метода на основе плоскостей. А для более сложных случаев может быть выполнено достаточно большое количество шагов, что приводит к падению производительности.

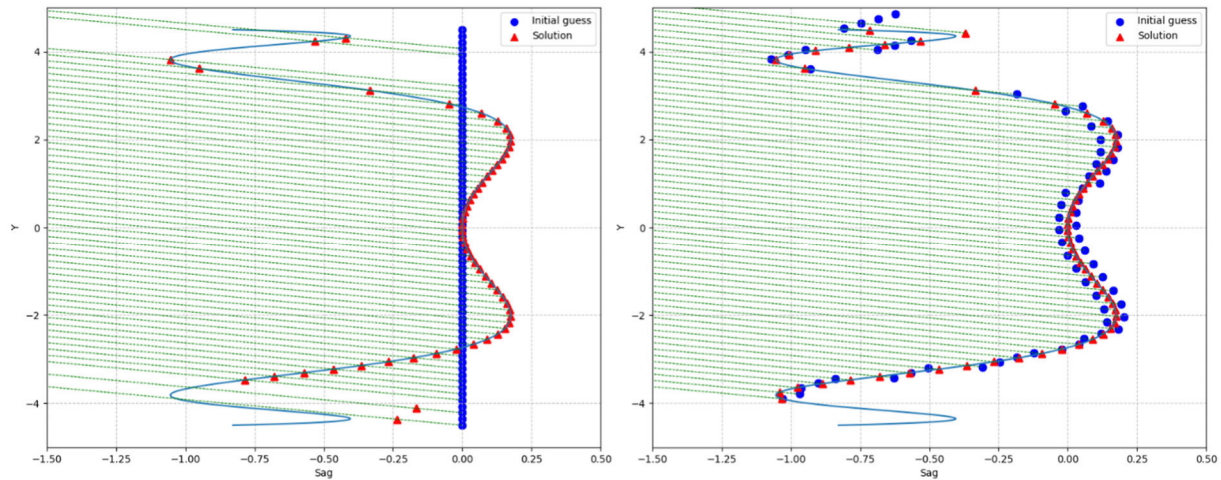


Рисунок 7. Визуализация начального приближения (синие точки) и найденного решения (красные треугольники): *слева* – метод на основе плоскости, *справа* – на основе опорных точек

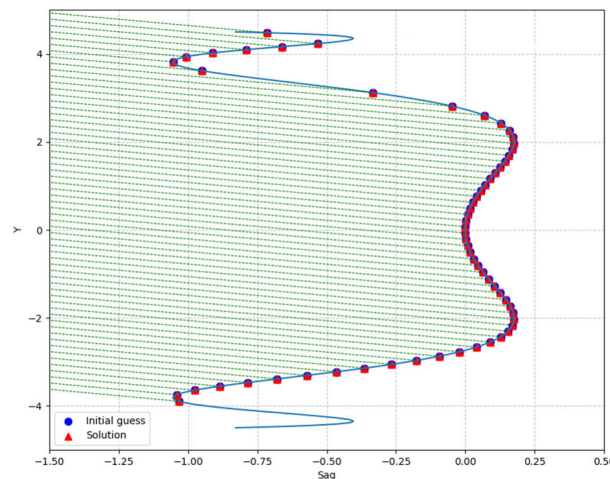


Рисунок 8. Визуализация начального приближения (синие точки) и найденного решения (красные треугольники) для предложенного метода

Таблица 2. Замеры производительности

Метод	Среднее время для пучка 0°, миллисекунды	Среднее время для пучка 20°, миллисекунды	Среднее время для пучка 40°, миллисекунды
Ограничивающие плоскости	28.39 ± 4.39	32.07 ± 6.46	36.31 ± 10.89
Опорные точки [Ren25]	446.21 ± 5.88	456.39 ± 4.96	463.26 ± 13.67
Предложенный метод	57.34 ± 20.48	353.86 ± 174.04	413.30 ± 261.82

Метод на основе опорных точек в проведенном эксперименте работает стабильно медленно, вне зависимости от угла наклона лучей. Это связано с тем, что в любом случае выполняется поиск ближайшей опорной точки к лучу. Вероятно, у этого метода имеется некоторый потенциал ускорения при реализации более эффективного метода поиска ближайшей опорной точки. Также следует отметить, что у этого метода имеется ещё один недостаток – необходимость вычислять и хранить опорные точки для поверхности. А в случае выполнения оптимизации оптической системы точки потребуются вычислять заново при каждом изменении поверхности.

## Закключение

Предложенный метод выбора начального приближения для задачи поиска пересечения луча и асферической поверхности позволяет обеспечить нахождение пересечения с высокой точностью в большем числе случаев, чем наиболее распространённый метод на основе ограничивающих плоскостей и метод на основе опорных точек [Ren25]. В разработанном тестовом сценарии предложенный метод смог найти на порядок большее количество пересечений, максимальное количество потерянных пересечений не превышает 0.51 %. При этом производительность предложенного метода сравнима с методом на основе ограничивающих плоскостей в простых случаях. В более сложных случаях эффективный поиск пересечений требует большего числа итераций и тем самым большего времени работы. При этом производительность предложенного метода сохраняется на уровне метода на основе опорных точек, который, судя по проведенным экспериментам, находит меньше пересечений. Таким образом, предложенный метод применим в задачах, где необходимо моделировать случаи сложного распространения света в оптических системах, что важно в таких задачах, как, например, анализ рассеянного света.

Также следует отметить, что разработанный метод требует лишь возможности вычисления функции поверхности и её производной и тем самым может быть потенциально использован не только для асферических поверхностей, заданных полиномами четных степеней, но и для других способов задания оптических поверхностей, к которым применим метод Ньютона.

## Источник финансирования

Исследование выполнено за счет гранта Российского научного фонда № 25-11-00054 «Спектральное моделирование распространения света и спектральный рендеринг в реальном времени на гибридных и мобильных платформах», <https://rscf.ru/project/25-11-00054>

## Список литературы

- [Wang22] Wang, Congli, Ni Chen, and Wolfgang Heidrich. "do: A differentiable engine for deep lens design of computational imaging systems". *IEEE Transactions on Computational Imaging* 8 (2022): 905-916.
- [Yang24] Yang, Xinge, Qiang Fu, and Wolfgang Heidrich. "Curriculum learning for ab initio deep learned refractive optics". *Nature communications* 15.1 (2024): 6572.
- [Heemels24] Heemels, Alexander, et al. "Optimizing freeform lenses for extended sources with algorithmic differentiable ray tracing and truncated hierarchical B-splines". *Optics Express* 32.6 (2024): 9730-9746.
- [Nie23] Nie, Y., Zhang, J., Su, R., & Ottevaere, H. (2023). Freeform optical system design with differentiable three-dimensional ray tracing and unsupervised learning. *Optics Express*, 31(5), 7450-7465.
- [Yang24Diff] Yang, Xinge, et al. "End-to-end hybrid refractive-diffractive lens design with differentiable ray-wave model". *SIGGRAPH Asia 2024 Conference Papers*. 2024.
- [Tang23] Tang, Haisong, et al. "Parallel ray tracing through freeform lenses with NURBS surfaces". *Chinese Optics Letters* 21.5 (2023): 052201.
- [Zhdanov22] Zhdanov, Dmitry, et al. "Use of two-level geometry representation for the ray tracing acceleration in optical systems with aspheric surfaces". *Optical Design and Testing XII*. Vol. 12315. SPIE, 2022.
- [Zhdanov24] Zhdanov, D. D., I. S. Potemin, and A. D. Zhdanov. "Dual Representation of Geometry for Ray Tracing Acceleration in Optical Systems with Freeform Surfaces". *Programming and Computer Software* 50.3 (2024): 273-282.
- [Martin00] Martin, W., Cohen, E., Fish, R., & Shirley, P. "Practical ray tracing of trimmed NURBS surfaces". *Journal of Graphics Tools* 5. 1. 2000: 27-52.
- [Joo16] Joo, Hyuntae, et al. "Efficient ray tracing through aspheric lenses and imperfect bokeh synthesis". *Computer Graphics Forum*. Vol. 35., no. 4. 2016.
- [Joy86] Joy, K. I., & Bhetanabhotla, M. N. (). Ray tracing parametric surface patches utilizing numerical techniques and ray coherence. *ACM SIGGRAPH Computer Graphics*, 20. 4. 1986: 279-285.
- [Ren25] Ren, Z., Zhou, J., Zhang, W., Yan, J., Chen, B., Feng, H., & Chen, S. Successive optimization of optics and post-processing with differentiable coherent psf operator and field information. *IEEE Transactions on Computational Imaging*. 2025.
- [Gillespie24] Gillespie, M., Yang, D., Botsch, M., & Crane, K. Ray Tracing Harmonic Functions. *ACM Trans. Graph* 43. 4. 2024.
- [Ban25] Ban, R., & Valasek, G. Generalized Lipschitz Tracing of Implicit Surfaces. In *Computer Graphics Forum*. 2025, February. Vol. 44, no. 1. P. e15202.

## Кроссплатформенный дифференцируемый рендеринг функций расстояний

А. С. Будаков<sup>1</sup>, А. Р. Гарифуллин<sup>2</sup>, В. А. Галактионов<sup>2</sup>, В. А. Фролов<sup>1</sup>

<sup>1</sup>Институт перспективных исследований проблем искусственного интеллекта  
и интеллектуальных систем МГУ им. М.В. Ломоносова, Москва, Россия

<sup>2</sup>Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В данной работе предлагаются пути улучшения предложенного в 2024 году метода дифференцируемого рендеринга функций расстояния со знаком, а также его кроссплатформенной реализации. Кроссплатформенность даёт возможность использовать широкий спектр графических ускорителей, обеспечивая применимость в различных аппаратных конфигурациях. Для базового метода предлагаются две модификации. Во-первых, используемый метод поиска пересечения луча и функции расстояния, sphere tracing, заменяется адаптированными под дифференцируемый рендеринг альтернативными подходами: методом Ньютона и аналитическим методом. Во-вторых, этап расчёта производных по текстурным и геометрическим параметрам сцены разбивается на два: для внутреннего и граничного интегралов. Это позволяет снизить число выборок метода Монте-Карло для расчёта градиентов по текстуре и разделить вычисление на два шейдера. В результате получен метод втрое быстрее базового, но сохраняющий его точность.

**Ключевые слова:** компьютерная графика, дифференцируемый рендеринг, трассировка лучей, кроссплатформенность

## Cross-platform differentiable signed distance function rendering

A. S. Budakov<sup>1</sup>, A. R. Garifullin<sup>2</sup>, V. A. Galaktionov<sup>2</sup>, V. A. Frolov<sup>1</sup>

<sup>1</sup>IAI Moscow State University, Moscow, Russia

<sup>2</sup>Keldysh Institute of Applied Mathematics, Moscow, Russia

**Abstract.** This work presents improvements to the differentiable signed distance function rendering method proposed in 2024, as well as its cross-platform implementation. Cross-platform nature makes possible the use of a wide range of GPUs, ensuring applicability in various hardware configurations. Two modifications are proposed for the base method. First, the method used to find the intersection of a ray and a distance function, sphere tracing, is replaced by alternative approaches adapted for differentiable rendering: Newton's method and an analytical method. Secondly, the gradient calculation stage with respect to both texture and geometric parameters of the scene is split into two, for internal and boundary integrals. This allows us to reduce the number of Monte Carlo samples for internal gradient estimation and perform calculations in two separate shaders. This results in a method three times faster than the base approach, while retaining its accuracy.

**Keywords:** computer graphics, differentiable rendering, ray tracing, cross-platform

### Введение

Обратный рендеринг – задача восстановления параметров сцены по одному или нескольким изображениям – является важным направлением в компьютерном зрении и графике. В последнее десятилетие, на волне активного внедрения методов машинного обучения, данное направление стремительно развивается в виде дифференцируемого рендеринга, применяющего градиентные методы оптимизации для решения задачи обратного рендеринга. Функции расстояния со знаком (signed distance function, SDF) – вид неявного представления поверхности объекта. Эта функция для точки пространства возвращает расстояние от данной точки до поверхности.

Код подавляющего большинства открытых методов дифференцируемого рендеринга накладывает ограничения на поддерживаемое аппаратное обеспечение – вычисления возможно проводить либо на центральном процессоре, что значительно замедляет процесс оптимизации, либо с использованием технологии CUDA, требующей для своей работы графические карты NVidia. При этом задача дифференцируемого рендеринга с учётом геометрических параметров в сотни раз вычислительно затратнее, чем прямой рендеринг, так как для оптимизации требуется провести большое число итераций, и на каждой, помимо отрисовки, нужно обновить параметры сцены. В связи с этим необходимо, чтобы программное решение не было привязано к определённому оборудованию, а производительность оставалась одним из первых приоритетов, наряду с точностью реконструкции.

## Обзор существующих работ

### 1. Дифференцируемый рендеринг

Все подходы, появившиеся с момента выхода первой статьи по дифференцируемому рендерингу [1], можно объединить в две группы, в зависимости от типа представления, используемого при оптимизации: две основные группы направлений представляют сцену объёмами или поверхностями. К группе объёмных представлений относятся такие методы, как нейронные поля освещённости (NeRF) [2], а также метод Gaussian splatting [3]. Методы, использующие поверхностные представления, характеризуются двумя основными направлениями. Первый – нейронный рендеринг функции расстояния [4, 5, 6], значения которой хранятся в виде векторов признаков для нейросети. Ранние методы [7, 4] использовали одну нейронную сеть для представления всей сцены, что сказывалось на скорости оптимизации. Эта проблема решалась разными способами: разбиением одной сети на несколько, каждая из которых представляла часть сцены [5, 8]; также модифицировался сам алгоритм рендеринга для существующих методов [9]. Второе направление – не-нейронные, аналитические методы, оптимизируют объекты, заданные мешем или функцией расстояния со знаком. Нас интересуют поверхностные методы, в частности аналитические, чтобы восстановленный объект можно было напрямую использовать в физически обоснованном рендеринге.

### 2. Методы аналитического рендеринга поверхностей

В 2018 году появился метод edge sampling [10], впервые позволивший получить аналитические производные по параметрам геометрии, в частности меша, в физически обоснованном рендеринге. В этих условиях цвет пикселя представляет собой интеграл. При дифференцировании по параметрам, задающим форму объекта, подынтегральное выражение терпит разрывы, положение которых зависит от этих параметров. Можно совершить переход к интегралу по области непрерывности подынтегральной функции – тогда от параметра будет зависеть область интегрирования. Для оценки производной этого интеграла необходимо учесть интеграл по разрывам, представляющим собой внешние и внутренние границы объекта. Этот интеграл называется граничным. Edge sampling требовал явного семплирования граничного интеграла, что для меша является нетривиальной задачей. Из-за этого отрисовка одного изображения занимала на порядок больше времени по сравнению с обычным рендерингом. При этом сам процесс оптимизации требует проведения сотен итераций, на каждой из которых нужно отрисовать одно или несколько изображений сцены с разных ракурсов. На практике это приводит к ощутимым затратам времени.

В 2019 и 2020 годах появились два связанных между собой метода, которые избавлялись от необходимости явного семплирования границ объекта при оценке производных. Первый подход, репараметризация [11], предлагал замену переменных, которая локально учитывала разрывы, что позволяло уйти от оценки граничного интеграла вовсе. Однако используемая оценка вносила дополнительные смещение и разброс. Метод 2020 года, warped area sampling (WAS) [12], строго показал, что граничный интеграл в формуле производной возникает согласно транспортной теореме Рейнольдса, и предложил продлить область определения граничного интеграла на пространство с помощью теоремы Остроградского-Гаусса. Для этого перехода строится специальное векторное поле. Этот метод даёт несмещённую оценку, и авторами было показано, как из построенного векторного поля можно получить репараметризацию, не вносящую смещение.

Три подхода к дифференцируемому рендерингу, описанные выше, использовали меши в практической реализации. Однако меши, с точки зрения задачи реконструкции поверхности, имеют ряд фундаментальных проблем, которые делают их использование нежелательным. Их недостатки хорошо описаны и проиллюстрированы в работе 2021 года [13]. Фундаментальной проблемой данного представления является невозможность менять топологию объекта (род поверхности), также во время оптимизации могут возникать необратимые самопересечения и неоптимальные распределения примитивов, из-за чего в местах, требующих высокой детализации, не будет достаточного числа примитивов для их представления. Модификации, представленные в работе, частично решают две последние проблемы, но самопересечения всё ещё возможны, и ограничение на топологию мешает свободному применению мешей для оптимизации. В связи с этим дальнейшие работы этой группы методов используют функции расстояния со знаком.

В 2022 году репараметризацию впервые применили для функций расстояния со знаком [14]: сравнивались два подхода – прямое применение метода WAS к функциям расстояния, а также адаптированную под SDF версию. В этой работе было предложено использовать алгоритм редистансинга. Редистансинг – задача пересчёта значений сетки, чтобы её вершины содержали корректные расстояния до поверхности. Для решения этой задачи применяются методы, вычисляющие на сетке уравнения в частных производных [15], например уравнение Эйконала [16]. В дифференцируемом рендеринге пересчёт нужен, поскольку после шага оптимизации и обновления расстояний в узлах сетки она, строго говоря, уже не является SDF. Наконец, в работе 2024 года [17] предлагается переход от граничного интеграла к объёмному без построения векторного поля. Метод использует свойства функций расстояния и оценивает граничный интеграл *интегралом по релаксированной границе* – тонкой линии вокруг силуэта объекта. Использование SDF позволяет находить точки, принадлежащие этой области, во время трассировки лучей: значение функции расстояния в них будет меньше порогового, определяемого гиперпараметром. Этот метод продемонстрировал лучшие результаты в сравнении с [14], поэтому был выбран в качестве базового. Авторами была опубликована реализация метода на Github [18].

С другой стороны, на методы дифференцируемого рендеринга можно смотреть с точки зрения типа восстанавливаемых параметров. Сложной задачей является одновременное восстановление материала и поверхности объекта [19]. Рассмотренные в этом подразделе методы, в том числе базовый, могут использоваться для её решения. Однако в данной работе преследовалась цель восстановления только поверхности объекта. Тем не менее, предложенный метод не вносит никаких ограничений в сравнении с базовым ни в задачу восстановления поверхности, ни в задачу одновременного восстановления поверхности и материала.

### 3. Мотивация построения нового метода

В [14, 17] SDF задана на регулярной сетке, и для поиска пересечения луча и функции расстояния применяется метод sphere tracing [20]. Этот итеративный метод, предложенный в 1995 году, вычисляет значение расстояния в точке на луче, а затем делает шаг на это расстояние, пока либо не пересечёт поверхность, либо не выйдет за пределы сцены. Sphere tracing является наиболее распространённым методом для поиска пересечения луча и SDF, однако существуют альтернативы. В [21] были предложены два метода, основанные на следующем факте: внутри вокселя функция расстояния вдоль луча представима полиномом третьей степени (при использовании трилинейной интерполяции). Искать нули этого полинома можно либо явно вычисляя вещественные корни – этот метод авторы называли *аналитическим*, – либо используя итерационный метод, в работе применяется метод Ньютона. По результатам сравнений, проведённых в [22], метод Ньютона и аналитический метод показали более высокую скорость в сравнении со sphere tracing в прямом рендеринге. Вероятно, они также способны повысить производительность дифференцируемого рендеринга. Помимо этого, в алгоритм базового метода можно внести изменения, которые ускорят семплирование.

### Предложенный метод

Метод, предлагаемый в данной статье, во-первых, адаптирует алгоритмы поиска пересечения луча и функции расстояния из [21] под дифференцируемый рендеринг, а во-вторых, для базового метода реализует явное семплирование границы. Разделение расчёта внутреннего и граничного интеграла позволит задать разное число семплов для их оценки и, в частности, уменьшить это число для внутреннего. Подробнее про них рассказано в подразделе, посвящённом явному семплированию. На рисунке 1 показана схема одной итерации базового алгоритма. Сперва выбирается ракурс, генерируются лучи с использованием параметров соответствующей камеры. Затем методом sphere tracing проводится рендеринг сцены: поиск пересечения лучей и поверхности объекта, а также поиск точек релаксированной границы; оцениваются внутренний и граничный интеграл. Наконец, считается функция потерь между референсным и полученным изображением, происходит обратное распространение ошибки. Красным цветом на рисунке 1 показаны предложенные модификации. Метод Ньютона [21] заменяет sphere tracing [20] в качестве метода поиска пересечения на этапе прямого рендеринга; на этапе явного семплирования идёт поиск точек для оценки граничного интеграла.



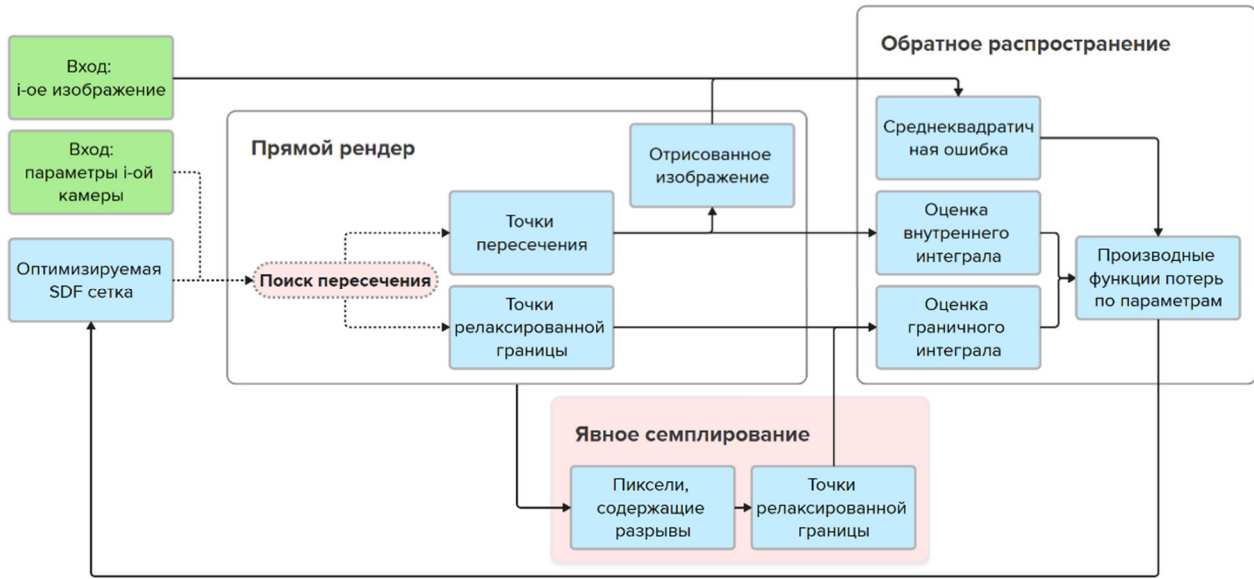


Рисунок 1. Схема одной итерации базового и предложенного методов. Зелёным цветом обозначены входные данные; синим – внутренние промежуточные. Стрелками показаны зависимости в их расчёте. Красным цветом выделены предложенные модификации

### 1. Явное семплирование

$$I = \int_{\Omega} f(\omega; \theta) d\omega. \quad (1)$$

В физически обоснованном рендеринге функция расчёта цвета пикселя представляет собой интеграл (1), где  $\theta$  – параметры сцены. При попытке дифференцирования по параметрам, задающим форму объекта, подынтегральная функция  $f$  терпит разрывы, и положение этих разрывов зависит от параметров, по которым проводится дифференцирование. Поэтому в [12] исходный интеграл разбивается на сумму интегралов по областям непрерывности подынтегрального выражения  $D_i$  (2), при этом область интегрирования полученных слагаемых будет зависеть от параметров сцены. Согласно обобщению формулы Лейбница для дифференцирования под знаком интеграла, названному транспортной теоремой Рейнольдса, в этом случае производная интеграла будет равна

$$\frac{\partial I}{\partial \theta} = \sum_i \int_{D_i} \frac{\partial f(\omega; \theta)}{\partial \theta} d\omega + \sum_i \oint_{\partial D_i} f(\omega; \theta) \langle \partial_{\theta} \omega, \hat{n} \rangle d\omega, \quad (2)$$

где  $\partial D_i$  – граница области  $D_i$ ,  $\hat{n}$  – нормаль, направленная наружу. Первое слагаемое – это *внутренний* интеграл, второе – *граничный*. Идея базового метода [17] заключается в приближении граничного интеграла интегралом по релаксированной границе  $A$ :

$$\oint_{\partial D} f(\omega; \theta) \langle \partial_{\theta} \omega, \hat{n} \rangle d\omega \approx I_A = \int_A \frac{1}{l(\omega)} f(\omega; \theta) \langle \partial_{\theta} \omega, \hat{n} \rangle d\omega, \quad (3)$$

$$l(\omega) = \frac{\varepsilon}{\|x - x^*\|}.$$

Здесь  $l(\omega)$  – ширина релаксированной границы, наблюдаемая из точки  $x$ , откуда пускаются лучи,  $x^*$  – точка релаксированной границы на луче с направлением  $\omega$ ,  $\varepsilon$  – гиперпараметр, пороговое значение для SDF.

В базовом методе внутренний и граничный интегралы оцениваются одновременно, с использованием одних и тех же семплов; для корректного восстановления поверхности авторы выбрали 64 семпла на пиксель. Однако для восстановления текстуры объекта можно использовать меньшее значение. Идея явного семплирования заключается в переносе расчёта граничного интеграла в отдельный шейдер (рис. 2). В отличие от основного, в нём не происходит обновление параметров цвета, поэтому отсутствует

сопутствующая логика. Благодаря разделению, новый шейдер позволяет сделать больше семплов для оценки граничного интеграла, уменьшив число семплов для основного шейдера. Если в оригинальной статье использовалось 64 семпла на пиксель, то в предложенном методе для основного шейдера это число снижается до 16. Для граничных пикселей выбрано 128 семплов.

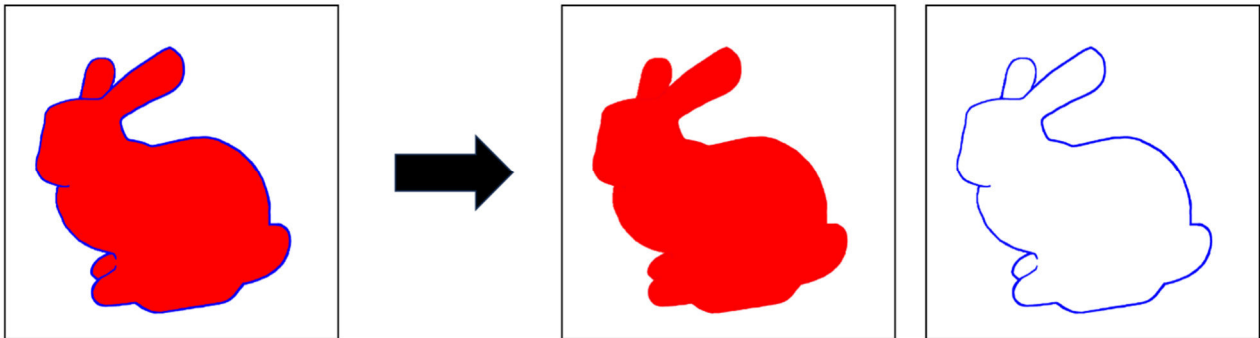


Рисунок 2. Визуализация явного семплирования. Вместо одновременного оценивания внутреннего и граничного интегралов (сэмплы красного и синего цвета соответственно) расчёт проводится сначала для внутреннего, затем для граничного

## 2. Альтернативные методы поиска пересечения

В качестве альтернативы sphere tracing добавлены метод Ньютона и аналитический метод. Эти методы обходят каждый воксель, в который попал луч, пока не будет найдено пересечение или луч не выйдет за пределы сцены. Для регулярной сетки, содержащей большое число пустых вокселей, эти методы были реализованы таким образом, чтобы во время поиска пересечения игнорировались воксели, если все его значения больше порогового значения, задающего релаксированную границу.

Также требовалось адаптировать для них поиск релаксированной границы. Точка, лежащая на луче, принадлежит релаксированной границе, если расстояние до поверхности в этой точке меньше порогового значения  $\varepsilon$ , а производная функции расстояния вдоль луча в ней равна нулю, и эта точка является точкой локального минимума (рис. 3). Оба метода основаны на том, что внутри вокселя функция расстояния вдоль луча представима в виде полинома третьей степени, для которого разными способами находятся корни. Таким образом, для поиска точек релаксированной границы в аналитическом методе и методе Ньютона нужно в каждом вокселе находить точки локального минимума полинома третьей степени, а затем проверять значение функции расстояния в них. Метод Ньютона уже находит нули производной для начального приближения, и поиск нужных значений внутри вокселя требует минимального усложнения логики алгоритма, поэтому он используется в сравнениях. Кроме того, при использовании трилинейной интерполяции нередки случаи, когда локальный минимум достигается на границе вокселей; этот случай обрабатывается отдельно.

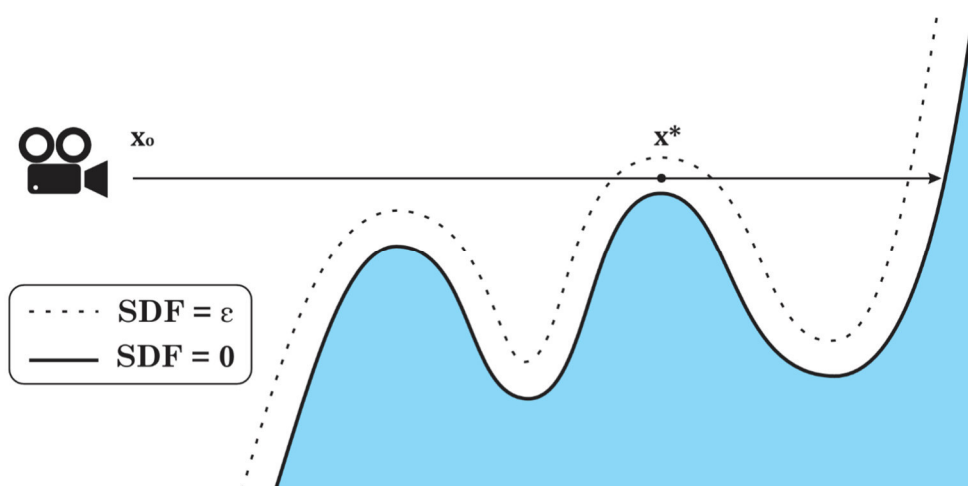


Рисунок 3. Визуализация пересечения луча, выпущенного из точки  $\mathbf{x}_0$ , с SDF. Точка  $\mathbf{x}^*$  принадлежит релаксированной границе, так как 1) SDF вдоль луча достигает в  $\mathbf{x}^*$  локального минимума, 2)  $SDF(\mathbf{x}^*) < \varepsilon$

### Детали реализации

В качестве инструментов разработки выбраны язык C++ и Vulkan для поддержки наиболее широкого спектра графических процессоров. Система Kernel Slicer [23] используется для переноса C++ кода на GPU. Разработка началась с реализации системы, которая может отрисовать объект, представленный функцией расстояния, методом sphere tracing, с поддержкой цвета и простой модели освещения. Следующим шагом была модификация функции рендеринга, которая, во-первых, добавляет поиск точек релаксированной границы в sphere tracing, а во-вторых, аналитически рассчитывает и распространяет производные по параметрам SDF – системы автоматического дифференцирования не используются. Для обеспечения параллелизма производные по параметрам записываются в отдельный массив. Затем, на этапе расчёта функции потерь, каждый элемент массива умножается на её производную, после чего применяется регуляризация, оптимизатор Adam и происходит обновление параметров. В конце для обновлённой сетки применяется алгоритм редистансинга для пересчёта и корректировки значений. Регуляризация имеет вид свёртки с дискретным оператором Лапласа, написана параллельная версия на CPU. Для редистансинга реализована параллельная версия метода fast sweeping [24], выполняется на CPU.

Эталонная и собственная реализации использовали одинаковую конфигурацию для экспериментов. Применяется модель освещения Ламберта с одним источником света. Значение гиперпараметра  $\varepsilon$  взято из оригинальной работы [17] и равно  $10^{-4}$ . Для восстановления используются изображения размером 512x512 пикселей, полученные отрисовкой модели-референса с 16 ракурсов. Их положение соответствует ракурсам эталонной реализации [18], полученным с параметрами  $\theta = 8$ ,  $\varphi = 4$ . На каждой итерации расчёт производных проводился для четырех ракурсов из 16. Оптимизация занимала 1000 итераций; каждые 200 итераций разрешение сетки увеличивалось в два раза, размер финальной сетки 512<sup>3</sup>.

### Результаты экспериментов

Эксперименты включают в себя сравнение средней скорости работы одной итерации базового и предложенного метода, а также валидацию, демонстрирующую сохранение качества реконструкции. В качестве GPU было выбрано Nvidia RTX 4090, так как эталонная реализация [18] написана с использованием CUDA. Эксперименты проводились для эталона и предложенных методов на идентичных сценах и ракурсах камер. Использовались шесть классических моделей: Bunny, Armadillo, Happy Buddha, Asian dragon – модели из Стэнфордского репозитория сканов [25]; Utah teapot – модель чайника за авторством Мартина Ньюэлла; Nefertiti – скан бюста Нефертити [26]. Модели содержат разные по степени детализированности поверхности, что подходит для валидации предлагаемого метода в сравнении со старым.

#### 1. Сравнение скорости

В проводимом сравнении производительности и в данной работе, и в эталонной реализации [18] замерялось время работы шейдеров, так как в обеих программах логика трассировки сцены, поиска точек релаксированной границы и оценки обоих интегралов выполняется полностью на GPU. Для тестов, использующих явное семплирование границы, время означает среднее суммарное время работы основного и нового шейдеров; в остальных случаях указано среднее время выполнения основного шейдера. На рисунке 4 приведён график сравнения скорости для рассматриваемых представлений. Тестировались эталонная реализация [18], собственная реализация базового метода, а также две предложенные модификации и их комбинация. Замерялось среднее время выполнения одной итерации алгоритма оптимизации на графическом процессоре для шести моделей. Из полученных результатов сравнения можно сделать следующие выводы:

1. Собственная реализация базового метода сопоставима с эталонной в скорости.
2. Каждая из предложенных модификаций ускоряет базовый метод.
3. Итоговый метод показывает лучший результат, ускоряя базовый метод в три раза.

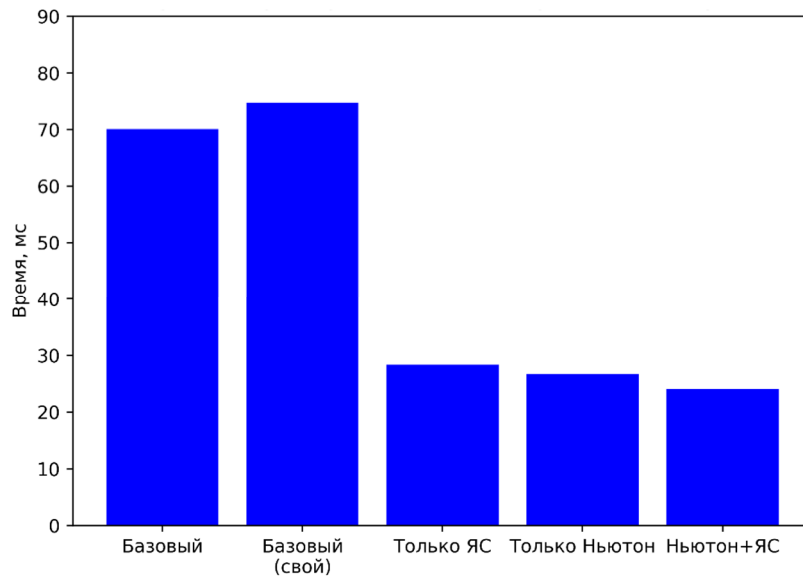


Рисунок 4. Среднее время работы одной итерации (GPU ядро)

Дополнительно проводилась оптимизация на других GPU. На рисунке 5 показано время работы на интегрированных графических ускорителях AMD Radeon Graphics (процессор AMD Ryzen 9 7950X) и Intel Iris Xe Graphics (процессор Intel Core i5-11300H). Базовый метод не включён в замеры, поскольку эталонная реализация не поддерживает GPU этих производителей.

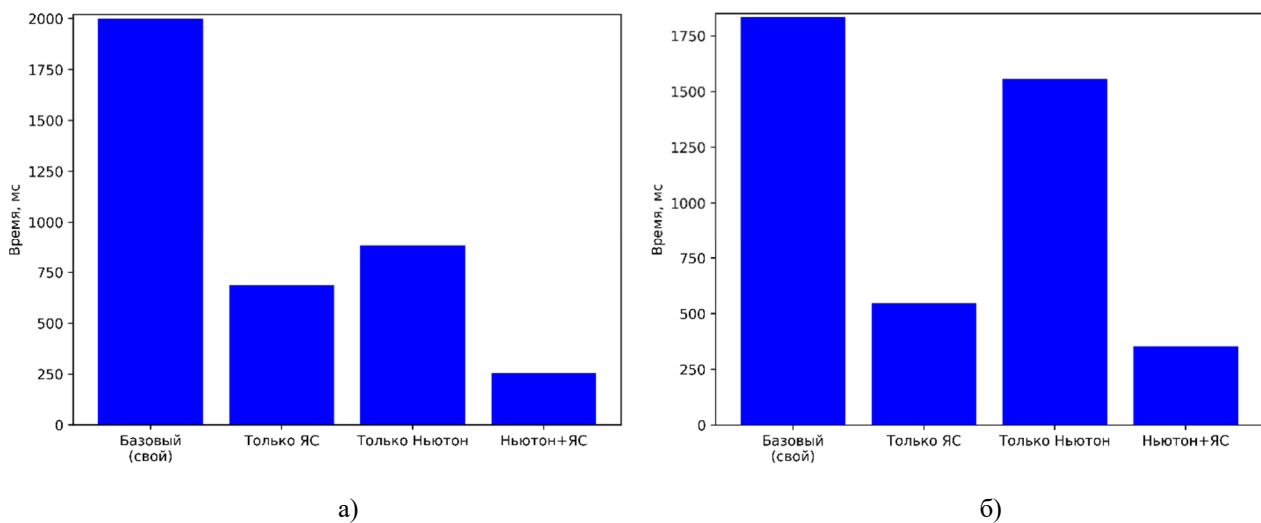


Рисунок 5. Среднее время работы одной итерации: а) на AMD Radeon Graphics; б) Intel Iris Xe Graphics

## 2. Валидация

Для валидации метода проведено сравнение результатов работы базового и итогового методов на рассматриваемых моделях с исходным мешем. После окончания процесса оптимизации сравнивалось среднее значение PSNR между восстановленной моделью и референсом на ракурсах, использованных в оптимизации.

На рисунке 6 приведено сравнение качества для эталона, своей реализации базового метода и метода со всеми модификациями. Очевидно, что качество реализованных методов сопоставимо с эталонной реализацией. Предложенный метод показывает немного лучшее качество в сравнении с базовым, так как он делает больше семплов для оценки интеграла, получая более точные градиенты. На рисунке 7 показаны рендеры исходного меша и восстановленной SDF-сетки для всех рассмотренных моделей. На рисунке 8 показаны различия между парами этих рендеров для каждой модели.

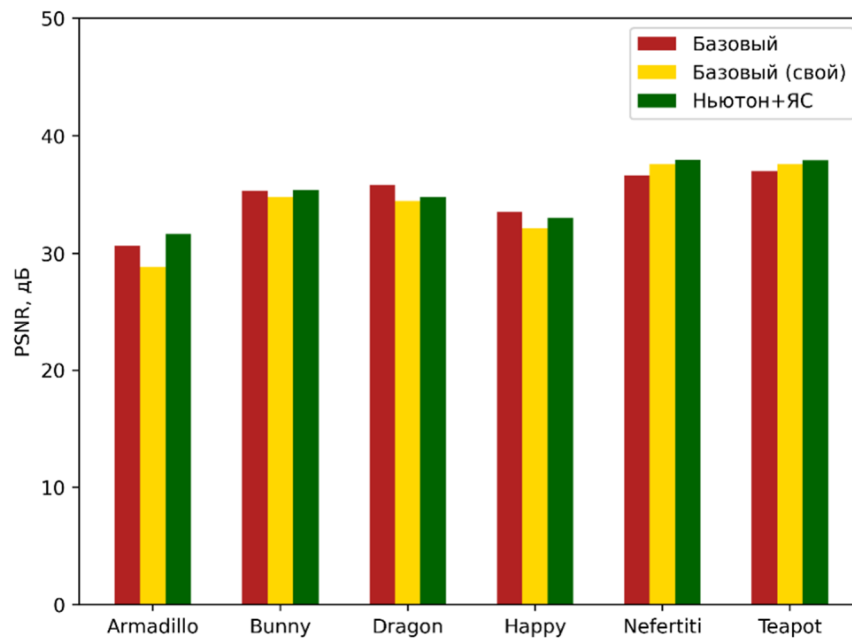


Рисунок 6. Среднее по ракурсам значение PSNR для восстановленных моделей

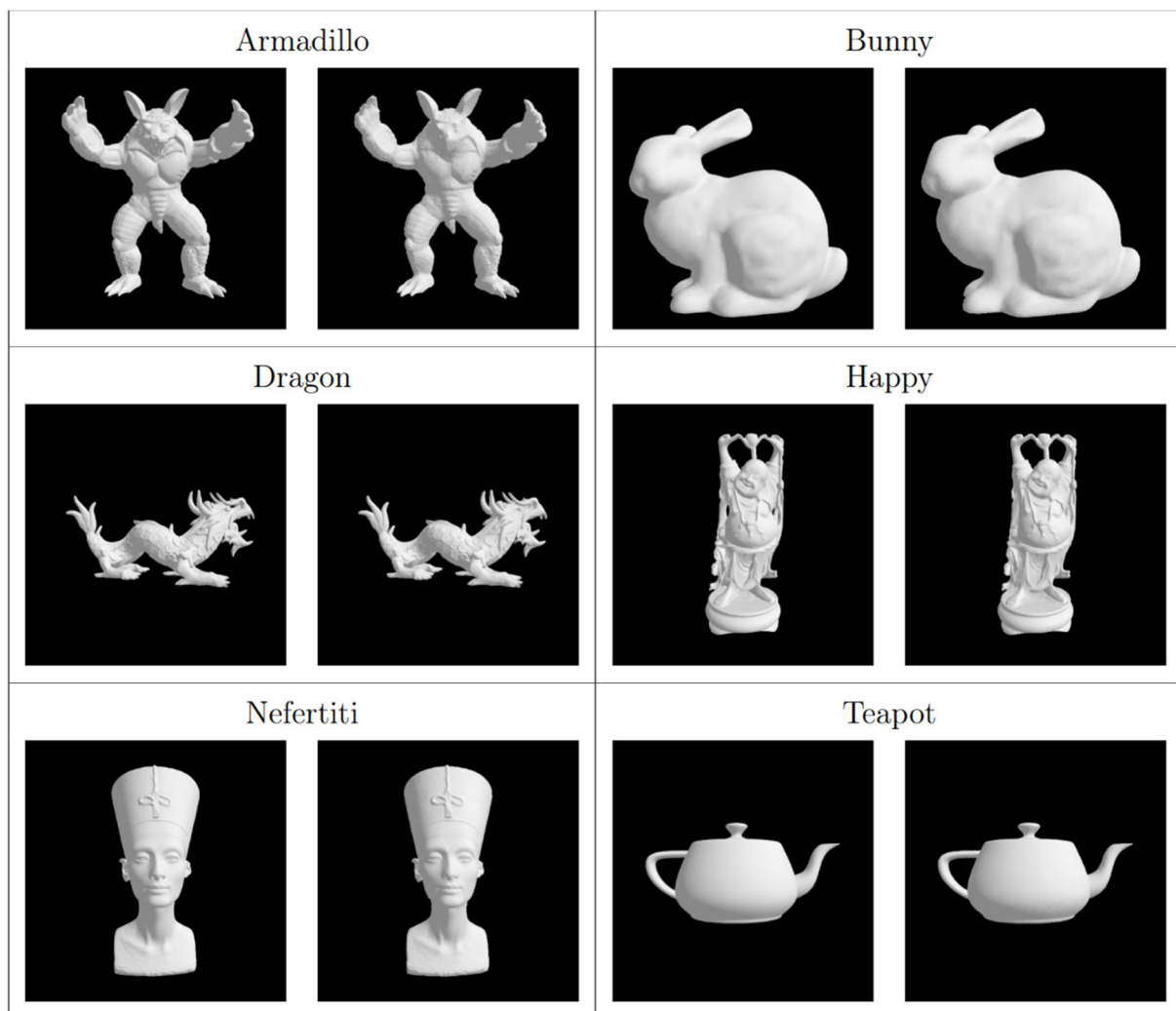


Рисунок 7. Сравнение исходного меша (слева) и его реконструкции (справа) для рассматриваемых моделей

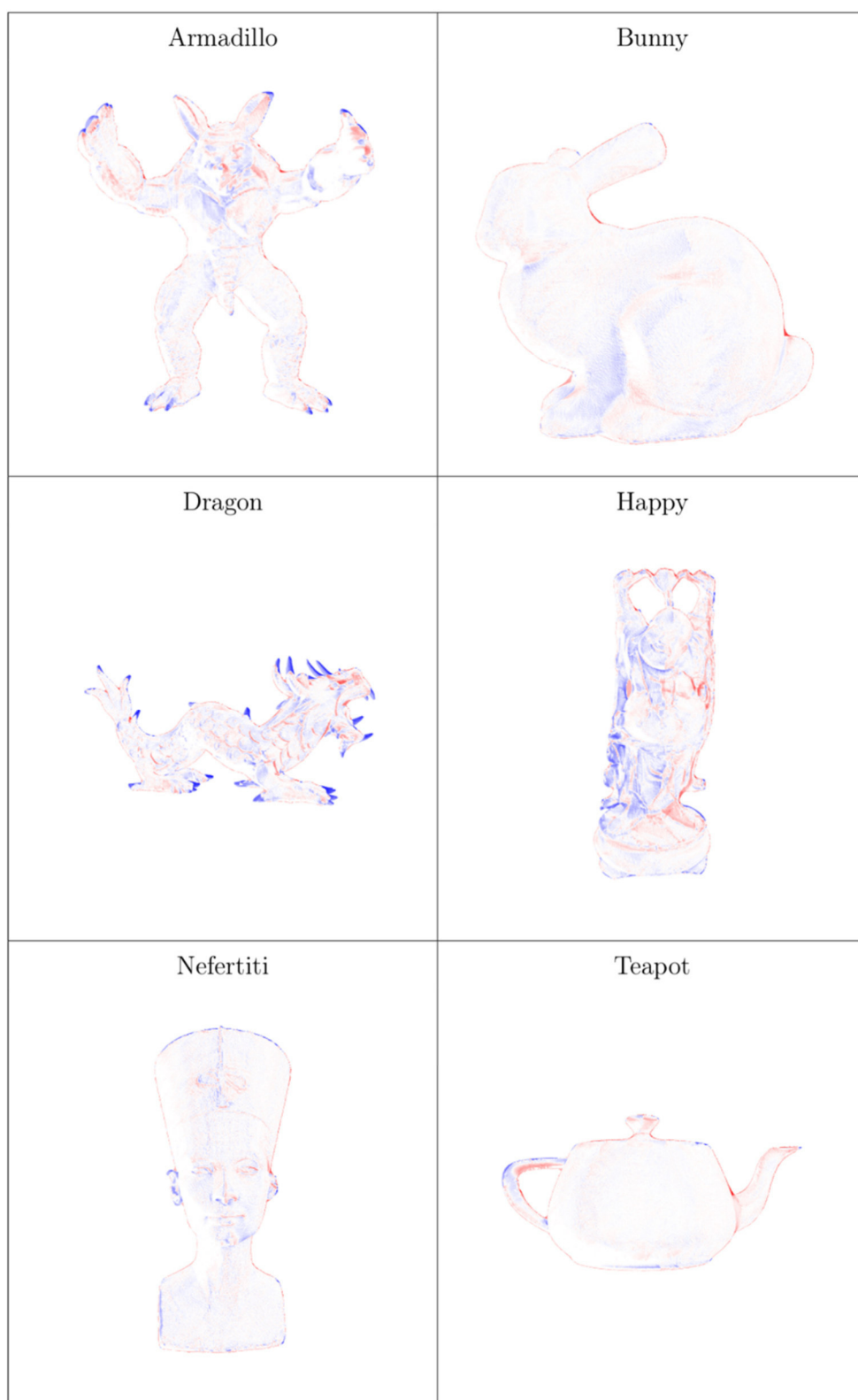


Рисунок 8. Различие между рендером восстановленной модели и исходным мешем (для лучшей видимости изменён контраст). Синим и красным цветом показаны отрицательное и положительное различия соответственно. Синие области на границах модели означают отсутствие у восстановленной модели поверхности там, где она есть у референса. Красные, наоборот, указывают на наличие поверхности там, где её нет у исходного меша



## Закключение

В данной работе реализована кроссплатформенная программная система дифференцируемого рендеринга SDF, основанная на подходе с использованием интеграла по релаксированной границе, с использованием C++ и Vulkan. Реализация базового метода сравнима с авторской реализацией по производительности в задаче реконструкции геометрии. Далее были введены модификации, значительно ускоряющие процесс оптимизации без потери качества. Реализованный метод не вводит новых ограничений, но содержит ограничение, унаследованное от используемого SDF-представления: для представления мелких деталей требуются воксели маленького размера. Для SDF-сетки это достигается исключительно повышением разрешения всей сетки, что значительно увеличивает размер модели. При этом большая часть вокселей представляет пустое пространство, а значит, информация о них не требует хранения. Дальнейшая работа будет посвящена адаптации реализованного метода под разреженные представления.

## Список литературы

1. M. M. Loper and M. J. Black, "OpenDR: An Approximate Differentiable Renderer", in *Computer Vision - ECCV* 2014, Zurich, Switzerland, 2014.
2. B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi and R. Ng, "NeRF: representing scenes as neural radiance fields for view synthesis", *Commun. ACM*, vol. 65, no. 1, p. 99–106, 2021.
3. B. Kerbl, G. Kopanas, T. Leimkuehler and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering", *ACM Trans. Graph.*, vol. 42, no. 4, pp. 1-14, 2023.
4. P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura and W. Wang, "NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction", *Advances in Neural Information Processing Systems*, 2021.
5. T. Müller, A. Evans, C. Schied and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding", *ACM Trans. Graph.*, vol. 41, no. 4, pp. 1-15, 2022.
6. Z. Li, T. Müller, A. Evans, R. H. Taylor, M. Unberath, M.-Y. Liu and C.-H. Lin, "Neuralangelo: High-Fidelity Neural Surface Reconstruction", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 2023.
7. J. J. Park, P. Florence, J. Straub, R. Newcombe and S. Lovegrove, "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, California, 2019.
8. Y. Wang, Q. Han, M. Habermann, K. Daniilidis, C. Theobalt and L. Liu, "NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction", *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 2023.
9. A. Dogaru, A. T. Ardelean, S. Ignatyev, E. Zakharov and E. Burnaev, "Sphere-Guided Training of Neural Implicit Surfaces", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, California, 2023.
10. T.-M. Li, M. Aittala, F. Durand and J. Lehtinen, "Differentiable Monte Carlo ray tracing through edge sampling", *ACM Trans. Graph.*, vol. 37, no. 6, pp. 1-11, 2018.
11. G. Loubet, N. Holzschuch and W. Jakob, "Reparameterizing discontinuous integrands for differentiable rendering", *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1-14, 2019.
12. S. P. Bangaru, T.-M. Li and F. Durand, "Unbiased warped-area sampling for differentiable rendering", *ACM Trans. Graph.*, vol. 39, no. 6, pp. 1-18, 2020.
13. B. Nicolet, A. Jacobson and W. Jakob, "Large steps in inverse rendering of geometry", *ACM Trans. Graph.*, vol. 40, no. 6, pp. 1-13, 2021.
14. D. Vicini, S. Speierer and W. Jakob, "Differentiable signed distance function rendering", *ACM Trans. Graph.*, vol. 41, no. 4, p. 18, 2022.
15. J. A. Sethian, "A Fast Marching Level Set Method for Monotonically Advancing Fronts", *National Academy of Sciences of the United States of America*, Washington, 1996.
16. H. Zhao, "A fast sweeping method for Eikonal equations", *Math. Comp.*, vol. 74, pp. 603-627, 2005.
17. Z. Wang, X. Deng, Z. Zhang, W. Jakob and S. Marschner, "A Simple Approach to Differentiable Rendering of SDFs", *SIGGRAPH Asia 2024 Conference Papers*, Tokyo, 2024.
18. Z. Wang, "The official implementation for 'A Simple Approach to Differentiable Rendering of SDFs'". Available: <https://github.com/zichenwang01/relaxed-boundary>. [Дата обращения: 07. 07. 2025].
19. F. Luan, S. Zhao, K. Bala and Z. Dong, "Unified Shape and SVBRDF Recovery using Differentiable Monte Carlo Rendering", *Computer Graphics Forum*, vol. 40, no. 4, pp. 101-113, 2021.

- 
20. J. Hart, "Sphere Tracing: A Geometric Method for the Antialiased Ray Tracing of Implicit Surfaces", *The Visual Computer*, vol. 12, p. 527–545, 1996.
  21. H. Hansson Söderlund, A. Evans and T. Akenine-Möller, "Ray Tracing of Signed Distance Function Grids", *Journal of Computer Graphics Techniques (JCGT)*, vol. 11, no. 3, pp. 94–113, 2022.
  22. A. S. Budak, A. R. Garifullin, V. A. Frolov and V. A. Galaktionov, "Study of Surface Representation Methods Based on Signed Distance Functions", *Programming and Computer Software*, vol. 51, p. 131–139, 2025.
  23. V. A. Frolov, V. V. Sanzharov and V. A. Galaktionov, "kernel slicer: high-level approach on top of GPU programming API", *Ivannikov Ispras Open Conference (ISPRAS)*, Moscow, 2022.
  24. M. Detrixhe, F. Gibou and C. Min, "A parallel fast sweeping method for the Eikonal equation", *Journal of Computational Physics*, vol. 237, pp. 46–55, 2013.
  25. Stanford Computer Graphics Laboratory, "The Stanford 3D Scanning Repository". Available: <https://graphics.stanford.edu/data/3Dscanrep/>. [Дата обращения: 07. 07. 2025].
  26. J. N. Nelles and N. Al-Badri, "Nefertiti scan", 2015. Available: <https://nefertitihack.alloversky.com/>. [Дата обращения: 07. 07. 2025].

## SDFMultiOctree: неявное представление поверхностей на основе функций расстояния со знаком

Н. Д. Майоров<sup>1</sup>, А. Р. Гарифуллин<sup>2</sup>, А. Г. Волобой<sup>2</sup>, В. А. Фролов<sup>1</sup>

<sup>1</sup>Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ им. М.В. Ломоносова, г. Москва, Россия

<sup>2</sup>Институт прикладной математики им. М.В. Келдыша РАН, г. Москва, Россия

**Аннотация.** Авторами предлагается метод представления поверхностей на основе октодеревьев и функций расстояния со знаком (Signed Distance Function, SDF). В основе данного метода лежат два принципиальных нововведения. Во-первых, мы комбинируем несколько независимых SDF в одном узле октодерева, что позволяет представлять более сложные элементы модели без углубления октодерева. Во-вторых, мы помечаем отдельные SDF как тонкие, что дает возможность представлять некоторые элементы моделей более точно, чем в существующих методах на основе SDF. Предложенный метод превосходит базовый по точности, сохраняя как скорость рендеринга, так и потребляемый объем памяти.

**Ключевые слова:** функция расстояния со знаком, трассировка лучей, неявное представление.

## SDFMultiOctree: implicit representation of surfaces based on signed distance functions

N. D. Mayorov<sup>1</sup>, A. R. Garifullin<sup>2</sup>, A. G. Voloboy<sup>2</sup>, V. A. Frolov<sup>1</sup>

<sup>1</sup>IAI Moscow State University, Moscow, Russia

<sup>2</sup>Keldysh Institute of Applied Mathematics, Moscow, Russia

**Abstract.** This paper proposes a method for surface representation based on octrees and signed distance functions (SDF). The method is founded on two key innovations. First, we combine multiple independent SDFs within a single octree node, enabling the representation of more complex model elements without deepening the octree. Second, we mark certain SDFs as thin, which allows us to represent specific model components more accurately than existing SDF-based methods. The proposed approach outperforms the baseline method in terms of accuracy while maintaining both rendering speed and memory efficiency.

**Keywords:** signed distance function, ray tracing, implicit representation.

### Введение

В последние годы наблюдается рост популярности функций расстояния со знаком (SDF) не только благодаря их применению в генеративных моделях [1–3], задачах 3D-реконструкции [4–6], но и за счет эффективной трассировки лучей по SDF-представлениям в различных приложениях компьютерной графики [7, 8]. SDF позволяют компактно и гибко описывать сложные формы и топологии, что особенно важно для современных приложений, требующих высокой детализации и реалистичности.

В качестве базового метода хранения и рендеринга разреженных SDF-данных широко используются структуры типа Sparse Voxel Octree (SVO) [9, 10]. В данной работе предлагается модификация SVO, предназначенная для хранения вокселей с интерполированной SDF внутри, что расширяет возможности базового метода при работе с моделями, не удовлетворяющими традиционным ограничениям: открытые поверхности, неидеальная топология, тонкие детали. Наш подход позволяет корректно описывать и визуализировать такие объекты, делая трассировку лучей по SDF более универсальным инструментом для рендеринга сложной геометрии в реальном времени.

### Обзор работ

Наиболее близкими к нашему методу являются структуры, сочетающие преимущества октодерева и локальных представлений SDF, а также современные нейросетевые методы.

#### 1. Классические структуры

Один из наиболее популярных подходов для хранения разреженных воксельных данных – Sparse Voxel Octree (SVO) [9, 10]. SVO хранит информацию только о занятых областях, а данные о вокселях часто агрегируются на уровне родителей для экономии памяти. Современные модификации SVO, например [11], позволяют хранить в листьях коэффициенты полиномиальной интерполяции SDF или увеличивать объем пустых областей на этапе предобработки. Такой подход дает возможность более

гибко подходить к хранению данных на границе объектов и использовать различные методы интерполяции, однако требует решения аналитических задач поиска корней, что может быть вычислительно затратно. Предложенные в данной статье модификации позволяют улучшить как раз подобные решения в ряде ситуаций, когда воксель не может содержать поверхность оригинальной модели. В ряде задач реконструкции геометрии, где модель постоянно меняется, для эффективного обновления данных вместо октодерева применяются хеш-таблицы [12, 13].

Другим заметным решением является Sparse Brick Set (SBS) [7], в котором используется комбинация BVH и регулярных воксельных решёток (bricks) внутри листьев дерева. Такой подход позволяет эффективно вычислять пересечения лучей с SDF с помощью аналитических и численных методов (например, метода Ньютона), используя значения SDF на углах вокселей, которые должны явно храниться в нужных точках.

Структура VDB [14] реализует иерархические многоуровневые решётки, которые обеспечивают балансировку дерева и ускоряют её обход. VDB хранит битовые маски активных вокселей, но не сжимает значения SDF, а сами иерархические решётки в целом характеризуются невысокой производительностью при рендеринге. Оптимизированная реализация трассировки лучей по VDB-структурам на GPU была предложена в [15], где для обхода используется 3D-дифференциальный анализатор.

## 2. Нейросетевые представления

В последние годы активно развиваются методы, использующие нейронные сети для моделирования SDF. Архитектура SIREN [16] использует периодические функции активации для улучшения гладкости и качества неявных представлений. DeepSDF [17] предлагает обучать нейронную сеть для восстановления непрерывной SDF по латентному вектору, кодирующему форму.

NGLOD [18] сочетает SVO и нейросетевой декодер: на каждом уровне октодерева хранятся векторные признаки на углах вокселей, которые вместе с координатами подаются в компактную нейросеть-декодер для оценки SDF. Это позволяет ускорять трассировку лучей без существенной потери качества и поддерживать несколько уровней детализации. Аналогичный подход реализован в CoFie [19], где форма разбивается на локальные патчи, каждый из которых моделируется отдельной нейросетью.

Интересное решение предложено в N-BVH [20], где нейросеть непосредственно предсказывает точку пересечения луча с поверхностью, используя хеш-энкодер. Такой подход обеспечивает высокую степень сжатия для сложных сцен, но уступает по качеству для простых объектов и требует гибридизации с классическими структурами для достижения баланса между качеством и эффективностью.

Также эффективным и популярным методом является Instant Neural Graphics Primitives (Instant NGP) [12], в котором используется многорезолюционный хеш-энкодер для компактного представления SDF и других неявных функций. Такой подход обеспечивает высокую скорость обучения и визуализации, а также масштабируемость для сложных сцен и различных графических задач.

## 3. Особенности и ограничения базового метода

В качестве базового метода для представления разреженных SDF-данных в настоящей работе выбран Sparse Voxel Octree (SVO) [9, 10]. SVO широко применяется благодаря своей компактности, способности адаптировать плотность разбиения к локальной сложности геометрии и эффективной поддержке операций поиска и трассировки лучей. Иерархическая структура октодерева позволяет автоматически повышать детализацию в областях с мелкими деталями, при этом экономя память за счет пропуска пустых регионов. Благодаря этим преимуществам, SVO стал де-факто стандартом для хранения и рендеринга сложных трёхмерных сцен с использованием SDF.

Однако классический подход к построению октодерева обладает рядом ограничений, которые становятся критичными в современных приложениях. Во-первых, для корректной работы требуется строго замкнутая поверхность модели. В противном случае в местах разрывов могут появляться заметные артефакты при визуализации. Во-вторых, для корректного отображения тонких элементов, ширина которых меньше размера вокселя, требуется локальное увеличение глубины дерева, что приводит к росту объёма данных и снижению производительности. Эта проблема также иллюстрируется на рисунке 1, а возникающие на практике артефакты заметны на рисунке 2. Кроме того, если модель

состоит из нескольких независимых частей (рис. 3), стандартный подход может некорректно заполнять пространство между ними, что ухудшает точность представления сложных объектов.

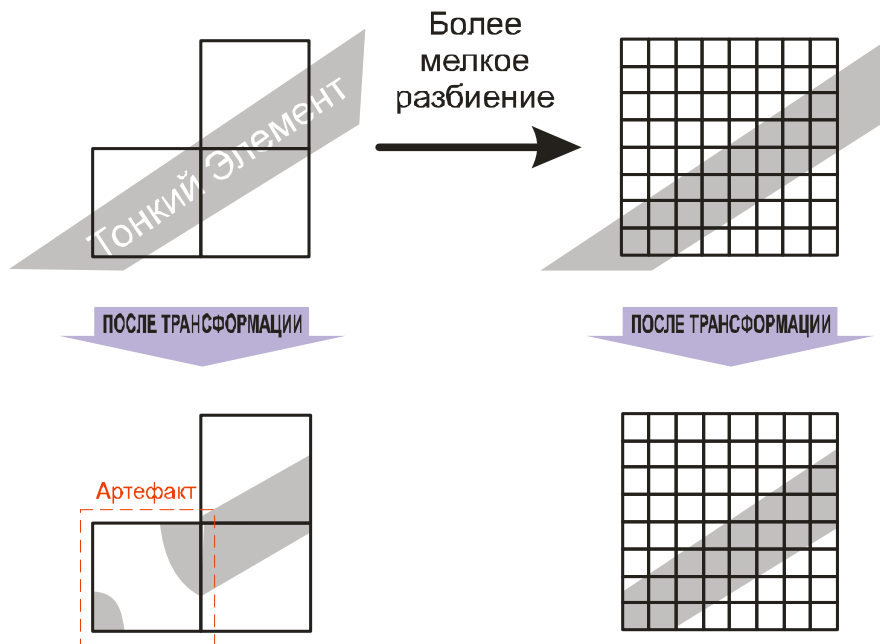


Рисунок 1. Слева: иллюстрация артефактов, возникающих при наличии тонких элементов у модели (из-за линейной интерполяции элемент модели может стать «дырявым» или вовсе пропасть). Справа: наивный способ решения этой проблемы – локальное увеличение глубины октодеревя

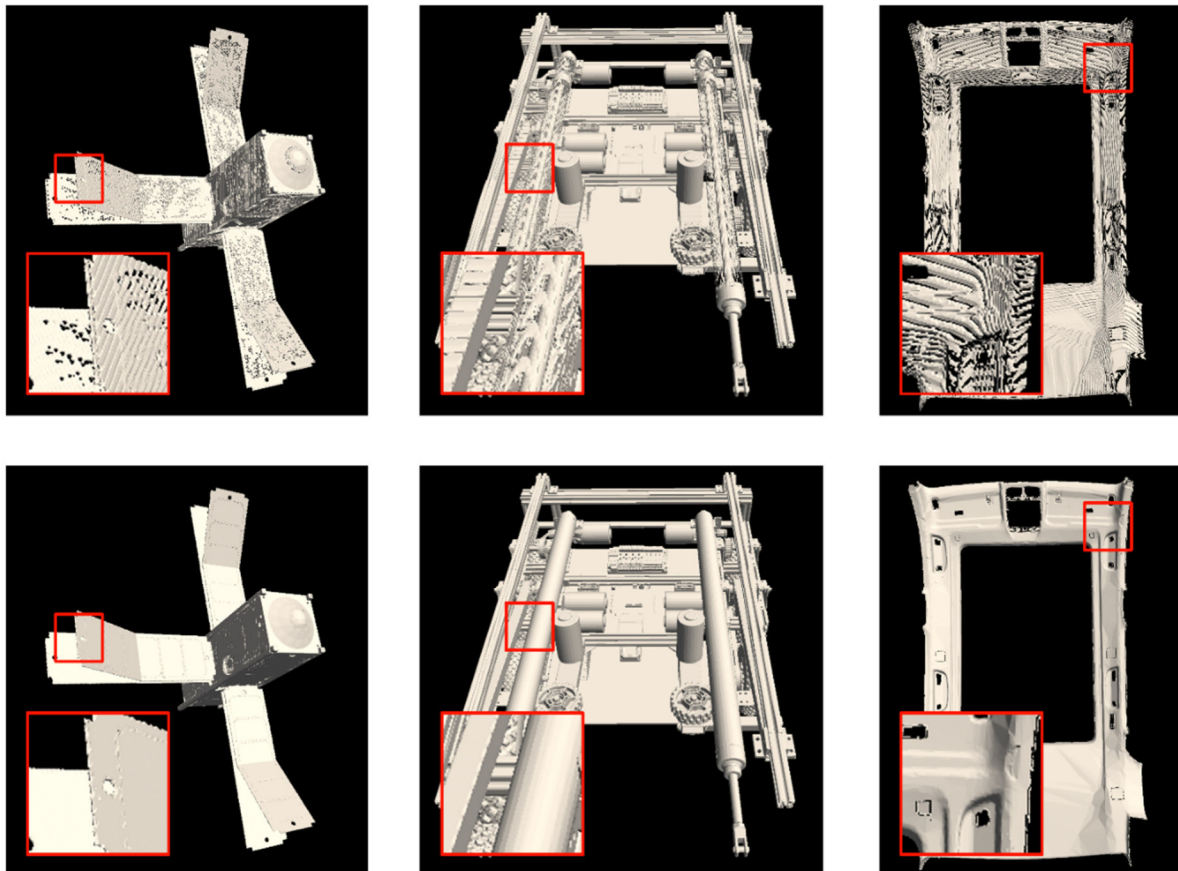


Рисунок 2. Сверху: октодеревья глубины 9, построенные базовым методом по моделям (слева направо) ABC 80006, ABC 83870 и ABC 515447. Снизу: октодеревья глубины 9, построенные нашим методом по тем же моделям. На этих моделях отчётливо видно артефакты, которые образуются при попытке приблизить тонкие элементы геометрии

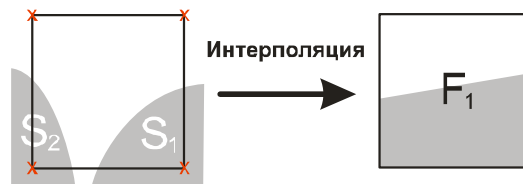


Рисунок 3. Иллюстрация артефактов, возникающих из-за близкого расположения разных частей модели (на расстоянии меньшем, чем размер вокселя)

### Предложенные модификации

Для устранения выявленных недостатков и повышения гибкости структуры были разработаны несколько модификаций базового алгоритма. Эти изменения направлены на улучшение качества аппроксимации поверхности, особенно в случаях наличия тонких элементов и незамкнутых поверхностей, однако предложенные изменения требуют добавления небольшого количества дополнительной информации в узлы октодеревя.

#### 1. Множественные узлы

Основная идея первой модификации заключается в том, что каждый узел октодеревя может содержать не один, а несколько вокселей данных, каждый из которых соответствует отдельной части модели, находящейся в пределах данного региона пространства, как на рисунке 4. Такой подход позволяет в рамках одного узла описывать сложные случаи, например, резкие стыки между поверхностями или несколько не связанных между собой фрагментов объекта, которые не могут быть корректно аппроксимированы одним вокселем. Для построения такой структуры необходимо определить, какие полигоны исходной модели относятся к какой компоненте внутри узла. В случае мешей это реализуется разбиением пересекающихся с узлом полигонов на связные компоненты, для каждой из которых строится отдельный воксель. Если в исходной модели присутствует информация о принадлежности полигонов к частям объекта, её также можно использовать для инициализации множественных вокселей, хотя на практике эта информация не всегда бывает полной или корректной.

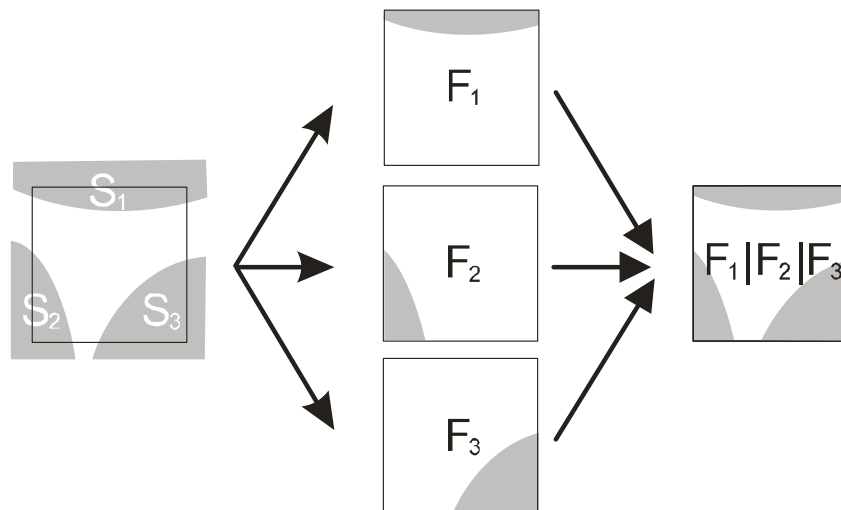


Рисунок 4. Идея «множественного узла»: *слева*: регион пространства, в котором находятся несколько независимых поверхностей; *посередине*: воксели, построенные по поверхностям, составляющие множественный узел; *справа*: множественный узел

Для эффективного хранения данных в каждом узле дополнительно сохраняется число, определяющее количество вокселей в данном регионе. На рисунке 5 показано соответствующее поле в структуре узла октодеревя. Для ограничения объёма памяти и предотвращения избыточного разбиения на практике мы ввели верхний предел на число вокселей в одном узле (например, 8–10). Превышение этого значения обычно связано с недостатками алгоритма выделения компонент, поэтому такие случаи желательно отслеживать и корректировать на этапе построения дерева. Для сокращения количества вокселей в узле нужно уметь объединять несколько вокселей в один и убирать воксели, которые не влияют на внешний вид модели. На рисунке 6 схематично изображено, как это происходит.



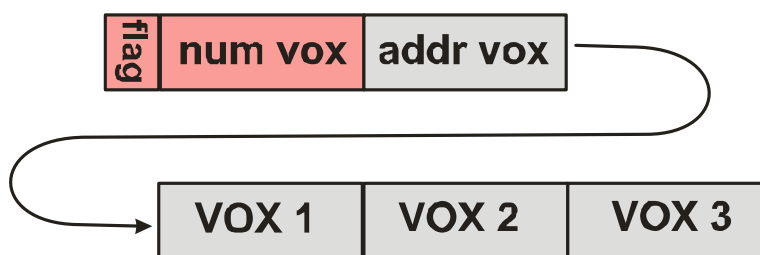


Рисунок 5. Модифицированная структура листа в октодереве. Красным выделены добавленные поля, необходимые для описания поверхностных представлений и множественных узлов

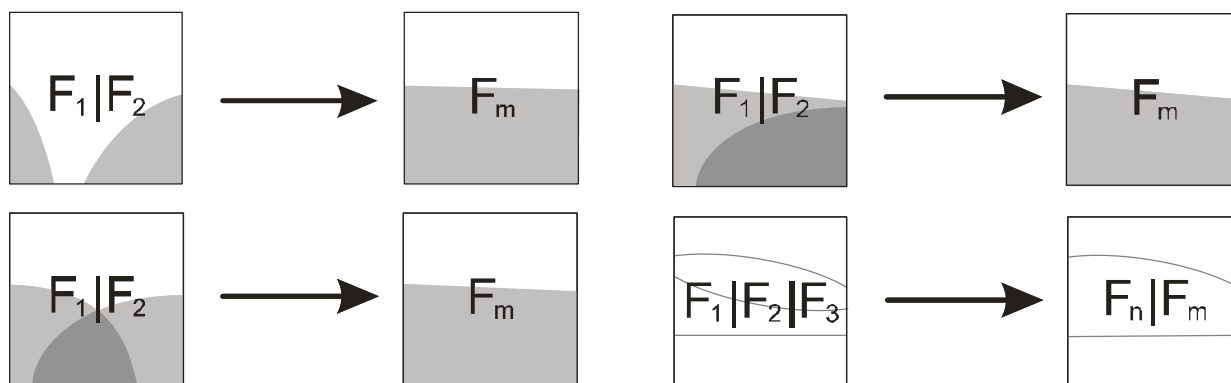


Рисунок 6. Алгоритм уменьшения количества вокселей, при необходимости уменьшения их количества:

*слева:* для объёмных вокселей с потерей качества;

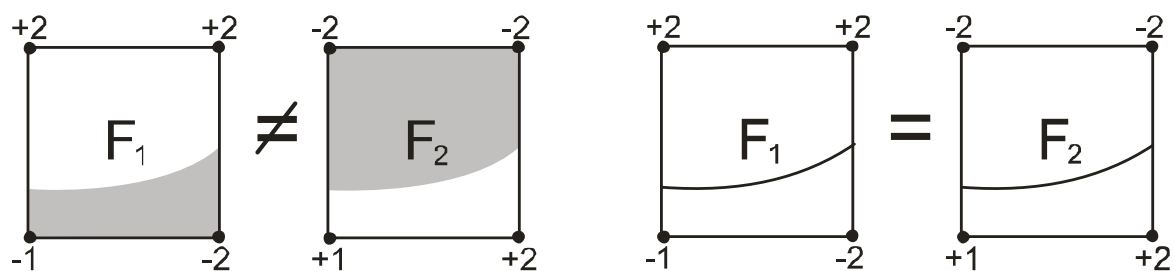
*справа наверху:* для объёмных вокселей без потери качества;

*справа внизу:* для поверхностных вокселей

## 2. Поверхностные узлы

Вторая модификация направлена на расширение возможностей представления незамкнутых поверхностей и тонких элементов модели. В рамках этого подхода допускается хранение вокселей, описывающих не объём, а только поверхность: значения функции расстояния в таком вокселе определяют лишь разделение пространства на два полупространства относительно поверхности, без явного выделения внутренней и внешней областей. На рисунке 7 продемонстрированы главные отличия таких узлов от обычных. Это особенно важно для моделей, не являющихся строго замкнутыми. А при комбинированной работе двух предложенных модификаций это может быть использовано для корректного отображения тонких элементов, ширина которых меньше размера вокселя. В базовой реализации такие элементы часто теряются, что приводит к появлению дыр в модели. Это можно понять по схеме, представленной на рисунке 7. В свою очередь, на рисунке 2 можно увидеть, как наш метод на практике справляется с такими артефактами.

$$F_1 = -F_2$$



**объёмное представление**

**поверхностное представление**

Рисунок 7. Демонстрация различий между стандартным объёмным представлением вокселя и поверхностным представлением

Однако полностью переходить к поверхностным узлам нецелесообразно: при недостаточной точности аппроксимации могут возникать щели между вокселями, как на рисунке 8. Поэтому в предложенной структуре поддерживается одновременное хранение как объёмных, так и поверхностных вокселей. Для этого в каждом узле дополнительно сохраняется флаг, определяющий тип вокселей, что позволяет гибко комбинировать оба подхода в рамках одной структуры.

На рисунке 5 можно увидеть соответствующую модификацию в структуре октодеревя. В результате достигается более корректное представление сложных и разнородных объектов, а также существенно расширяется область применимости метода для моделей с дефектами топологии и открытыми поверхностями.

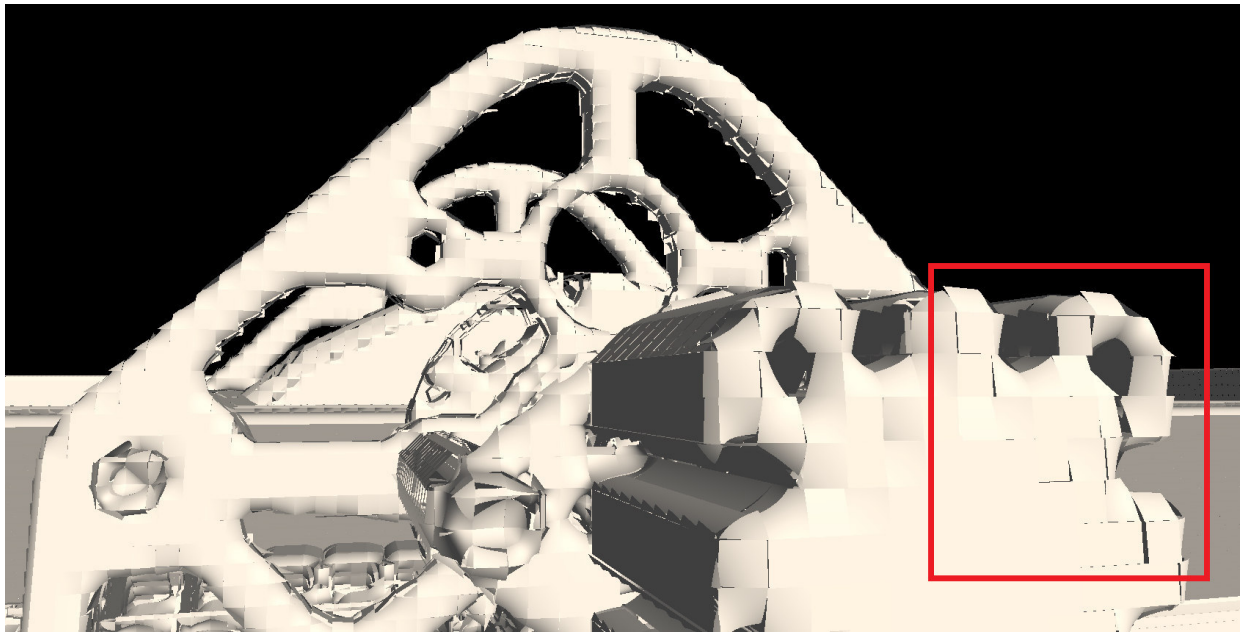


Рисунок 8. Пример модели, где все воксели считаются поверхностными, но была немного уменьшена точность вычисления значений в них. За счёт этого возникают артефакты в виде «щелей» между вокселями

## Сравнения

### 1. Сравнение с октодеревьями

В этом разделе будет произведено сравнение с методом, который строит обычные разреженные октодеревья [9, 10]. Мы сравнили базовый метод с предложенным на нескольких моделях из ABC датасета [21, 22], используя разные параметры глубины. На рисунках 9 и 10 продемонстрированы графики, иллюстрирующие зависимость среднего FLIP и PSNR для нескольких ракурсов от размера полученного октодеревя для конкретных моделей из датасета. Чем больше PSNR и чем меньше размер октодеревя, тем, соответственно, лучше результат. Как видно из графиков, предложенный метод даёт лучший результат и позволяет получить больший PSNR, используя меньшую глубину октодеревя и получая при этом меньшую модель по размеру. В таблице 1 тоже представлены данные для конкретной модели с октодеревьями разной глубины.

В таблице 2 приведены данные для нескольких моделей с зафиксированной глубиной, равной 9. Как видим, предложенные модификации повышают качество модели, но при этом и увеличивают её размер. Однако, как показывает анализ представленных выше графиков, увеличение глубины октодеревя для базового метода даёт меньший прирост в качестве и больший прирост в размере, что доказывает преимущество нашего метода.

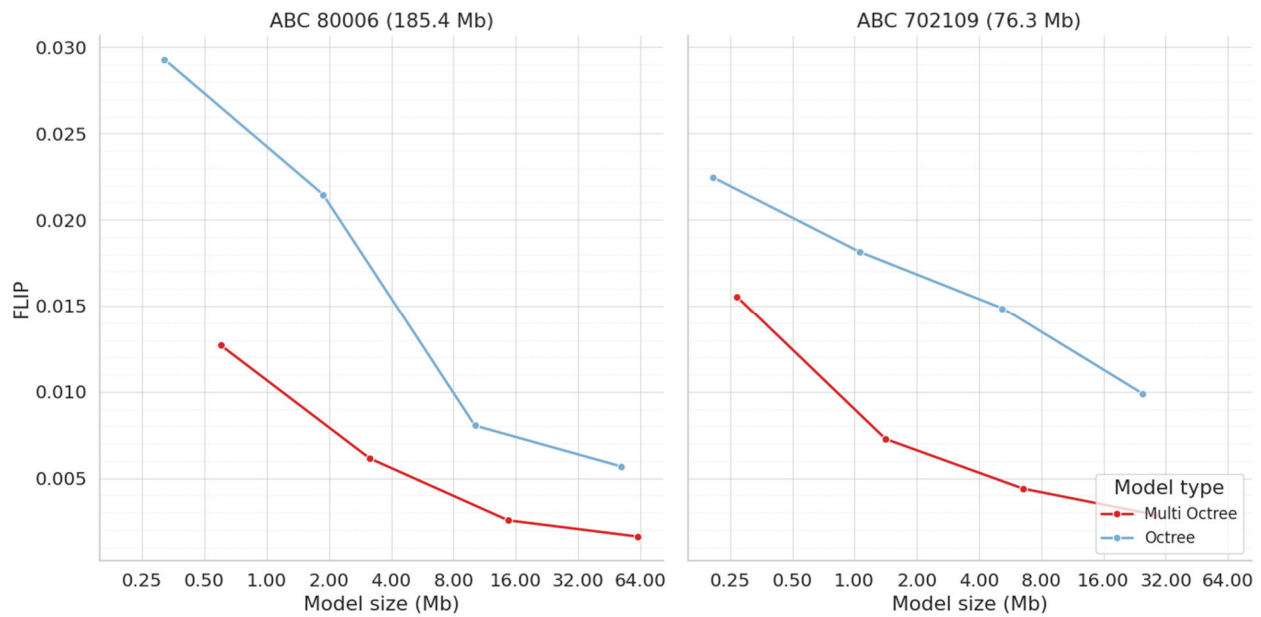


Рис. 9. График сравнения базового октодеревя с нашим методом по FLIP и размеру октодеревьев разной глубины. Чем график ближе к левому нижнему углу, тем метод более качественный

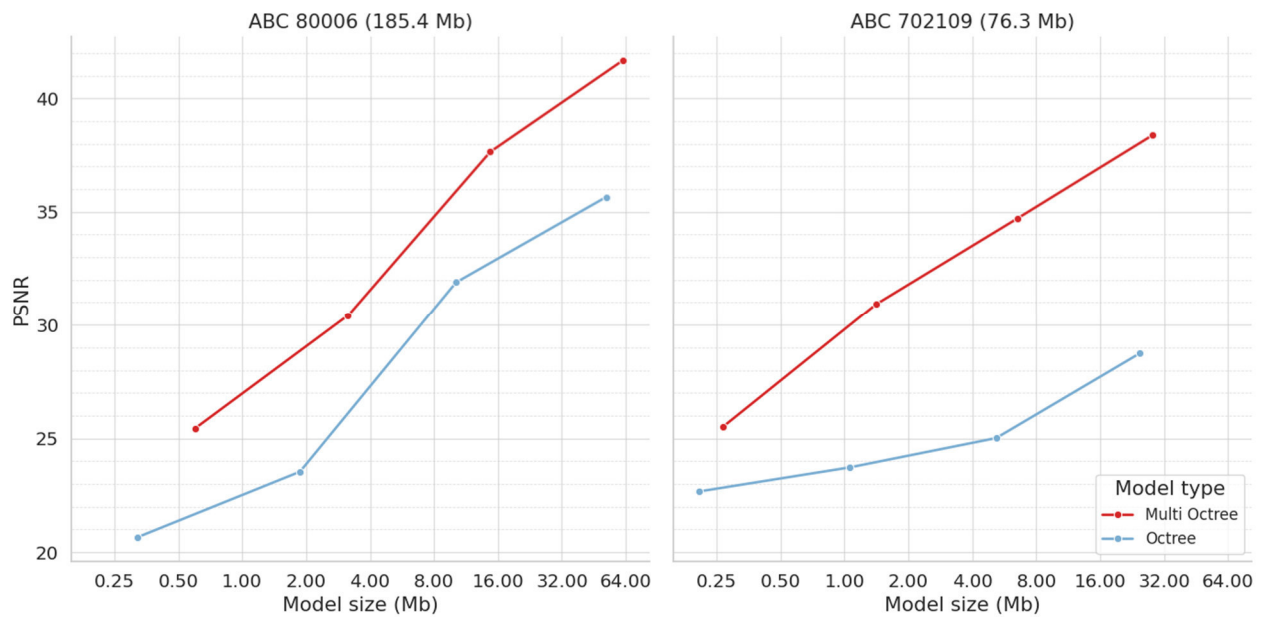


Рис. 10. График сравнения базового октодеревя с нашим методом по PSNR и размеру октодеревьев разной глубины. Чем график ближе к левому верхнему углу, тем метод более качественный

Таблица 1. Сравнение PSNR, FLIP, размеров октодеревьев и их времени рендеринга на модели ABC 88060 (71.06 MB) для разных настроек глубины

CPU: Intel Core i5-8600 @ 3.10 GHz, 16 GB RAM. GPU не использовался.

Depth	Sparse Voxel Octree				Ours			
	PSNR	FLIP	Size MB	Time ms	PSNR	FLIP	Size MB	Time ms
7	29.19	0.0087	<b>0.36</b>	<b>96.3</b>	<b>31.51</b>	<b>0.0072</b>	0.42	97.2
8	32.47	0.0066	<b>1.69</b>	104.2	<b>35.36</b>	<b>0.0051</b>	1.96	<b>103.2</b>
9	38.38	0.0040	<b>7.83</b>	<b>108.8</b>	<b>39.59</b>	<b>0.0036</b>	8.10	109.2
10	41.94	0.0030	<b>33.46</b>	<b>114.7</b>	<b>42.82</b>	<b>0.0028</b>	34.10	115.5

Таблица 2. Сравнение PSNR, FLIP, размеров октодеревьев глубины 9 и их времени рендеринга для разных моделей из датасета, указанных на рисунке 11

CPU: Intel Core i5-8600 @ 3.10 GHz, 16 GB RAM. GPU не использовался.

Model	Sparse Voxel Octree				Ours			
	PSNR	FLIP	Size MB	Time ms	PSNR	FLIP	Size MB	Time ms
ABC 80006 (185.36)	31.89	0.0080	<b>10.16</b>	<b>61.2</b>	<b>37.65</b>	<b>0.0026</b>	14.71	64.1
ABC 88060 (71.06)	38.38	0.0040	<b>7.83</b>	<b>108.8</b>	<b>39.59</b>	<b>0.0036</b>	8.10	109.2
ABC 88828 (14.43)	22.73	0.0262	<b>3.10</b>	<b>76.2</b>	<b>29.47</b>	<b>0.0093</b>	6.92	85.2
ABC 83870 (149.35)	28.14	0.0150	<b>22.93</b>	84.9	<b>33.04</b>	<b>0.0086</b>	25.02	<b>80.6</b>
ABC 515447 (57.43)	28.52	0.0120	<b>4.38</b>	96.4	<b>33.79</b>	<b>0.0055</b>	5.47	<b>90.0</b>
ABC 687231 (76.30)	30.31	0.0101	<b>17.19</b>	101.6	<b>30.78</b>	<b>0.0097</b>	22.58	<b>97.9</b>
ABC 701584 (55.15)	34.74	0.0050	<b>7.47</b>	70.9	<b>35.23</b>	<b>0.0041</b>	8.94	<b>70.2</b>
ABC 702109 (76.27)	25.01	0.0149	<b>5.18</b>	93.1	<b>34.71</b>	<b>0.0044</b>	6.53	<b>83.6</b>

## 2. Сравнение с нейросетевым методом

Проведем сравнение с методом NGLOD [18], так как это интересный современный метод, сочетающий в себе идею октодеревьев и нейросетевой декодер. Произведём сравнение на тех же моделях из того же датасета [21, 22], используя разные параметры глубины для нашего метода и разные параметры детализации для NGLOD. В таблице 3 представлены данные для конкретной модели с разными настройками. Очевидно, что при увеличении размера нейросетевой модели, качество увеличивается недостаточно. Модель меньшего размера, построенная нашим методом, показывает более качественный результат.

Таблица 3. Сравнение PSNR, FLIP, размеров и времени рендеринга моделей нашего метода с моделями нейросетевого метода NGLOD [18] на модели ABC 88060 (71.06 MB) для разных настроек методов

CPU: Intel Core i5-8600 @ 3.10 GHz, 16 GB RAM. GPU не использовался.

Depth / Detail	NGLOD				Ours			
	PSNR	FLIP	Size MB	Time ms	PSNR	FLIP	Size MB	Time ms
7 / 2	27.28	0.0152	<b>0.14</b>	719.2	<b>31.51</b>	<b>0.0072</b>	0.42	<b>97.2</b>
8 / 3	30.06	0.0104	<b>0.76</b>	1014.1	<b>35.36</b>	<b>0.0051</b>	1.96	<b>103.2</b>
9 / 4	29.52	0.0105	<b>5.16</b>	1242.4	<b>39.59</b>	<b>0.0036</b>	8.10	<b>109.2</b>
10 / 5	33.21	0.0069	38.71	1623.3	<b>42.82</b>	<b>0.0028</b>	<b>34.10</b>	<b>115.5</b>

В таблицах 4 и 5 представлены данные для нескольких моделей с зафиксированной глубиной, равной 9, и зафиксированной детализацией, так чтобы размер нейросетевой модели не был значительно больше наших моделей. Как можно наблюдать из этой таблицы, предложенный метод даёт лучшее качество в сравнении с NGLOD, при том имея меньший размер. Скорость рендеринга также значительно выше у нашего метода. Из анализа данных таблицы 3 можно сделать вывод, что нейросетевой метод даёт довольно хорошие результаты при необходимости экстремального сжатия.

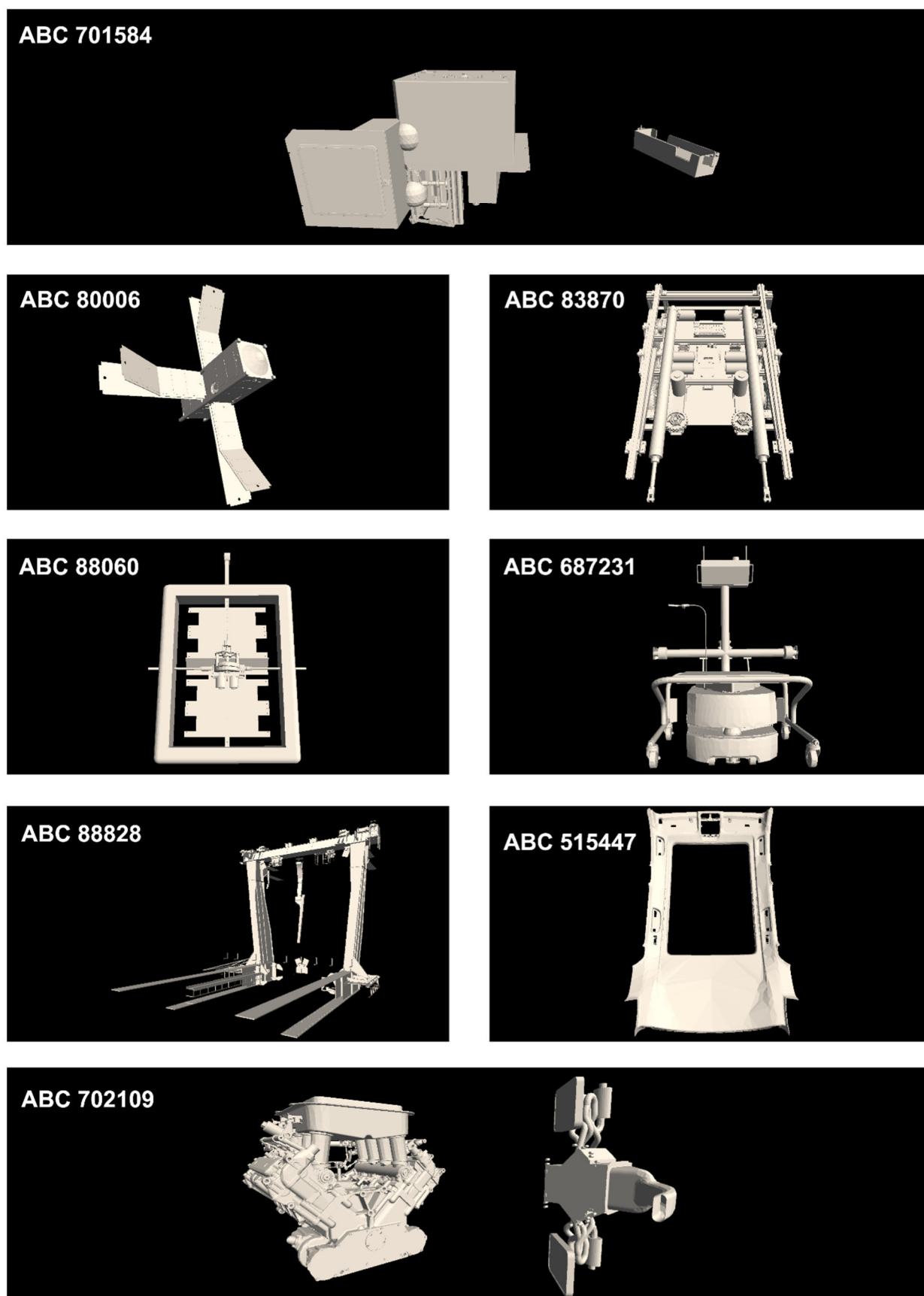


Рис. 11. Модели из датасета [21, 22], на которых производились сравнения

Таблица 4. Сравнение PSNR, FLIP, размеров и времени рендеринга октодеревьев глубины 9 и моделей, полученных с помощью метода NGLOD [18] со сравнимым размером, для разных моделей из датасета [21, 22], указанных на рисунке 11

CPU: Intel Core i5-8600 @ 3.10 GHz, 16 GB RAM. GPU не использовался.

Model	NGLOD				Ours			
	PSNR	FLIP	Size MB	Time ms	PSNR	FLIP	Size MB	Time ms
ABC 80006 (185.36)	25.47	0.0128	<b>5.163</b>	942.3	<b>37.65</b>	<b>0.0026</b>	14.71	<b>64.1</b>
ABC 88060 (71.06)	29.52	0.0105	<b>5.163</b>	1242.4	<b>39.59</b>	<b>0.0036</b>	8.10	<b>109.2</b>
ABC 88828 (14.43)	23.46	0.0209	<b>5.163</b>	1561.1	<b>29.47</b>	<b>0.0093</b>	6.92	<b>85.2</b>
ABC 83870 (149.35)	22.41	0.0322	<b>5.163</b>	1623.1	<b>33.04</b>	<b>0.0086</b>	25.02	<b>80.6</b>
ABC 515447 (57.43)	29.91	0.0090	<b>5.163</b>	1528.1	<b>33.79</b>	<b>0.0055</b>	5.47	<b>90.0</b>
ABC 687231 (76.30)	27.42	0.0141	<b>5.163</b>	1461.1	<b>30.78</b>	<b>0.0097</b>	22.58	<b>97.9</b>
ABC 701584 (55.15)	27.25	0.0135	<b>5.163</b>	1311.2	<b>35.23</b>	<b>0.0041</b>	8.94	<b>70.2</b>
ABC 702109 (76.27)	28.69	0.0097	<b>5.163</b>	1817.7	<b>34.71</b>	<b>0.0044</b>	6.53	<b>83.6</b>

Таблица 5. Сравнение PSNR, FLIP, размеров и времени рендеринга октодеревьев глубины 9 и моделей, полученных с помощью метода NGLOD [18] с не меньшим размером, для разных моделей из датасета [21, 22], указанных на рисунке 11

CPU: Intel Core i5-8600 @ 3.10 GHz, 16 GB RAM. GPU не использовался.

Model	NGLOD				Ours			
	PSNR	FLIP	Size MB	Time ms	PSNR	FLIP	Size MB	Time ms
ABC 80006 (185.36)	26.69	0.0110	38.71	1104.7	<b>37.65</b>	<b>0.0026</b>	<b>14.71</b>	<b>64.1</b>
ABC 88060 (71.06)	33.21	0.0069	38.71	1623.3	<b>39.59</b>	<b>0.0036</b>	<b>8.10</b>	<b>109.2</b>
ABC 88828 (14.43)	25.65	0.0163	38.71	1758.7	<b>29.47</b>	<b>0.0093</b>	<b>6.92</b>	<b>85.2</b>
ABC 83870 (149.35)	24.70	0.0243	38.71	2083.5	<b>33.04</b>	<b>0.0086</b>	<b>25.02</b>	<b>80.6</b>
ABC 515447 (57.43)	30.12	0.0087	38.71	1723.8	<b>33.79</b>	<b>0.0055</b>	<b>5.47</b>	<b>90.0</b>
ABC 687231 (76.30)	28.35	0.0132	38.71	1735.6	<b>30.78</b>	<b>0.0097</b>	<b>22.58</b>	<b>97.9</b>
ABC 701584 (55.15)	28.53	0.0121	38.71	1412.4	<b>35.23</b>	<b>0.0041</b>	<b>8.94</b>	<b>70.2</b>
ABC 702109 (76.27)	29.41	0.0090	38.71	1897.1	<b>34.71</b>	<b>0.0044</b>	<b>6.53</b>	<b>83.6</b>

## Выводы

Авторы данной статьи предлагают модификации базового метода разреженных октодеревьев. Эти изменения позволяют хранить в каждом узле дерева несколько поверхностей вместо одной в базовом алгоритме. Также появляется возможность представлять элементы модели не только как объёмные узлы, но и как поверхности. Эксперименты показывают, что такие модификации немного повышают размер итоговой модели, однако при этом значительно повышают их качество, а время рендеринга изменённых моделей при этом сравнимо с рендерингом базовых моделей.

## Список литературы

1. G. Chou, Y. Bahat and F. Heide, "Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)", Diffusion-SDF: Conditional Generative Modeling of Signed Distance Functions, Paris, France, 2023.
2. L. Yariv, O. Puny, N. Neverova, O. Gafni and Y. Lipman, "Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)", Mosaic-SDF for 3D Generative Models, Seattle, WA, USA, 2024.
3. J. Shim, C. Kang and K. Joo, "Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)", Diffusion-Based Signed Distance Fields for 3D Shape Generation, Vancouver, BC, Canada, 2023.
4. C.-H. Lin, C. Wang and S. Lucey, "SDF-SRN: Learning Signed Distance 3D Object Reconstruction from Single-View Images", arXiv preprint arXiv:2010.10505, 2020.
5. C. Sommer, L. Sang, D. Schubert and D. Cremers, "Gradient-SDF: A Semi-Implicit Surface Representation for 3D Reconstruction", arXiv preprint arXiv:2111.13652, 2022.



6. M. Slavcheva, W. Kehl, N. Navab and S. Ilic, "Proceedings of the European Conference on Computer Vision (ECCV)", SDF-2-SDF: Highly Accurate 3D Object Reconstruction, Amsterdam, The Netherlands, 2016.
7. H. Hansson-Söderlund, A. Evans and T. Akenine-Möller, "Ray Tracing of Signed Distance Function Grids", *Journal of Computer Graphics Techniques*, vol. 11, no. 3, p. 94–113, 2022.
8. A. R. Garifullin, V. A. Frolov, A. S. Budak and V. A. Galaktionov, "Study of Surface Representation Methods Based on Signed Distance Functions", *Programming and Computer Software*, vol. 51, no. 1, pp. 131-139, 2025.
9. C. Crassin, F. Neyret, S. Lefebvre and E. Eisemann, "Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games (I3D '09)", *GigaVoxels: ray-guided streaming for efficient and detailed voxel rendering*, Boston, Massachusetts, 2009.
10. S. Laine and T. Karras, "Proceedings of the 2010 Symposium on Interactive 3D Graphics and Games (I3D '10)", *Efficient sparse voxel octrees*, New York, NY, USA, 2010.
11. E. Pujol and A. Chica, "Rendering piecewise approximations of SDFs through analytic intersections", *Computers & Graphics*, vol. 122, no. C, p. 9, November 2024.
12. T. Müller, A. Evans, C. Schied and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding", *ACM Transactions on Graphics*, vol. 41, no. 4, pp. Article 102, 1–15, July 2022.
13. M. Nießner, M. Zollhöfer, S. Izadi and M. Stamminger, "Real-time 3D reconstruction at scale using voxel hashing", *ACM Transactions on Graphics*, vol. 32, no. 6, pp. Article 169, 1–11, November 2013.
14. K. Museth, "VDB: High-resolution sparse volumes with dynamic topology", *ACM Transactions on Graphics*, vol. 32, no. 3, pp. Article 27, 1–22, July 2013.
15. R. K. Hoetzlein, "Proceedings of High Performance Graphics (HPG '16)", *GVDB: raytracing sparse voxel database structures on the GPU*, Dublin, Ireland, 2016.
16. V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell and G. Wetzstein, "Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)", *Implicit neural representations with periodic activation functions*, Red Hook, NY, USA, 2020.
17. J. J. Park, P. R. Florence, J. Straub, R. A. Newcombe and S. Lovegrove, "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", *DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation*, 2019.
18. T. Takikawa, J. Litalien, K. Yin, K. Kreis, C. Loop, D. Nowrouzezahrai, A. Jacobson, M. McGuire and S. Fidler, "Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)", *Neural Geometric Level of Detail: Real-Time Rendering with Implicit 3D Shapes*, 2021.
19. H. Jiang, H. Yang, G. Pavlakos and Q. Huang, "CoFie: Learning Compact Neural Surface Representations with Coordinate Fields", *arXiv preprint arXiv:2406.03417*, 2024.
20. P. Weier, A. Rath, É. Michel, I. Georgiev, P. Slusallek and T. Boubekeur, "ACM SIGGRAPH 2024 Conference Papers (SIGGRAPH '24)", *N-BVH: Neural ray queries with bounding volume hierarchies*, Denver, CO, USA, 2024.
21. S. Koch, A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin and D. Panozzo, "ABC Dataset", *Доступно под MIT лицензией*, 2019.
22. S. Koch, A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin and D. Panozzo, "ABC: A Big CAD Model Dataset For Geometric Deep Learning", *arXiv preprint arXiv:1812.06216*, 2019.

## Кодирование глобального освещения гауссианами в приложениях реального времени

К. С. Зубатов, А. С. Щербаков, В. А. Фролов, В. В. Санжаров

Московский государственный университет им. М. В. Ломоносова, Москва, Россия

**Аннотация.** В графике реального времени необходимо рассчитывать глобальное (вторичное) освещение – компоненту освещенности, возникающую в результате многократного переотражения света от поверхностей сцены. Вторичное освещение напрямую зависит от материалов сцены, однако его точность в графике реального времени ограничена производительностью пользовательского устройства и требованиями приложения. В данной работе исследуется применение метода гауссовых сплатов (Gaussian splatting) для хранения вторичного освещения сцены. Этот метод реконструирует материалы сцены и позволяет изменять сложность представления сцены в зависимости от требуемой точности. Предложенный метод использует этап предобработки, в ходе которого по набору изображений с диффузным вторичным освещением сцены создается множество гауссиан, аппроксимирующих световую информацию. В результате был получен метод, имеющий высокую точность при сравнении с эталонным методом трассировки путей и позволяющий регулировать точность для изменения скорости рендеринга.

**Ключевые слова:** глобальное освещение, вторичное освещение, гауссианы, рендеринг

## Gaussians encoding of global illumination in real time applications

K. S. Zubatov, A. S. Shcherbakov, V. A. Frolov, V. V. Sanzharov

Lomonosov Moscow State University, Moscow, Russia

**Abstract.** In real-time graphics, it is necessary to calculate global (secondary) illumination – a component of illumination that occurs as a result of multiple reflections of light from scene surfaces. Secondary illumination directly depends on scene materials, but its accuracy in real-time graphics is limited by the performance of the user's device and application requirements. This paper investigates the use of Gaussian splatting to store secondary scene lighting. This method reconstructs scene materials and allows changing the complexity of scene representation depending on the required accuracy. The proposed method uses a preprocessing stage, during which a set of images with diffuse secondary scene lighting is used to create a set of Gaussians that approximate the light information. The result is a method that is highly accurate when compared to the reference path tracing method and allows changing the accuracy to vary the rendering speed.

**Keywords:** global illumination, secondary illumination, gaussians, rendering

### Введение

Освещение играет ключевую роль в ощущении реалистичности сцен. Свет, взаимодействуя с поверхностями объектов, отражается, преломляется, рассеивается или поглощается в зависимости от материалов. Для первичного освещения было создано множество математических моделей, позволяющих достаточно точно описать взаимодействие света с материалом [1]. Однако вторичное (глобальное) освещение, являющееся результатом множественного переотражения света от объектов сцены, представляет собой более сложную задачу. В условиях графики реального времени вычисление глобального освещения особенно затруднено тем, что на большинстве пользовательских устройств не получится решать многомерное интегральное уравнение, описывающее взаимодействие света со сложными материалами, с достаточной скоростью. В связи с этим используются различные аппроксимации, позволяющие получить удовлетворительный результат при имеющихся у пользователя ограниченных вычислительных ресурсах. Это создает необходимость балансировать между точностью освещения и количеством потребляемых ресурсов при выборе метода расчета глобального освещения. В данной работе предлагается метод представления вторичного освещения в виде гауссиан, который позволяет регулировать точность представления материалов, используемых для создания вторичного освещения, с целью получения различной скорости рендеринга.

### Обзор существующих решений

Существует множество методов аппроксимации глобального освещения, используемых в графике реального времени, отличающихся по точности, скорости работы, поддержке динамических объектов,

способу хранения информации и способности балансировать между точностью результата и скоростью работы.

*Карты освещения (Light maps)* [2] используют текстурный атлас для информации об освещенности объектов. Каждому текселю атласа соответствует трехмерная точка объекта, для которой вычисляется вторичное освещение на этапе предобработки. Вычисление глобального освещения во время рендеринга сводится к простой операции выборки из текстуры. Метод широко применяется на практике [3, 4] из-за высокой точности итогового освещения и минимальной вычислительной нагрузки при рендеринге, однако данный метод неприменим для сцен с динамическим освещением. Также при масштабировании требуется полный перерасчет освещения, а при высокой точности – большое количество памяти для текстур.

*Сферические гармоники* [5] хранят информацию об освещении в виде позиции гармоники и набора коэффициентов ортонормированных сферических функций, которыми кодируется освещение, входящее под определенным направлением в гармонику с заданной позицией. Во время рендеринга глобальное освещение объекта вычисляется с помощью интерполяции освещения ближайших гармоник. Метод позволяет компактно хранить информацию о диффузном освещении и поддерживает динамические объекты, однако требует большого количества памяти для хранения спекулярного освещения, не поддерживает динамические источники света, а также имеет низкую точность в закрытых сценах.

*Метод излучательности (Radiosity)* [6] для каждой пары полигонов сцены вычисляет (на этапе предобработки) долю света (форм-фактор), которая попадает с одного полигона на другой, формируя матрицу форм-факторов  $N \times N$ , где  $N$  – число полигонов в сцене (ввиду геометрической сложности большинства сцен обычно используют упрощенное представление сцены). Во время рендеринга вычисление вторичного освещения полигонов сводится к умножению матрицы форм-факторов на вектор их первичной освещенности. Метод является физически корректным и вычислительно простым, отчего нашел широкое применение в графике реального времени [7], однако имеет квадратичную сложность от числа полигонов  $N$  по количеству арифметических операций и по памяти.

*Кэш освещенности (Irradiance Cache)* [8] хранит информацию об освещении в виде трехмерной сетки, элементы которой содержат сферические гармоники. Также каждой сферической гармонике соответствует октаэдральная текстура глубины, определяющая зону действия её излучения. Вычисление освещения заключается в интерполяции излучения от ближайших гармоник для заданной точки сцены.

*Метод трассировки вокселей конусами (Voxel Cone Tracing)* [9] использует упрощенное представление сцены в виде разреженной воксельной сетки. В каждом кадре источники освещения обновляют информацию о входящем излучении для видимых вокселей, а в конце кадра вычисляется итоговое вторичное освещение при помощи трассировки конусами (т.е. упрощенной трассировки лучей) по воксельной сетке. Метод позволяет вычислять множественные переотражения света для динамических сцен, но из-за высокой вычислительной нагрузки и нагрузки на память обычно ограничиваются вычислением одного переотражения. Также для алгоритма характерны «протечки» света сквозь тонкие объекты сцены (особенно при малом разбиении сцены на воксели).

*Метод объемов распространения света (Light Propagation Volumes)* [10] вычисляет глобальное освещение сцены с помощью численного решения дифференциального уравнения на трехмерной сетке. В сетке хранятся сферические гармоники, кодирующие входящее со всех направлений излучение. Начальное приближение задается виртуальными источниками света [11]. Для расчета затенения при вычислении светопереноса используется дополнительная трехмерная сетка с приближенной вокселизацией поверхностей сцены. Данный метод применим для полностью динамических сцен, так как вычисляет распространение света по сетке каждого кадра, однако требует большого количества памяти для хранения регулярных трехмерных сеток (для решения этой проблемы разработана модификация с вложенными сетками [12]), имеет низкую точность освещения и подвержен «протечкам» света из-за недостаточной дискретизации сцены.

У всех перечисленных методов возникают проблемы с масштабированием, когда в зависимости от имеющихся вычислительных ресурсов необходимо либо уменьшить точность освещения для экономии

ресурсов, либо получить более точное освещение в сложных участках сцены при наличии свободных ресурсов.

- Методы, использующие трехмерные сетки (трассировка вокселей конусами, объемы распространения света, кэш освещенности), при изменении размера вокселей либо теряют информацию в областях со сложным освещением, либо перерасходуют память для хранения освещения в однородных областях сцены.

- Для карт освещения проблема аналогична методам с трехмерными сетками – с ростом размера текселя снижается точность освещения, а при малых размерах текселя множество текселов содержит одинаковую информацию для участков сцены с однородным освещением. На практике проблема частично решается применением блочного сжатия текстур. Сжатие изображений (например, в формате png) не представляется возможным из-за длительного декодирования на gpu.

- Метод излучательности имеет квадратичную сложность по памяти и вычислениям от числа полигонов сцены, что затрудняет его использование для сцен с высокой детализацией. Использование сильно упрощенной сцены для расчета форм-факторов приведет к потере точности или некорректному переотражению света.

- Изменение числа сферических гармоник требует их ручной расстановки или использования эвристических алгоритмов. Также потребуется дополнительная информация, ограничивающая зону влияния гармоник, для предотвращения просвечивания света сквозь геометрию сцены.

### Описание представления сцены гауссианами

Для хранения глобального освещения предложенный метод использует представление сцены в виде гауссиан (Gaussian splatting) [13]. Гауссиана  $g$  описывается пятеркой  $g = (\mu, s, r, h, o)$ :

- $\mu \in \mathbb{R}^3$  – координаты центра гауссианы в мировом пространстве;
  - $s \in \mathbb{R}_+^3$  – коэффициенты масштаба гауссианы по осям  $x, y, z$ ;
  - $r = (r_0, r_i, r_j, r_k)^T \in \mathbb{R}^4$  – коэффициенты кватерниона, описывающего поворот гауссианы в мировом пространстве;
  - $h \in \mathbb{R}^{(n+1)^2 \times 3}$  – коэффициенты сферических гармоник порядка  $n$ ;
  - $o \in [0, 1]$  – коэффициент прозрачности гауссианы,
- и функцией  $G(x)$  вида:

$$G(x) = \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right).$$

Матрица ковариации  $\Sigma$  в функции (1) задается через матрицы поворота  $R$  и масштаба  $S$ :

$$\Sigma = RSS^T R^T,$$

$$R = \begin{pmatrix} 1 - 2(r_j^2 + r_k^2) & 2(r_i r_j - r_k r_0) & 2(r_i r_k + r_j r_0) \\ 2(r_i r_j + r_k r_0) & 1 - 2(r_i^2 + r_k^2) & 2(r_j r_k - r_i r_0) \\ 2(r_i r_k - r_j r_0) & 2(r_j r_k + r_i r_0) & 1 - 2(r_i^2 + r_j^2) \end{pmatrix},$$

$$S = Es,$$

где  $E$  обозначает единичную матрицу.

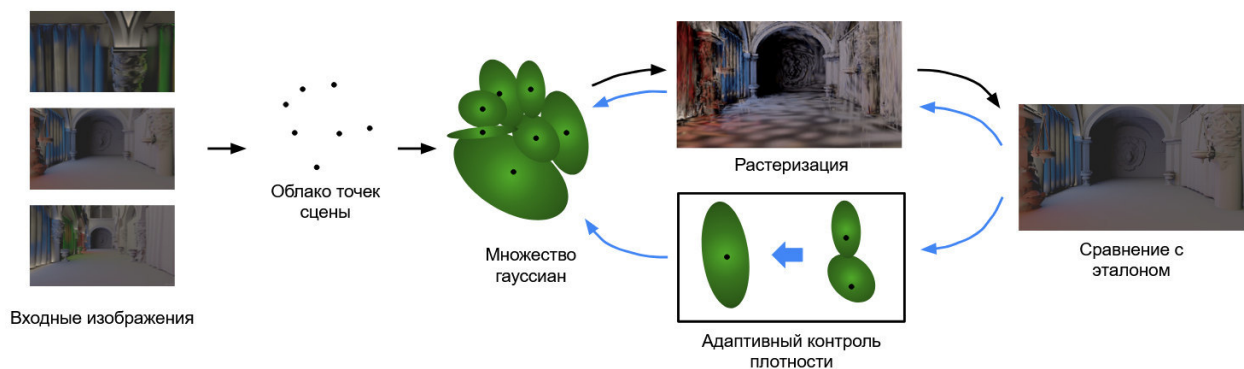


Рисунок 1. Схема построения гауссиан по изображениям сцены

Построение множества гауссиан (рис. 1) для трехмерной сцены по набору изображений заключается:

- в создании облака точек  $X$  по входным изображениям;
- задании начального множества гауссиан. Центрами гауссиан считаются элементы облака точек  $x \in X$ , а начальные размеры каждой гауссианы определяются на основе усредненного расстояния до трех ближайших точек из облака  $X$ ;
- оптимизации параметров гауссиан (рис. 2) с помощью итеративного процесса растеризации гауссиан и сравнения с входными изображениями. Для оптимизации размера, позиции, цвета и прозрачности используется стохастический градиентный спуск;
- адаптивном контроле плотности, регулирующем количество гауссиан через разбиение больших или слияние слишком малых гауссиан (выполняется один раз в несколько итераций).

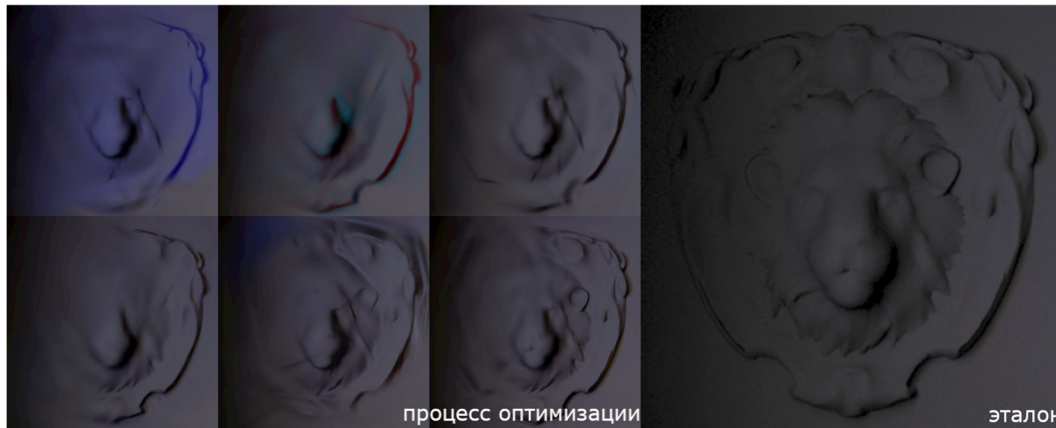


Рисунок 2. Процесс оптимизации сцены. Представление в виде гауссиан сравнивается с эталоном

Для растеризации полученного множества необходимо:

1) выполнить предобработку гауссиан с использованием матриц перехода в пространство камеры и перспективной проекции. В результате данного этапа будут отсечены гауссианы (рис. 3), не попавшие в усеченную пирамиду видимости, а для видимых гауссиан будут вычислены цвет  $C_i$ , глубина  $d_i$ , двумерная матрица ковариации  $\Sigma_i$  и прозрачность  $\alpha_i$ ;

2) отсортировать видимые гауссианы по глубине;

3) растеризовать видимые гауссианы в порядке возрастания глубины. Для этого в каждом пикселе задается переменная  $C := (0, 0, 0)$ , которая будет хранить итоговый цвет, и переменная  $T := 1$ , описывающая вклад текущей гауссианы в итоговый цвет. Гауссианы, двумерные проекции которых пересекают данный пиксель, обрабатываются в порядке возрастания глубины  $d_i$ . Для проекции гауссианы с центром в точке экрана  $\mu_i$  и пикселя с координатами  $p$  вычисляется прозрачность  $\alpha_0 = \alpha_i \exp((\mu_i - p)\Sigma_i(\mu_i - p)^T)$ . Если прозрачность в пикселе больше  $1/255$ , то обновляются итоговый цвет  $C := C + C_i T \alpha_0$  и вклад следующей гауссианы  $T := T(1 - \alpha_0)$ . Если новое значение  $T$  близко к нулю, то обработка гауссиан прекращается.

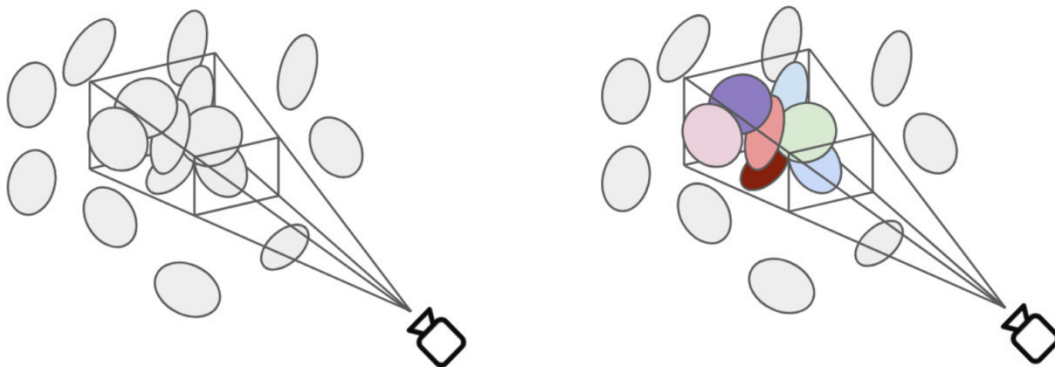


Рисунок 3. Отсечение невидимых гауссиан (серые). Далее будут обработаны и отрисованы только видимые (цветные) гауссианы

## Предложенный метод

### 1. Предобработка сцены

Предложенный метод использует этап предобработки для создания множества гауссиан по набору изображений сцены. Однако, в отличие от оригинального метода, который в основном используется для реконструкции сцен по наборам фотографий или видеозаписям, в данной работе будет реконструироваться только информация о вторичном освещении. Вычисление вторичного освещения, как и в картах освещения [2], основывается на использовании стохастической трассировки путей.

Полное освещение сцены описывается интегральным уравнением [14]:

$$L_o(x, \omega) = L_e(x, \omega) + \int_{\Omega} f_r(x, \omega, \omega') L_i(x, \omega') (-\omega' \cdot n) d\omega', \quad (1)$$

где  $L_e(x, \omega)$  – количество собственного излучения, исходящего из точки  $x$  вдоль направления  $\omega$ ;

$f_r(x, \omega, \omega')$  – двулучевая функция отражательной способности;

$L_i(x, \omega')$  – количество излучения, входящего в точку  $x$  вдоль направления  $\omega'$ ;

$\Omega$  – единичная полусфера для точки  $x$  и нормали  $n$ ;

точка  $x$  и направление  $\omega$  определяются камерой.

Данный интеграл описывает полное освещение  $L_o(x, \omega)$  сцены для точки  $x$  и направления  $\omega$ . Получение вторичного освещения  $L_s(x, \omega)$  обеспечивается игнорированием вклада от лучей, отразившихся один раз, и собственного излучения  $L_e(x, \omega)$ , исходящего из точки  $x$  вдоль направления  $\omega$ , при вычислении интеграла (2). Для описания влияния луча, имеющего направление  $\omega$  и попавшего в точку  $x$ , на вторичное освещение введем функцию  $1(x, \omega)$ :

$$1(x, \omega) = \begin{cases} 1, & \text{луч не попал в камеру после первого отражения} \\ 0, & \text{иначе} \end{cases}$$

Тогда  $L_s(x, \omega)$  имеет вид

$$L_s(x, \omega) = \int_{\Omega} 1(x, \omega) f_r(x, \omega, \omega') L_i(x, \omega') (-\omega' \cdot n) d\omega'. \quad (2)$$

Для вторичного освещения рассматриваются ламбертовские материалы с  $f_r(x, \omega, \omega') = \frac{C}{\pi}$ , где  $C$  обозначает цвет объекта в точке  $x$ . Тогда уравнение (2) можно переписать в следующем виде:

$$L_s(x, \omega) = C \int_{\Omega} \frac{1}{\pi} 1(x, \omega) L_i(x, \omega') (-\omega' \cdot n) d\omega'. \quad (3)$$

Таким образом, при вычислении вторичного диффузного освещения можно вынести цвет  $C$  точки  $x$ , получив тем самым более низкочастотное изображение с входящим в точку вторичным освещением (рис. 4).



Рисунок 4. Сравнение вторичного исходящего и входящего освещения

Данный подход призван снизить количество гауссиан, необходимых для хранения глобального освещения. Его использование не имеет визуального влияния на конечный результат.



## 2. Масштабирование

При оптимизации параметров гауссиан на этапе предобработки для функции потерь используется следующая комбинация метрик  $\mathcal{L}_1$  (манхэттенское расстояние) и  $\mathcal{L}_{D-SSIM}$ :

$$\mathcal{L} = 0.2\mathcal{L}_1 + 0.8\mathcal{L}_{D-SSIM}.$$

$\mathcal{L}_{D-SSIM}(x, y) = \frac{1-SSIM(x,y)}{2}$  выражает структурные отличия двух изображений через метрику SSIM.

Ограничивая минимальное значение функции потерь, можно достичь разного количества гауссиан, необходимых для приближенного представления вторичного освещения. Таким образом, предложенный метод позволяет плавно выбирать между необходимой точностью освещения и скоростью визуализации, которая зависит от числа гауссиан.

## 3. Рендеринг сцены

На этапе рендеринга для получения изображения с полным освещением сцены необходимо:

1) вычислить первичное освещение сцены:

$$L_p(x, \omega) = L_e(x, \omega) + \int_{\Omega} (1 - \mathbb{1}(x, \omega)) f_r(x, \omega, \omega') L_i(x, \omega') (-\omega' \cdot n) d\omega';$$

2) растеризовать гауссианы для получения вторичного освещения  $L_s(x, \omega)$  вида (3);

3) совместить первичное и вторичное освещение для получения полного освещения  $L_o(x, \omega)$ :

$$L_o(x, \omega) = L_p(x, \omega) + L_s(x, \omega) = L_e(x, \omega) + \int_{\Omega} f_r(x, \omega, \omega') L_i(x, \omega') (-\omega' \cdot n) d\omega'.$$

Первичное освещение вычисляется с помощью техники отложенного освещения, в результате которой сохраняется текстура цвета объектов сцены, которая будет использована для третьего пункта. Он заключается в композиции первичного освещения и вторичного входящего освещения, умноженного (используется произведение Адамара) на текстуру с цветом объектов сцены (рис. 5).

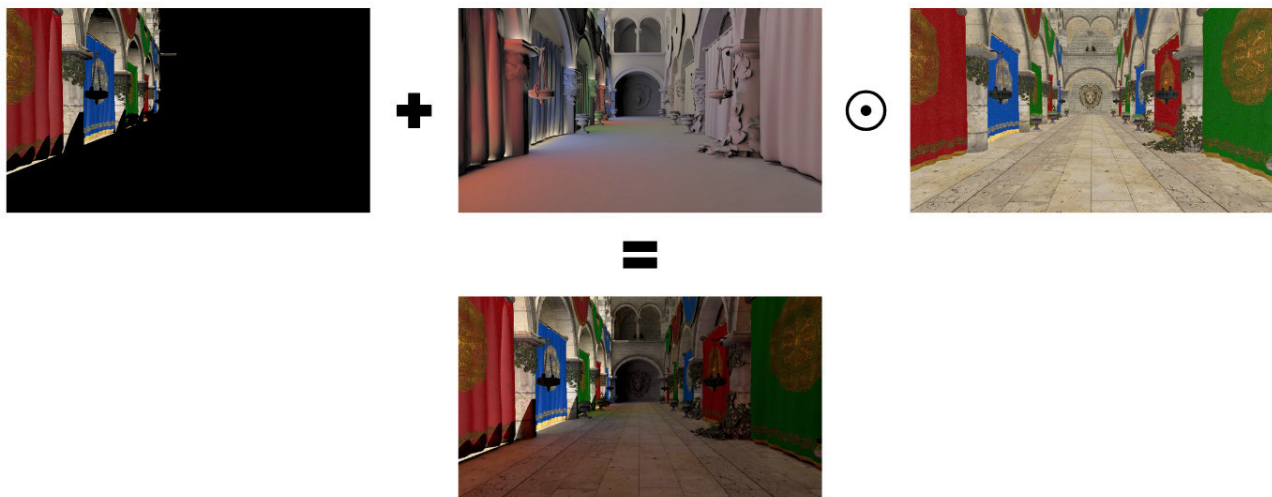


Рисунок 5. Композиция первичного освещения, вторичного освещения и буфера цвета

## Результаты экспериментов

### 1. Сравнение с существующими решениями

Из существующих решений был выбран метод карт освещения, так как он имеет наивысшую точность среди методов реального времени. Сравнение включало анализ потребляемой памяти и времени рендеринга (рис. 6). Тестирование предложенного метода проводилось на сцене *Crytek Sponza* в разрешении  $1920 \times 1080$ , использовалась видеокарта *Nvidia RTX 3060*.

Результаты показали, что предложенный метод обладает высокой точностью и потребляет меньше памяти, чем метод карт освещения. Однако растеризация гауссиан в предложенном методе требует больше вычислительных ресурсов, чем выборка из текстурного атласа в картах освещения. Для оценки точности метода было проведено сравнение с трассировкой путей, являющейся эталонным методом, по метрике SSIM (рис. 7). Метод продемонстрировал высокое сходство с эталонными изображениями.

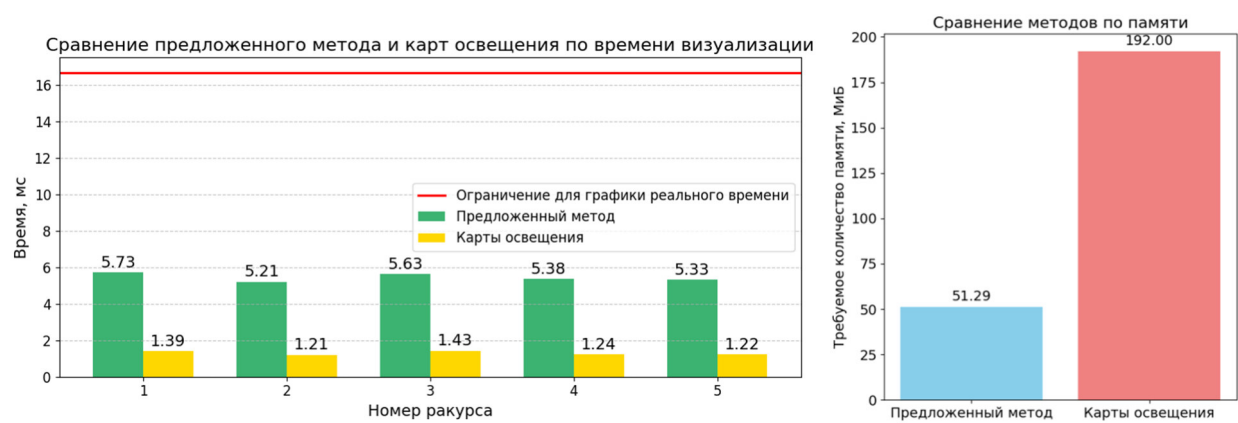


Рисунок 6. Сравнение предложенного метода и карт освещения по времени и памяти



Рисунок 7. Сравнение предложенного метода и трассировки путей (эталона) по метрике SSIM

## 2. Исследование масштабируемости метода

Также было исследовано влияние количества используемых гауссиан на производительность и точность. Эксперименты показали (рис. 8), что возможно достичь существенного ускорения рендеринга при умеренном снижении качества визуализации, что демонстрирует масштабируемость метода под конкретные требования к скорости и качеству визуализации.

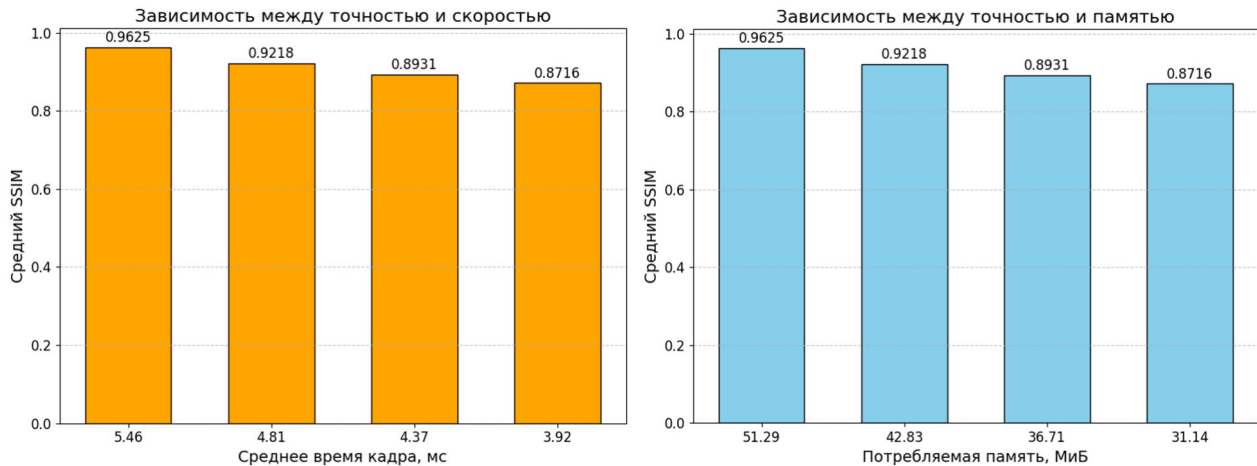


Рисунок 8. Сравнение точности и времени кадра предложенного метода при разном числе гауссиан

## Выводы

В работе представлен метод хранения вторичного освещения на основе гауссиан. Подход позволяет реконструировать материалы сцены и управлять детализацией освещения, тем самым обеспечивая баланс между точностью визуализации и производительностью. Результаты исследования показали, что метод способен значительно снизить объем используемой памяти по сравнению с традиционными картами освещения при сохранении высокой точности, сопоставимой с трассировкой путей. Также показана масштабируемость метода: настройка порогового значения функции потерь позволяет балансировать между скоростью визуализации и качеством.

## Список литературы

1. Pharr M., Jakob W., Physically based rendering: From theory to implementation. MIT Press, 2023.
2. Quake's lighting model: Surface caching / M. Abrash, et al. // Graphic programming black book. 2000. Vol. 2.
3. Unity Technologies. Unity User Manual, Lightmapping. URL: <https://docs.unity3d.com/2019.1/Documentation/Manual/Lightmappers.html>.
4. Epic Games. Understanding Lightmapping in Unreal Engine. URL: <https://dev.epicgames.com/documentation/en-us/unreal-engine/understanding-lightmapping-in-unreal-engine>.
5. Sloan P. P., Kautz J., Snyder J. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments // Seminal Graphics Papers: Pushing the Boundaries. 2023. Vol. 2. P. 339-348.
6. Cohen M. F., Greenberg D. P. The hemi-cube: A radiosity solution for complex environments // ACM Siggraph Computer Graphics. 1985. Vol. 19, no. 3. P. 31-40.
7. Martin S. Enlighten real-time radiosity // ACM SIGGRAPH 2011. Computer Animation Festival. 2011. P. 97-97.
8. McGuire M., et al. Real-time global illumination using precomputed light field probes // Proceedings of the 21st ACM SIGGRAPH symposium on interactive 3D graphics and games. 2017. P. 1-11.
9. Interactive indirect illumination using voxel-based cone tracing: An insight / C. Crassin, et al. // ACM SIGGRAPH 2011 Talks. 2011. P. 1-1.
10. Kaplanyan A. Light propagation volumes in cryengine 3 // ACM SIGGRAPH Courses. 2009. Vol. 7, no. 2.
11. Keller A. Instant radiosity // Proceedings of the 24th annual conference on Computer graphics and interactive techniques. 1997. P. 49-56.
12. Kaplanyan A., Dachsbacher C. Cascaded light propagation volumes for real-time indirect illumination // Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games. 2010. P. 99-107.
13. 3D Gaussian splatting for real-time radiance field rendering / B. Kerbl, et al. // ACM Trans. Graph. 2023. Vol. 42, no. 4. P. 139:1-139:14.
14. Kajiya J. T. The rendering equation // Proceedings of the 13th annual conference on Computer graphics and interactive techniques. 1986. P. 143-150.

## Enhancing 3D Gaussian Splatting with diffusion models: a survey

**Maksim Raenchuk**

MSU Institute for Artificial Intelligence, Moscow, Russia  
Lomonosov Moscow State University, Moscow, Russia

**Abstract.** 3D Gaussian Splatting (3DGS) has emerged as a technique for real-time novel view synthesis, offering explicit scene representation and efficient rendering. Concurrently, diffusion models have demonstrated unprecedented capabilities in generating and manipulating complex, high-fidelity data distributions. This survey explores the rapidly evolving intersection of these two powerful paradigms: the enhancement of 3DGS using diffusion models. It systematically categorizes recent research that uses diffusion priors to overcome key challenges in 3DGS, specifically examining how these models are integrated into different stages of the pipeline. These integrations include generating Gaussian parameters, providing optimization regularization, refining outputs, and enabling generative capabilities like text-to-3DGS. Existing approaches are analyzed and compared based on their technical innovations, strengths, and limitations. Furthermore, open challenges, such as computational efficiency, multi-view consistency, and controllability, are identified, and promising future research directions are outlined. This survey aims to provide researchers and practitioners with a structured understanding of how diffusion models are advancing the state-of-the-art (SOTA) in 3DGS, fostering further innovation in efficient and generative 3D scene representation.

**Keywords:** 3D Gaussian Splatting, diffusion models, novel view synthesis, 3D scene representation, generative 3D reconstruction, denoising diffusion probabilistic models, sparse view reconstruction, semantic scene editing, text-to-3D generation.

### Introduction

The accelerating demand for immersive technologies—spanning augmented reality (AR), virtual production, and autonomous robotics—has intensified the need for efficient, high-fidelity 3D scene reconstruction and rendering. 3D Gaussian Splatting (3DGS) (Kerbl, et al. 2023) has revolutionized this domain with its explicit, point-based scene representation, enabling real-time photorealistic novel view synthesis while dramatically reducing computational costs compared to neural radiance fields (NeRFs) (Mildenhall, et al. 2020). Despite its impact, 3DGS faces persistent challenges: sensitivity to initialization, artifacts under sparse inputs, and limited geometric coherence. These limitations hinder its deployment in applications requiring robustness to imperfect data or creative control, such as dynamic scene editing and generative content creation.

Concurrently, diffusion models have emerged as a paradigm-shifting force in generative artificial intelligence (AI), demonstrating unprecedented capabilities in synthesizing complex, high-dimensional data distributions through iterative denoising. By learning rich priors from vast datasets, these models excel at hallucinating plausible structures from partial inputs, refining noisy observations, and enabling semantic manipulation via natural language guidance (Rombach, et al. 2022; Blattmann, et al. 2023). Their probabilistic framework offers a natural mechanism to address the ill-posed nature of 3D reconstruction, where multiple valid solutions may explain sparse visual evidence.

Diffusion models enhance 3DGS by providing priors that improve optimization, mitigate artifacts through training regularization, refine results via post-processing, and unlock generative capabilities such as text-to-3D synthesis. This convergence bridges the gap between efficient rendering and generative intelligence, empowering applications from interactive scene editing to on-demand virtual world creation. Yet, research remains fragmented across disparate methodologies with no unified analysis of their trade-offs, scalability, or fundamental limitations.

This survey provides an examination of diffusion-enhanced 3DGS. Cutting-edge approaches are systematized into a coherent taxonomy, their technical innovations analyzed, and performance evaluated across key metrics.

### Structure of the Survey

The remainder of this article is organized as follows: Section 2 reviews foundational concepts in 3DGS and diffusion models. Section 3 categorizes diffusion-enhanced 3DGS methodologies by integration strategy (diffusion-guided 3DGS optimization, direct 3DGS generation from latent space, diffusion for consistent novel view synthesis). Section 4 analyzes fundamental challenges and limitations in benchmarking diffusion-enhanced 3DGS methods. Section 5 discusses unresolved challenges and emerging paradigms. Section 6 concludes with reflections on the future of generative 3D scene representation.



### Foundational Concepts

This section establishes the mathematical foundations of 3DGS and diffusion models, providing the theoretical underpinnings for their integration. The representation, optimization, and rendering processes of 3DGS are formalized (figure 1 provides an overview of the 3DGS framework), followed by a rigorous treatment of diffusion probabilistic models (the forward and reverse processes of a diffusion model are illustrated in figure 3) and their connection to score-based generative frameworks.

#### 3D Gaussian Splatting (Kerbl, et al. 2023)

##### Primitive Representation

A 3D scene is fundamentally represented as a collection of anisotropic Gaussian primitives  $\mathbf{G} = \{G_i\}_{i=1}^N$ , where each primitive  $G_i$  constitutes a learnable volumetric entity parameterized by its spatial position  $\boldsymbol{\mu}_i \in \mathbb{R}^3$ , covariance matrix  $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3 \times 3}$ , opacity  $\alpha_i \in \mathbb{R} \cap [0,1]$ , and view-dependent appearance modeled through spherical harmonics (SH) coefficients  $\boldsymbol{\psi}_i \in \mathbb{R}^k$ . The Gaussian function defines a radially decaying density field centered at  $\boldsymbol{\mu}_i$ :

$$G_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right).$$

To maintain physical validity during optimization,  $\boldsymbol{\Sigma}_i$  is constrained to positive semi-definite configurations through a differentiable factorization into rotational and scaling components:

$$\boldsymbol{\Sigma}_i = \mathbf{R}_i \mathbf{S}_i \mathbf{S}_i^\top \mathbf{R}_i^\top.$$

Here,  $\mathbf{R}_i \in \text{SO}(3)$  represents a rotation matrix derived from a normalized quaternion  $\mathbf{q}_i \in \mathbb{R}^4$  to avoid gimbal lock, while  $\mathbf{S}_i = \text{diag}(\mathbf{s}_i)$  constitutes a scaling matrix with  $\mathbf{s}_i \in \mathbb{R}_+^3$  enforcing anisotropic stretching along principal axes. This parameterization ensures both numerical stability and gradient tractability during optimization.

##### View-Dependent Appearance

The appearance model captures complex light transport effects by encoding color  $\mathbf{c}_i$  as a function of viewing direction  $\mathbf{d}$  via spherical harmonics basis expansion:

$$\mathbf{c}_i(\mathbf{d}) = \sum_{l=0}^L \sum_{m=-l}^l \boldsymbol{\psi}_i^{(l,m)} Y_l^m(\mathbf{d}),$$

where  $Y_l^m$  denotes the spherical harmonics basis function of degree  $l$  and order  $m$ , and  $\boldsymbol{\psi}_i^{(l,m)} \in \mathbb{R}^3$  contains RGB coefficients per band.

The maximum band  $L$  governs representational capacity, with higher bands capturing finer specular details at the cost of increased parameter dimensionality. Direction  $\mathbf{d}$  is typically derived from the normalized vector between the Gaussian mean  $\boldsymbol{\mu}_i$  and the camera origin. This frequency-based decomposition provides rotationally invariant directional representation while maintaining differentiability essential for gradient-based optimization.

##### Differentiable Rendering

The image synthesis process employs a differentiable rasterizer that aggregates contributions from  $K$  depth-ordered Gaussians per pixel  $\mathbf{p}$ . The composited color follows alpha blending with transmittance accumulation:

$$\mathbf{c}(\mathbf{p}) = \sum_{k=1}^K \mathbf{c}_k \hat{\alpha}_k \prod_{j=1}^{k-1} (1 - \hat{\alpha}_j),$$

where  $\hat{\alpha}_k = \alpha_k \cdot G_k(\mathbf{x}_k)$  represents the projected opacity modulated by the Gaussian's evaluation at its projected screen-space position  $\mathbf{x}_k$ .

The elliptical weighted average (EWA) splatting technique projects 3D covariance  $\boldsymbol{\Sigma}_i$  to 2D screen space via the Jacobian  $\mathbf{J}$  of the projective transformation:

$$\hat{\boldsymbol{\Sigma}}_i = \mathbf{J} \boldsymbol{\Sigma}_i \mathbf{J}^\top.$$

This projection accounts for perspective distortion during rasterization, ensuring accurate shape preservation across view changes. The sorting operation employs a tile-based parallelization strategy where Gaussians are binned into screen-space tiles prior to depth ordering, enabling efficient GPU implementation.

## Optimization

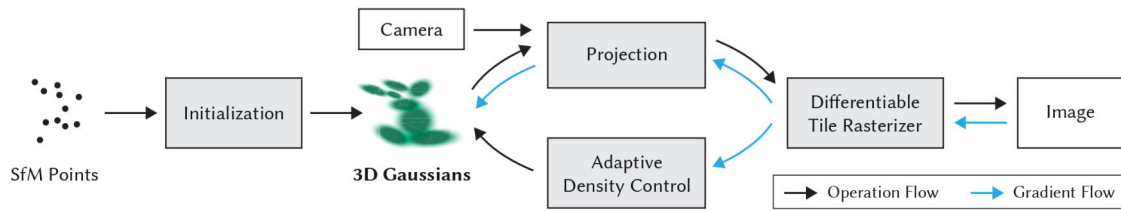


Figure 1. 3DGS framework overview. Optimization starts with the sparse SfM point cloud and creates a set of 3D Gaussians. Subsequent steps optimize and adaptively control the density of this Gaussian set. A fast tile-based renderer is utilized during optimization, enabling competitive training times compared to SOTA fast radiance field methods. After training, the renderer permits real-time navigation across a wide variety of scenes

Parameter optimization minimizes the photometric loss between rendered and ground truth images:

$$\mathcal{L} = \lambda \cdot \text{L1}(\text{render}, \text{target}) + (1 - \lambda) \cdot \text{DSSIM}(\text{render}, \text{target}),$$

where  $\text{DSSIM}(\cdot, \cdot)$  enhances structural similarity through a dissimilarity term based on the SSIM metric.

Parameters  $\theta_i = \{\mu_i, \mathbf{q}_i, \mathbf{s}_i, \alpha_i, \Psi_i\}$  are updated via stochastic gradient descent with adaptive learning rates, where positional parameters  $\mu_i$  typically employ higher rates than rotational  $\mathbf{q}_i$  or appearance  $\Psi_i$  components.

Adaptive densification dynamically regulates primitive density based on spatial gradient analysis (as shown in figure 2). Regions exhibiting significant view-space position gradients are identified as under-reconstructed, triggering Gaussian cloning with positional perturbation to increase coverage. Conversely, large Gaussians in high-frequency regions are split along principal axes to resolve fine details. Pruning mechanisms periodically remove primitives with opacity  $\alpha_i$  below a threshold  $\tau$  or those residing in low-density regions, maintaining computational efficiency. The optimization alternates between geometric refinement and appearance adjustment, progressively enhancing scene representation fidelity across iterations.

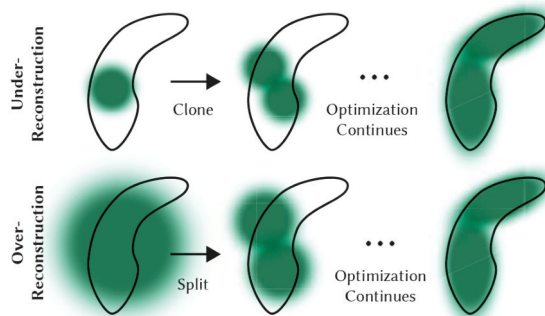


Figure 2. 3DGS framework overview. The adaptive Gaussian densification scheme operates through two primary mechanisms. For under-reconstruction cases shown in the top row, insufficient coverage of small-scale geometry (black outline) triggers cloning of the respective Gaussian. For over-reconstruction cases depicted in the bottom row, small-scale geometry represented by a single large splat undergoes splitting into two components

Recent advancements in adaptive densification (Kheradmand, et al. 2025; Bulò, et al. 2024; Grubert, et al. 2025; Deng, et al. 2025) have demonstrated superior alternatives to conventional heuristic approaches for regulating primitive density. Rather than relying solely on spatial gradient analysis to identify under-reconstructed regions, enhanced methodologies implement structured refinements.

Several studies propose replacing cloning and splitting operations with long-axis splitting strategies, which strategically position new Gaussians to minimize overlap while preserving density distributions. Complementary adaptive pruning techniques significantly reduce redundant primitives by dynamically eliminating low-opacity Gaussians based on iterative opacity thresholds or significance-aware metrics that evaluate rendering impact.

Further improvements incorporate dynamic adjustments to densification thresholds, including exponential scheduling protocols that progressively modulate gradient sensitivity during optimization. To promote efficient resource utilization, regularization terms encourage sparsity in opacity and scale parameters, implicitly pruning underutilized primitives.



Certain frameworks reinterpret Gaussian placement as Markov Chain Monte Carlo (MCMC) sampling, where principled relocation strategies maintain rendering consistency while redistributing Gaussians. The optimization alternates between these enhanced geometric refinements and appearance adjustments, progressively enhancing scene representation fidelity across iterations.

These collective advances address fundamental limitations in gradient-based densification, yielding improved robustness to initialization and more efficient Gaussian utilization without compromising reconstruction quality.

### Diffusion Models (Ho, et al. 2020)

Diffusion models constitute a prominent family of generative approaches that learn complex data distributions through a dual process of iterative noise corruption and denoising. These models derive their theoretical foundation from non-equilibrium thermodynamics, where a forward diffusion process progressively injects Gaussian noise into data samples over multiple timesteps, transforming structured data into pure noise. Conversely, a learned reverse process systematically removes noise to recover coherent data structures from random noise inputs. The method's efficacy stems from its stable training dynamics compared to adversarial approaches and its capacity to model complex distributions without restrictive assumptions about data topology.

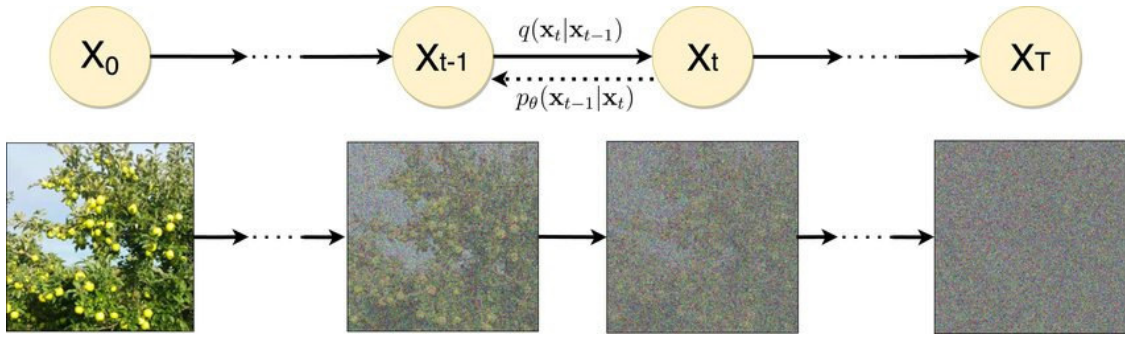


Figure 3. Diffusion models overview. The graphical illustration of diffusion models with their fixed forward process  $q(\mathbf{x}_t|\mathbf{x}_{t-1})$  and learnt backward process  $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$

The forward diffusion trajectory operates as a Markov chain that incrementally adds noise to an initial data sample  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ . At each timestep  $t$ , the conditional distribution  $q(\mathbf{x}_t|\mathbf{x}_{t-1})$  is governed by a Gaussian transition parameterized by a pre-determined variance schedule  $\beta_t \in (0,1)$ . This schedule controls the rate of information degradation, typically following a monotonically increasing function that accelerates noise addition in later timesteps. The cumulative effect of  $T$  transitions admits a closed-form expression due to the Gaussian nature of the process:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

where  $\alpha_t = 1 - \beta_t$ , and  $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ .

This reparameterization enables efficient sampling of  $\mathbf{x}_t$  for arbitrary  $t$  without simulating the entire Markov chain, significantly accelerating training.

The generative reverse process is parameterized by a neural network that learns to invert the diffusion trajectory. Starting from isotropic Gaussian noise  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ , the model iteratively refines the sample through transitions:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)).$$

Common implementations fix the covariance  $\boldsymbol{\Sigma}_\theta$  to time-dependent constants, while the mean  $\boldsymbol{\mu}_\theta$  is reparameterized to predict the injected noise:

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right).$$

Here  $\boldsymbol{\epsilon}_\theta$  denotes a deep neural network, typically a U-Net architecture with residual blocks and self-attention mechanisms, conditioned on timestep embeddings. The training objective simplifies to a denoising score matching loss that minimizes the discrepancy between true and predicted noise:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t, \epsilon} [\|\epsilon - \epsilon_{\theta}(\mathbf{x}_t, t)\|^2],$$

where  $t \sim \mathcal{U}(\mathbf{1}, \mathbf{T})$ , and  $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ .

This objective exhibits superior training stability compared to the full variational lower bound, as it circumvents compounding prediction errors across timesteps. During inference, sampling initiates from  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  and applies the learned reverse transitions for  $T$  iterations. Advanced samplers such as denoising diffusion implicit models (DDIMs) further accelerate generation by leveraging non-Markovian trajectories that reduce required steps.

Conditional generation extends the framework by augmenting the noise predictor  $\epsilon_{\theta}$  with contextual information  $\mathbf{y}$  (e.g., class labels or text embeddings), yielding  $\epsilon_{\theta}(\mathbf{x}_t, t, \mathbf{y})$ . Guidance techniques like classifier-free diffusion employ stochastic conditioning to enhance sample fidelity without auxiliary models. Recent innovations integrate diffusion with latent spaces, hierarchical refinement, and hybrid architectures, enabling applications spanning high-resolution image synthesis, molecular design, and video generation. The model's inherent flexibility and strong mode coverage continue to drive empirical advances in generative modeling.

### Methodological Taxonomy

This section synthesizes contemporary research integrating diffusion models with 3DGS, categorized by three synergistic paradigms: diffusion-guided 3DGS optimization, direct 3DGS generation from latent space, and diffusion for consistent novel view synthesis. Each approach addresses specific limitations in the 3DGS pipeline—such as sparse view reconstruction, geometric inconsistencies, and generative capability gaps—by leveraging diffusion priors at distinct stages of the scene representation workflow. The three synergistic paradigms discussed in this section are summarized in figure 4.

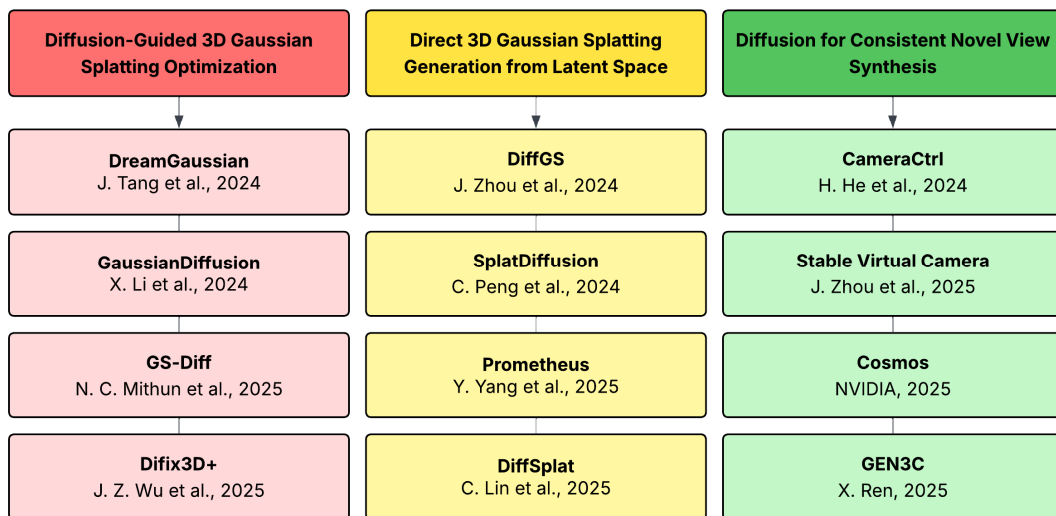


Figure 4. Three synergistic paradigms: diffusion-guided 3DGS optimization, direct 3DGS generation from latent space, and diffusion for consistent novel view synthesis

### Diffusion-Guided 3D Gaussian Splatting Optimization

*DreamGaussian (Tang, et al. 2024)*

DreamGaussian pioneers the integration of diffusion priors with 3DGS to address efficiency bottlenecks in optimization-based 3D generation. The framework employs a two-stage pipeline for both image-to-3D and text-to-3D tasks. An overview of the DreamGaussian framework is shown in figure 5. In the first stage, 3D Gaussians are optimized via score distillation sampling (SDS) (Poole, et al. 2022) from 2D diffusion models, leveraging progressive densification to accelerate convergence compared to NeRF-based approaches. This stage generates coarse geometry and appearance within seconds but suffers from texture blurriness due to SDS ambiguity. The second stage addresses this limitation through novel mesh extraction and UV space refinement: an efficient local density query algorithm converts Gaussians to textured meshes, followed by a diffusion-guided fine-tuning that minimizes pixel-wise mean squared error (MSE) losses between rendered

views and diffusion-refined images. This refinement explicitly disentangles texture details from geometric ambiguities while avoiding the over-saturation artifacts typical of direct SDS application in UV space.

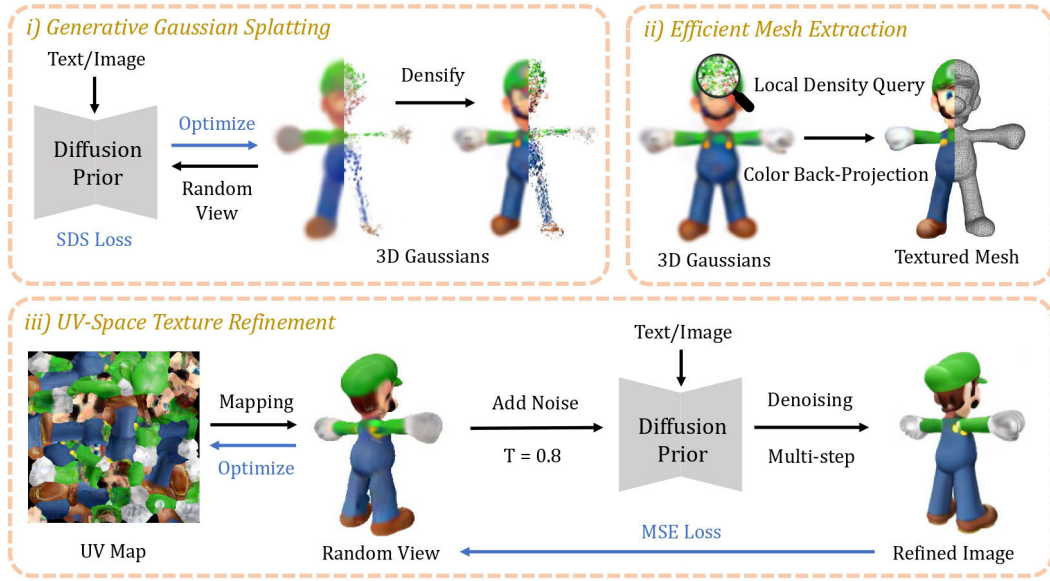


Figure 5. DreamGaussian framework overview

Despite achieving a  $10 \times$  speedup over prior methods, DreamGaussian remains susceptible to the multi-face Janus problem (see figure 6 for an example). This artifact, characterized by semantically incoherent surfaces featuring duplicated frontal features (e.g., multiple faces), stems from inconsistent 3D guidance provided by 2D diffusion priors, highlighting inherent cross-view consistency challenges.



Figure 6. The Janus problem in distilling 3D knowledge from 2D diffusion models refers to a failure mode where generated 3D objects exhibit multiple, inconsistent faces (e.g., two front-facing views) on a single object

The Janus problem—manifesting as geometrically inconsistent surfaces—arises from fundamental limitations in text-to-3D optimization. Distilling knowledge from 2D diffusion models into 3D representations introduces geometric contradictions because these models generate view-specific outputs optimized for local photorealism without explicit cross-view constraints during optimization. Consequently, adjacent viewpoints often synthesize conflicting surface orientations or textures, violating 3D consistency. This issue is exacerbated early in optimization, where high noise levels amplify stochastic variations across views, and in regions with sparse supervision, leading to unconstrained extrapolation and hallucinated structures.

Compounding this issue, the non-convex optimization landscape traps parameters in view-dependent local minima. These minima satisfy individual view losses but collectively produce degenerate geometry incompatible with physical coherence.

Consequently, the Janus artifact represents not merely a rendering flaw but an optimization pathology inherent to disentangled view-wise distillation.

*GaussianDiffusion (Li, et al. 2024)*

To overcome the core challenge of multi-face Janus artifacts, GaussianDiffusion introduces structured 3D noise injection. Unlike methods sampling independent 2D noise per viewpoint, this approach generates structured noise by rendering randomized 3D Gaussians into view-dependent 2D distributions, ensuring multi-view coherence. Consequently, the denoising process inherently satisfies cross-view geometric constraints without requiring diffusion model fine-tuning, preserving its realism. The GaussianDiffusion framework is depicted in figure 7.

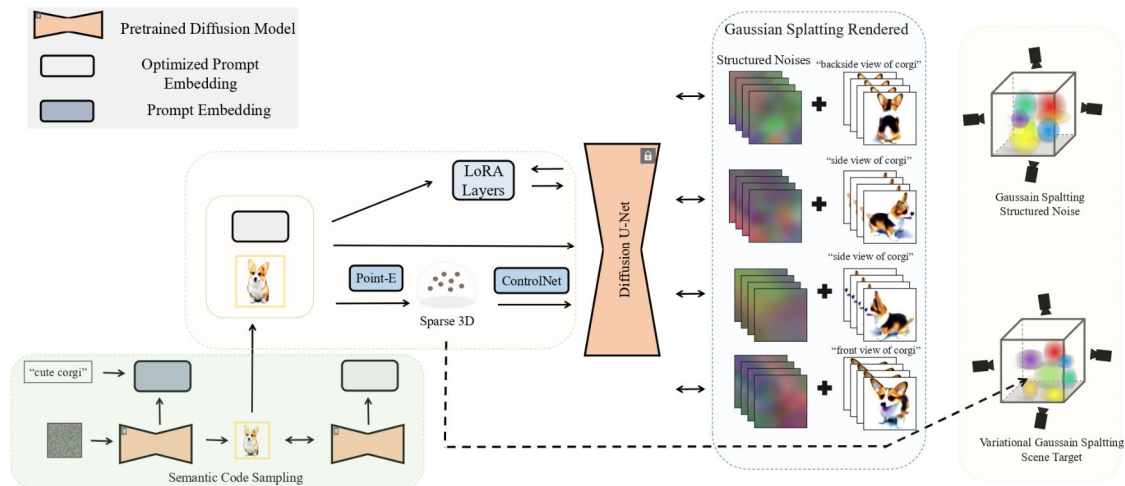


Figure 7. GaussianDiffusion framework overview. A sparse point cloud generated from an image using Point-E undergoes pose projection to create a depth map. This depth map functions as a geometric constraint for ControlNet. Simultaneously, low-rank adaptation (LoRA) provides additional optimization through fine-tuning of the diffusion model. The original sparse point cloud from Point-E serves as the initial input to the 3DGS process. Gradients from the diffusion model are conveyed to 3DGS via SDS. Mitigating issues related to multi-view consistency and artifacts involves the introduction of structured noise and a variational 3DGS approach, resulting in a realistic 3D appearance

Complementing this, variational 3DGS addresses local minima by modeling Gaussian parameters (e.g., position, scale) as distributions rather than point estimates. Perturbing these parameters with noise scaled to the diffusion timestep encourages coarse-to-fine optimization, expands the convergence domain, and enhances stability.

Quantitatively, GaussianDiffusion demonstrates superior geometric consistency through reduced pose-prediction variance in COLMAP evaluations, though it incurs longer training times than highly optimized baselines like DreamGaussian.

The methodologies diverge in their handling of diffusion priors and geometric constraints. DreamGaussian relies on conventional SDS for initial optimization but circumvents its limitations via separate mesh refinement. GaussianDiffusion rethinks noise perturbation at its core, using shared 3D noise sources and variational distributions to embed multi-view consistency directly into the gradient estimation process. Consequently, DreamGaussian excels in rapid asset generation for downstream applications, while GaussianDiffusion advances robustness and geometric fidelity for complex text prompts, albeit at a higher computational cost.

*GS-Diff (Mithun, et al. 2025)*

The GS-Diff framework shares foundational principles with prior diffusion-guided 3D reconstruction methods in its utilization of generative priors to regularize underconstrained optimization. Similar to DreamGaussian, GS-Diff leverages diffusion models to mitigate ambiguities arising from sparse or inconsistent inputs. However, while DreamGaussian employs SDS from 2D diffusion models for geometry optimization, GS-Diff diverges significantly by integrating a multi-view diffusion model to synthesize geometrically consistent pseudo-observations. This approach aligns more closely with GaussianDiffusion's philosophy of enforcing cross-view consistency through structured 3D constraints but operates within an explicit reconstruction paradigm rather than generative synthesis. The GS-Diff framework overview is presented in figure 8.



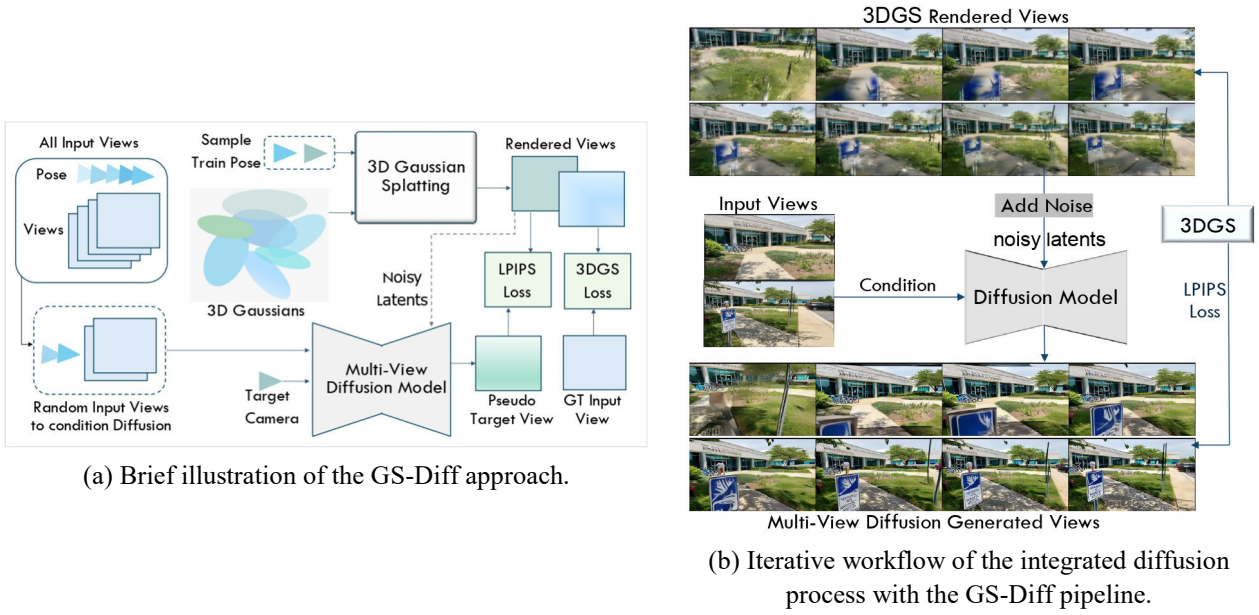


Figure 8. GS-Diff framework overview: (a) conceptual illustration and (b) iterative diffusion process workflow.

GS-Diff adopts an iterative optimization strategy that interleaves diffusion guidance with 3D Gaussian updates at regular intervals. This stands in contrast to DreamGaussian’s sequential pipeline, which separates geometry optimization from texture refinement. By conditioning pseudo-view generation on neighboring input views via camera trajectory interpolation, GS-Diff implicitly enforces multi-view coherence—addressing the Janus problem without requiring explicit 3D noise projection. The method further shares GaussianDiffusion’s emphasis on stability through the introduction of an LPIPS (Zhang et al. 2018) loss threshold, which dynamically excludes inconsistent pseudo-views during training. This integration mitigates the risk of hallucination inherent in diffusion models.

#### *Difix3D+* (Wu, et al. 2025)

The Difix3D+ framework shares fundamental operational principles with contemporary approaches that leverage diffusion models for 3D reconstruction enhancement, particularly in its core strategy of distilling diffusion-refined pseudo-views into the underlying 3D representation. Similar to Deceptive-NeRF (Xinhang Liu, et al. 2024) and 3DGS-Enhancer (Xi Liu, et al. 2024), Difix3D+ employs a diffusion model to correct artifacts in rendered views, subsequently using these enhanced images as pseudo-ground truth to update the 3D model parameters. This shared methodology circumvents the computational burden of per-optimization-step diffusion queries by treating the diffusion model as an offline enhancer that generates improved training data. The distillation process aligns with established paradigms where generative priors are transferred to the 3D representation through iterative refinement of pseudo-observations, thereby addressing underconstrained regions while maintaining multi-view consistency. The architecture of the Difix3D+ framework is shown in figure 9.

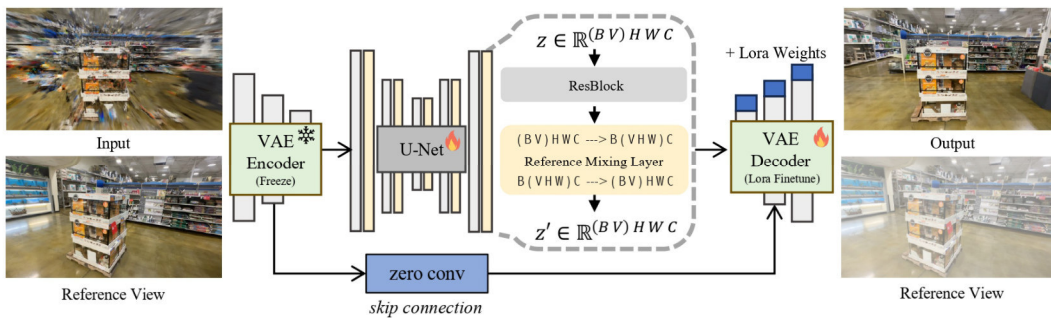


Figure 9. Difix3D+ framework overview. Difix processes a noisy rendered image alongside reference views as input (left), producing an enhanced output image with reduced artifacts (right). The system also generates identical reference views, discarded during practical application and consequently depicted as transparent. The architecture employs a U-Net structure featuring a cross-view reference mixing layer to ensure consistency across reference views. Difix is fine-tuned from SD-Turbo (Sauer et al. 2023, 2024), utilizing a frozen variational autoencoder (VAE) and a LoRA-adapted decoder

Difix3D+ further echoes prior works in its dual-phase utilization of diffusion guidance: during optimization to improve the 3D representation and during inference to refine rendered outputs. This bifurcated approach mirrors architectures like GANeRF (Roessle, et al. 2023) and NeRFLiX (Zhou, Li, Wang, et al. 2023; Zhou, Li, Jiang, et al. 2023), where generative models enhance both the reconstruction process and final rendering quality. The conditioning mechanism on reference views—implemented via cross-attention layers—draws inspiration from multi-view diffusion models (e.g., MVDream (Shi, et al. 2024), SyncDreamer (Y. Liu, et al. 2024)) that aggregate information from input perspectives to maintain contextual coherence. By inheriting these design principles, Difix3D+ operates within the broader trend of integrating 2D generative priors to compensate for 3D reconstruction limitations, particularly in sparse-view scenarios where geometric ambiguities persist.

However, Difix3D+ introduces critical innovations to this shared foundation. The progressive 3D update pipeline—inspired by Instruct-NeRF2NeRF (Haque, et al. 2023)—iteratively refines camera trajectories and augments training data, enabling consistent artifact correction even in extreme novel views. This contrasts with single-step distillation in prior methods, which often struggle with long-range consistency. Additionally, the adoption of a single-step diffusion model (SD-Turbo (Sauer, et al. 2023, 2024)) significantly accelerates both training and inference, addressing efficiency limitations in earlier diffusion-guided frameworks. While maintaining conceptual alignment with existing paradigms, these advancements position Difix3D+ as a scalable solution for large-scale scenes where computational overhead traditionally impedes deployment.

Strong conditioning of the diffusion model on rendered novel views and reference views is crucial for achieving multi-view consistency and high fidelity to input perspectives. When the target trajectory is distant from input views, the conditioning signal weakens, forcing the diffusion model to hallucinate excessively. To mitigate this, an iterative training scheme analogous to Instruct-NeRF2NeRF (Haque, et al. 2023) is adopted, progressively expanding the set of 3D cues rendered to novel viewpoints. Starting with reference views, the 3D representation undergoes optimization with periodic perturbation of ground-truth camera poses toward target views. The resulting novel views are rendered, refined by Difix, and added to the training set for further optimization cycles. This gradual pose perturbation and data augmentation enhance 3D consistency in challenging regions.

Despite distillation, slight multi-view inconsistencies and residual blur persist due to limitations in reconstruction capacity. Difix3D+ addresses this by applying Difix as a final post-processing step during inference, removing artifacts while preserving coherence. Leveraging SD-Turbo's (Sauer, et al. 2023, 2024) single-step architecture, this adds only 76 ms per frame on an NVIDIA A100 GPU—significantly faster than multi-step diffusion models. This dual optimization-inference application of diffusion guidance extends prior frameworks where generative enhancement was typically restricted to training.

### **Direct 3D Gaussian Splatting Generation from Latent Space**

These approaches train diffusion models to directly generate the 3DGS representation itself, often by learning a compact latent space of the Gaussians or functions defining them.

*DiffGS (Zhou, et al. 2024)*

DiffGS exemplifies this paradigm by reformulating the unstructured nature of 3DGS representation into a continuous functional representation, enabling latent diffusion over Gaussian attributes. Similar to concurrent methods like GaussianCube (Zhang, et al. 2024) and GVGEN (He, et al. 2024), which structure Gaussians into volumetric grids for tractable generation, DiffGS circumvents the discrete challenges of 3DGS representation by introducing three disentangled continuous functions: Gaussian probability function (GauPF), Gaussian color function (GauCF), and Gaussian transform function (GauTF). These functions collectively parameterize the geometry, appearance, and transformations of the Gaussians, effectively distilling the unstructured 3DGS representation into a compact, generative latent space. The DiffGS framework is illustrated in figure 10.

A Gaussian VAE compresses these functions into a low-dimensional latent vector, regularizing the space for stable diffusion training. The latent diffusion model (LDM) then operates in this space, denoising samples to generate novel functional representations conditioned on inputs like text, images, or partial 3DGS. This mirrors the latent-space generation strategies of triplane-based LDMs (e.g., Rodin (Wang, et al. 2022) and 3DGen (Gupta, et al. 2023)), but avoids their reliance on grid structures by directly modeling Gaussian properties through neural predictors. After generation, an octree-guided discretization algorithm extracts discrete Gaussians from the continuous functions. This process optimizes proxy points toward high-probability regions defined by the GauPF, then queries the GauCF and GauTF to assign attributes. Analogous to marching



cubes for implicit fields—but tailored to Gaussian primitives—this approach ensures scalable extraction of Gaussians at arbitrary densities, free from voxel-resolution constraints.

While methods like DreamGaussian optimize Gaussians via SDS, DiffGS directly generates the 3DGS representation end-to-end via latent diffusion, eliminating per-scene optimization. This functional approach enhances efficiency and generality, though it inherits challenges in color consistency and geometric fidelity under sparse conditioning, as seen in analogous latent-space generative models.

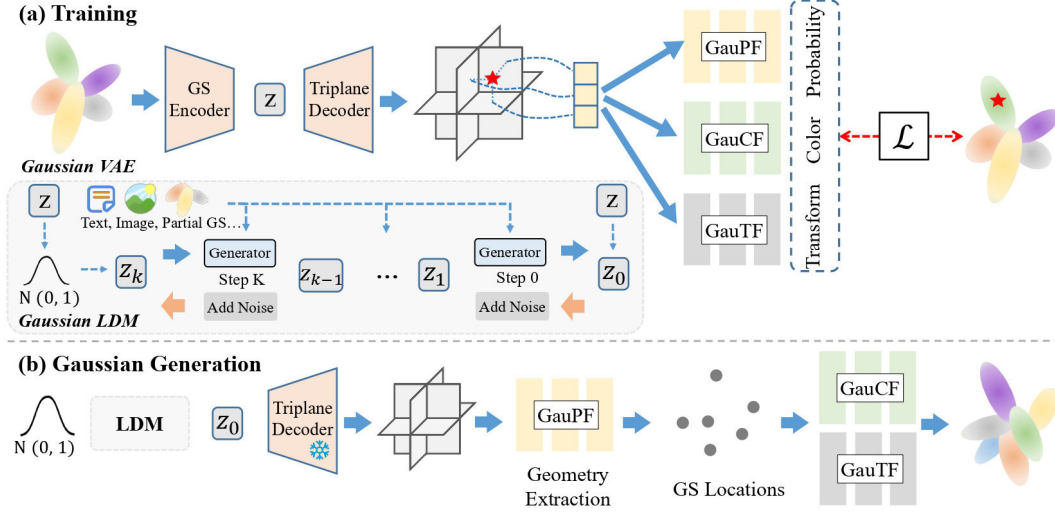


Figure 10. DiffGS framework overview. (a) The fitted 3DGS is disentangled into three 3DGS functions to model Gaussian probabilities (GauPF), Gaussian colors (GauCF), and Gaussian transforms (GauTF). A Gaussian VAE is then trained with a conditional latent diffusion model (LDM) to generate these functions. (b) During generation, Gaussian geometry is first extracted from the generated GauPF, followed by applying GauCF and GauTF to obtain Gaussian attributes

#### *SplatDiffusion (Peng, et al. 2024)*

SplatDiffusion addresses the fundamental challenge of modality mismatch in diffusion-based 3D generation by introducing a teacher-guided framework that decouples the denoised signal domain from the supervision domain. Unlike conventional diffusion paradigms requiring aligned 3D supervision, SplatDiffusion leverages pre-trained deterministic predictors (e.g., Splatter Image (Szymanowicz, et al. 2024), Flash3D (Szymanowicz, et al. 2025)) as "noisy teachers" to synthesize corrupted 3DGS samples. This strategy circumvents the scarcity of 3D ground truth by utilizing imperfect teacher predictions as pseudo-targets for diffusion training. Crucially, at noise levels beyond a critical timestep  $t^*$ , the distribution of teacher-generated noisy samples approximates that of forward-noised ground truth, inspired by SDEdit (Meng, et al. 2022) principles.

The framework operates in two synergistic stages. The two-stage SplatDiffusion framework is overviewed in figure 11. First, during bootstrapping, the diffusion model learns single-step denoising using direct 3D supervision from the teacher's predictions combined with 2D rendering losses. This initial phase ensures computational efficiency while aligning the model with the teacher's capabilities. Second, multi-step denoising fine-tuning replaces single-step prediction with iterative refinement across multiple timesteps. Here, the model progressively denoises from  $t > t^*$  to  $t = 0$ , supervised exclusively by image reconstruction losses on rendered views. This sequence mirrors inference dynamics, enabling gradient propagation through the full denoising trajectory and significantly enhancing detail recovery.

Additionally, SplatDiffusion integrates a cycle consistency loss, where predicted novel views drive secondary 3D reconstructions that are rendered back to the source view. This regularizes spatial coherence and mitigates geometric drift. For multi-view inputs, the model incorporates guidance gradients during sampling—modulating noise estimates using discrepancies between rendered and target views—which further elevates reconstruction quality without architectural modifications. By operating directly on 3DGS parameters and using only 2D supervision, SplatDiffusion achieves SOTA novel view synthesis while sidestepping the limitations of 3D data dependency. Its flexibility allows seamless integration with diverse teacher models, demonstrating consistent gains across both object-level and scene-level benchmarks.

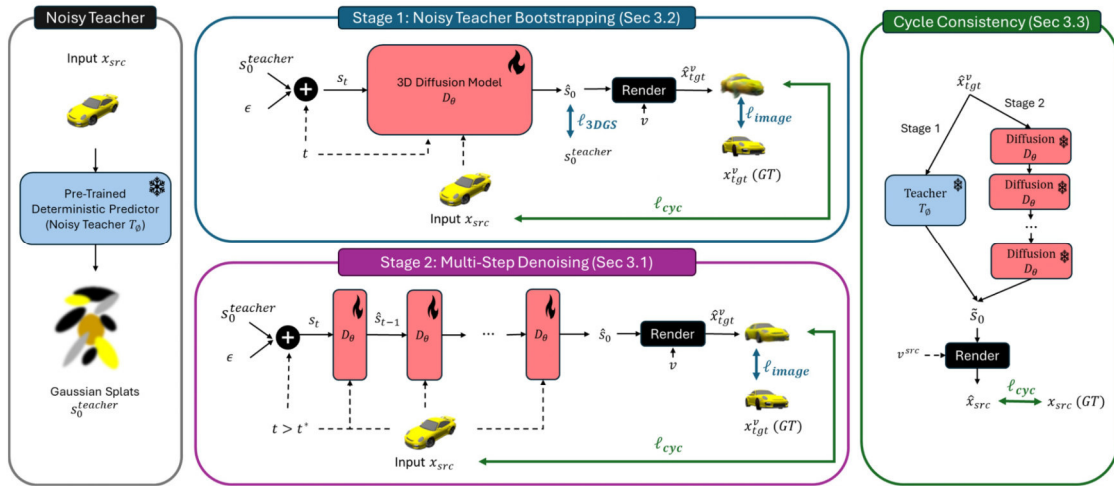


Figure 11. SplatDiffusion framework overview. This framework utilizes a pre-trained deterministic predictor network for 3DGS, termed the "noisy teacher" (left). In stage 1 (top), sampled views are lifted to generate an imperfect 3DGS prediction. This prediction provides noisy samples and supervision for the diffusion denoiser in 3DGS, augmented with additional image supervision. In stage 2 (bottom), the noisy samples are decoupled from supervision. Instead, the noisy teacher generates noisy samples at noise levels  $t > t^*$ , while a multi-step denoising strategy generates high-quality predictions to facilitate image-only supervision. Both stages incorporate cycle consistency regularization.

### Prometheus (Yang, et al. 2025)

Prometheus aligns with emerging paradigms that leverage LDMs (Rombach, et al. 2022) to directly synthesize 3DGS representations, circumventing intermediate optimization or explicit multi-view image synthesis. Similar to DiffGS, Prometheus employs a functional compression strategy: it distills complex 3DGS attributes into a structured latent space amenable to diffusion modeling. However, while DiffGS decomposes Gaussians into geometry, color, and transformation functions, Prometheus adopts a multi-view latent conditioning approach. It encodes RGB-D image observations into a joint latent space using a frozen Stable Diffusion encoder, then decodes these latents into pixel-aligned 3D Gaussians via a modified Stable Diffusion decoder. This formulation implicitly embeds Gaussian properties (position, rotation, scale, opacity, spherical harmonics) within the latent-to-3DGS mapping, effectively structuring the unstructured nature of 3DGS through the lens of 2D priors. The Prometheus training pipeline is shown in figure 12.

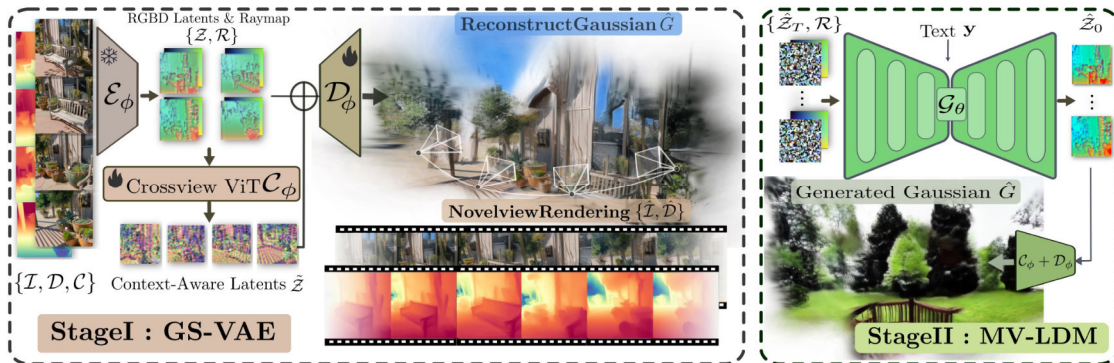


Figure 12. Prometheus framework overview. The training process is divided into two stages. In stage 1, the objective is to train a GS-VAE. Utilizing multi-view images along with their corresponding pseudo-depth maps and camera poses, the GS-VAE is designed to encode these multi-view RGB-D images, integrate cross-view information, and ultimately decode them into pixel-aligned 3DGS. In stage 2, the focus shifts to training an MV-LDM. Multi-view RGB-D latents can be generated by sampling from randomly-sampled noise with the trained MV-LDM.

The diffusion process in Prometheus operates on multi-view RGB-D latents rather than explicit Gaussian parameters. MV-LDM denoises these latents conditioned on camera poses and text prompts, analogous to how DiffGS denoises functional representations. Crucially, Prometheus introduces cross-view transformers to fuse latent information across camera views, ensuring spatial coherence in the generated 3DGS—addressing a key

challenge in multi-view generative models. This fusion mechanism, coupled with depth-aware latent conditioning, disentangles geometric and appearance attributes, enhancing fidelity. Post-diffusion, the GS-VAE decoder reconstructs the 3DGS scene in a feed-forward manner, paralleling DiffGS's function-to-Gaussian extraction but optimized for scene-level generalization.

While both methods avoid per-scene optimization, Prometheus uniquely harnesses hybrid training data that combines single-view and multi-view inputs, along with a pre-aligned RGB-D latent space, to exploit massive 2D datasets. This contrasts with DiffGS's reliance on explicit 3D function learning. The resulting model achieves rapid generation with robustness to diverse prompts, though it inherits latent diffusion challenges such as sensitivity to noise schedules and guidance trade-offs.

*DiffSplat (Lin, et al. 2025)*

DiffSplat aligns with emerging methodologies that harness pre-trained 2D diffusion models for direct 3DGS generation, circumventing the limitations of intermediate multi-view synthesis or explicit 3D dataset dependencies. Unlike approaches that structure Gaussians into volumetric grids or functional representations, DiffSplat reinterprets multi-view 3DGS property grids as stylized images, enabling the repurposing of large-scale image diffusion priors. This strategy leverages the inherent capability of image diffusion models to infer 3D geometry—such as depth and surface normals—from web-scale 2D data, thereby bridging the domain gap between image generation and 3D content creation. The DiffSplat framework is summarized in figure 13.

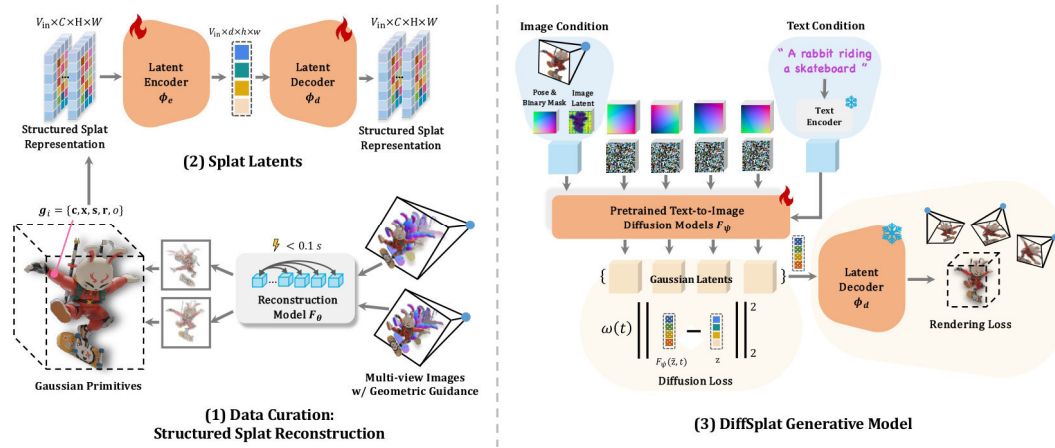


Figure 13. DiffSplat framework overview. (1) A lightweight reconstruction model creates high-quality structured representations for synthetic dataset curation. (2) An image VAE is fine-tuned to encode 3DGS properties into a shared latent space. (3) DiffSplat leverages 2D priors from text-to-image diffusion models to natively generate 3D content from image or text prompts.

Central to DiffSplat is a lightweight reconstruction model that regresses pixel-aligned Gaussian attributes (position, rotation, scale, color, opacity) from multi-view images in under 0.1 seconds. These attributes are structured into 2D grids, analogous to multi-channel images, which serve as scalable pseudo-ground truth data for training. To align these grids with the latent space of image diffusion models, the framework fine-tunes the VAE of pre-trained LDMs (e.g., Stable Diffusion, SDXL), compressing 3DGS property grids into 3DGS latents. This adaptation ensures compatibility with the input expectations of 2D denoising networks while preserving the spatial structure of Gaussian attributes.

The generative backbone of DiffSplat operates by denoising 3DGS latents conditioned on text or single-view images. Crucially, DiffSplat incorporates a 3D rendering loss alongside the standard diffusion loss. This dual-loss mechanism—trained end-to-end—ensures that the model adheres to 3D geometric constraints while leveraging the rich priors embedded in image diffusion architectures. For multi-view generation, DiffSplat explores view-concat and spatial-concat paradigms, where 3DGS latents are fused along view or spatial dimensions, augmented with Plücker embeddings to encode relative camera poses. Minimal architectural modifications (e.g., zero-initialized layers for pose embeddings) ensure seamless integration with diverse base models, enabling techniques like ControlNet to be adapted for controllable 3D generation.

By treating 3DGS property grids as a stylized image modality, DiffSplat effectively "fine-tunes" image diffusion models to generate 3D-consistent Gaussian primitives, eliminating the need for two-stage pipelines

or explicit 3D supervision. This approach not only capitalizes on the scalability of 2D diffusion models but also establishes a unified framework for text-to-3D and image-to-3D tasks, achieving SOTA fidelity and generalizability across benchmarks.

### Diffusion for Consistent Novel View Synthesis

This approach primarily focuses on generating high-quality, 3D-consistent 2D novel views using diffusion models. While it does not explicitly build or refine a 3DGS representation internally for its primary function, the consistency and quality of its generated views are sufficient to enable the reconstruction of a coherent 3DGS model as a downstream task.

*CameraCtrl (He, et al. 2025)*

CameraCtrl represents a significant advancement in integrating explicit camera control within video diffusion models, addressing a critical gap in cinematic controllability for generative video synthesis. Unlike conventional approaches that lack precise viewpoint manipulation, CameraCtrl introduces a plug-and-play module that enables accurate trajectory control without modifying core model parameters. The methodology employs Plücker embeddings as the primary camera representation, encoding intrinsic and extrinsic parameters into pixel-wise ray descriptors that provide geometrically interpretable conditioning signals. This formulation captures the epipolar geometry underlying camera movements, offering richer spatial cues than numerical parameter inputs. The camera control module is deliberately decoupled from appearance learning; it processes only Plücker embeddings, ensuring domain-agnostic generalization across diverse visual styles. This design prevents appearance leakage from training data, a common limitation in conditioning frameworks that fuse image latents with control signals. The CameraCtrl framework overview is provided in figure 14.

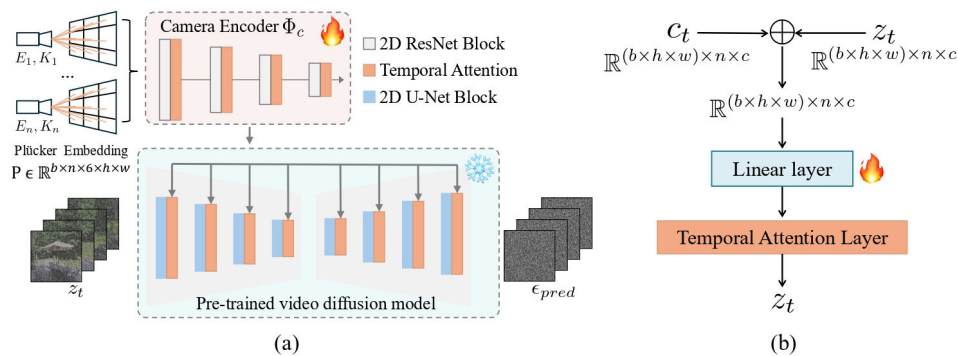


Figure 14. CameraCtrl framework overview. (a) Given a pre-trained video diffusion model, CameraCtrl trains a camera encoder on it, which takes the Plücker embedding as input and outputs multi-scale camera representations. These features are then integrated into the temporal attention layers of the U-Net at their respective scales to control the video generation process. (b) Details of the camera injection process. The camera features  $c_t$  and the latent features  $z_t$  are first combined through the element-wise addition. A learnable linear layer is adopted to further fuse two representations which are then fed into the first temporal attention layer of each temporal block

Integration occurs through injection into temporal attention layers of the U-Net backbone, aligning camera dynamics with the sequential nature of video generation. The features derived from Plücker sequences are fused via element-wise addition and linear transformation before modulating temporal self-attention. Empirical analysis reveals that this temporal integration outperforms spatial attention fusion, as camera trajectories inherently exhibit causal dependencies across frames. For training, RealEstate10K is selected due to its diverse camera distributions and appearance alignment with common video diffusion datasets. This choice balances trajectory complexity and visual realism, mitigating geometric ambiguities that arise from synthetic or narrow-distribution data. Quantitative metrics confirm CameraCtrl's superiority in trajectory adherence (RotErr, TransErr) over alternatives like MotionCtrl (Wang, et al. 2024), while maintaining baseline visual quality (FVD, CLIPSIm).

The framework demonstrates versatility across text-to-video (e.g., AnimateDiff (Guo, et al. 2024)) and image-to-video (e.g., Stable Video Diffusion (Blattmann, et al. 2023)) pipelines, and seamlessly combines with structural controllers like SparseCtrl (Guo, et al. 2023) for multimodal conditioning. Its plug-and-play nature enables rapid adoption in production workflows, though limitations persist in extreme rotational regimes due to dataset biases. Conceptually, CameraCtrl bridges cinematographic principles with generative AI, enabling dynamic storytelling through controlled viewpoint dynamics.



### Stable Virtual Camera (SEVA) (Zhou, et al. 2025)

Similarly, SEVA leverages diffusion models for 3D-consistent view synthesis but operates without explicit intermediate 3D representations. While methods like DreamGaussian and GaussianDiffusion directly optimize 3D Gaussians using diffusion priors, SEVA focuses on synthesizing photorealistic 2D novel views that implicitly satisfy 3D consistency. This approach conditions the diffusion process on input views and their camera poses, utilizing a latent denoising U-Net enhanced with 3D self-attention and temporal operators. The model inherits strong priors from pre-trained 2D diffusion architectures, enabling it to generate large viewpoint changes and smooth interpolations without relying on explicit geometric scaffolds. The SEVA framework is illustrated in figure 15.

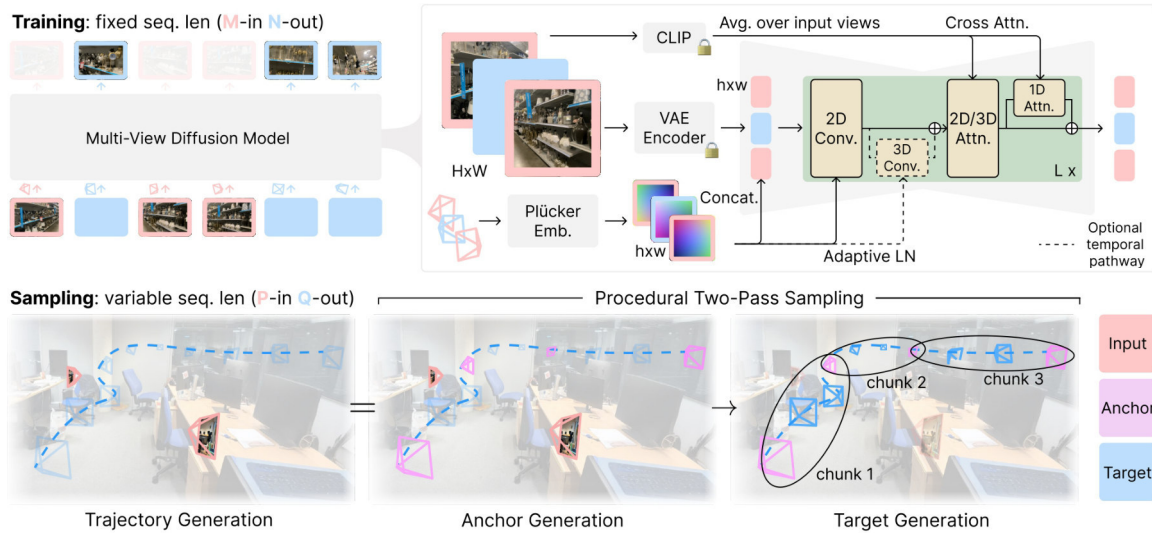


Figure 15. SEVA framework overview. SEVA is trained with a fixed sequence length as an "M-in N-out" multi-view diffusion model using a standard architecture. It conditions on CLIP embeddings, VAE latents of the input views, and their corresponding camera poses. During sampling, SEVA can be cast as a generative "P-in Q-out" renderer operating with variable sequence length, where P and Q need not equal M and N. To enhance temporal and 3D consistency across generated views, particularly when generating along a trajectory, procedural two-pass sampling is presented as a general strategy

Crucially, SEVA's procedural two-pass sampling strategy ensures temporal and spatial coherence across generated views. For long trajectories, a memory bank of anchor views maintains long-range consistency by retrieving spatially nearest neighbors during autoregressive generation. The resulting views exhibit sufficient geometric consistency to enable high-quality 3DGS reconstruction as a downstream task, as validated by distillation experiments. Unlike optimization-based methods, SEVA sidesteps the Janus problem during synthesis by avoiding entangled 3D parameter optimization, instead relying on diffusion's generative capacity to hallucinate plausible yet consistent unseen regions. However, it shares the challenge of handling dynamic textures and ambiguous scenes, where generation uncertainty may necessitate careful guidance scale tuning to balance detail fidelity and coherence. Thus, while SEVA diverges from explicit 3DGS optimization, its output quality and consistency functionally enable comparable downstream 3D reconstruction, streamlining the pipeline through decoupled view synthesis and geometry distillation.

### Cosmos (Agarwal, et al. 2025; Alhaija, et al. 2025)

Similarly, the Cosmos world foundation model (WFM) employs diffusion-based architectures for generating 3D-consistent novel views without explicit intermediate geometric representations. Like SEVA, it conditions the diffusion process on camera parameters—specifically leveraging Plücker coordinate embeddings to encode relative camera poses—alongside visual observations from preceding frames. This approach integrates temporal and spatial conditioning through a modified DiT backbone, utilizing hybrid positional embeddings with FPS-aware 3D RoPE (Su, et al. 2023) and learnable absolute embeddings to maintain coherence across arbitrary viewpoints and trajectory lengths. The model's latent space, constructed via a causal tokenizer, compresses high-dimensional visual inputs while preserving critical spatial and temporal relationships, enabling efficient denoising in a reduced-dimensional domain.

For extended sequences, Cosmos adopts a progressive training strategy that incrementally increases resolution and context length, coupled with multi-aspect bucketing to handle diverse scene compositions. This design mirrors SEVA's emphasis on cross-view consistency but extends it through joint image-video training, where domain-specific normalization aligns latent distributions across modalities. The resulting outputs exhibit sufficient geometric fidelity to facilitate downstream 3D reconstruction tasks, as validated by metrics like Sampson error and novel-view synthesis quality. However, Cosmos diverges by directly supporting dynamic perturbations (e.g., robotic actions or autonomous vehicle controls), broadening its applicability beyond static scene synthesis. This flexibility introduces challenges in physics adherence, particularly for contact-rich interactions, which remain an active area of refinement.

*GEN3C (Ren, et al. 2025)*

GEN3C represents a significant advancement in 3D-consistent video generation by integrating explicit geometric scaffolding into a diffusion-based framework. Unlike purely latent-space approaches, GEN3C constructs a spatiotemporal 3D cache—a collection of colored point clouds generated by unprojecting depth estimates from input images or pre-generated video frames. This cache serves as an approximate 3D scene representation, rendered into 2D videos conditioned on user-specified camera trajectories. The core innovation lies in leveraging these renderings as strong geometric priors for a video diffusion model (e.g., Stable Video Diffusion (Blattmann, et al. 2023) or Cosmos (Agarwal, et al. 2025; Alhaija, et al. 2025)), which is fine-tuned to translate imperfect cache renderings into high-fidelity, temporally coherent videos. The GEN3C framework overview is shown in figure (Ren, et al. 2025).

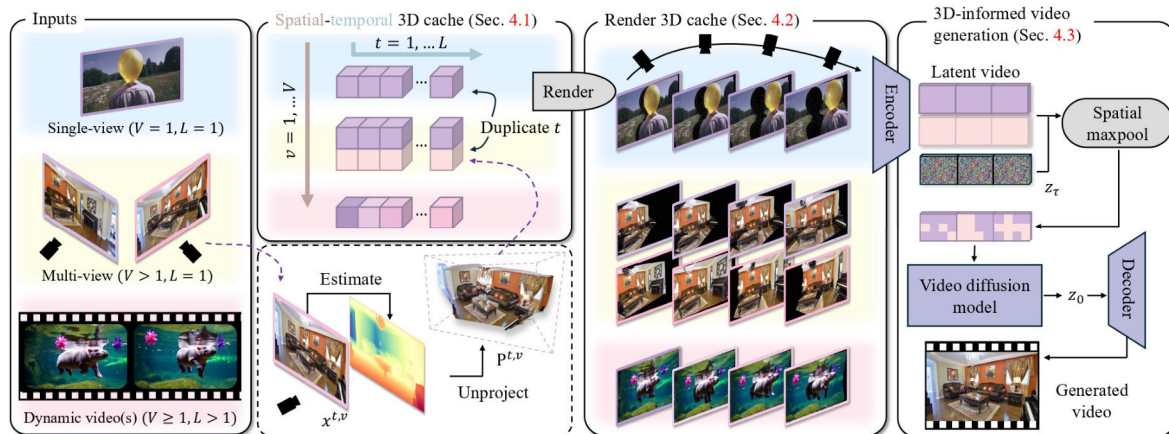


Figure 16. GEN3C framework overview. Given user input—which can be a single-view image, multi-view images, or dynamic videos—a spatiotemporal 3D cache is constructed through depth prediction for each image and subsequent unprojection into 3D. Using camera poses provided by the user, the cache is rendered into videos. These rendered videos are then fed into a video diffusion model to generate a photorealistic video aligned with the target camera poses

The 3D cache dynamically adapts to application contexts: for single-image inputs, it duplicates the initial point cloud across frames; for multi-view static scenes, it aggregates point clouds from each view; and for dynamic scenes, it integrates per-frame depth predictions from source videos. During diffusion, rendered cache frames and disocclusion masks are encoded into latent space, masked to exclude uncovered regions, and fused via viewpoint-invariant max-pooling before conditioning the denoising U-Net. This design ensures the diffusion model focuses generative capacity on disoccluded areas and dynamic refinements rather than inferring scene geometry from camera parameters alone.

GEN3C addresses long-range consistency through autoregressive chunk generation with cache updates. For extended sequences, previously generated frames are depth-estimated, scale-aligned to the existing cache via reprojection error minimization, and unprojected to augment the 3D representation iteratively. This enables applications like driving scene simulation, where horizontal camera shifts beyond training trajectories are synthesized plausibly. Crucially, the explicit cache permits direct 3D editing (e.g., object removal or trajectory modification) by manipulating point clouds before re-rendering and re-generation.

Quantitatively, GEN3C achieves SOTA performance in sparse-view novel view synthesis. Its conditioning strategy yields superior camera controllability over CameraCtrl, particularly under out-of-domain trajectories,



while maintaining efficiency. Limitations include dependence on pre-computed dynamics for dynamic scenes and sensitivity to extreme depth noise. By bridging explicit 3D proxies with generative priors, GEN3C establishes a scalable paradigm for cinematic control and 3D-consistent synthesis in sparse-view settings.

### Comparative Evaluation Limitations

The rapid proliferation of diffusion-based methodologies for enhancing 3DGS optimization presents significant challenges for systematic benchmarking and comparative evaluation. These limitations stem from fundamental methodological inconsistencies, resource constraints inherent to academic research, and heterogeneous practices in code sharing within the computer vision and graphics communities, as summarized in table 1. This section delineates the structural impediments that preclude direct quantitative comparison across SOTA techniques, contextualizing these challenges within broader research practices while maintaining scholarly rigor.

Table 1. **Code availability of diffusion-enhanced 3DGS methods. The table shows the availability of training and inference code for each method. The table is based on the information provided in the original papers and the availability of the code on the authors’ GitHub repositories**

Method	Training code available	Inference code available
<b>Diffusion-Guided 3D Gaussian Splatting Optimization</b>		
DreamGaussian (Tang et al. 2024)	Uses third-party diffusion models	Yes
GaussianDiffusion (Li et al. 2024)	No	No
GS-Diff (Mithun et al. 2025)	No	No
Difix3D+ (Wu et al. 2025)	Yes	Yes
<b>Direct 3D Gaussian Splatting Generation from Latent Space</b>		
DiffGS (Zhou et al. 2024)	Yes	Yes
SplatDiffusion (Peng et al. 2024)	No	No
Prometheus (Yang et al. 2025)	Partially available	Yes
DiffSplat (Lin et al. 2025)	Yes	Yes
<b>Diffusion for Consistent Novel View Synthesis</b>		
CameraCtrl (He et al. 2025)	Yes	Yes
SEVA (Zhou et al. 2025)	No	Yes
Cosmos (NVIDIA, :, Agarwal, et al. 2025; NVIDIA, :, Alhajja, et al. 2025)	Yes	Yes
GEN3C (Ren et al. 2025)	No	Yes

### Heterogeneity of Training Data and Experimental Conditions

A primary barrier to comparative analysis lies in the substantial divergence in training datasets and experimental protocols employed across existing studies. Research efforts in diffusion-enhanced 3DGS optimization exhibit pronounced dataset specificity, with methodologies developed and evaluated against bespoke collections of scenes that vary dramatically in content complexity, image resolution, capture conditions, and scene diversity. This dataset fragmentation creates an irreconcilable normalization challenge: performance metrics including PSNR, SSIM, and LPIPS become fundamentally incomparable when derived from non-overlapping visual domains with varying texture richness, lighting conditions, and structural complexity. Consequently, any cross-study quantitative comparison risks confounding methodological efficacy with dataset-specific characteristics. Moreover, the absence of standardized evaluation suites tailored to diffusion-enhanced 3DGS exacerbates this issue, as researchers select validation scenes that optimally demonstrate their method’s strengths while potentially obscuring limitations observable under alternative data regimes.

### Resource Disparities and Training Inaccessibility

The computational burden associated with retraining diffusion models for 3DGS guidance constitutes a prohibitive barrier to controlled evaluation. Contemporary diffusion frameworks require extensive computational resources—typically hundreds of GPU hours—and specialized hyperparameter tuning to achieve published performance levels. The majority of surveyed works provide only inference code and pretrained

checkpoints, omitting critical implementation details necessary for replication of training pipelines. This practice renders method retraining practically infeasible for researchers lacking institutional-scale computational infrastructure, effectively preventing standardized evaluation on common datasets. Even when partial training code exists, documentation gaps regarding data preprocessing, augmentation strategies, and optimization schedules introduce significant reproducibility uncertainties. The challenge intensifies for approaches incorporating custom diffusion architectures or novel conditioning mechanisms, where architectural details are frequently under-specified in publications and absent from repositories. Consequently, the research community faces an asymmetrical landscape where novel methodologies can be empirically validated only against their authors' predetermined baselines under non-reproducible conditions, rather than through independent comparative assessment.

### **Implementation Inconsistencies and Code Availability Gaps**

The absence of standardized implementations across diffusion-enhanced 3DGS frameworks presents another critical barrier to fair benchmarking. Many influential papers in this domain either provide no public implementation at all or release only partial inference codebases lacking training infrastructure. This necessitates arduous reimplementation efforts that introduce significant fidelity risks—particularly when reconstructing complex diffusion-3DGS interaction mechanisms described ambiguously in prose.

Furthermore, the rapid publication tempo in this field outpaces the community's ability to establish standardized evaluation baselines or reproduce existing methods before newer approaches displace them.

### **Metric Limitations and Qualitative Performance Dimensions**

Established quantitative metrics for novel view synthesis inadequately capture the perceptual and structural nuances critical to diffusion-enhanced 3DGS outcomes. Traditional measures like PSNR and SSIM prioritize pixel-level fidelity over geometric coherence, failing to penalize artifacts specific to Gaussian representations such as floaters, over-saturation in high-frequency regions, or topological inconsistencies. These metrics prove particularly inadequate for evaluating diffusion guidance contributions, which often target higher-order scene attributes including material consistency, illumination stability, and view-consistent texture synthesis—dimensions poorly quantified by current measures. The problem intensifies for editing applications, where no standardized metrics exist to assess how diffusion enhancement facilitates semantic manipulation while preserving geometric integrity. Consequently, qualitative assessment remains indispensable yet methodologically problematic: visual comparisons suffer from selective scene presentation and rendering parameter variations across studies. Moreover, the most significant advantages of diffusion-enhanced approaches—such as reduced artifacts in sparse-view reconstruction or enhanced detail recovery in textureless regions—manifest under specific challenging conditions rarely represented in standardized test suites.

These challenges do not diminish the substantive advances in diffusion-enhanced 3DGS optimization but rather highlight the growing pains of an explosively evolving field. As the community coalesces around standardized evaluation practices, the current methodological heterogeneity may ultimately yield robust hybrid approaches synthesizing the strengths of diverse guidance strategies. However, present constraints necessitate scholarly candor regarding the inferential boundaries of cross-method assessment.

### **Open Challenges and Future Directions**

Despite significant advancements in diffusion-enhanced 3DGS, several fundamental challenges persist. Computational efficiency remains a critical bottleneck, as integrating iterative diffusion sampling with 3DGS optimization creates prohibitive training and inference costs—particularly for high-resolution or dynamic scenes. This is exacerbated by the memory-intensive nature of denoising processes, limiting real-time deployment in applications like augmented reality. Multi-view consistency also presents unresolved difficulties, as evidenced by persistent "Janus artifacts" in generative pipelines like DreamGaussian and GS-Diff, where 2D diffusion priors fail to enforce coherent 3D geometry across viewpoints. Such inconsistencies are amplified in sparse-view reconstruction scenarios, where geometric ambiguities lead to artifacts.

Controllability constitutes another significant gap. While methods like CameraCtrl enable trajectory control via Plücker embeddings, they struggle with extreme camera rotations or complex object-centric manipulations. Semantic editing—such as material or lighting changes—often requires per-scene optimization rather than flexible, high-level guidance. Furthermore, dynamic scene modeling remains largely underdeveloped; most frameworks prioritize static scenes, with nascent approaches like GEN3C relying heavily on pre-computed

motion data rather than generative dynamics. Physics-aware synthesis (e.g., collision responses or fluid interactions) is almost entirely unexplored within the 3DGS paradigm.

Scalability is equally pressing: current datasets like RealEstate10K or Objaverse lack the diversity needed for robust generalization, particularly for uncommon objects or cultural contexts. This data scarcity intensifies biases and limits cross-domain adaptability.

For dynamic scenes, integrating neural physics engines with 3DGS optimization could synthesize plausible object interactions without pre-specified motion data.

## Conclusion

This survey has examined the rapidly evolving synergy between diffusion models and 3DGS, highlighting a paradigm shift in generative 3D scene representation. Diffusion priors effectively address critical limitations inherent in conventional 3DGS pipelines, including sensitivity to initialization, geometric inconsistencies under sparse inputs, artifacts in novel view synthesis, and the lack of semantic controllability. Methodologies integrating these two powerful paradigms have been systematically categorized, revealing distinct yet complementary strategies: diffusion-guided 3DGS optimization leverages 2D or multi-view diffusion models to regularize training and refine geometry with textures; direct latent generation of 3DGS representations utilizes diffusion models operating on compact functional or parameter spaces derived from disentangled Gaussian attributes; and diffusion for consistent novel view synthesis generates multi-view coherent imagery subsequently distilled into high-fidelity 3DGS models.

Key innovations such as Plücker embeddings for geometrically meaningful camera conditioning, variational 3DGS formulations mitigating local minima, structured 3D noise injection ensuring multi-view consistency, and hybrid distillation-rendering losses demonstrate significant advancements. These approaches yield substantial improvements in rendering quality, geometric fidelity, training stability, and generative flexibility across diverse tasks—text-to-3D, image-to-3D, sparse-view reconstruction, and dynamic scene synthesis—while often preserving the real-time rendering advantage of 3DGS. However, fundamental challenges persist, including computational overhead during diffusion-guided optimization, inherent tensions between 2D perceptual quality and strict 3D consistency, limitations in handling complex material properties and lighting, and scalability to large-scale or dynamic scenes.

Future research directions are poised to focus on improving physical plausibility via material-aware and physics-informed modeling, and enabling finer-grained semantic and dynamic control through multimodal conditioning. The convergence of efficient explicit scene representation with powerful generative priors marks a transformative step towards interactive, photorealistic, and controllable 3D content creation, promising significant impact across domains such as augmented reality, virtual production, robotics, and digital twins.

## References

1. Blattmann, Andreas, Tim Dockhorn, Sumith Kulal, et al. 2023. *Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets*. <https://arxiv.org/abs/2311.15127>.
2. Bulò, Samuel Rota, Lorenzo Porzi, and Peter Kontschieder. 2024. *Revising Densification in Gaussian Splatting*. <https://arxiv.org/abs/2404.06109>.
3. Deng, Xiaobin, Changyu Diao, Min Li, Ruohan Yu, and Duanqing Xu. 2025. *Efficient Density Control for 3D Gaussian Splatting*. <https://arxiv.org/abs/2411.10133>.
4. Grubert, Glenn, Florian Barthel, Anna Hilsman, and Peter Eisert. 2025. “Improving Adaptive Density Control for 3D Gaussian Splatting.” *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 610–21. <https://doi.org/10.5220/0013308500003912>.
5. Guo, Yuwei, Ceyuan Yang, Anyi Rao, et al. 2024. *AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models Without Specific Tuning*. <https://arxiv.org/abs/2307.04725>.
6. Guo, Yuwei, Ceyuan Yang, Anyi Rao, Maneesh Agrawala, Dahua Lin, and Bo Dai. 2023. *SparseCtrl: Adding Sparse Controls to Text-to-Video Diffusion Models*. <https://arxiv.org/abs/2311.16933>.
7. Gupta, Anchit, Wenhan Xiong, Yixin Nie, Ian Jones, and Barlas Oğuz. 2023. *3DGen: Triplane Latent Diffusion for Textured Mesh Generation*. <https://arxiv.org/abs/2303.05371>.
8. Haque, Ayaan, Matthew Tancik, Alexei A. Efros, Aleksander Holynski, and Angjoo Kanazawa. 2023. *Instruct-NeRF2NeRF: Editing 3D Scenes with Instructions*. <https://arxiv.org/abs/2303.12789>.
9. He, Hao, Yinghao Xu, Yuwei Guo, et al. 2025. *CameraCtrl: Enabling Camera Control for Text-to-Video Generation*. <https://arxiv.org/abs/2404.02101>.

10. He, Xianglong, Junyi Chen, Sida Peng, et al. 2024. *GVGEN: Text-to-3D Generation with Volumetric Representation*. <https://arxiv.org/abs/2403.12957>.
11. Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. *Denoising Diffusion Probabilistic Models*. <https://arxiv.org/abs/2006.11239>.
12. Kerbl, Bernhard, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. *3D Gaussian Splatting for Real-Time Radiance Field Rendering*. <https://arxiv.org/abs/2308.04079>.
13. Kheradmand, Shakiba, Daniel Rebain, Gopal Sharma, et al. 2025. *3D Gaussian Splatting as Markov Chain Monte Carlo*. <https://arxiv.org/abs/2404.09591>.
14. Li, Xinhai, Huaibin Wang, and Kuo-Kun Tseng. 2024. *GaussianDiffusion: 3D Gaussian Splatting for Denoising Diffusion Probabilistic Models with Structured Noise*. <https://arxiv.org/abs/2311.11221>.
15. Lin, Chenguo, Panwang Pan, Bangbang Yang, Zeming Li, and Yadong Mu. 2025. *DiffSplat: Repurposing Image Diffusion Models for Scalable Gaussian Splat Generation*. <https://arxiv.org/abs/2501.16764>.
16. Liu, Xinhang, Jiaben Chen, Shiu-hong Kao, Yu-Wing Tai, and Chi-Keung Tang. 2024. *Deceptive-NeRF/3DGS: Diffusion-Generated Pseudo-Observations for High-Quality Sparse-View Reconstruction*. <https://arxiv.org/abs/2305.15171>.
17. Liu, Xi, Chaoyi Zhou, and Siyu Huang. 2024. *3DGS-Enhancer: Enhancing Unbounded 3D Gaussian Splatting with View-Consistent 2D Diffusion Priors*. <https://arxiv.org/abs/2410.16266>.
18. Liu, Yuan, Cheng Lin, Zijiao Zeng, et al. 2024. *SyncDreamer: Generating Multiview-Consistent Images from a Single-View Image*. <https://arxiv.org/abs/2309.03453>.
19. Meng, Chenlin, Yutong He, Yang Song, et al. 2022. *SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations*. <https://arxiv.org/abs/2108.01073>.
20. Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. <https://arxiv.org/abs/2003.08934>.
21. Mithun, Niluthpol Chowdhury, Tuan Pham, Qiao Wang, et al. 2025. *Diffusion-Guided Gaussian Splatting for Large-Scale Unconstrained 3D Reconstruction and Novel View Synthesis*. <https://arxiv.org/abs/2504.01960>.
22. Niket Agarwal, et al. 2025. *Cosmos World Foundation Model Platform for Physical AI*. <https://arxiv.org/abs/2501.03575>.
23. Hassan Abu Alhaija, et al. 2025. *Cosmos-Transfer1: Conditional World Generation with Adaptive Multimodal Control*. <https://arxiv.org/abs/2503.14492>.
24. Peng, Chensheng, Ido Sobol, Masayoshi Tomizuka, Kurt Keutzer, Chenfeng Xu, and Or Litany. 2024. *A Lesson in Splats: Teacher-Guided Diffusion for 3D Gaussian Splats Generation with 2D Supervision*. <https://arxiv.org/abs/2412.00623>.
25. Poole, Ben, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2022. *DreamFusion: Text-to-3D Using 2D Diffusion*. <https://arxiv.org/abs/2209.14988>.
26. Ren, Xuanchi, Tianchang Shen, Jiahui Huang, et al. 2025. *GEN3C: 3D-Informed World-Consistent Video Generation with Precise Camera Control*. <https://arxiv.org/abs/2503.03751>.
27. Roessle, Barbara, Norman Müller, Lorenzo Porzi, Samuel Rota Bulò, Peter Kontschieder, and Matthias Niessner. 2023. "GANeRF: Leveraging Discriminators to Optimize Neural Radiance Fields." *ACM Transactions on Graphics* 42 (6): 1–14. <https://doi.org/10.1145/3618402>.
28. Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. *High-Resolution Image Synthesis with Latent Diffusion Models*. <https://arxiv.org/abs/2112.10752>.
29. Sauer, Axel, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. 2024. *Fast High-Resolution Image Synthesis with Latent Adversarial Diffusion Distillation*. <https://arxiv.org/abs/2403.12015>.
30. Sauer, Axel, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. 2023. *Adversarial Diffusion Distillation*. <https://arxiv.org/abs/2311.17042>.
31. Shi, Yichun, Peng Wang, Jianglong Ye, Mai Long, Kejie Li, and Xiao Yang. 2024. *MVDream: Multi-View Diffusion for 3D Generation*. <https://arxiv.org/abs/2308.16512>.
32. Su, Jianlin, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. 2023. *RoFormer: Enhanced Transformer with Rotary Position Embedding*. <https://arxiv.org/abs/2104.09864>.
33. Szymanowicz, Stanislaw, Eldar Insafutdinov, Chuanxia Zheng, et al. 2025. *Flash3D: Feed-Forward Generalisable 3D Scene Reconstruction from a Single Image*. <https://arxiv.org/abs/2406.04343>.
34. Szymanowicz, Stanislaw, Christian Rupprecht, and Andrea Vedaldi. 2024. *Splatter Image: Ultra-Fast Single-View 3D Reconstruction*. <https://arxiv.org/abs/2312.13150>.
35. Tang, Jiaxiang, Jiawei Ren, Hang Zhou, Ziwei Liu, and Gang Zeng. 2024. *DreamGaussian: Generative Gaussian Splatting for Efficient 3D Content Creation*. <https://arxiv.org/abs/2309.16653>.
36. Wang, Tengfei, Bo Zhang, Ting Zhang, et al. 2022. *Rodin: A Generative Model for Sculpting 3D Digital Avatars Using Diffusion*. <https://arxiv.org/abs/2212.06135>.

- 
37. Wang, Zhouxia, Ziyang Yuan, Xintao Wang, et al. 2024. *MotionCtrl: A Unified and Flexible Motion Controller for Video Generation*. <https://arxiv.org/abs/2312.03641>.
  38. Wu, Jay Zhangjie, Yuxuan Zhang, Haithem Turki, et al. 2025. *Difix3D+: Improving 3D Reconstructions with Single-Step Diffusion Models*. <https://arxiv.org/abs/2503.01774>.
  39. Yang, Yuanbo, Jiahao Shao, Xinyang Li, Yujun Shen, Andreas Geiger, and Yiyi Liao. 2025. *Prometheus: 3D-Aware Latent Diffusion Models for Feed-Forward Text-to-3D Scene Generation*. <https://arxiv.org/abs/2412.21117>.
  40. Zhang, Bowen, Yiji Cheng, Jiaolong Yang, et al. 2024. *GaussianCube: A Structured and Explicit Radiance Representation for 3D Generative Modeling*. <https://arxiv.org/abs/2403.19655>.
  41. Zhang, Richard, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*. <https://arxiv.org/abs/1801.03924>.
  42. Zhou, Jensen, Hang Gao, Vikram Voleti, et al. 2025. *Stable Virtual Camera: Generative View Synthesis with Diffusion Models*. <https://arxiv.org/abs/2503.14489>.
  43. Zhou, Junsheng, Weiqi Zhang, and Yu-Shen Liu. 2024. *DiffGS: Functional Gaussian Splatting Diffusion*. <https://arxiv.org/abs/2410.19657>.
  44. Zhou, Kun, Wenbo Li, Nianjuan Jiang, Xiaoguang Han, and Jiangbo Lu. 2023. *From NeRFLiX to NeRFLiX++: A General NeRF-Agnostic Restorer Paradigm*. <https://arxiv.org/abs/2306.06388>.
  45. Zhou, Kun, Wenbo Li, Yi Wang, et al. 2023. *NeRFLiX: High-Quality Neural View Synthesis by Learning a Degradation-Driven Inter-Viewpoint MiXer*. <https://arxiv.org/abs/2303.06919>.

## T3D: Модульный программно-аппаратный комплекс реконструкции 3D-объектов

А. О. Жирков<sup>1</sup>, Р. О. Родионов<sup>1</sup>, П. Д. Копанев<sup>2</sup>, В. А. Фролов<sup>1</sup>

<sup>1</sup>Институт перспективных исследований проблем искусственного интеллекта  
и интеллектуальных систем МГУ им. М.В. Ломоносова, Москва, Россия

<sup>2</sup> Институт технологических инноваций, Абу Даби, ОАЭ

**Аннотация.** В данной работе предложен модульный программно-аппаратный комплекс для создания цифровых копий объектов с поверхностями, трудновоспроизводимыми при помощи существующих решений на основе фотограмметрии и активных трехмерных сканеров. Комплекс включает: (1) специализированную студию с вращающимся подиумом, антибликовой прозрачной платформой, системой фона, контролируемым LED-освещением и маркерной системой; (2) ПО на базе интеграции COLMAP и OpenMVS; (3) модуль оптимизации SVBRDF на основе обратного рендеринга при помощи рендер-системы Mitsuba3. Предложенный комплекс решает проблему создания цифровых копий объектов с проблемными поверхностями: полупрозрачными, зеркально отражающими, однотонными и тонкими. Это достигается за счёт введения классификации объектов и выбора метода реконструкции под каждый класс: например, при необходимости может быть подключен программный модуль 3DGS для реконструкции особо тонких и полупрозрачных поверхностей. Ключевой результат нашей работы – возможность автоматизированного создания фотореалистичной цифровой копии (геометрия, параметры материалов SVBRDF, текстура в высоком разрешении) за одно сканирование без перемещения объекта. Система ориентирована на относительно бюджетное и быстрое создание цифровых копий, но имеет возможность подключения аппаратного модуля оптического сканера в случае необходимости увеличения точности реконструкции геометрии. Экспериментально достигнуты метрики качества PSNR > 35 дБ, SSIM > 0.98, IoU > 98 % при времени сканирования менее 10 минут на объект. Система ориентирована на области применения, где важно достижение фотореалистичного визуального совпадения с оригиналом, а также требуется возможность помещения цифровой копии в другую виртуальную сцену (с другим освещением).

**Ключевые слова:** 3D-реконструкция, фотограмметрия, нейронный рендеринг, обратный рендеринг, цифровая копия, SVBRDF, материалы, автоматизация, компьютерное зрение, поворотный стол, антибликовое стекло, HDR-панорама.

## T3D: Modular hardware and software complex for reconstruction of 3D objects

A. O. Zhirkov<sup>1</sup>, R. O. Rodionov<sup>1</sup>, P. D. Kopanov<sup>2</sup>, V. A. Frolov<sup>1</sup>

<sup>1</sup>IAI Moscow State University, Moscow

<sup>2</sup> Technology Innovation Institute, Abu Dhabi, UAE

**Abstract.** In this paper, we propose a modular hardware and software package for creating digital copies of objects with surfaces that are difficult to reproduce using existing solutions based on photogrammetry and active three-dimensional scanners. The complex includes: (1) a specialized studio with a rotating podium, an anti-glare transparent platform, a background system controlled by LED lighting and a marker system; (2) Software based on the integration of COLMAP and OpenMVS; (3) SVBRDF optimization module based on reverse rendering using the Mitsuba3 rendering system. The proposed complex solves the problem of creating digital copies of objects with problematic surfaces: translucent, specularly reflecting, monophonic and thin. This is achieved by introducing a classification of objects and choosing a reconstruction method for each class: for example, if necessary, the 3DGS software module can be connected to reconstruct particularly thin and translucent surfaces. The key result of our work is the possibility of automated creation of a photorealistic digital copy (geometry, SVBRDF material parameters, texture in high resolution) in one scan without moving the object. The system is focused on relatively low-cost and fast creation of digital copies, but it has the ability to connect the optical scanner hardware module if it is necessary to increase the accuracy of geometry reconstruction. The quality metrics PSNR > 35 dB, SSIM > 0.98, and IoU > 98% have been experimentally achieved with a scan time of less than 10 minutes per object. The system is focused on applications where it is important to achieve a photorealistic visual match with the original, and also requires the ability to place a digital copy in another virtual scene (with different lighting).

**Keywords:** 3D reconstruction, photogrammetry, neural rendering, reverse rendering, digital copy, SVBRDF, materials, automation, computer vision, turntable, anti-glare glass, HDR panorama.

### Введение

Актуальность разработки систем 3D-реконструкции обусловлена растущим спросом на цифровые копии для различных задач, таких как Web/AR/VR, виртуальные музеи и электронная коммерция [Petrelli19]. Это чрезвычайно обширная область, существующие работы в которой можно условно разделить на три класса: (1) методы и алгоритмы реконструкции [Luan21; Huang24; Gao24];



(2) программно-аппаратные комплексы [Menna, et al., 2017; Apollonio F. I., et al., 2021; Karami, et al., 22]; (3) работы, которые ориентируются не на процесс реконструкции как таковой, а на создание набора данных (датасетов) для решения определенной задачи [Kuang23; Ummenhofer24; Dong25]. Методы обычно сфокусированы на решении некоторой выделенной проблемы реконструкции в условиях, близких к идеальным. Они часто игнорируют проблемы в исходных данных, порожденные неточностью измерительной и съемочной аппаратуры, неидеальными условиями съемки и т.д. Программно-аппаратные комплексы много внимания уделяют построению и функционированию самой аппаратуры с целью создания некоторой универсальной системы. При этом, как правило, используются существующие методы, которые комбинируются нетривиальным образом. Работы по датасетам уделяют больше внимания самому датасету в некоторой ограниченной постановке и, как правило, не претендуют ни на универсальность или новизну используемого программно-аппаратного комплекса, ни на создание новых методов.

Все три класса задач могут быть направлены на достижение различных целей: (А) решение базовой задачи реконструкции, которая заключается в синтезе новых видов объекта для альтернативного ракурса камеры [Mildenhall21; Kerbl23]; (В) решение задачи восстановления геометрии трехмерной модели или геометрии и цветовой текстуры [Laine20; Nicolet21; Huang24]; (С) решение комплексной задачи восстановления геометрии и материалов [Luan21; Gao24; Wang24]. Последний тип задачи, как правило, является наиболее сложным, поскольку он требует решения проблемы с неоднозначным решением: множество различных моделей могут корректно отображаться в виртуальной сцене, в которой проводилась реконструкция, но только некоторые из них сохраняют корректное отображение при переносе в другую виртуальную сцену с измененными условиями освещения и окружения. К этой категории относятся такие проблемы, как классическая задача запекания бликов в диффузных текстурах при использовании метода обратного рендеринга [Luan21; Wang24] и ограничения на реконструкцию зеркальных объектов с использованием методов 3DGS [Kerbl23].

В соответствии с предложенной классификацией данная работа относится к категории 2-С, что указывает на ее принадлежность к исследованиям, связанным с разработкой программно-аппаратного комплекса, решающего задачу одновременной реконструкции 3D геометрии и материалов. Комплекс является модульным в том смысле, что он позволяет нам относительно легко менять как отдельные аппаратные составляющие, так и программные компоненты. Это позволяет воспроизводить конвейер 3D-реконструкции в различных вариантах, в зависимости от особенности снимаемых объектов и требований к качеству цифровой копии.

### **Обзор предыдущих работ**

Существующие работы по дифференцированному рендерингу ([Luan21; Wang24] и др.) демонстрируют, что решение задачи (С) возможно с неплохой точностью в условиях, близких к идеальным. Однако подобные работы реконструируют объекты относительно простой формы, без характерных невыпуклых элементов поверхности и мелких деталей. Кроме того, существует большое число специализированных алгоритмов реконструкции, способных восстанавливать лишь определенные типы объектов, например, тела вращения или здания, состоящие из прямоугольных коробок [Ryabogin24; Garifullin24]. Однако 3D-реконструкция по реальным данным съемки и измерений сталкивается с большим числом неточностей, неизвестных данных и неопределенностей, которые не позволяют применить существующие методы как есть. Некоторые работы пытаются решать эту проблему «додумыванием» неизвестной информации на основе имеющейся обучающей выборки [Szymanowicz24; Zou24].

Один из наиболее подробных обзоров систем реконструкции можно найти в [Zhou 24]. Мы не будем повторять его, однако отметим наиболее близкие и важные к нашей работе системы и комплексы. Многие существующие решения, например Fraunhofer Cultlab3D [cultlab3d] и Witikon [witikon], используют лазерное сканирование и фотограмметрию. Они решают задачи восстановления геометрии и текстур достаточно качественно для определённых типов объектов, например музейных экспонатов, но достигается это созданием больших дорогостоящих комплексов с применением промышленных активных 3D-сканеров. Есть и более доступные решения [Petrelli19; Menna, et al., 2017]. Однако на

практике вместо это часто используется ручное моделирование и обратное проектирование, когда 3D-художники фактически вручную воссоздают 3D-объект со всеми его материалами, что является примером наиболее трудоемкого, но максимально гибкого способа. Кроме того, задача реконструкции всегда сталкивается с рядом проблем, порожденных неточностью измерительной и съемочной аппаратуры, что в большинстве научных работ попросту игнорируют; однако это легко приводит к потере точности и неприемлемым артефактам. Например, в работе [Apollonio F. I., et al., 2021] авторы предлагают метод комбинирования фотометрии и стерео различия для достижения точной и полной 3D-реконструкции для поверхностей с блестящими текстурными поверхностями, но обходят стороной проблемы с одноцветными поверхностями.

Из работ, ориентированных на создание датасета, можно выделить Stanford-ORB [Kuang23], Objects With Lighting [Ummenhofer24] и DTC [Dong25]. Stanford-ORB позиционирует себя как одну из первых работ, предлагающих набор данных для соревнования систем обратного рендеринга. Геометрия реконструировалась при помощи оптического сканера с высокой точностью. Затем объекты помещались в центр комнаты, для которой с использованием известного метода [Yu99] реконструировалась HDR-панорама. Эта панорама использовалась в качестве окружения, чтобы различные исследователи могли тестировать свои решения задачи восстановления материалов при известной геометрии либо решать совместную задачу восстановления материалов и текстур. Слабым местом этого набора данных является отсутствие нормальных эталонов текстур. В действительности они были получены очень приближенным образом, используя достаточно простой дифференцируемый рендер (nvdiffrast [Laine20]) без поддержки глобального освещения, и Stanford-ORB не предоставляет какой-либо оценки точности этих эталонов.

Набор данных DTC [Dong25] был создан с той же целью, что и Stanford-ORB. Однако он значительно более обширен (содержит 2000 объектов), а также имеет отдельно реконструированные SVBRDF для материалов в виде набора PBR текстур: альbedo (диффузная текстура), металличность, шероховатость, карта нормалей. Для реконструкции геометрии использовался коммерческий оптический сканер, для реконструкции материалов – «проприетарный» [Dong25] конвейер реконструкции на основе обратного рендеринга. Окружение, как и в Stanford-ORB, представлено панорамной текстурой. Валидация производилась путем сопоставления рендера реконструированной модели в синтетической сцене и фотографии. В целом данная работа выглядит как одна из наиболее завершенных и результативных. Однако одна из сложностей, с которой столкнулись авторы [Dong25], заключалась в том, что оптимизация материалов объектов с зеркальными свойствами работала с низкой точностью. В связи с этим авторы [Dong25] прибегли к услугам профессиональных 3D-художников для тщательной доработки материалов, используемых в создании глянцевых и отражающих объектов. Несмотря на обширный каталог объектов, отметим, что в этой работе отсутствуют объекты, содержащие полупрозрачные элементы и объекты с мелкими деталями, также отсутствуют высоко-зеркальные объекты и объекты, сделанные из мягких материалов. Это обусловлено ограничением технологической базы.

В качестве базовых методов в нашей работе были использованы библиотеки с открытым исходным кодом COLMAP [COLMAP] и MVS. COLMAP представляет собой комплексную систему, которая обеспечивает полный цикл обработки данных Structure-from-Motion (SfM) и Multi-View Stereo (MVS) [MVS; MSV2]. Эта библиотека позволяет автоматически создавать как разреженные, так и плотные облака точек на основе множества изображений. OpenMVG, в свою очередь, ориентирован на этап SfM. Он выполняет извлечение характерных точек, оценку параметров камер и формирование начального разреженного облака точек. OpenMVS использует результаты, полученные OpenMVG или частью COLMAP, Mitsuba3 [Mitsuba3] представляет собой физически обоснованный рендерер (PBR), который предоставляет возможности для оптимизации/подгонки материалов и освещения при помощи дифференцируемого рендеринга.

Просуммируем ключевые проблемы, возникающих при создании цифровых копий посредством 3D-сканирования. Во-первых, ограничения, связанные с характеристиками поверхностей: лазерные и оптические сканеры испытывают трудности с высокоотражающими материалами из-за бликов и отражений, снижая точность измерений до неприемлемых уровней, а полупрозрачные материалы, такие

как стекло и пластик, вызывают ошибки в определении глубины; фотограмметрия плохо работает с однородными поверхностями. Во-вторых, ограничена область сканирования: например, системы с поворотными столами не способны захватить нижнюю часть объекта, что усложняет процесс, а другие методы требуют переворачивания и объединения полученных с разных позиций объекта данных. Третий момент – большинство технологий фиксируют лишь базовую геометрию и диффузную альбедо-текстуру, не предоставляя параметров BRDF/PBR (шероховатость, металличность) для физически корректного рендеринга. Эти факторы неизбежно ведут к значительной ручной постобработке для устранения артефактов, ремоделинга геометрии и ретекстурирования, что существенно увеличивает временные и ресурсные затраты.

### **Предложенный программно-аппаратный комплекс**

Целью предлагаемого подхода является разработка системы, обеспечивающей создание цифровых копий физических объектов с требуемой точностью сохранения геометрии, текстуры и оптических свойств материалов со всех сторон объекта без перемещения, за один сеанс сканирования и без необходимости ручного ремоделирования и редактирования полученной модели. Важно отметить, что данная система должна быть способна обрабатывать объекты различных габаритов и с разнообразными характеристиками поверхностей, включая полупрозрачные, однотонные, отражающие и тонкие структуры.

Для достижения данной цели, предлагается модульная система, состоящая из трёх базовых компонент:

1. Студия на базе вращающейся платформы с метками со стеклом повышенной прозрачности, системой нанесения текстурированного покрытия, однородным фоном с меткой, управляемой системой освещения и фотографирования. Такая конструкция позволяет получать полный набор изображений объекта, включая нижнюю часть, без необходимости его перемещения или переворота. Использование текстурных покрытий должно обеспечить стабильность процесса фотограмметрии даже для объектов с однотонными или отражающими поверхностями. Фон и платформа с метками позволит определять взаимное расположения платформы, фона и освещения.

2. Двухэтапная методика съемки и обработки:

- 2.1. Фотограмметрическая реконструкция геометрии. На первом этапе выполняется фотограмметрическая реконструкция объекта на основе серии фотографий, снятых с применением маркеров и временных текстурирующих покрытий. Такой подход позволяет существенно повысить точность и детализацию геометрии, обеспечивая корректное выравнивание и сопоставление изображений.

- 2.2. На втором этапе осуществляются восстановление цветовой информации и оптимизация параметров материалов (SVBRDF), используя методы обратного рендеринга. Для этого используются фотографии объекта без текстурирующих покрытий, что позволяет получить чистую цветовую компоненту. В качестве ПО для реконструкции геометрии и базового текстурирования предлагается использовать классические алгоритмы 3D-реконструкции, например, COLMAP[COLMAP] и OpenMVS[MVS]. Важно, чтобы в них можно было сделать модификацию, в которой калибровка и определение положения камер было общим модулем, а реконструкция геометрии и реконструкция текстур использовала разные исходные фотографии. В качестве ПО для реконструкции материалов, предлагается использовать пайплайн обратного рендеринга Mitsuba3 [Mitsuba3].

3. Альтернативные варианты сканирования и обработки данных. В зависимости от свойств объекта и требований к цифровой копии системе могут быть подключены модули активного 3D-сканера и модуль реконструкции в альтернативное полигональному представлению, например, в представления 3D Гауссиановским сплаттингом 3DGS (3D Gaussian Splatting) [3DGS] или посредством неявного представления сцены объемным полем излучательности NeRF (Neural Radiance Fields) [NeRF]. Эти технологии повышают стоимость системы и ограничивают универсальность, поэтому их следует использовать только там, где недостаточно точности базовых методов.

Структура работы данных модулей показана на рисунке 1 в зависимости от классов обработки. На выходе работы системы на рисунке представлены результаты работы рендера соответствующим полученным классом обработки.



Рисунок 1. Блок-схема работы классов конвейеров обработки реконструкции объекта с предложенными модулями с отметкой «Т3Д».

На рисунке 2 представлено сравнение различных методов обработки данных съемки для идентичных исходных данных. Анализ показал, что реконструкция с использованием 3DGS (класс 1) демонстрирует высокую эффективность при работе с полупрозрачными моделями, однако отмечается значительное снижение четкости текстур и геометрии по сравнению с методами, использующими полигональные модели (другие классы).



Рисунок 2. Реконструкция объекта, содержащего полупрозрачные и тонкие поверхностные:  
1 – конвейером 1-го класса; 2 – конвейером 2-го класса; 0 – фотография объекта из исходных данных для реконструкции

На рис. 3. метод фотограмметрии с применением текстурного покрытия (класс 4) демонстрирует существенное улучшение результатов на объектах с однородным фоном по сравнению с классической фотограмметрией (класс 2). Применение оптического сканера для реконструкции геометрии значительно повышает ее точность в классе 3, однако при переосвещении и наложении текстур данное улучшение становится менее заметным по сравнению с предложенным методом обработки по классу 4.



Рисунок 3. Реконструкция объекта, содержащего чёткую геометрию (чёткие углы) и чёткую текстуру (надписи, мелкие детали): 1 – конвейером 1-го класса; 2 – конвейером 2-го класса; 3 – конвейером 3-го класса; 3\* – конвейером 3-го класса, с визуализацией модели, полученной непосредственно после оптического сканера без модулей T3D-текстурирования и восстановления материалов модели

На рисунке 4 показано, что различные методы классической фотограмметрии демонстрируют неприемлемое качество реконструкции объектов с одноцветными гладкими поверхностями. Предложенный конвейер по классу 4 справляется с задачей обработки объектов с однородными поверхностями.



Рисунке 4. Реконструкция объекта, содержащего одноцветные гладкие поверхности конвейерами обработки: 2 – конвейером 2-го класса; 4 – конвейером 4-го класса; 2\* – конвейера 2-го класса, где вместо базовой фотограмметрии на основе COLMAP и OpenMVS, использовалась платная версия модуля фотограмметрии Polycam.Inc. [Polycam]

В настоящей работе был разработан модуль реконструкции параметров материалов реальных объектов, представленных с помощью пространственно варьирующей функции распределения отражения (рис. 5). Основой для реализации послужил физически обоснованный рендер-движок Mitsuba3 и его Python-API, что позволило организовать дифференцируемый рендеринг с высоким



контролем над процессом оптимизации. В качестве исходных данных используется геометрическая сетка и диффузная текстура с запечённым освещением, полученные на этапе реконструкции поверхности объекта. Освещение моделировалось с помощью HDR-панорам реальной комнаты. Для оптимизации использовались эталонные фотографии объекта на однородном фоне.

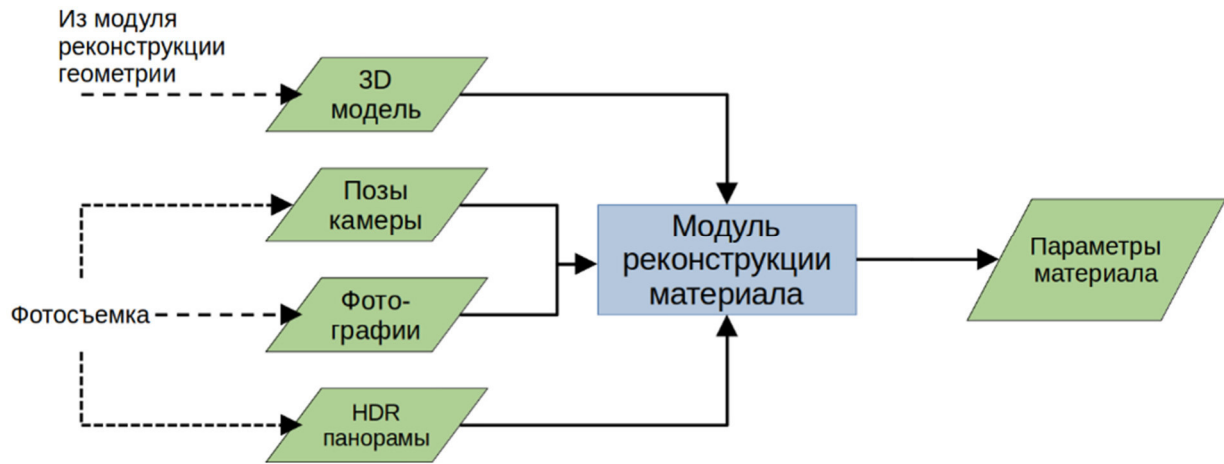


Рисунок 5. Входные и выходные данные для модуля реконструкции материалов

В качестве модели материала была выбрана составная BSDF-модель «Principled», способная описывать широкий спектр поверхностей – от чистых металлов до шероховатых диэлектриков. Для обеспечения совместимости результатов с форматом glTF 2.0 и снижения вычислительной нагрузки оптимизация велась лишь по трём ключевым текстурам: базового (диффузного) цвета, шероховатости и металличности. Тестировалось также использование текстуры для дополнительного прозрачного покрытия (clearcoat) – присутствующего в специальном расширении формата glTF 2.0.

Процесс съёмки HDR-панорамы включал многокадровую экспозиционную серию 360° камерой и последующий блендинг в специальном ПО. Такой подход вызывает ряд трудностей: необходимо восстановить положение окружения относительно снимаемого объекта, а также исправить различия в яркости, цветовой температуре и других параметрах, возникающие из-за использования разных камер.

Для реконструкции положения меша относительно окружения на фон наносятся маркеры, затем с помощью ключевых точек на маркере определяется положение камеры. Поскольку положение камеры относительно меша известно с прошлого этапа, можно получить положение объекта относительно окружающего освещения. Дополнительно восстанавливались параметры вращающейся платформы – её ось и угол поворота при каждом кадре – чтобы обеспечить согласованность между всеми матрицами трансформации.

В процессе оптимизации материалы и параметры яркости и цветовой температуры освещения реконструировались одновременно. Используется взвешенная функция потерь, в которой ошибка внутри маски объекта и за её пределами учитываются с разными коэффициентами, поскольку реконструкция окружения по освещению реконструируемого объекта в ряде случаев может вносить сильную неоднозначность.

Процесс восстановления включал адаптивное разрешение текстур – низкое разрешение в начале оптимизации, с автоматическим удвоением разрешения каждые несколько итераций – и при необходимости сглаживающую регуляризацию, помогающую избавиться от шума в текстурах, а также сгладить границы фрагментов текстур.

Получаемая в результате работы модуля реконструкции геометрии диффузная текстура в ряде случаев подходит для начальной инициализации PBR текстуры диффузного альбеда, что позволяет получить высокоточные детали в начале оптимизации.

Диффузная текстура, полученная в результате работы модуля реконструкции геометрии подходит для начальной инициализации PBR-текстуры диффузного альбеда, что полезно для обеспечения высокой детализации на начальных этапах оптимизации. Процесс оптимизации материала и освещения на сцене из нашего набора данных показан на рисунке 6.



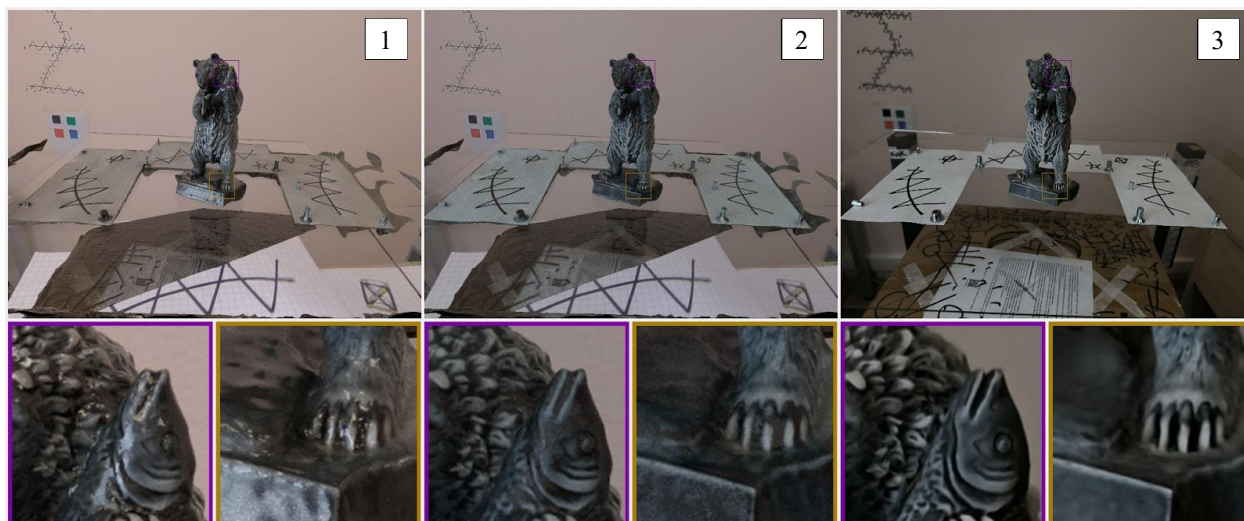


Рисунок 6. Оптимизация параметров материала и освещения на сцене bear: 1 – панорама и объект в начале оптимизации; 2 – через 15 шагов оптимизатора; 3 – референсное изображение

### Результаты экспериментов

Модуль реконструкции материалов был протестирован на собственном наборе из 8 объектов (matryoshka, hedgehog, router, bear, dragon, thermos, colored\_cup, white\_cup). Время реконструкции материалов составляло от 8 до 15 минут. Тестовая выборка составляет 20 % от обучающей. По итогу замеров модуль реконструкции на большинстве сцен набирает  $PSNR > 31$  и  $IoU > 0.97$ . Однако возникают трудности с отдельными сценами. Так, сцена thermos представляет собой объект с металлическим анизотропным материалом, что плохо описывается используемой моделью. Сцена router показала проблемы с согласованием положения камеры и панорамы, из-за чего для остальных замеров был использован обновленный маркер. В таблице 1 указаны усреднённые на тестовой выборке данные PSNR (с указанием стандартного отклонения) для синтезированных и эталонных изображений, а также значения метрики IoU для маски объекта (эталонные маски получены вручную).

Переосвещение полученных моделей и материалов в условиях искусственно смоделированного окружения (cornell box) демонстрирует их пригодность для PBR-рендеринга: блики и тени ведут себя физически корректно, а реконструированный объект реалистично взаимодействует со сценой, что крайне важно для задач дополненной и виртуальной реальности (AR и VR) (рис. 7).



Рисунок 7. Сцены hedgehog и bear из собственного набора, реконструированные и помещенные в cornell box

Таблица 1. Оценка качества реконструкции на собственном наборе данных

Параметры	Эталонное изображение	Реконструкция (оригинальное освещение)	Реконструкция (новое освещение)
Сцена: matryoshka			
PSNR (среднее): 31.317+/-0.866			
IoU (среднее): 0.9813			
Сцена: hedgehog			
PSNR (среднее): 34.572+/-1.359			
IoU (среднее): 0.9794			
Сцена: router			
PSNR (среднее): 27.646+/-1.286			
IoU (среднее): 0.9857			
Сцена: bear			
PSNR (среднее): 33.666+/-0.739			
IoU (среднее): 0.9785			
Сцена: dragon			
PSNR (среднее): 35.689+/-1.326			
IoU (среднее): 0.9736			



Окончание таблицы 1

Параметры	Эталонное изображение	Реконструкция (оригинальное освещение)	Реконструкция (новое освещение)
Сцена: thermos			
PSNR (среднее): 25.867+/-1.151			
IoU (среднее): 0.9836			
Сцена: colored_cup			
PSNR (среднее): 31.396+/-1.381			
IoU (среднее): 0.9775			
Сцена: white_cup			
PSNR (среднее): 33.104+/-0.651			
IoU (среднее): 0.9707			

### Обсуждение результатов

Эксперименты показали, что предложенная система с разным конвейером выполнения модулей позволяет получать результаты лучшие, чем один определённый вариант выполнения.

Стоит отметить сложность корректного сравнения с другими существующими программно-аппаратными системами ввиду отсутствия доступа как к самим подобным системам, так и снятым с помощью их объектам. К объективным преимуществам можно отнести лишь существенно более низкую стоимость комплекса по сравнению с аналогами, её масштабируемость и потенциальную воспроизводимость. Более полный сравнительный качественный анализ приведён в таблице 2.

Таблица 2. Сравнительный анализ методов получения цифровой копии

Метод получения цифровой копии	Время оператора	Стоимость оборудования, Р	Универсальность поверхностей	Геометрическая точность	Текстурная точность	Точность материалов	Полный охват объекта 360°x360°
Оптический ручной сканер со спреем для сложных материалов	От 10 минут до 10 часов (зависит от поверхностей)	~ 1 млн	+ -	+	- +	-	- +

Окончание таблицы 2

Метод получения цифровой копии	Время оператора	Стоимость оборудования, Р	Универсальность поверхностей	Геометрическая точность	Текстурная точность	Точность материалов	Полный охват объекта 360°x360°
Стационарный оптический сканер вместе с фотостудией [Dong25]	~20 минут скан + ручная коррекция сложных поверхностей	>10 млн	+ -	+	+	+	- +
AI генерация	~10 минут	~100 тыс.	+ -	-	-	-	+
3DGS [Kerbl23]	~10 минут	~200 тыс.	+	- +	- +	+ -	- +
Классическая фотограмметрия	~10 минут	~300 тыс.	- +	- +	+	- +	- +
T3D: фотограмметрия + текстурное окрашивание	~20 минут	~500 тыс.	+ -	+ -	+	+ -	+ -
T3D: фотограмметрия + текстурное окрашивание + ручной сканер	От 20 минут до 10 часов (зависит от поверхностей)	~1.5 млн	+ -	+	+	+ -	+ -

Среди выявленных недостатков разработанной системы следует отметить сложность процесса нанесения текстурирующего покрытия как с использованием меловых аэрозольных баллончиков, так и смываемых маркеров, требующую навыка для предотвращения смещения объекта и предотвращения окрашивания самой студии.

Текстурирующие покрытия поддаются удалению посредством применения обычной воды, но их использование может быть ограничено в отношении определенных типов объектов. В частности, для музейных экспонатов, требующих особой сохранности, рекомендуется применять бесконтактные методы 3D-сканирования. На впитывающие поверхности объектов в качестве текстурирующего покрытия можно наносить стикеры, но не маркеры и баллончики с краской.

Несмотря на заявленную масштабируемость системы в зависимости от размеров и веса объектов, в текущей реализации лучшая конфигурация используется для объектов менее 40 сантиметров и массой до 10 килограммов, для которых в качестве антибликовой прозрачной платформы используется специальное музейное стекло с высокой прозрачностью. Для более крупных или тяжелых объектов применяется стандартное осветленное стекло. Музейное стекло обладает светопропускающей способностью в диапазоне 97-99 % и коэффициентом светотражения менее 1 %. Осветленное стекло обладает существенно худшими характеристиками прозрачности: имеет светопропускание только около 90 % и коэффициент светотражения примерно 20 %.

Использование разработанного модуля реконструкции материалов на реальных данных показало целесообразность применения обратного рендеринга для данной задачи, но выявило ограничения, вызывающие артефакты и неточности. К ограничениям относится необходимость в тщательном подборе гиперпараметров, неоднозначность значений шероховатости и металличности для сложных материалов, а также необходимость в качественном моделировании источников освещения. Как показано в работе [Ling24], использование environment map может не подходить для реконструкции материалов с пространственно варьирующимися параметрами. Недостаточная информация об окружении может приводить к наблюдаемым на реконструированных моделях артефактам (запечённые блики, некорректные тени и др.). Перспективным направлением видится объединение дифференцируемого рендеринга с методом NeRF для моделирования окружения и подбор более эффективных регуляризаций для более надёжной оптимизации параметров.

## Выводы

Программно-аппаратный комплекс «Т3D» предлагает модульную систему, интегрирующую существующие и ряд инновационных решений для задачи реконструкции 3D-копий объектов, в особенности для объектов, традиционно сложных для систем реконструкции. Предложенный алгоритм вариантов использования модулей и соответствующих классификации объектов для их применения может помочь для создания цифровых копий с достаточным качеством при минимизации трудозатрат. Так, объект при одноэтапной реконструкции требует в среднем порядка 5 минут трудозатрат оператора, а при двухэтапной с нанесением покрытия – порядка 20 минут.

## Список литературы

- [Luan21] Unified shape and svbrdf recovery using differentiable monte carlo rendering / Luan F., et al. // Computer Graphics Forum. 2021. Vol. 40. No. 4. P. 101-113.
- [Huang24] 2d gaussian splatting for geometrically accurate radiance fields / Huang B., et al. // ACM SIGGRAPH 2024 conference papers. 2024. P. 1-11.
- [Gao24] Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing / Gao Jian, et al. // *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024.
- [Kuang23] Kuang Z., et al. Stanford-orb: a real-world 3d object inverse rendering benchmark // *Advances in Neural Information Processing Systems*. 2023. Vol. 36. P. 46938-46957.
- [Petrelli19] Petrelli, D. Making virtual reconstructions part of the visit: An exploratory study. *Digit. Appl. Archaeol. Cult. Herit*. 2019.
- [cultlab3d] Available online: <https://www.cultlab3d.de/> (accessed on 21 November 2020).
- [witikon] Available online: <http://witikon.eu/> (accessed on 21 November 2020).
- [sketchfab] Available online: <https://sketchfab.com/> (accessed on 13 June 2020).
- [Menna, et al. 2017] An open source low-cost automatic system for image-based 3D digitization / Menna F., et al. // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2017. Vol. 42. P. 155-162.
- [Apollonio F. I., et al. 2021] A photogrammetry-based workflow for the accurate 3D construction and visualization of museums assets / Apollonio F. I., et al. // *Remote Sensing*. 2021. Vol. 13. No. 3. P. 486.
- [Karami et al 22] Combining photogrammetry and photometric stereo to achieve precise and complete 3D reconstruction / Karami A., Menna F., Remondino F. // *Sensors*. 2022. Vol. 22. No. 21. P. 8172.
- [Zhou 24] A comprehensive review of vision-based 3d reconstruction methods / Zhou L., et al. // *Sensors*. 2024. Vol. 24. No. 7. P. 2314.
- [Dong25] Digital twin catalog: A large-scale photorealistic 3d object digital twin dataset / Dong Z., et al. // *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025. P. 753-763.
- [Ummerhofer24] Objects with lighting: A real-world dataset for evaluating reconstruction and rendering for object relighting / Ummerhofer B., et al. // *2024 International Conference on 3D Vision (3DV)*. IEEE, 2024. P. 137-147.
- [Wang24] A simple approach to differentiable rendering of sdfs / Wang Z., et al. // *SIGGRAPH Asia 2024 Conference Papers*. 2024. P. 1-11.
- [Nicolet21] Large steps in inverse rendering of geometry / Nicolet B., Jacobson A., Jakob W. // *ACM Transactions on Graphics (TOG)*. 2021. Vol. 40. No. 6. P. 1-13.
- [Mildenhall21] Nerf: Representing scenes as neural radiance fields for view synthesis / Mildenhall B., et al. // *Communications of the ACM*. 2021. Vol. 65. No. 1. P. 99-106.
- [Kerbl23] 3D Gaussian splatting for real-time radiance field rendering / Kerbl B., et al. // *ACM Trans. Graph.* 2023. Vol. 42. No. 4. P. 139:1-139:14.
- [Laine20] Modular primitives for high-performance differentiable rendering / Laine S., et al. // *ACM Transactions on Graphics (ToG)*. 2020. Vol. 39. No. 6. P. 1-14.
- [Ling24] NeRF as a Non-Distant Environment Emitter in Physics-based Inverse Rendering / Ling J., et al. // *ACM SIGGRAPH 2024 Conference Papers*. 2024. P. 1-12.
- [Ryabogin24] Reconstruction of convex bodies of revolution from the areas of their shadows / Ryabogin D., Zvavitch A. // *Archiv der Mathematik*. 2004. Vol. 83. No. 5. P. 450-460.
- [Garifullin24] Single-view 3d reconstruction via differentiable rendering and inverse procedural modeling / Garifullin A., et al. // *Symmetry*. 2024. Vol. 16. No. 2. P. 184.
- [Szymanowicz24] Splatler image: Ultra-fast single-view 3d reconstruction / Szymanowicz S., Rupprecht C., Vedaldi A. // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2024. P. 10208-10217.
- [Zou24] Triplane meets gaussian splatting: Fast and generalizable single-view 3d reconstruction with transformers / Zou Z. X., et al. // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2024. P. 10324-10335.

- 
25. [Yu99] Inverse global illumination: Recovering reflectance models of real scenes from photographs / Yu Y., et al. // Proceedings of the 26th annual conference on Computer graphics and interactive techniques. 1999. P. 215-224.
  26. [Mitsuba3] Mitsuba 2: A retargetable forward and inverse renderer / Nimier-David M., et al. // ACM Transactions on Graphics (ToG). 2019. Vol. 38. No. 6. P. 1-17.
  27. [COLMAP] COLMAP. General-purpose Structure-from-Motion (SfM) and Multi-View Stereo (MVS) pipeline. <https://colmap.github.io/>
  28. [MVS] OpenMVS: open Multi-View Stereo reconstruction library. <https://github.com/cdcseacave/openMVS>
  29. [MVS2] Structure-from-motion revisited / Schonberger J. L., Frahm J. M. // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 4104-4113.
  30. [Polycam] Polycam Inc. (2025). Polycam: 3D Capture & Photogrammetry Platform. Доступно по адресу: <https://poly.cam>



## Темпоральный и пространственный ресэмплинг для отрисовки с использованием физического переноса света

Я. Н. Дав

Университет ИТМО, Санкт-Петербург, Россия

**Аннотация.** С увеличением вычислительных мощностей современных графических процессоров отрисовка с использованием трассировки лучей становится все более важным направлением исследований в компьютерной графике, в том числе компьютерной графике реального времени. Трассировка лучей позволяет достоверно воспроизводить многие оптические эффекты, наблюдаемые в реальном мире, однако моделирование эффектов волновой оптики, таких как дифракция и интерференция представляет особые сложности, так как в волновой оптике отсутствует концепция луча света. В последнее время стали появляться подходы, позволяющие восстановить некоторое подобие концепции лучей для моделирования волновой оптики (например, в работах Штейнберга и других авторов). Однако в таких методах актуальной задачей остается повышение скорости расчета и улучшение сходимости при помощи использования более удачных стратегий выборки. В данной работе предложен подход к темпоральному и пространственному ресэмплингу для отрисовки с учетом волновой природы света, основанный на широко используемой стратегии ресэмплинга ReSTIR. Отрисовка с использованием ресэмплинга приближает результаты интерактивной отрисовки к результатам, полученным с большим объемом выборки. Несмотря на то, что алгоритм, предложенный в этой работе, не достигает производительности реального времени на современных графических процессорах, он является важным шагом к достижению этой цели.

**Ключевые слова:** компьютерная графика реального времени, трассировка лучей, физически достоверная отрисовка, волновая оптика, ресэмплинг.

## Spatiotemporal resampling for physical light transport rendering

Ya. N. Dav

ITMO University, Saint Petersburg, Russian Federation

**Abstract.** With increasing computational power of modern GPUs, ray tracing is becoming an increasingly important area of research in computer graphics, including real-time computer graphics. Ray tracing can accurately reproduce many optical effects observed in the real world, but wave optics effects such as diffraction and interference are particularly challenging because wave optics lacks the concept of a light ray. Recently, approaches have begun to appear that allow some semblance of the ray concept to be restored for wave optics effects simulation, for example in the works of Steinberg et al. However, in such methods increasing the computational efficiency and improving convergence with better sampling strategies remains a relevant problem. In this paper, I propose an approach to temporal and spatial resampling for physical light transport rendering based on the widely used ReSTIR strategy. Resampling brings the results of interactive rendering closer to the reference obtained with a large sample size. Although the algorithm proposed in this work does not achieve real-time performance on modern GPUs, it is an important step towards achieving this goal.

**Keywords:** real-time computer graphics, ray tracing, physically-based rendering, wave optics, resampling.

### Введение

Исторически большая часть алгоритмов отрисовки в компьютерной графике была построена на допущениях геометрической оптики, так как геометрическая оптика позволяет пользоваться моделью луча света. Эта модель чрезвычайно удобна для расчета переноса света, поскольку обладает важными свойствами локальности (т.е. существование в бесконечно малой точке пространства и распространение в конкретном направлении в пространстве) и линейности (т.е. лучи не влияют друг на друга и их совместная интенсивность равна сумме их интенсивностей). Предположения геометрической оптики достаточно хорошо выполняются во многих ситуациях, однако при использовании когерентных или частично когерентных источников света или при взаимодействии с некоторыми типами материалов проявляются эффекты волновой оптики – дифракция и интерференция, которые невозможно симулировать, используя эти предположения.

За годы развития области компьютерной графики появлялось достаточно большое количество работ, посвященных различным подходам к моделированию эффектов волновой оптики. Отметить можно, например, работы, использующие функцию распределения Вигнера для описания излучения

и материалов, с которыми излучение взаимодействует. Такой подход позволяет расширить концепцию световых полей на случай отрицательной интенсивности света и помогает достаточно эффективно рассчитывать сложные волновые эффекты. Пример такого подхода можно найти в работе [1]. К сожалению, несмотря на то, что итоговая интенсивность в каждом пикселе при бесконечно большом объеме выборки гарантированно неотрицательная, в реальности при ограниченном объеме выборки отрицательная суммарная интенсивность может быть результатом и ведет к артефактам при отрисовке. Кроме того, наличие у лучей отрицательной интенсивности несовместимо со многими популярными физически достоверными техниками отрисовки. Среди прочих методов моделирования волновой оптики можно отметить, например, подходы на основе матриц переноса [2]. Но эти методы сопряжены с большими вычислительными сложностями, которые делают их применение в компьютерной графике неоправданным.

Перспективный метод отрисовки волновых эффектов, называемый физическим переносом света, предложили Штейнберг и другие в работе [3]. Их подход основан на замене метрики интенсивности света на функцию взаимной спектральной интенсивности, характеризующую помимо интенсивности корреляцию фаз волн в двух точках в некой окрестности. Это позволяет задать луч света как некий малый объем пространства, распространяющийся в малый телесный угол. Конкретные характеристики луча зависят от его когерентности и подобраны так, чтобы два таких луча не были когерентны друг другу, то есть обладали свойством линейности. Этот метод хорош тем, что предлагает глобальный фреймворк для симуляции волновой оптики, схожий с фреймворками переноса света в геометрической оптике, но обладает проблемой медленной сходимости и большой вычислительной сложности, что делает его применение в графике реального времени практически невозможным. Даже последняя итерация подхода авторов в работе [4], использующая обратную трассировку лучей, несколько стратегий выборки по значимости и аппаратное ускорение трассировки лучей, не соответствует ограничениям реального времени на современных графических ускорителях даже при выборке объема 1 элемент на пиксель. Поэтому адаптация такого фреймворка для графики реального времени потребует уменьшения вычислительной сложности и увеличение скорости сходимости. Стоит отметить, что использование подходов к расчету глобального освещения в геометрической оптике, таких как фотонные карты [5], для облегчения вычислений в волновой оптике затруднителен в силу необходимости сохранять волновые характеристики луча при его переносе. Это означает, что вместо сохранения интенсивности «фотонов», при прямой трассировке и заполнении фотонных карт нужно сохранять обобщенные параметры Стокса и флаги о состоянии лучей, что означает быстрое разрастание затрат памяти при увеличении количества лучей.

В данной работе будет предложен способ увеличения скорости сходимости алгоритма с помощью темпорального и пространственного ресэмплинга и проанализированы результаты его применения.

### Постановка задачи

Чтобы алгоритм физического переноса света (PLT – Physical Light Transport) [4] можно было использовать в компьютерной графике реального времени, необходимо улучшить сходимость интегралов при расчете освещенности пикселя детектора на малых объемах выборки интерактивной отрисовки (далее под интерактивной отрисовкой подразумевается выборка объема 1 элемент на пиксель). Для лучшей сходимости интеграла освещения на малой выборке в компьютерной графике часто применяется стратегия пространственного и темпорального ресэмплинга. При таком подходе эффективное количество элементов выборки увеличивается за счет использования элементов предыдущих кадров и соседних пикселей. Задача данного исследования заключается в разработке подхода к темпоральному и пространственному ресэмплингу для отрисовки методом PLT, создании его реализации в реальном фреймворке и анализе результатов.

### Теория

Для приближенного вычисления интеграла функции в компьютерной графике обычно используется оценка интеграла методом Монте-Карло, которая выглядит следующим образом для интеграла функции  $f(x)$ :

$$\langle I \rangle = \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{p(x_i)}, \quad (1)$$

где  $p(x)$  – функция плотности вероятности выбора элементов  $\{x_i\}$ ;  $N$  – количество элементов в выборке.

Оценка (1) сходится к реальному значению интеграла, если  $p(x) > 0$  там, где  $f(x) \neq 0$ , и сходится она тем быстрее, чем больше корреляция между функциями  $f(x)$  и  $p(x)$ . В случае, когда хорошо коррелирующая с  $f(x)$  функция  $\hat{p}(x)$  неизвестна в явном виде (но ее значение в конкретной точке можно получить), для более быстрой сходимости часто используется подход ресэмплинга. В этом подходе начальная выборка  $\{x_i\}$  получается с помощью неоптимальной, но простой для генерации элементов функции  $p(x)$  (представляющей, например, равномерное распределение). Далее из этой выборки производится повторная выборка одного или более элементов с дискретными вероятностями

$$p(z|x) = \frac{w(x_z)}{\sum_{i=1}^M w(x_i)}, w(x) = \frac{\hat{p}(x)}{p(x)},$$

где  $z$  – индекс конкретного элемента выборки. В случае одного выбранного элемента у оценка интеграла приобретает следующий вид:

$$\langle I \rangle = \frac{f(y)}{\hat{p}(y)} \left( \sum_{i=1}^N w(x_i) \right).$$

В работе [6] была впервые предложена модификация оценки интеграла с помощью ресэмплинга для расчета прямого освещения, названная авторами ReSTIR (Reservoir-based Spatio-Temporal Importance Resampling). Суть этой модификации заключается в использовании выборки с помощью резервуаров для уменьшения затрат памяти на хранение изначальной выборки, а также подхода множественной выборки по значимости для повторного использования элементов выборки соседних пикселей с разными областями интегрирования. Выборка с использованием резервуаров осуществляется следующим образом: в так называемом резервуаре хранится текущий выбранный элемент, его вес, равный дискретной вероятности его выбора, и сумма весов всех обработанных элементов. Далее каждый следующий элемент выборки  $x_m$  заменяет текущий с вероятностью

$$\frac{w(x_m)}{\sum_{j=1}^m w(x_j)}.$$

Такой подход позволяет легко добавлять новые элементы из разных источников, например, генерируя несколько исходных, а затем комбинируя резервуары текущего пикселя на текущем кадре с резервуаром этого пикселя на предыдущем кадре и резервуарами соседних пикселей. Поскольку соседние пиксели могут иметь разную область интегрирования, наивное комбинирование их резервуаров может приводить к артефактам, особенно вблизи геометрических разрывов. Поэтому для их комбинации используется подход множественной выборки, который позволяет правильно взвесить вклады от разных функций плотности вероятности  $p_i(x)$ , используя эвристику

$$w_s(x) = \frac{N_s p_s(x)}{\sum_j N_j p_j(x)}.$$

Более приближенное к цели данной работы применение ReSTIR было предложено в работе [7] для ресэмплинга глобального освещения (так называемый ReSTIR GI). В работе авторы использовали генерацию случайного направления из видимой точки для получения ближайшей точки сцены в этом направлении. Для этой точки рассчитывалось исходящее освещение после трассировки дальнейшего пути. Это освещение использовалось в качестве меры значимости этой точки, а далее применялся ReSTIR для получения более значимой точки с исходящим освещением. Для ресэмплинга между соседними пикселями авторы получили формулу якобиана перехода из одной системы координат видимой точки в другую, поскольку стратегия выборки зависит от положения и нормали поверхности в видимой точке. Якобиан перехода имеет вид

$$|J_{q \rightarrow r}| = \frac{|\cos(\phi_2^r)|}{|\cos(\phi_2^q)|} \cdot \frac{\|x_1^q - x_2^q\|^2}{\|x_1^r - x_2^q\|^2}, \quad (2)$$

где  $x_1^q, x_2^q$  – первые две вершины пути с соседнего пикселя;  $x_1^r$  – видимая точка текущего пикселя;  $\phi_2^q, \phi_2^r$  – углы векторов  $x_1^q - x_2^q$  и  $x_1^r - x_2^q$  соответственно с нормалью поверхности в точке  $x_2^q$ .

В работе [8] была предложена дополнительная модификация ReSTIR GI, повышающая точность пространственного ресэмплинга с помощью использования хэш-таблицы для хранения исходящего освещения от путей со схожими координатами и нормальными начальными. При этом таблица построена авторами таким образом, чтобы ее ячейки увеличивались в размере (в мировом пространстве) по мере удаления от камеры. В таблице хранятся индексы из полноэкранного буфера, в котором хранятся начальные точки и исходящий свет от сохраненных путей. Пути учитываются в генерации и ресэмплинге, если они начинаются от точки на материале с шероховатостью больше порогового значения и заканчиваются на источнике света. Пространственный ресэмплинг, построенный на таких принципах, учитывает схожесть нормалей и позиций точек при повторном использовании путей, поэтому его результаты имеют меньшее количество шума, особенно вблизи геометрических разрывов.

Для алгоритма PLT из работы [4] можно применить схожую стратегию ресэмплинга с алгоритмами ReSTIR GI. Версия алгоритма из работы действует следующим образом: на первой стадии, называемой стадией выборки, параметры пикселя на детекторе распространяются обратной трассировкой, пока либо не встретят источник света, либо не достигнут максимального количества участков пути. Далее, на второй стадии, называемой стадией решения, свет от источника света, найденного на предыдущей стадии, а также от других источников, найденных с помощью предсказания следующего события [9], распространяется вдоль найденного пути до детектора и вклад его интенсивности добавляется к цвету пикселя.

Важные отличия данного алгоритма от обычных алгоритмов переноса света без учета волновых эффектов в том, что вклады от нескольких лучей нельзя просуммировать до взаимодействия с детектором, так как это приведет к потерям информации о поляризации и когерентности, а также то, что лучи несут информацию лишь о нескольких выбранных длинах волн, не характеризуя интенсивность на всем видимом спектре. Выбор алгоритма ресэмплинга слабо может повлиять на вторую особенность, и ниже будет предложен потенциальный путь минимизации возникающих отрицательных эффектов на этапе сглаживания шума.

Отсутствие возможности просуммировать вклады от нескольких лучей приводит к необходимости сохранять лучи на этапе генерации. Рационально сохранять их на этапе до переноса до последнего рассеяния, затем производить темпоральный и пространственный ресэмплинг, используя в качестве меры важности суммарную интенсивность всех входящих в него длин волн (что для используемого в PLT представления лучей означает сумму нулевых параметров Стокса для каждой длины волны). На этапе пространственного ресэмплинга используется тот же якобиан (2), что и в работе [7], так как выборка направлений из каждой точки пути света в PLT осуществляется тем же образом, что и в работе [7]. Затем, после получения финального набора лучей, они доводятся до детектора, и производится измерение финальной интенсивности. Схематично работа метода показана на рисунке 1. Улучшение из работы [8] не было использовано в данной работе, так как построение хэш-таблицы занимает дополнительное вычислительное время, однако в последующих работах, возможно, стоит рассмотреть его применение и оценить его эффект на качество отрисовки.

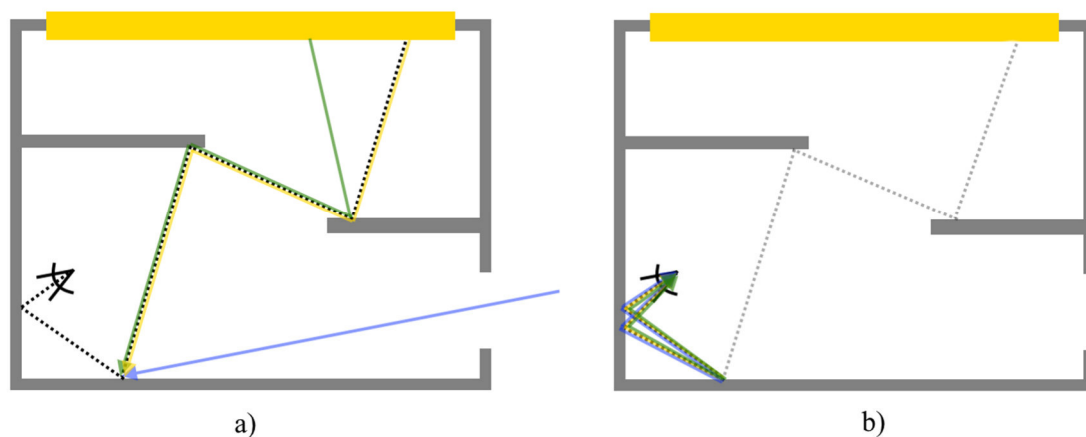


Рисунок 1. Пространственное повторное использование пути света: а) этап генерации лучей для одного пикселя; б) повторное использование этих же лучей для соседнего пикселя после ресэмплинга

## Результаты экспериментов

Алгоритм был реализован в качестве модификации к реализации алгоритма PLT из работы [4] в среде NVIDIA Falcor 6.0. Блок-схема проходов алгоритма показана на рисунке 2. Первые две ступени с обратной трассировкой обобщенных лучей параметров детектора (так называемая стадия выборки) и прямой трассировки лучей света (так называемая стадия решения) претерпели следующие изменения: в стадии выборки информация о геометрии видимых точек записывается в буфер для сравнения пикселей в стадиях ресэмплинга, а в стадии решения лучи переносятся на момент до последнего рассеяния о материал и сохраняются в буфер. Буфер содержит фиксированное количество сохраняемых лучей на пиксель (в дальнейших тестах это количество равно 4, это значение было выбрано исходя из тестов для конкретной сцены, в резервуарах практически каждый раз оказывалось от 1 до 4 лучей). Обе ступени работают по блокам пикселей (тайлам) фиксированного размера, как и в изначальной реализации.

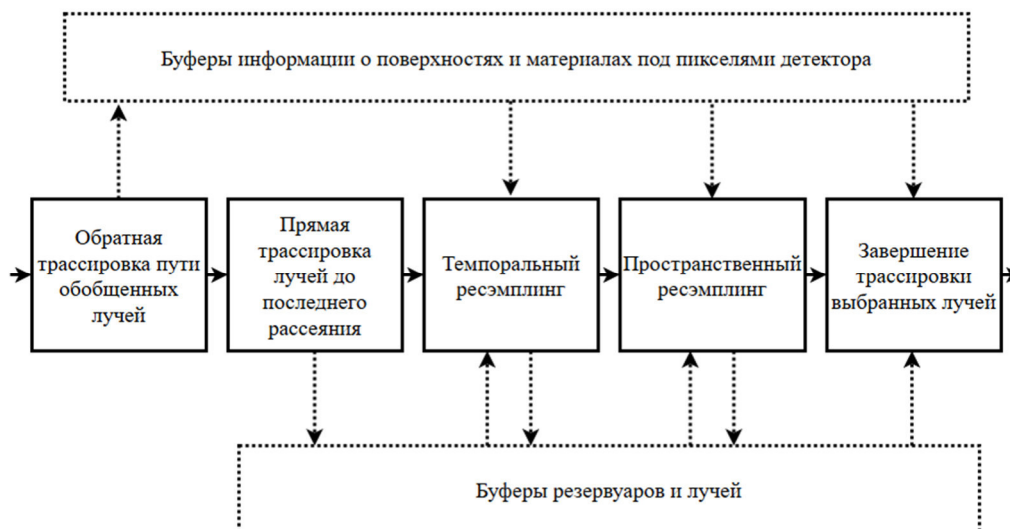


Рисунок 2. Блок-схема алгоритма пространственного и темпорального ресэмплинга (пунктирными стрелками показаны операции чтения и записи)

Для оценки улучшения сходимости при использовании полученного алгоритма были использованы метрики пикового отношения сигнала к шуму (PSNR), индекса структурного сходства (SSIM) и относительной средней ошибки (RMAE). Метрики определялись сравнением результатов отрисовки с образцом (результатом отрисовки изначальным алгоритмом с использованием большого объема выборки) при помощи библиотеки OpenCV. Для определения характера изменений этих метрик при увеличении сходимости они были получены для результатов отрисовки оригинальным алгоритмом из работы [4] с объемом выборки от 1 до 8 элементов. Результаты этих измерений представлены на рисунке 3. Из анализа этих результатов можно сделать вывод, что SSIM и RMAE достаточно быстро меняются на малых объемах выборки и при оценке можно полагаться на них, в то время как PSNR растет достаточно медленно (кроме того, как отмечено в работе [6], среднеквадратичное отклонение, на котором основана эта метрика, довольно сильно зависит от случайных выбросов и менее достоверно показывает реальную сходимость при отрисовке).

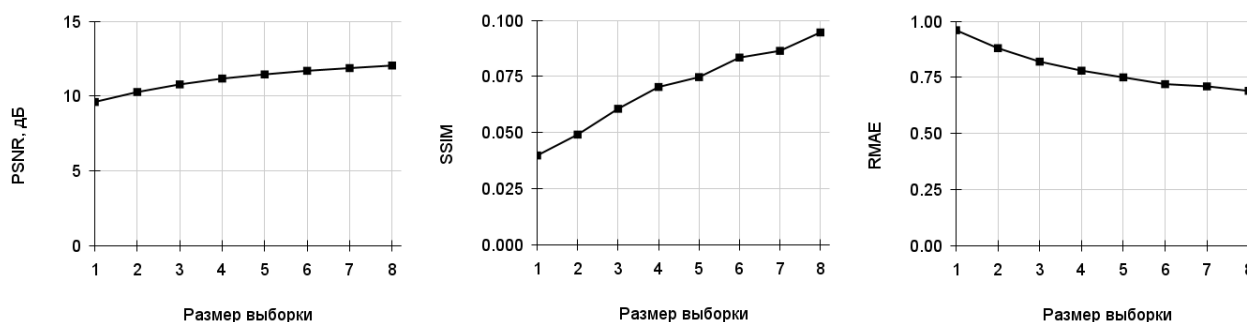


Рисунок 3. Изменение метрик PSNR, SSIM и RMAE при увеличении объема выборки



Результаты работы алгоритма получены на системе с NVIDIA RTX 5070 Ti, производительность дополнительно измерена на системе с NVIDIA RTX 3080. Результаты отрисовки изначальным и модифицированным алгоритмом представлены на рисунке 4. Очевидно, что темпоральный ресэмплинг увеличивает сходимость по всем метрикам при интерактивной отрисовке, по сравнению с изначальным алгоритмом. Добавление пространственного ресэмплинга увеличивает сходимость по SSIM, но уменьшает по другим метрикам, что связано с распространением выбросов большей интенсивности между соседними пикселями. Однако дополнительный пространственный ресэмплинг визуально помогает различать больше деталей на сцене, как видно, к примеру, на выделенном участке, где освещение и затенение более различимы и соответствуют образцу.

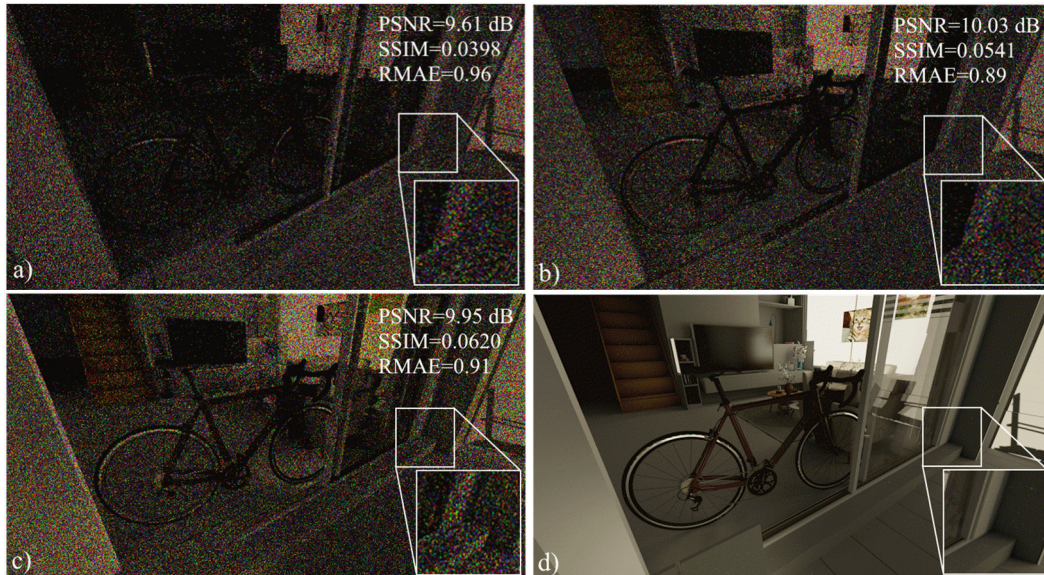


Рисунок 4. Результаты отрисовки: *a)* без использования ресэмплинга; *b)* с использованием темпорального ресэмплинга; *c)* с использованием темпорального и пространственного ресэмплинга; *d)* образец, отрисованный с использованием выборки объема 16 000

Производительность и качество изображения при использовании ресэмплинга также дополнительно были оценены по сравнению с производительностью оригинального алгоритма при объеме выборки в 2 элемента на пиксель. Целью этого было определение целесообразности применения ресэмплинга по сравнению с простым увеличением объема выборки. На рисунке 5 показаны результаты этого сравнения.

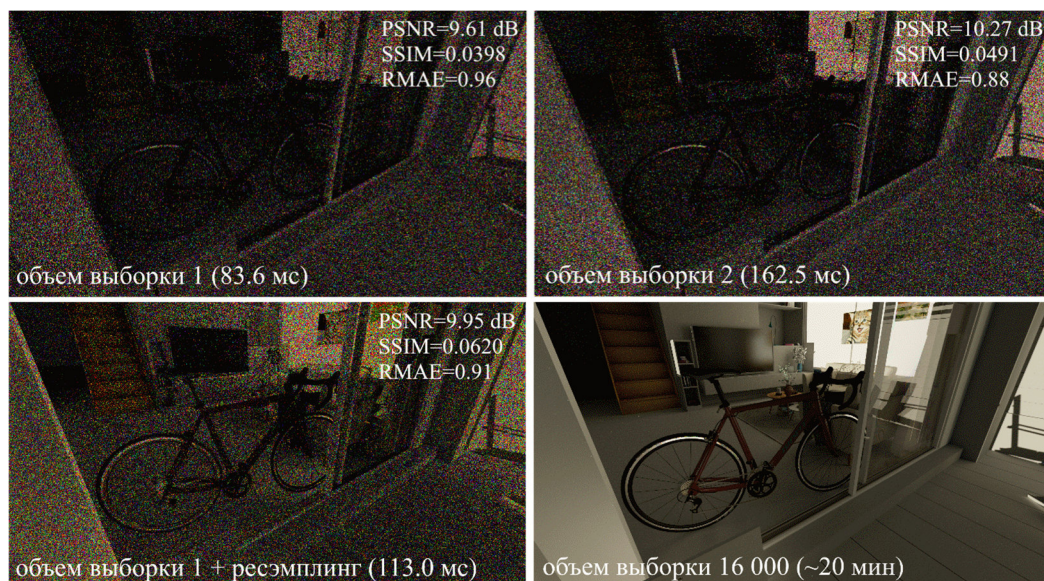


Рисунок 5. Сравнение производительности алгоритмов: оригинального алгоритма с выборками объема 1 и 2 элемента на пиксель, модификацией алгоритма с использованием пространственного и темпорального ресэмплинга и алгоритмом PLT с большим объемом выборки



Сравнение по метрикам SSIM и RMAE позволяет сделать вывод, что использование ресэмплинга дает возможность получить изображение схожего или чуть лучшего качества при меньшем времени на отрисовку, чем использование выборки большего объема. Производительность каждой из стадий оригинального и модифицированного алгоритмов на двух разных GPU представлена в таблицах 1 и 2.

Таблица 1. Производительность и влияние на память алгоритмов на RTX 5070 TI

Стадия	Время, мс	
	Оригинальный алгоритм	Модифицированный алгоритм (+ 750 МБ)
Выборка-решение	83.6	107.3
Темпоральный ресэмплинг	-	1.6
Пространственный ресэмплинг	-	3.2
Финальная стадия	-	0.9

Таблица 2. Производительность и влияние на память алгоритмов на RTX 3080

Стадия	Время, мс	
	Оригинальный алгоритм	Модифицированный алгоритм (+ 750 МБ)
Выборка-решение	88.8	123.91
Темпоральный ресэмплинг	-	2.41
Пространственный ресэмплинг	-	5.9
Финальная стадия	-	1.2

### Обсуждение результатов

С помощью использования темпорального и пространственного ресэмплинга удалось добиться более стабильного во времени результата с большим количеством различных деталей. Производительность по сравнению с алгоритмом из работы [4] ухудшилась, однако при объеме выборки 1 элемент время на отрисовку модифицированным алгоритмом значительно меньше, чем при отрисовке оригинальным алгоритмом с объемом выборки 2 элемента, при значительно большем количестве различных деталей (рис. 4). Время выполнения этапов выборки-решения значительно возросло по сравнению с оригинальным алгоритмом в основном в связи с большим количеством операций записи данных о лучах в память. Потребление памяти при использовании ресэмплинга также значительно возрастает вследствие необходимости сохранения большого количества информации о лучах между стадиями.

Для разрешения отрисовки 1920 на 1080 пикселей и 4 сохраняемых лучах в каждом пикселе буферы под резервуары для ресэмплинга добавляют примерно 750 МБ к используемой видеопамяти (вследствие тайлового подхода оригинальный алгоритм использует 1 буфер размером около 330 МБ для хранения информации о пути света между стадиями алгоритма). Возможно, в дальнейших работах имеет смысл сократить количество сохраняемой информации с помощью алгоритмов сжатия или понижения точности чисел с плавающей точкой. Предложенный алгоритм так же, как и оригинальный, не удовлетворяет требованиям отрисовки в реальном времени, так как время отрисовки даже на достаточно производительном современном графическом ускорителе выходит за рамки бюджета в 33.3 мс на один кадр. Тем не менее, данный результат полезен для достижения цели использования PLT в графике реального времени, так как использование ресэмплинга улучшает сходимость алгоритма на малых выборках.

Для дальнейшего уменьшения шума в результате отрисовки была предпринята попытка использовать денойзер NVIDIA OptiX, результаты приведены на рисунке 6.

Как можно увидеть, обыкновенные денойзеры достаточно плохо подходят для сглаживания при отрисовке физическим переносом света и приводят к возникновению нестабильных во времени цветных пятен и других артефактов. Это связано с уже упомянутой особенностью данного вида отрисовки – лучи в нем не несут информацию о всем спектре, а лишь о нескольких случайно выбранных длинах волн. Возможный путь дальнейшего улучшения данного вида отрисовки для интерактивных приложений

в создании специально адаптированного денойзера, который будет учитывать эту особенность и создавать меньшее количество артефактов.

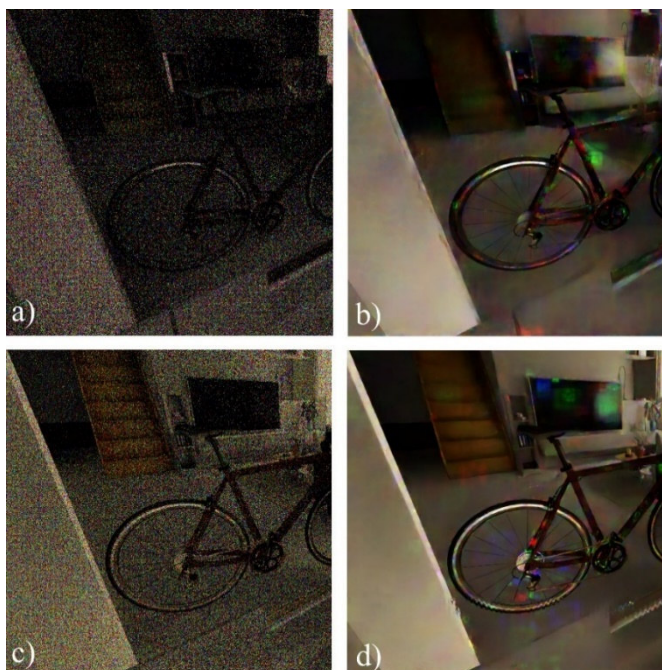


Рисунок 6. Результаты интерактивной отрисовки: a) без ресэмплинга; b) без ресэмплинга с денойзером; c) с ресэмплингом; d) с ресэмплингом и денойзером

### Выводы

В работе был представлен алгоритм отрисовки физическим переносом света с помощью пространственного и темпорального ресэмплинга. Реализация алгоритма протестирована на качество и производительность интерактивной отрисовки на современном графическом процессоре. Была подтверждена значительно улучшенная сходимость результата отрисовки по сравнению с оригинальным алгоритмом из работы [4]. Производительность интерактивной отрисовки ухудшилась примерно на 35% по сравнению с оригинальным алгоритмом.

В дальнейшем планируется уменьшить вычислительную сложность алгоритма и оптимизировать предложенные в данной работе модификации для достижения цели – использования физического переноса света для отрисовки в реальном времени.

### Благодарность

Автор благодарит научного руководителя Андрея Сергеевича Карсакова (НИУ ИТМО).

### Список литературы

1. Reflectance model for diffraction / Cuypers T., et al. // ACM Trans Graph. 2012. Vol. 31, no. 5.
2. Saleh B.E.A., Teich M.C. Fundamentals of photonics. John Wiley & Sons Inc., 1991.
3. Steinberg S., Yan L.-Q. A generic framework for physical light transport // ACM Trans. Graph. New York, NY, USA: Association for Computing Machinery, 2021. Vol. 40, no. 4.
4. A Generalized Ray Formulation For Wave-Optical Light Transport / Steinberg S., et al. // ACM Trans. Graph. New York, NY, USA: Association for Computing Machinery, 2024. Vol. 43, no. 6.
5. Jensen H.W. Global Illumination using Photon Maps // Rendering Techniques '96 (Proceedings of the Seventh Eurographics Workshop on Rendering). 1996.
6. Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting / Bitterli B., et al. // ACM Trans. Graph. New York, NY, USA: Association for Computing Machinery, 2020. Vol. 39, no. 4.
7. ReSTIR GI: Path Resampling for Real-Time Path Tracing / Ouyang Y., et al. // Computer Graphics Forum. 2021. Vol. 40, no. 8. P. 17–29.
8. Chaine R., Deng Z., Kim M.H. World-Space Spatiotemporal Path Resampling for Path Tracing // Computer Graphics Forum (Proceedings of PG 2023). 2023. Vol. 42, no. 7.
9. Pantelev A. Light at the End of the Ray // SIGGRAPH 2019. 2019.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-154-163

## Гибридный метод отображения теней в реальном времени для систем с несколькими видеокартами

Р. Р. Султанов, Г. С. Лях, М. К. Богданов, А. М. Суворов  
Университет ИТМО, г. Санкт-Петербург, Россия

**Аннотация.** Карты теней – наиболее распространенный способ расчета теней в приложениях реального времени, таких как компьютерные игры. В данной работе рассматривается применимость использования алгоритмов разделения задач между графическими процессорами на гибридных системах с несколькими видеокартами для увеличения производительности системы в целом. В исследовании предложен и протестирован алгоритм, разделяющий графический конвейер между двумя графическими процессорами с вынесением расчета теней на дополнительную видеокарту, в то время как на основной выполняются другие расчеты с последующим копированием данных перед отрисовкой освещения сцены. Для реализации использовался API DirectX 12, позволяющий реализовывать кросс-адаптерное взаимодействие. В результате разработанный алгоритм позволяет эффективно задействовать все доступные аппаратные ресурсы системы, обеспечивая при этом высокое качество и точность получаемых теней. Предложенное решение легко встраивается в существующие графические конвейеры и способствует увеличению производительности гибридных графических систем без значительных дополнительных затрат.

**Ключевые слова:** карты теней, каскадные карты теней, DirectX 12, Multi-GPU, компьютерная графика

## Hybrid method for displaying shadows in real-time in systems with multiple GPUs

R. R. Sultanov, G. S. Lyakh, M. K. Bogdanov, A. M. Suvorov  
ITMO University, Saint Petersburg, Russia

**Abstract.** Shadow maps are the most common method for calculating shadows in real-time applications such as video games. This work explores the applicability of task-splitting algorithms across multiple graphics processors in hybrid systems with several GPUs to improve overall system performance. The study proposes and evaluates an algorithm that distributes the graphics pipeline between two GPUs, offloading shadow computations to the secondary graphics card while the primary one handles other rendering tasks, followed by data copying before the final lighting pass. The implementation was based on the DirectX 12 API, which enables cross-adapter communication. As a result, the developed algorithm efficiently utilizes all available hardware resources while maintaining high quality and accuracy of the computed shadows. The proposed solution integrates seamlessly into existing rendering pipelines and contributes to the performance increase of hybrid graphics systems without significant additional costs.

**Keywords:** Shadow Maps, Cascaded Shadow Maps, DirectX 12, Multi-GPU, computer graphics

### Введение

Современные компьютерные игры и графические приложения предъявляют всё более высокие требования к производительности графических систем. Реалистичная визуализация сцены включает множество параллельных алгоритмов: глобальное освещение, эффекты постобработки, моделирование сложных материалов. Особое место среди этих задач занимают тени в реальном времени, поскольку они одновременно повышают визуальную достоверность сцены и являются одним из наиболее ресурсоёмких этапов рендеринга.

Традиционные методы мульти-GPU рендеринга, такие как Alternate Frame Rendering (AFR) и Split Frame Rendering (SFR), хорошо изучены и применялись в графических системах прошлого поколения. Однако их использование в потребительских условиях ограничено: AFR подвержен проблеме межкадровых зависимостей и высокой задержке ввода, а SFR требует значительных затрат на синхронизацию и передачу данных. Поэтому возникает необходимость в исследовании более гибких подходов, в частности – гибридного рендеринга, при котором различные этапы графического конвейера распределяются между несколькими GPU.

Цель данной работы – исследовать возможность вынесения построения карт теней на дополнительный графический процессор (в том числе интегрированный) и определить эффективность подобного распределения в условиях домашних систем с ограниченной пропускной способностью межсоединения (PCI Express).

В этой работе:

- 1) предложена и реализована схема гибридного рендеринга, при которой вычисление карт теней выполняется на вспомогательном GPU, а основной адаптер обрабатывает остальные стадии конвейера;
- 2) показана применимость данного подхода для конфигураций с интегрированным и дискретным GPU без использования специализированных высокоскоростных интерфейсов (NVLink, XGMI);
- 3) проведён анализ производительности и выявлены условия, при которых межадapterный обмен не нивелирует выигрыш от вынесения расчёта теней.

## Обзор

Для систем с несколькими GPU существуют три базовых класса распределения работы: межкадровое (AFR), экранно-пространственное (SFR) и функционально-задачное (гибридное, по стадиям конвейера). Такая классификация восходит к классическим работам по параллельному рендерингу [1,2] и сегодня реализуема средствами современных графических API. В ряде обзоров подчеркивается, что выбор схемы распределения вычислений определяется архитектурой соединений и характером решаемых задач [3,4].

1. Alternate Frame Rendering (AFR). Каждый адаптер рендерит свой кадр (например, чётные/нечётные). Преимущества: простая декомпозиция и низкая связность между исполнителями [5,6]. Недостатки: рост полной задержки ввода-вывода на один кадр, «микростаттер» [5] при неоднородности производительности адаптеров, необходимость полной репликации ресурсов (сцены и состояний) на каждом GPU [7], а также ухудшение эффективности при наличии межкадровых зависимостей (TAA, TSR, SSR, временные тени и накопительные фильтры), характерных для современного рендеринга. На практике теоретическая близкая к линейной масштабируемость достигается редко.

Режим AFR исторически поддерживался драйверными технологиями SLI/CrossFire и специализированными режимами вроде VR SLI [8–10], однако в современных драйверах поддержка постепенно сокращается [8,9]. Современные работы также отмечают низкую эффективность AFR в условиях широкого использования темпоральных алгоритмов [6,11].

2. Split Frame Rendering (SFR). Кадр делится на регионы экрана (или изображение собирается композицией частичных результатов), каждый регион/подкадр отрисовывается на своём GPU [12,13]. Данный подход активно исследовался начиная с 1990-х годов [2,13]. Преимущества: потенциально лучшая балансировка нагрузки по сценно-зависимой сложности регионов, пониженная задержка по сравнению с AFR. Недостатки: необходимость дублирования/распределения геометрии, сложная синхронизация при прозрачностях и пересечениях регионов, высокие требования к пропускной способности межсоединения при композиции итогового изображения.

Более поздние работы предложили оптимизации: сокращение числа передаваемых примитивов [14], ассоциативная композиция в CHOPIN [15], использование иерархических структур для композиции [16]. Однако все эти решения предъявляют высокие требования к пропускной способности межсоединения и памяти [17].

Современные системы масштабируют SFR до десятков плат за счёт параллельной композиции: CHOPIN [15], MCM-GPU [18], работы по распределённым фреймбуферам [12,19]. Они демонстрируют хорошую масштабируемость, но требуют либо специализированных межсоединений (NVLink, XGMI) [20,21], либо высокопроизводительных сетевых технологий (Infiniband, RDMA) [13].

3. Гибридный Multi-GPU рендеринг. Разные стадии конвейера закрепляются за разными GPU: один считает тени/геометрию/постобработку, второй – основной кадр, и т.д. [22] Преимущества: снижение нагрузки на «главный» GPU и высокое перекрытие по времени, особенно для стадий с низкими межкадровыми зависимостями и компактным выходом (например, shadow maps, Hi-Z, SSAO). Недостатки: необходимость явного управления обменом данными и синхронизацией, а выигрыш ограничен полосой пропускания межсоединения и затратами на копирование.

Метод особенно эффективен там, где  $T_{stage@GPU} \gg T_{copy} + T_{overheads}$  [23], то есть если длительность стадии на основном GPU значительно превышает сумму времени передачи данных и накладных расходов, то ее выгодно вынести на второй GPU. В таком случае копирование можно полностью скрыть под другими стадиями [23].

В отличие от CHOPIN и аналогичных решений [15,18], в данной работе рассматриваются сценарии домашних ПК, где:

- взаимодействие ограничено PCIe (Gen3/Gen4/Gen5) [24, 25];
- отсутствует NVLink/XGMI [20, 21];
- возможны гетерогенные пары GPU (например, NVIDIA + Intel или NVIDIA + AMD), что подчёркивается в ряде исследований по гетерогенным вычислениям [26, 27].

API-аспекты (Direct3D 12, Vulkan). В Direct3D 12 реализованы два сценария: linked-node для «связанных» устройств и explicit multi-adapter для разнородных (включая связки «дискретная + интегрированная»). Поддерживаются кросс-адаптерные ресурсы и heap'ы и межадаптерные копии; синхронизация – через межадаптерные fences. В Vulkan аналогичная функциональность предоставляется через device groups и внешнюю память/семафоры. Эти механизмы являются базой для гибридных схем с переносом отдельных стадий на второй GPU.

В потребительских ПК обмен между адаптерами чаще всего идёт по PCIe; его пиковая пропускная способность ограничивает размер и количество межадаптерных копий. В рабочих станциях и HPC доступны NVLink и XGMI/IF Link с существенно большими скоростями и более низкими задержками [18,19,26], однако такие конфигурации нехарактерны для «домашних» систем. Выбор стратегии обмена данными и частоты синхронизаций должен учитывать именно эти аппаратные границы.

Метод Shadow Maps широко используется для реалистичного отображения теней в трёхмерной графике. Он был впервые предложен в 1978 году [28] и с тех пор применяется как в офлайн-рендеринге, так и в реальном времени. Идея заключается в том, что сцена рендерится с позиции источника света в буфер глубины (shadow map), после чего основной кадр отрисовывается из камеры с сравнением глубины каждого пикселя со значением из shadow map. Этот метод позволяет быстро получать как жёсткие, так и мягкие тени, поддерживает любые источники света и не требует сложных геометрических вычислений (в отличие от stencil-теней) [29].

Однако классические Shadow Maps имеют ограниченное разрешение текстуры глубины, из-за чего при большом расстоянии или большом угле обзора тени сильно искажаются (возникает перспективный «алиасинг», см. рис. 1). При удалении камеры от затенённых объектов несколько пикселей экрана могут проецироваться на один пиксел shadow map, приводя к размытию и «ступенчатым» артефактам. Проблема перспективного алиасинга считается одной из самых серьёзных при наложении теней в реальном времени.

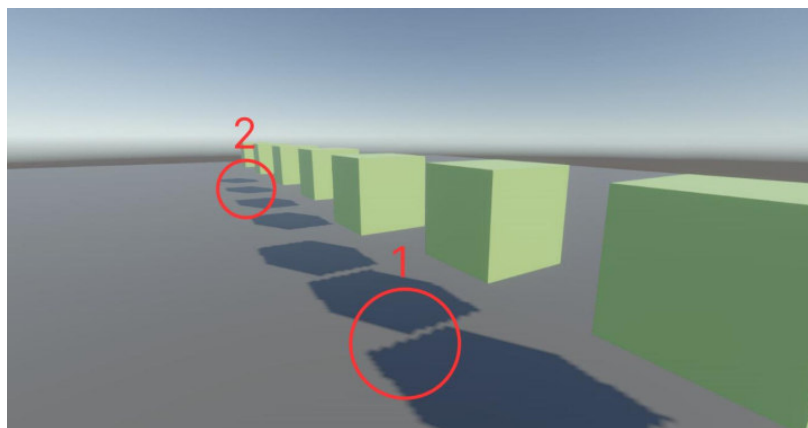


Рисунок 1. Демонстрация проблемы алиасинга

Perspective Shadow Maps (PSM) [30] использует перспективную проекцию от камеры, что повышает плотность теней на ближних расстояниях, но приводит к искажениям на дальних. Метод Light-space Perspective Shadow Maps (LiSPSM) [31] улучшает эту идею, учитывая геометрию сцены и направление света при формировании проекции, что обеспечивает лучшее распределение плотности теней. Также был предложен метод Trapezoidal Shadow Maps (TSM) [32], в котором форма shadow map деформируется в трапецию, приближенную к области обзора камеры, что уменьшает aliasing. Другие методы, такие как Sub-pixel Shadow Mapping [33], направлены на повышение качества теней за счёт

более точной фильтрации границ, а Irregular Z-Buffer [34] использует лучевое построение буфера глубины, полностью избавляясь от регулярной сетки shadow map. Однако оба эти метода требуют высокой вычислительной мощности и сложны в реализации.

Также для улучшения распределения разрешения по глубине сцены были предложены Parallel-Split Shadow Maps [35] и Cascaded Shadow Maps (CSM) [36]. В этих методах пирамида видимости камеры (фрустум) разбивается на несколько диапазонов (каскадов), каждый из которых покрывает свою часть сцены и имеет свою shadow map. Это позволяет выделить большую текстуру и более высокое разрешение теней для ближнего диапазона, где качество особенно важно, и меньше – для дальних объектов.

Современные графические движки также применяют более сложные техники: виртуальные карты теней (Virtual Shadow Maps) [36] благодаря концепции виртуальных текстур [37] позволяют создать карты теней очень высокого виртуального разрешения (например,  $16384 \times 16384$  в UE5) с динамической загрузкой страниц.

Для данного исследования были выбраны методы Shadow Mapping и Cascaded Shadow Maps по нескольким причинам. Во-первых, они являются наиболее широко применяемыми в современной индустрии интерактивной графики, включая игровые движки и системы визуализации в реальном времени, что обеспечивает их практическую актуальность. Во-вторых, эти алгоритмы отличаются разумным балансом между качеством теней и вычислительной сложностью, что особенно важно при построении гибридных мульти-GPU систем. В отличие от более продвинутых методов, выбранные алгоритмы обладают более простой архитектурой и легче масштабируются в условиях нескольких видеокарт, не требуя сложной синхронизации или нестандартных этапов рендеринга. Кроме того, Virtual Shadow Maps, хотя и обеспечивают очень высокое качество, пока остаются слишком ресурсоемкими и требуют тесной интеграции с кастомными рендерерами, что затрудняет их применение в гибридных системах без глубокого вмешательства в движок. Таким образом, выбор Shadow Mapping и CSM позволяет исследовать производительность и масштабируемость классических алгоритмов в условиях мультиадаптерной архитектуры, сохраняя применимость полученных решений в реальных проектах.

### Постановка задачи

Цель данного исследования – реализовать и сравнить методы Shadow Mapping и Cascaded Shadow Maps в контексте гибридных графических систем с несколькими видеокартами. Основное внимание уделяется разработке и тестированию мульти-GPU версий этих алгоритмов, в которых расчёт карт теней переносится на дополнительный графический процессор, в то время как основной GPU выполняет остальные этапы рендеринга. Таким образом, задача состоит в том, чтобы:

- 1) реализовать оба алгоритма (Shadow Maps и Cascaded Shadow Maps) как в однопроцессорной конфигурации, так и в конфигурации с несколькими GPU;
- 2) измерить и сравнить производительность (время рендеринга кадра, FPS) в обоих режимах работы для разных сцен и наборов видеокарт;
- 3) оценить, даёт ли multi-GPU реализация преимущество в скорости по сравнению с классической однопроцессорной схемой;
- 4) провести анализ полученных данных, чтобы определить, в каких условиях использование второй видеокарты наиболее эффективно.

### Теория

В данной работе расчёт карт теней полностью вынесен на вспомогательный адаптер (в т.ч. интегрированный GPU), в то время как основной GPU параллельно выполняет G-pass и эффекты, не зависящие от теней. По готовности shadow maps выполняется кросс-адаптерное копирование в локальную память дискретного GPU, после чего запускается проход освещения. Декомпозиция выбрана как слабо зависящая от данных предыдущего кадра и имеющая компактные выходные ресурсы, что позволяет перекрыть обмен с основными стадиями и минимизировать простои. Реализация опирается на механизмы Direct3D 12 explicit multi-adapter и кросс-адаптерные ресурсы [38]. Схема алгоритма представлена на диаграмме (рис. 2).



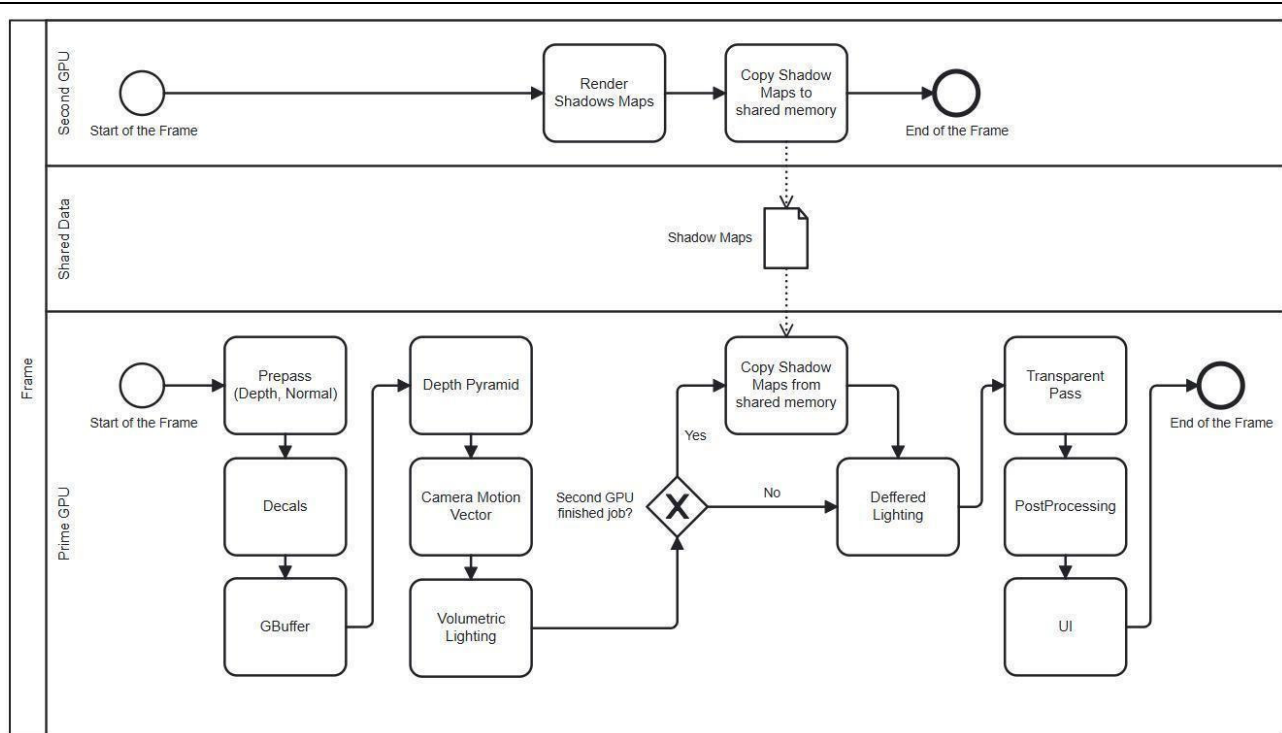


Рисунок 2. Диаграмма активности multi-GPU Shadow Mapping алгоритма

Классический алгоритм Shadow Maps базируется на рендеринге сцены из точки зрения источника света и сравнении глубин. Его этапы обычно таковы:

1. Создание shadow map: сцена рендерится с позиции источника света (обычно с использованием ортогографической или перспективной проекции, в зависимости от типа света), при этом пишется лишь буфер глубины (рис. 3). Полученная карта глубины хранит расстояния до ближайших поверхностей относительно света.

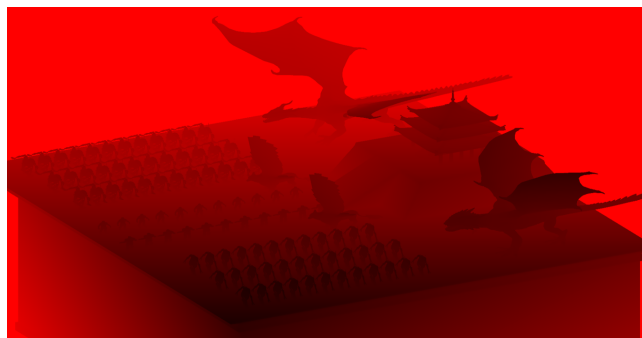


Рисунок 3. Пример shadow map

2. Отрисовка сцены камерой: затем выполняется обычная отрисовка сцены с точки зрения камеры, но при вычислении освещения для каждого фрагмента дополнительно определяется его координата в системе координат света и сравнивается его глубина с соответствующим значением из shadow map. Если глубина фрагмента больше, чем в shadow map, то этот фрагмент невидим для света и должен быть затенён.

Таким образом, Shadow Maps позволяют быстро определить теневые участки без сложных вычислений в пространстве изображения. Недостатком метода является ограниченность разрешения карты глубины. При необходимости покрыть большой объём сцены одна текстура shadow map не содержит достаточно пикселей, чтобы обеспечить плавные тени на больших расстояниях. Это приводит к артефактам перспективного искажения теней (blocky shadows, перспективный «aliasing»).

Метод каскадных теневых карт (CSM) решает эту проблему за счёт адаптивного распределения разрешения по расстоянию. Его суть заключается в разделении пирамиды видимости камеры на несколько диапазонов (сплитов) по глубине (рис. 4). Ближайшие к камере объекты попадают в первый

сплит и затеняются с помощью первой shadow map, второй диапазон – с помощью второй карты глубины и т.д. При этом ближним диапазонам камеры выделяется меньшая область пространства (но на той же текстуре shadow map), что повышает плотность пикселей теней там, где требуется больше деталей. Для дальнего диапазона используется отдельная карта теней, покрывающая больший объём, но его потеря в качестве менее заметна на удалении.

На рисунке 4 изображено типичное разбиение пирамиды видимости на три каскада при использовании CSM: ближний каскад (самый маленький объём) получает самую крупную часть разрешения, дальний – меньшую. Для каждого сплита вычисляются собственные матрицы вида и проекции света, что даёт возможность более точно подобрать область рендеринга shadow map под этот диапазон. Благодаря такому разбиению теней повышается качество вблизи камеры, а дальние области при необходимости рисуются более грубо.

Алгоритм CSM можно формализовать следующими шагами:

1. Разбить frustum камеры на  $n$  диапазонов по глубине.
2. При рендеринге shadow map: для каждого диапазона вычислить матрицы вида и проекции источника света и отрисовать сцену в отдельную текстуру глубины.
3. При финальной отрисовке сцены: для каждого фрагмента определить, к какому диапазону он принадлежит, и использовать соответствующую карту глубины для проверки тени.

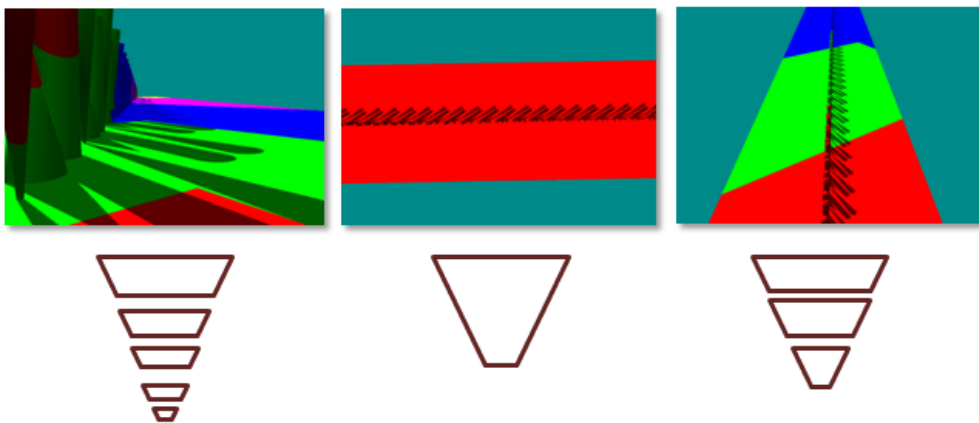


Рисунок 4. Разделение пространства на каскады, отмеченные разным цветом

Основной недостаток CSM – увеличение числа проходов рендеринга теневой карты. Если в обычном алгоритме достаточно одного рендера сцены от света, то в CSM таких проходов  $n$  (по числу каскадов). Это требует дополнительных вычислений и может снижать FPS. Тем не менее в задачах больших открытых миров и при дальних планах эта затрата оправдана за счёт видимого улучшения качества теней (уменьшения артефактов).

### Результаты экспериментов

В рамках эксперимента были проведены замеры времени рендеринга и FPS на различных конфигурациях оборудования. Были протестированы две реализации алгоритмов Shadow Mapping и Cascaded Shadow Maps: стандартная (одна видеокарта) и гибридная (две видеокарты, с расчётом теней на вторичном GPU) при различных параметрах SSAA (SuperSample Anti-Aliasing).

#### 1. Shadow Mapping

Наиболее существенное уменьшение времени  $\varepsilon = (t_1 - t_2)/t_1$  на кадр, где  $t_1$  – время рендеринга с использованием одной видеокарты, а  $t_2$  – время рендеринга с использованием двух видеокарт, зафиксировано на системе с NVIDIA GeForce GTX 1650 и AMD Radeon(TM) Vega 8 Graphics – уменьшение составило 10 % (табл. 1). Среднее уменьшение для остальных систем также варьировалось от 5 до 10 %, демонстрируя положительный эффект от вынесения расчёта теней на вторую видеокарту. Для данных результатов использовалась сцена "Bistro" (2 829 238 треугольников)

На рисунке 5 представлены графики времени рендеринга при различных множителях SSAA. По результатам видно, что при увеличении разрешения (множители 2, 4 и 8) относительный выигрыш

от использования multi-GPU снижается. При множителе 8 выигрыш практически исчезает, что связано с тем, что основная нагрузка в этом случае смещается с расчёта теней на рендеринг основного кадра.

Таблица 1. **Время обработки кадра при использовании обычного метода Shadow Map на сцене "Bistro" (2 829 238 треугольников)**

Тестируемая система	SSAA x1			SSAA x3		
	t <sub>1</sub> , мс	t <sub>2</sub> , мс	ε	t <sub>1</sub> , мс	t <sub>2</sub> , мс	ε
NVIDIA GeForce GTX 980 + Intel(R) HD Graphics 630	9,26	8,85	4%	16,67	16,67	0%
NVIDIA GeForce GTX 1060 + Intel(R) UHD Graphics 630	12,20	10,99	10%	22,73	20,83	8%
NVIDIA GeForce GTX 1650 + AMD Radeon(TM) Vega 8 Graphics	15,38	14,08	8%	27,03	24,39	10%
NVIDIA GeForce GTX 1070 + Intel(R) HD Graphics 630	7,94	7,30	8%	14,29	13,33	7%
NVIDIA GeForce GTX 1080 + Intel(R) HD Graphics 630	6,10	5,46	10%	10,87	10,20	6%
NVIDIA GeForce GTX 1080 Ti + Intel(R) HD Graphics 630	4,98	4,52	9%	8,93	8,47	5%
NVIDIA GeForce RTX 2080 + Intel(R) UHD Graphics 630	4,55	4,88	3%	7,41	6,76	2%

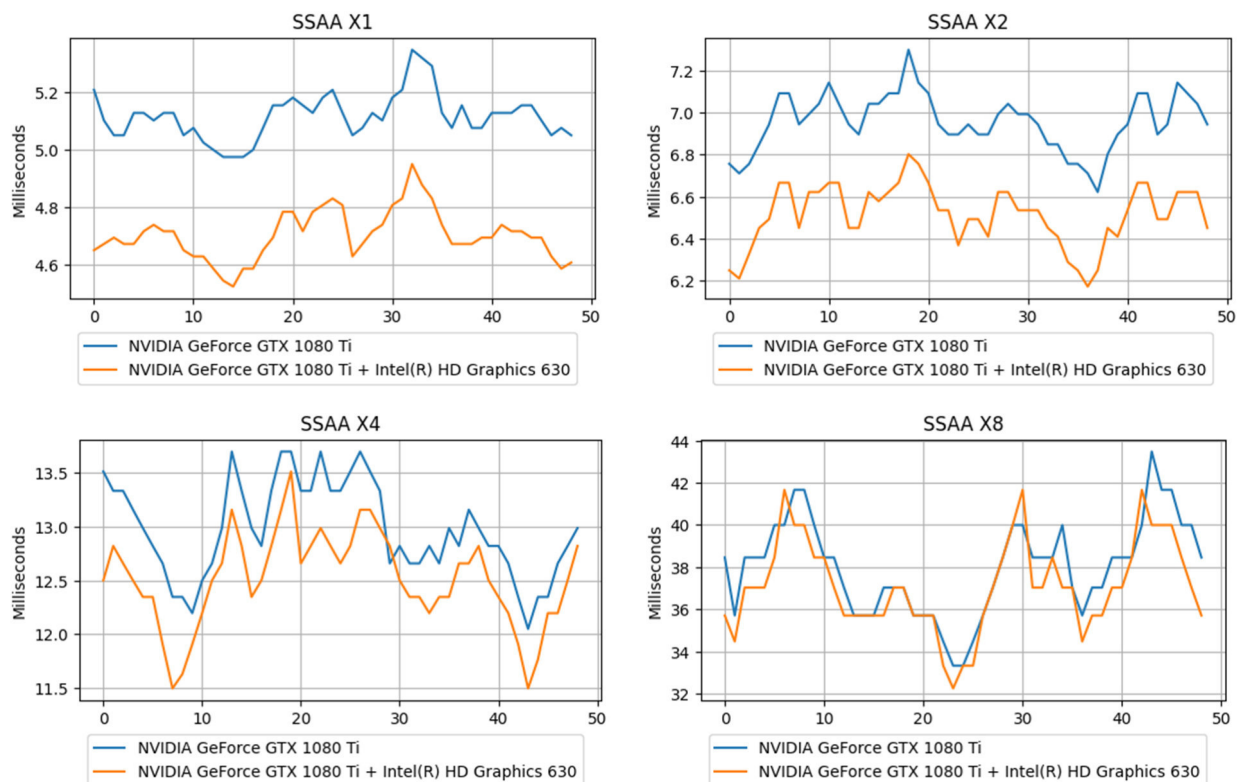


Рисунок 5. Графики времени рендеринга при различных множителях SSAA

### 1. Cascaded Shadow Maps

В трёх сценах разной сложности – "Pica Pica", "The Town on Capital Isle" и "Bistro" – фиксировалось заметное сокращение времени рендеринга в мульти-GPU варианте. Например, система с GTX 1650 Ti и UHD Graphics показала сокращение времени до 49 % на сцене "Bistro" (табл. 2–4). Система с RTX 3060 и Radeon Graphics показала от 20 до 44 % сокращение времени.

Видно, что мульти-GPU эффективно масштабирует именно те части рендеринга, которые можно выполнять независимо, – такие как расчёт карт теней. Однако общая эффективность зависит от архитектурной совместимости GPU, скорости передачи данных и объёма общей памяти.

**Таблица 2. Сокращение времени рендеринга для сцены "Pica Pica" (76 274 треугольников) при использовании CSM**

Тестируемая система	$t_1$ , мс	$t_2$ , мс	Сокращение времени рендера
NVIDIA GeForce GTX 1050 + Intel(R) UHD Graphics 630	2,73	1,97	28%
NVIDIA GeForce GTX 1650 Ti + Intel(R) UHD Graphics	1,41	1,08	23%
NVIDIA GeForce RTX 3060 + AMD Radeon(TM) Graphics	0,7	0,56	20%
NVIDIA GeForce RTX 4070 Laptop GPU + Intel(R) Iris(R) Xe Graphics	0,93	0,86	8%

**Таблица 3. Сокращение времени рендеринга для сцены "The Town on Capital Isle" (193 485 треугольников) при использовании CSM**

Тестируемая система	$t_1$ , мс	$t_2$ , мс	Сокращение времени рендера
NVIDIA GeForce GTX 1050 + Intel(R) UHD Graphics 630	3,61	2,66	26%
NVIDIA GeForce GTX 1650 Ti + Intel(R) UHD Graphics	2,16	1,5	31%
NVIDIA GeForce RTX 3060 + AMD Radeon(TM) Graphics	0,96	0,69	28%
NVIDIA GeForce RTX 4070 Laptop GPU + Intel(R) Iris(R) Xe Graphics	1,28	1,1	14%

**Таблица 4. Сокращение времени рендеринга для сцены "Bistro" (2 829 238 треугольников) при использовании CSM**

Тестируемая система	$t_1$ , мс	$t_2$ , мс	Сокращение времени рендера
NVIDIA GeForce GTX 1050 + Intel(R) UHD Graphics 630	55,74	42,09	24%
NVIDIA GeForce GTX 1650 Ti + Intel(R) UHD Graphics	91,52	46,42	49%
NVIDIA GeForce RTX 3060 + AMD Radeon(TM) Graphics	3,67	2,05	44%
NVIDIA GeForce RTX 4070 Laptop GPU + Intel(R) Iris(R) Xe Graphics	4,27	2,52	41%

### Обсуждение результатов

Полученные данные демонстрируют, что мульти-GPU подход способен значительно улучшить производительность при расчёте теней. Особенно это заметно на сценах с высокой геометрической сложностью и большим количеством теневых проходов, где использование CSM даёт до 49 % выигрыша по времени.

При этом простой метод Shadow Mapping также выигрывает от распределения нагрузки, но его ускорение ограничивается примерно 10–16 % в зависимости от конкретной конфигурации.

### Направления дальнейших исследований

Перспективным направлением развития предложенного подхода является более тонкое распределение задач между адаптерами. В частности, возможна дифференциация каскадов карт теней: ближние каскады, требующие минимальной задержки и высокой точности, могут вычисляться на основном дискретном GPU, в то время как дальние каскады, обладающие меньшими требованиями к качеству и скорости, могут быть вынесены на вспомогательный GPU. Подобное разделение позволит дополнительно снизить нагрузку на основной адаптер и более гибко использовать доступные аппаратные ресурсы.

Другим направлением может стать интеграция гибридного мульти-GPU подхода с другими алгоритмами, которые могут быть вынесены на вторичный GPU и при этом требуют доступа к данным о тенях. К таким задачам относятся, например, процедурная симуляция растительности (листья, травы), расчёты водных поверхностей, а также дополнительные эффекты глобального освещения. Совмещение этих алгоритмов с вынесенным построением карт теней позволит повысить степень параллелизма и снизить задержку в критических стадиях конвейера.

### Выводы

В работе была реализована мульти-GPU архитектура рендеринга теней, в которой расчёт shadow map переносится на вспомогательный графический адаптер. На её основе протестированы методы Shadow Mapping и Cascaded Shadow Maps в различных конфигурациях. Эксперименты показали, что такой подход позволяет достичь прироста производительности до 49 % на сценах с высокой детализацией.

Таким образом, построенный алгоритм эффективно использует доступные аппаратные ресурсы домашних ПК, позволяя разгрузить основной GPU. Разработка хорошо интегрируется в существующие графические конвейеры и применима в широком спектре визуализаторов реального времени, включая игровые движки и системы архитектурной визуализации.

### Список литературы

1. A sorting classification of parallel rendering / Molnar S., et al. // IEEE Comput Graph Appl. 1994. Vol. 14, no. 4. P. 23–32.
2. Sort-first, sort-middle, sort-last parallel rendering / Humphreys G., Hanrahan P. // ACM SIGGRAPH Course Notes. 2002.
3. A Survey of General-Purpose Computation on Graphics Hardware / Owens J.D., et al. // Computer Graphics Forum. 2007. Vol. 26, no. 1. P. 80–113.
4. Pharr M., Fernando R., Humphreys G. GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation. Addison-Wesley, 2005.
5. Krawczyk G. Microstuttering in AFR-based multi-GPU rendering // Computer Graphics Forum. 2009. Vol. 28, no. 2. P. 367–374.
6. Haines E. Temporal reprojection and frame dependencies in rendering // ACM Queue. 2018. Vol. 16, no. 6. P. 40–61.
7. Interactive visibility ordering and transparency computations on GPUs / Govindaraju N.K., et al. // Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games. 2005. P. 49–56.
8. NVIDIA. SLI Best Practices Guide. 2015.
9. VR SLI – Scalable Per-Eye Rendering for VR // NVIDIA Developer. URL: <https://developer.nvidia.com/vrworks/graphics/vrsli> (дата обращения: 01.09.2025).
10. SLI Best Practices and Rendering Modes // NVIDIA Developer. URL: <https://docs.nvidia.com/gameworks/content/technologies/desktop/sli.htm> (дата обращения: 01.09.2025).
11. Karis B. High-quality temporal supersampling // ACM SIGGRAPH Courses. 2014.
12. Moreland K., Angel E. The Distributed Framebuffer // Proceedings of the IEEE Symposium on Parallel Visualization and Graphics. 2001. P. 61–68.
13. Parallel volume rendering using binary-swap compositing / Ma K.-L., et al. // IEEE Comput Graph Appl. 1994. Vol. 14, no. 4. P. 59–68.
14. GPUd: A fast and scalable multi-GPU architecture using cooperative projection and distribution / Kim Y., et al. // Proceedings of the Annual International Symposium on Microarchitecture, MICRO. IEEE Computer Society, 2017. Vol. Part F131207. P. 574–586.

15. Ren X., Lis M. CHOPIN: Scalable Graphics Rendering in Multi-GPU Systems via Parallel Image Composition // Proceedings - International Symposium on High-Performance Computer Architecture. IEEE Computer Society, 2021. Vol. 2021-February. P. 709–722.
16. Load balancing for sort-first parallel rendering / Samanta R., et al. // IEEE Trans Vis Comput Graph. 2001. Vol. 7, no. 4. P. 353–368.
17. Mueller C., Moorhead R., Moreland K. Evaluating bandwidth requirements for parallel rendering systems // IEEE Symposium on Parallel Visualization and Graphics. 2002. P. 103–110.
18. MCM-GPU: Multi-Chip-Module GPUs for Continued Performance Scaling / Jain A., et al. // Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA). 2017. P. 320–332.
19. Moreland K. Scalable Rendering on HPC Systems // Computer Graphics Forum. 2016. Vol. 35, no. 3. P. 485–494.
20. NVLink High-Speed GPU Interconnect // NVIDIA. URL: [www.nvidia.com/en-us/data-center/nvlink](http://www.nvidia.com/en-us/data-center/nvlink) (дата обращения: 01.09.2025).
21. Infinity Fabric Link // AMD. URL: [www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/white-papers/overview-amd-epyc7003-series-processors-microarchitecture.pdf](http://www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/white-papers/overview-amd-epyc7003-series-processors-microarchitecture.pdf) (дата обращения: 01.09.2025).
22. Computer Graphics: Principles and Practice / Foley J. et al. // 3rd изд. Addison-Wesley, 2014.
23. Aila T., Laine S. Understanding the efficiency of ray traversal on GPUs // Proceedings of the Conference on High Performance Graphics (HPG). 2009. P. 145–149.
24. PCI-SIG. PCI Express Base Specification 5.0. 2019.
25. Intel. PCI Express Gen5 Performance Overview. 2021.
26. Qilin: Exploiting parallelism on heterogeneous multiprocessors with adaptive mapping / Luk C.-K., et al. // Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO). 2009. P. 45–58.
27. Diamos G.F., Yalamanchili S. Harmony: An execution model and runtime for heterogeneous many core systems // 2008 IEEE 15th International Symposium on High Performance Computer Architecture. 2008. P. 356–367.
28. Williams L. Casting curved shadows on curved surfaces. 1978.
29. Chapter 12. Omnidirectional Shadow Mapping // NVIDIA Developer. URL: <https://developer.nvidia.com/gpugems/gpugems/part-ii-lighting-and-shadows/chapter-12-omnidirectional-shadow-mapping> (дата обращения: 01.09.2025).
30. Stamminger M., Drettakis G. Perspective shadow maps // Proceedings of the 29th annual conference on Computer graphics and interactive techniques. New York, NY, USA: ACM, 2002. P. 557–562.
31. Wimmer M., Scherzer D., Purgathofer W. Light Space Perspective Shadow Maps. 2004.
32. Anti-aliasing and Continuity with Trapezoidal Shadow Maps / Jensen H.W., et al. // Eurographics Symposium on Rendering. 2004.
33. Sub-pixel shadow mapping / Lecocq P., et al. // Proceedings of the 18th meeting of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games. New York, NY, USA: ACM, 2014. P. 103–110.
34. Johnson G.S., Mark W.R., Burns C.A. The Irregular Z-Buffer and its Application to Shadow Mapping. 2004.
35. Parallel-split shadow maps on programmable GPUs. URL: <https://developer.nvidia.com/gpugems/gpugems3/part-ii-light-and-shadows/chapter-10-parallel-split-shadow-maps-programmable-gpus> (дата обращения: 01.09.2025).
36. Dimitrov Rouslan. Cascaded Shadow Maps. 2007.
37. Van J.M.P., Software W.-I., Hart -Nvidia E. Using Virtual Texturing to Handle Massive Using Virtual Texturing to Handle Massive Texture Data Texture Data. 2010.
38. Multi-adapter systems and Direct3D 12; Shared resources across adapters // Microsoft Learn. URL: <https://learn.microsoft.com/en-us/windows/win32/direct3d12/multi-engine> (дата обращения: 01.09.2025).



## Гибридный алгоритм отображения облаков для систем с несколькими графическими адаптерами

М. К. Богданов, А. М. Суворов, А. П. Булаев

Университет ИТМО, г. Санкт-Петербург, Россия

**Аннотация.** Параллельный рендеринг представляет собой широко распространенное решение для повышения производительности в компьютерной графике, при котором задействуется несколько графических процессоров (GPU) для одного приложения. Такие методы позволяют эффективно балансировать нагрузку и уменьшать временные простои при рендеринге, которые могут ограничивать частоту и детализацию итогового изображения. Однако современные реализации в основном ориентированы на серверные решения с аппаратной поддержкой высокоскоростных шин передачи данных и не учитывают особенности игровых платформ, таких как ноутбуки и персональные компьютеры, оснащенные как интегрированной, так и дискретной видеокартами. В этой работе основной фокус ставится именно на пользовательские конфигурации [1], которые чаще всего оснащены двумя графическими адаптерами. Визуализация реалистичных облаков в компьютерной графике является сложной задачей, связанной с большим количеством вычислений. В работе представлен гибридный алгоритм отображения реалистичных облаков, распределяющий графические задачи между GPU на основе отдельных техник, которые выполняются на основном графическом процессоре. Такой подход позволяет использовать DirectX 12 в режиме EMA, не зависящем от аппаратных интерфейсов и специфики производителей видеоадаптеров. Разработанное решение использует дополнительный графический процессор для генерации текстур облаков, что обеспечивает рациональное использование всех доступных аппаратных ресурсов системы без потери качества итогового изображения. Предложенный метод легко интегрируется в существующие конвейеры рендеринга и позволяет достичь значительного прироста производительности в пользовательских гетерогенных мульти-GPU-системах при минимальных затратах на реализацию.

**Ключевые слова:** компьютерная графика, адаптивные алгоритмы, облака, DirectX, GPU-симуляция, параллельный рендеринг.

## Hybrid cloud simulation algorithm for multi-GPU systems

M. K. Bogdanov, A. M. Suvorov, A. P. Bulaev

ITMO University, Saint-Petersburg, Russia

**Abstract.** Parallel rendering represents a widely adopted solution for enhancing performance in computer graphics, utilizing multiple graphics processing units (GPUs) within a single application. Such methods enable efficient load balancing and reduce rendering latency, which may otherwise constrain frame rates and the level of detail in the final image. However, contemporary implementations primarily target server-based solutions equipped with hardware support for high-speed data buses and fail to address the specificities of gaming platforms, such as laptops and personal computers featuring both integrated and discrete graphics cards. The primary focus of this work is on user systems, which are most commonly equipped with two graphics adapters. Rendering realistic clouds in computer graphics poses a significant challenge due to its computationally intensive nature. This paper presents a hybrid algorithm for realistic cloud visualization, which distributes graphics tasks across GPUs by delegating specific rendering techniques to the primary graphics processor. This approach facilitates the use of DirectX 12 in Explicit Multi-Adapter (EMA) mode, eliminating dependencies on proprietary hardware interfaces and vendor-specific limitations. The proposed solution employs an additional GPU for cloud texture generation, ensuring optimal utilization of all available hardware resources without compromising output quality. The method is designed for seamless integration into existing rendering pipelines and demonstrates substantial performance improvements in heterogeneous multi-GPU systems with minimal implementation overhead.

**Keywords:** computer graphics, adaptive algorithms, clouds, DirectX, GPU-simulation, parallel rendering.

### Введение

Компьютерная графика является важной частью визуализации физических процессов, интерактивных развлечений и современной инженерии и дизайна. Степень реалистичности итогового рендера определяется корректностью используемых методов просчёта освещения и рендеринга. Эти факторы напрямую влияют на то, как пользователь воспринимает изображение. Особые трудности возникают при визуализации облаков, так как необходимо каждый кадр просчитывать многократное рассеивание света, динамически изменяемые вольюметрические (объёмные) структуры, а также поддерживать производительность на комфортном для пользователя уровне.

Применяемые сегодня алгоритмы, использующие растеризацию и трассировку лучей, демонстрируют компромисс между скоростью и точностью. Физически корректные методы обладают высокой вычислительной сложностью. В то же время упрощенные методы приводят к упрощению визуализации и худшему качеству изображения, все эти решения не подходят для гибридных систем. Растущая популярность игровых ноутбуков, оснащённых несколькими графическими адаптерами [1], подчеркивает важность разработки алгоритмов, оптимизированных для подобных конфигураций.

Цель работы – исследование возможности вынесения генерации текстур облаков на дополнительный графический процессор (включая интегрированные решения) и оценка эффективности такого распределения вычислений в условиях потребительских систем с ограниченной пропускной способностью шины PCI Express.

В данной работе:

- 1) предложен и разработан алгоритм гибридного рендеринга облаков, распределяющий нагрузку между графическими адаптерами и предназначенный для использования в приложениях, где требуется реалистичная визуализация неба без необходимости поддержки произвольных углов обзора;
- 2) показана применимость такого подхода для конфигураций с интегрированным и дискретным GPU без использования специализированных интерфейсов по типу NVLink и XGMI;
- 3) проведён анализ производительности и выявлены условия, в которых обмен данными между адаптерами не даёт прироста производительности системе.

### Постановка задачи

Цель данного исследования – реализовать метод гибридного рендеринга облаков для multi-GPU-систем и провести сравнение с подобным методом, использующим один графический адаптер. Основное внимание уделяется разработке и тестированию алгоритма, в котором текстуры облаков вычисляются на вспомогательном GPU и передаются на основной GPU, на котором обрабатываются все остальные стадии графического конвейера.

Таким образом, задача состоит в том, чтобы:

- 1) написать алгоритмы рендеринга облаков, использующие как один графический адаптер, так и несколько;
- 2) измерить производительность и провести сравнение времени рендеринга кадра и FPS с этими алгоритмами на разных системах;
- 3) оценить преимущества алгоритма гибридного рендеринга облаков перед однопроцессорной конфигурацией;
- 4) провести анализ полученных данных, чтобы определить, в каких условиях использование второй видеокарты наиболее эффективно.

### Теория

Рассмотрим основные подходы к созданию двумерных структур в виде облаков. Широкое распространение в компьютерной графике получила техника Billboard [2], где в сцену интегрируется плоское изображение, постоянно направленное к камере пользователя (рис. 1).

Другой популярный метод основан на наложении нескольких слоёв текстур с варьируемыми атрибутами прозрачности и позиции [3], (рис. 2). Оба подхода страдают от статичности результата, поскольку опираются на статичные изображения, что исключает динамическую трансформацию форм.

Алгоритм процедурной генерации с применением математических шумовых функций выступает альтернативой. Технология обеспечивает создание уникальных анимированных текстур, однако при повышении детализации или выборе сложных математических функций вычислительные затраты резко возрастают (рис. 3). Также недостатком является потеря объёмности – критически важного аспекта в визуализации облаков.

Для преодоления указанных проблем разработан комбинированный алгоритм, объединяющий метод процедурной генерации шума с методом многослойного наложения текстур. Динамическое создание каждого слоя отдельно обеспечивает объёмность, визуальную динамику и возможности имитации различных конфигураций облаков. Для симуляции неба используется техника Skybox, для верхней части которой каждый кадр процедурно вычисляется облачная текстура. Пример представлен на рисунке 4.

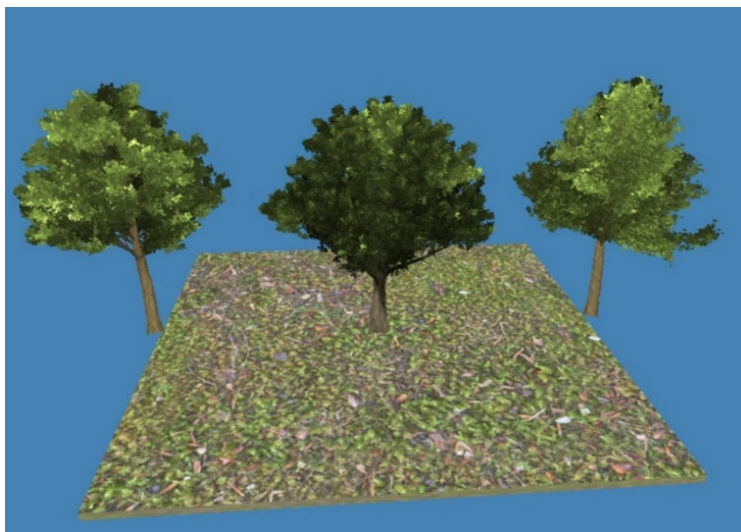


Рисунок 1. Демонстрация работы метода Billboard



Рисунок 2. Демонстрация работы метода Layered Texture

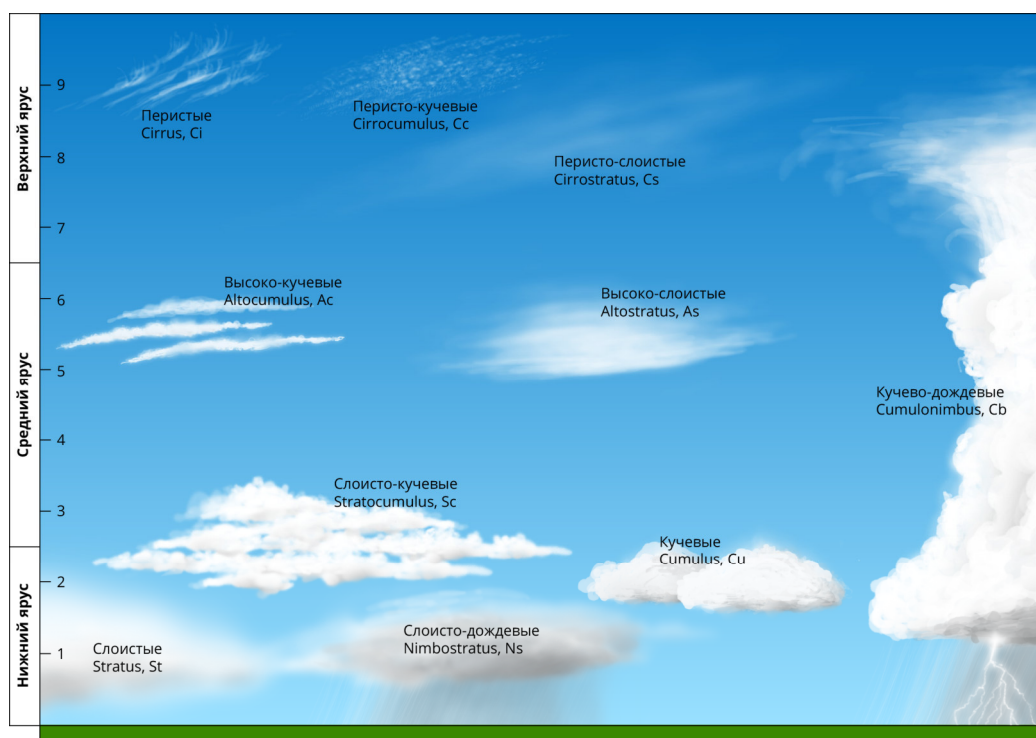


Рисунок 3. Типы облаков



Рисунок 4. Визуализация работы алгоритма

Рассмотрим способы работы с несколькими GPU [4,5]. В современных системах с несколькими графическими процессорами применяются три основных подхода к распределению вычислительной нагрузки.

1. *Межкадровый параллелизм.* Это технологии наподобие AFR (Alternate frame rendering) [6,7], при использовании которой каждый GPU рендерит отдельные кадры, чередующиеся с кадрами с другой видеокарты. Например, при двух видеокартах одна рендерит чётные кадры, а вторая – нечётные. Основные преимущества – простая реализация и минимальная связность между устройствами. Недостатки – повышенная задержка вывода, необходимость полного дублирования ресурсов на каждом адаптере, а также проблемы совместимости с темпоральными алгоритмами современного рендеринга (например TAA, TSR, SSR) [8]. Данный подход использовался в технологиях NVIDIA SLI и AMD CrossFire, но в настоящее время их поддержка сокращается [9,10].

2. *Пространственное разделение.* Технологии наподобие SFR (Split-frame rendering) [11,12] разделяют экран на отдельные области, которые обрабатываются разными GPU. У решений с таким подходом (например CHOPIN [13], MCM-GPU [14]) присутствуют сложность синхронизации при наложении регионов друг на друга, высокие требования к пропускной способности интерфейса и необходимость распределения геометрических данных. Современные реализации требуют специализированных высокоскоростных интерконнектов (NVLink, XGMI) [15,16] или высокопроизводительных сетевых технологий (Infiniband, RDMA) [12] и подходят в основном для серверных решений.

3. *Функциональное разделение (гибридный подход).* При таком подходе разные стадии графического конвейера закрепляются за разными GPU [17]. Например, один адаптер генерирует текстуры облаков, в то время как другой занимается основным рендерингом финального кадра со всеми стадиями. Подход эффективен, когда время выполнения конкретной стадии значительно превышает затраты на передачу данных. Это применимо ко многим генеративным алгоритмам и симуляциям в компьютерной графике. Ключевым ограничением является пропускная способность шины PCI Express, что особенно актуально для потребительских систем с гетерогенными конфигурациями GPU, например ноутбуками.

Современные графические API (DirectX 12, Vulkan) предоставляют необходимые механизмы для реализации гибридного подхода, включая кросс-адаптерные ресурсы и синхронизацию. Однако эффективность решения напрямую зависит от оптимизации межадаптерного обмена данными в условиях ограниченной пропускной способности PCIe.

В DirectX12, который мы используем в своей работе, за работу с несколькими GPU отвечает режим Explicit Multi-Adapter (EMA), у которого есть два варианта работы:

- **Linked Mode.** Несколько физических адаптеров работает как единое устройство с общей памятью, что требует однородности GPU и специализированных высокоскоростных интерконнектов (NVLink/XGMI);

- **Unlinked Mode.** Каждый GPU рассматривается как независимое устройство с отдельной памятью, позволяя работать с гетерогенными конфигурациями.

Для нашего гибридного алгоритма рендеринга облаков был выбран Unlinked Mode по следующим причинам:

- 1) подход к гетерогенным системам. Unlinked Mode позволяет задействовать разнородные GPU (дискретный + интегрированный) без требований к идентичности архитектуры;
- 2) отсутствие зависимости от специализированных интерконнектов. Unlinked Mode работает через стандартные PCIe-соединения, что соответствует стандартам потребительских систем;
- 3) гибкость распределения задач. Unlinked Mode предоставляет точный контроль над назначением стадий рендеринга конкретным адаптерам.

После создания алгоритма симуляции облаков была реализована его гибридная версия, алгоритм показан на рисунке 5. Данный подход, в отличие от последовательного расчета на одном GPU, организует параллельные вычисления на нескольких GPU, оптимизируя загрузку вычислительных ресурсов системы. На дополнительном GPU происходит генерация текстур облаков, после чего они копируются на другую видеокарту. Основной GPU использует уже созданную другим адаптером текстуру облаков при рендере кадра, а не вычисляет, в отличие от версии алгоритма для одного адаптера. Таким образом мы полностью снимаем достаточно затратную по вычислениям задачу генерации облаков с основного GPU.

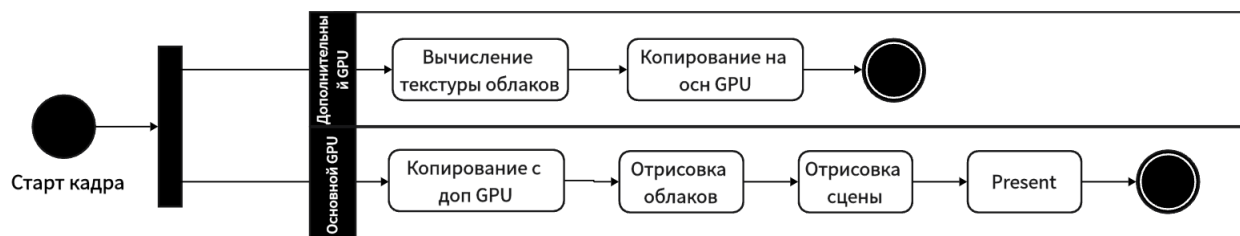


Рисунок 5. Гибридный алгоритм

## Результаты экспериментов

Для обеспечения значимости результатов выполнено по 10 прогонов на различных тестовых стендах и замеры данные. Для примера на представленных графиках (рис. 6 и 7) показаны усредненные значения времени рендера кадра на двух стендах. Каждый запуск включал 30-секундную фазу инициализации и прогрева кэшей, после чего в течение 120 секунд осуществлялись основные расчеты и замеры производительности. Для каждого кадра регистрировалось время, затраченное на его рендеринг.

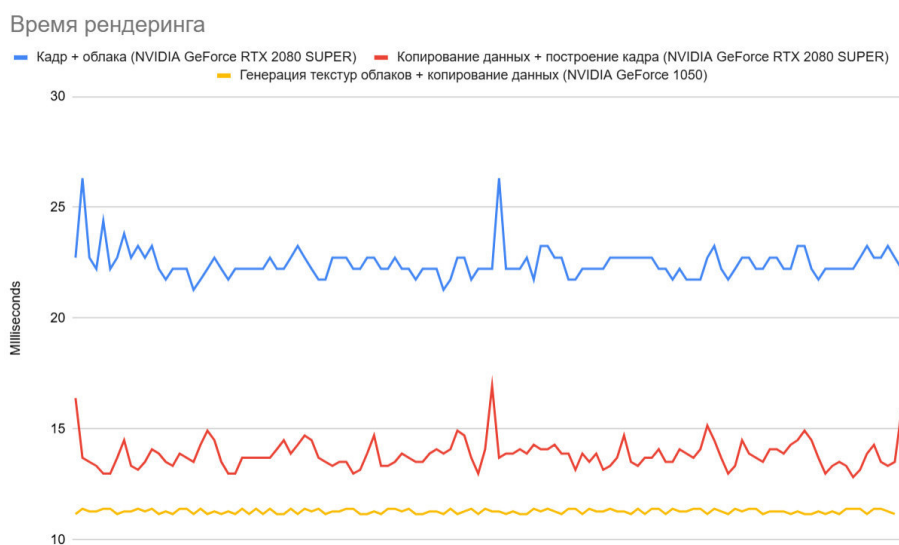


Рисунок 6. Результаты тестирования RTX 2080 SUPER + NVIDIA GeForce 1050 (Процессор AMD Ryzen 2800X)

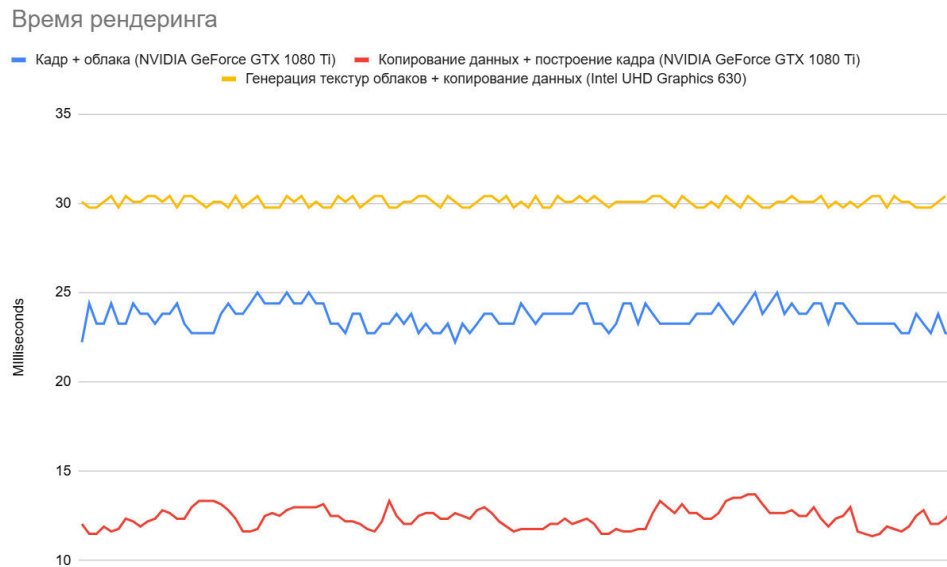


Рисунок 7. Результаты тестирования GTX 1080 Ti + Intel UHD Graphics 630 (Процессор Intel Core i7 9700)

Сцена для теста состояла из вращающейся вокруг центра мира камеры, статических объектов, системы частиц с симуляцией 20000 полупрозрачных частиц и пяти источников освещения, один из которых представлял собой Directional light, эмулирующий солнечный свет. Количество отображаемых полигонов составляло примерно два миллиона. Динамические тени от всех источников освещения рассчитывались для каждого кадра. Для симуляции облачного неба использовалась техника Skybox, для верхней части которой в каждом кадре процедурно вычислялась текстура облаков. Пример можно наблюдать на рисунке 8. Все эксперименты проводились в разрешении Full HD (1920×1080 пикселей) с использованием последних стабильных версий драйверов с официальных сайтов производителей, доступных на момент тестирования.

Анализ данных показал прирост суммарной производительности рендеринга в реальном времени на 40–60 % в зависимости от конфигурации системы.

Однако тесты показали возникновение визуальных артефактов при значительном отрыве в производительности GPU. Это обусловлено существенным отставанием частоты обновления данных на дополнительном адаптере от частоты рендера кадра на основном. Для устранения данного недостатка в алгоритм внедрён коэффициент масштабирования, изменяющий размер текстуры облаков.

При запуске приложения замеряются время генерации текстуры (на дополнительном адаптере, iGPU) и время построения кадра (на основном адаптере, dGPU) для одного кадра.

Если дополнительный адаптер завершил кадр быстрее основного или за то же время, то разрешение текстуры не меняется (в нашем примере изначально 4096×4096 пикселей). Если же основной адаптер рендерит один кадр быстрее дополнительного, то разрешение уменьшается в 2 раза по каждой стороне и начинается цикл изменения разрешения для поиска подходящего; если при замере времени видеокарт в следующем кадре время обеих видеокарт примерно совпадает, происходит выход из цикла и текущее разрешение сохраняется на всё время работы приложения. Если время iGPU меньше времени dGPU, разрешение увеличивается на 50 % от прошлого диапазона, если больше – уменьшается на 50 % от прошлого диапазона. После этого цикл начинается заново с измерения времени рендера на видеокартах.

После добавления алгоритма изменения размера текстуры скорости рендеринга на обоих GPU стали примерно одинаковыми и артефакты, вызванные сильным отставанием дополнительного адаптера от основного, пропали.



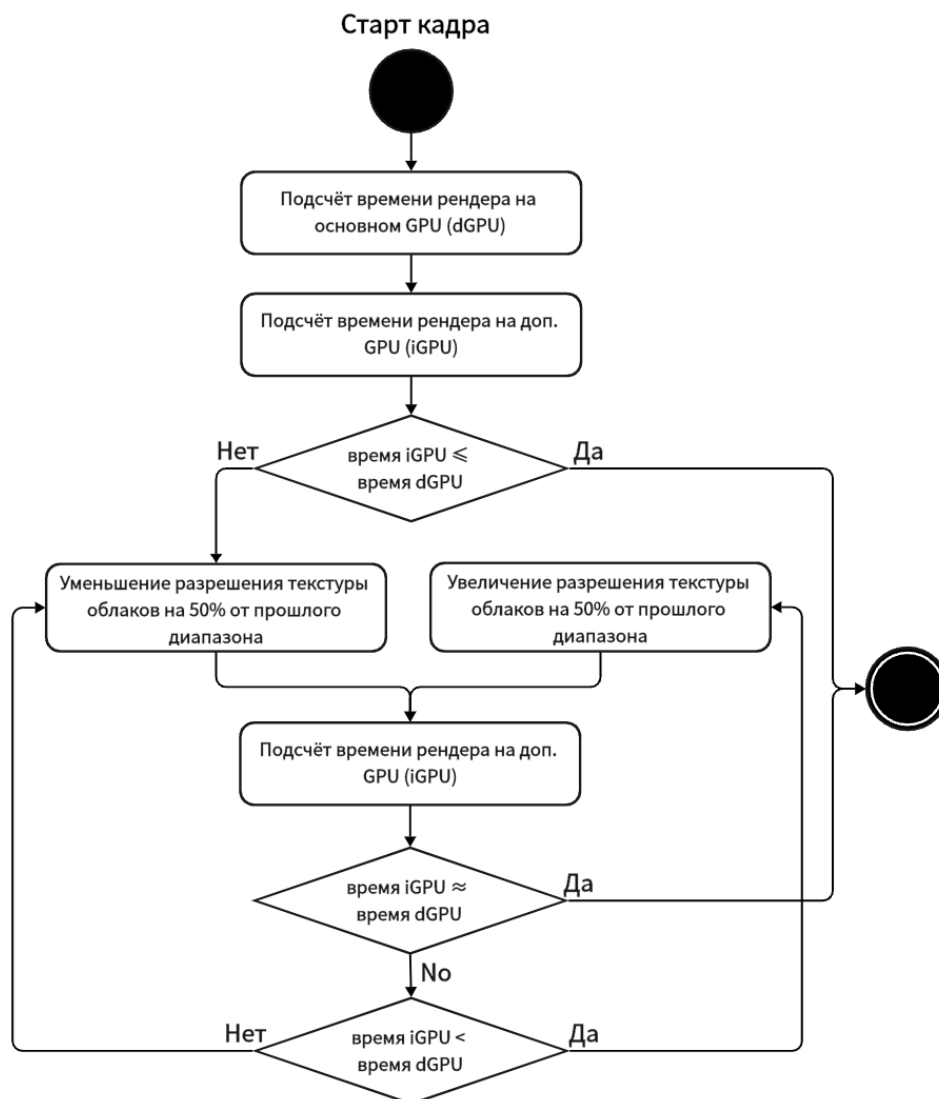


Рисунок 8. Алгоритм расчета коэффициента масштабирования текстуры облаков

## Время рендеринга с изменением размера текстуры

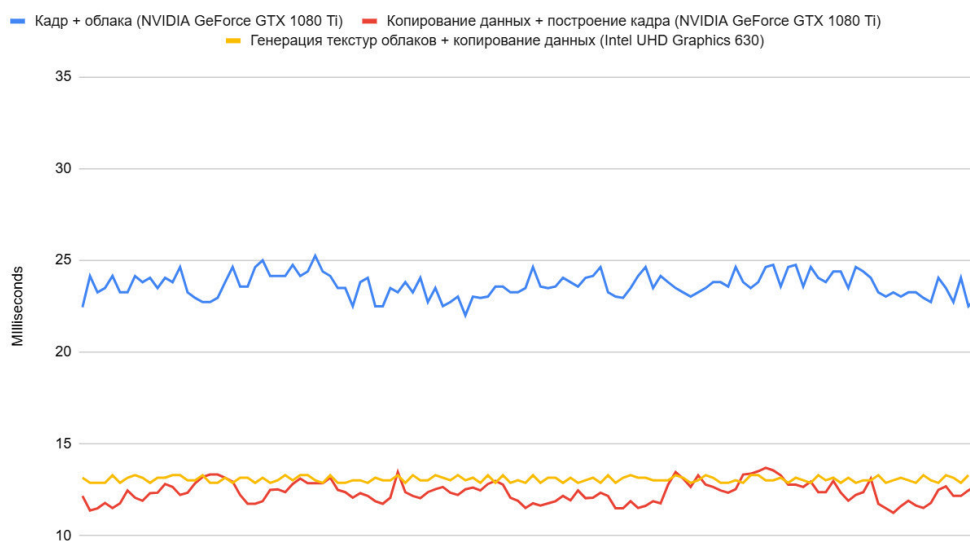


Рисунок 9. Результаты тестирования GTX 1080 Ti + Intel UHD Graphics 630 после добавления коэффициента масштабирования текстуры (процессор Intel Core i7 9700)

## Выводы

Результаты данного исследования демонстрируют обеспечение прироста производительности графической части системы с несколькими GPU благодаря разработанному методу отображения облаков, а также создают перспективы для дальнейших исследований в данной области. Было выявлено, что гибридные вычисления сильно зависят от конфигурации системы. В ходе работы были созданы подходы для решения проблем в системах с кардинально различающимися по мощности видеокартами.

В перспективе планируется решить проблему с уменьшением детализации неба и исследовать эффективность распределения вычислительных алгоритмов на несколько видеоадаптеров для объёмных 3D-облаков, а также разработать методы компенсации качества генерируемой текстуры при её использовании в системах с GPU разной производительности.

## Список литературы

1. Steam Данные об оборудовании пользователей за апрель. URL: <https://store.steampowered.com/hwsurvey/video-card> (дата обращения: 19.05.2025).
2. Tomas Akenine-Miller, Eric Haines Billboarding. URL: [https://www.flipcode.com/archives/Billboarding-Excerpt\\_From\\_iReal-Time\\_Rendering\\_2E.shtml](https://www.flipcode.com/archives/Billboarding-Excerpt_From_iReal-Time_Rendering_2E.shtml) (дата обращения: 14.05.2025).
3. Roden T., Parberry I. Clouds and stars. New York, NY, USA: ACM, 2005. С. 434–437.
4. A Sorting Classification of Parallel Rendering / Molnar S., et al. 1994.
5. Humphreys G., Hanrahan P. Sort-first, sort-middle, sort-last parallel rendering // ACM SIGGRAPH Course Notes. 2002.
6. Krawczyk G. Microstuttering in AFR-based multi-GPU rendering // Computer Graphics Forum. 2009. Vol. 28, no. 2. P. 367–374.
7. Haines E. Temporal reprojection and frame dependencies in rendering // ACM Queue. 2018. Vol. 16, no. 6. P. 40–61.
8. Karis B. High-quality temporal supersampling // ACM SIGGRAPH Courses. 2014.
9. VR SLI – Scalable Per-Eye Rendering for VR // NVIDIA Developer. URL: <https://developer.nvidia.com/vrworks/graphics/vrslr> (дата обращения: 01.09.2025).
10. NVIDIA. SLI Best Practices Guide. 2015.
11. Moreland K., Angel E. The Distributed Framebuffer // Proceedings of the IEEE Symposium on Parallel Visualization and Graphics. 2001. P. 61–68.
12. Parallel volume rendering using binary-swap compositing / Ma K.-L., et al. // IEEE Comput Graph Appl. 1994. Vol. 14, no. 4. P. 59–68.
13. Ren X., Lis M. CHOPIN: Scalable Graphics Rendering in Multi-GPU Systems via Parallel Image Composition // Proceedings – International Symposium on High-Performance Computer Architecture. IEEE Computer Society, 2021. Vol. 2021-February. P. 709–722.
14. MCM-GPU: Multi-Chip-Module GPUs for Continued Performance Scaling / Jain A., et al. // Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA). 2017. P. 320–332.
15. NVLink High-Speed GPU Interconnect // NVIDIA. URL: [www.nvidia.com/en-us/data-center/nvlink](http://www.nvidia.com/en-us/data-center/nvlink) (дата обращения: 01.09.2025).
16. Infinity Fabric Link // AMD. URL: [www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/white-papers/overview-amd-epyc7003-series-processors-microarchitecture.pdf](http://www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/white-papers/overview-amd-epyc7003-series-processors-microarchitecture.pdf) (дата обращения: 01.09.2025).
17. Computer Graphics: Principles and Practice / Foley J., et al. 3rd изд. Addison-Wesley, 2014.

## Гибридный метод расчета окружающего затенения для систем с несколькими видеоадаптерами

М. К. Богданов, М. Е. Ивашечкина, А. М. Суворов, А. П. Булаев  
Университет ИТМО, г. Санкт-Петербург, Россия

**Аннотация.** Параллельный рендеринг – это популярный подход, используемый в компьютерной графике для повышения производительности системы за счет использования связки видеокарт. Однако большинство современных реализаций полагается на серверные версии видеокарт, у которых есть аппаратная поддержка связи по скоростной шине передачи данных, а также полностью игнорирует рынок игровых устройств, таких как ноутбуки и десктопные компьютеры, в которых есть дискретная видеокарта и интегрированный в процессор видеочип. Мы решили разработать механизм разделения графических задач не на основе данных, а на основе техник, выполняемых на основной видеокарте; это позволит нам использовать DirectX 12 в режиме ЕМА и не полагаться на аппаратные интерфейсы, а также избавиться от привязки к производителю видеоадаптеров. Как идеальный кандидат для оценки эффективности такого разделения нами был выбран алгоритм расчета окружающего затенения. Наш алгоритм использует дополнительную видеокарту для вычисления карты затенения, тем самым эффективно применяя все доступные в системе аппаратные ресурсы, сохраняя при этом качество и точность рассчитываемых данных. Такой подход легко интегрируется в существующие конвейеры рендеринга и позволяет без особых затрат увеличить производительность рендеринга на системах с несколькими GPU.

**Ключевые слова:** Ambient Occlusion, SSAO, DirectX 12, Multi-GPU, компьютерная графика, HBAO

## Hybrid ambient occlusion method for systems with multiple video adapters

M. K. Bogdanov, M. E. Ivashechkina, A. M. Suvorov, A. P. Bulaev  
ITMO University, Saint-Petersburg, Russia

**Abstract.** Parallel rendering is a popular approach used in computer graphics to enhance system performance by utilizing multiple graphics cards. However, most of all these approaches rely on server-grade GPUs, which feature hardware support for high-speed data bus communication, and completely ignore the consumer device market, such as laptops and desktop systems that include both a discrete GPU and an integrated graphics chip within the processor. We have developed a mechanism for distributing graphics tasks not based on data, but rather on techniques executed on the primary GPU. This allows us to use DirectX 12 in Explicit Multiadapter (EMA) mode, avoiding dependence on specific hardware interfaces and eliminating vendor lock-in for graphics adapters. As an ideal candidate for evaluating the efficiency of such distribution, we selected the ambient occlusion calculation algorithm, which takes into account multiple GPUs in the system. Our algorithm utilizes an additional GPU to compute the shading map, thereby efficiently leveraging all available hardware resources in the system while maintaining the quality and accuracy of the computed data. This approach can be easily integrated into existing rendering pipelines and enables a cost-effective performance boost for rendering on systems with multiple GPU.

**Keywords:** Ambient Occlusion, SSAO, DirectX 12, Multi-GPU, computer graphics, HBAO

### Введение

Графические процессоры (GPU) разрабатывались для ускорения обработки графики – процесса генерации 2D-изображений из 3D-моделей [1]. Несмотря на развитие видеокарт и исследование методов по их использованию для универсальных вычислений, высокопроизводительная обработка графики все так же составляет одну из основных долей спроса на GPU.

Более того, графика остается доминирующим источником дохода для поставщиков GPU: например, доходы от рынка графики для NVIDIA за 2024 год принесли примерно 44 % от суммарной годовой прибыли и исчисляются миллиардами долларов [2]. Это обусловлено тем, что многим приложениям, включая игры, научную визуализацию данных, компьютерное проектирование, виртуальную реальность (VR), дополненную реальность (AR) и так далее, необходимы все более производительные графические адаптеры. Сама игровая индустрия продолжает развиваться: современные игры в 4K и VR требуют производительности в 4x и 7x больше, чем игры в формате 1080p HD соответственно, в то время как сами игры содержат миллионы или миллиарды треугольников, часто меньших, чем пиксель [3].

Постоянные нововведения увеличивают потребность в повышении графической производительности, однако ее становится все труднее удовлетворить с помощью традиционных одиночных графических

адаптеров. Чтобы продолжить масштабирование производительности GPU, поставщики GPU недавно создали продвинутые аппаратные механизмы передачи данных и синхронизации видеокарт [4, 5], на использование которых полагаются различные распределенные архитектуры, такие как Multi-Chip-Module GPU (MCM-GPU) MCM [6].

В теории эти платформы могут предложить значительные возможности для повышения производительности; на практике же их решения касательно производительности для обработки графики отличаются от однокристальных GPU, и полная реализация преимуществ требует использования распределенных алгоритмов рендеринга.

Параллельный рендеринг не является чем-то новым; разработчики видеокарт уже давно объединяют от двух до четырех видеоадаптеров с помощью аппаратных интерфейсов, таких как SLI и Crossfire [5, 7, 8]. Благодаря этим интерфейсам можно распределять работу, необходимую для отображения кадра либо с помощью альтернативного рендеринга (AFR), где разные связанные видеокарты последовательно обрабатывают кадры, либо с помощью разделенного рендеринга (SFR), который делит отображаемый кадр на части, а обработка каждой непересекающейся области осуществляется отдельной видеокартой. В AFR каждая видеокарта в системе обрабатывает кадры независимо, что увеличивает частоту отображения кадров в системе, но это справедливо только для двух идентичных видеокарт: при подключении двух и более видеокарт с разной производительностью возникает проблема бутылочного горлышка и суммарная производительность системы ограничивается самой «слабой» видеокартой [9]. В свою очередь, SFR может улучшить как частоту кадров, так и уменьшить рендер одиночного кадра [10–13]. Из-за этого SFR чаще используется на практике, и, как следствие, существует огромное количество различных модификаций данного подхода к рендерингу с несколькими видеокартами.

Компромисс состоит в том, что SFR требует, чтобы GPU обменивались как данными, необходимыми для создания кадра, так и результатами расчетов, что создает значительные проблемы пропускной способности и задержки из-за синхронизации [5]. Хотя недавнее появление высокопроизводительных соединений, таких как NVLink и XGMI, направлено на уменьшение ограничений пропускной способности между различными GPU, определенные аппаратные ограничения все так же остаются.

К сожалению, все популярные и актуальные алгоритмы рендеринга на нескольких видеокартах акцентируются только на системах, в которых есть аппаратная возможность соединить между собой графические адаптеры. Такие подходы не допускают возможность использования видеокарт разных производителей и к тому же полностью игнорируют состав отображаемого кадра с точки зрения используемых графических техник.

Эти ограничения полностью исключают возможность использования существующих подходов к параллельному рендерингу на самых распространённых игровых платформах, таких как ноутбуки и персональные компьютеры, в которых есть одна дискретная видеокарта и интегрированное в процессор видео ядро [14].

В данной статье мы предлагаем новый гибридный метод расчета окружающего затенения, который использует в своей работе несколько видеокарт, вне зависимости от их производителя. Для этого мы используем разделение отрисовки не по данным, а по задачам, которые выполняются на каждой видеокарте. Такой подход позволил снизить время построения кадра, а также уменьшить требования к видеопамати, необходимой для построения кадра по сравнению с актуальными алгоритмами.

В рамках работы были выполнены следующие задачи:

- рассмотрены основные популярные алгоритмы рендеринга с использованием нескольких видеокарт и выявлены их эксплуатационные затраты и ограничения, которые мешают использовать данные подходы на распространённых гибридных игровых конфигурациях;
- на основе анализа графа построения кадра в современных игровых движках и играх предложена гибридная реализация техники окружающего затенения, которую потенциально можно эффективно перенести для параллельного вычисления;
- разработан графический бенчмарк и протестированы разные реализации алгоритмов окружающего затенения на разных аппаратных платформах с учетом разделения вычислений на разные графические адаптеры.

Исходя из анализа полученных тестовых данных, разработанный нами гибридный подход к расчету локального затенения, позволяет уменьшить время построения кадра до 20 % в зависимости от конфигурации системы. Он избавлен от необходимости дублирования всех данных между устройствами и не привязан к одному производителю аппаратного обеспечения, а также его можно интегрировать в существующие игровые и графические решения реального времени.

### Предпосылки и мотивации

#### А. Ограничения существующих подходов к параллельному рендерингу

Для удовлетворения растущих требований к качеству графики и суммарной производительности системы разрабатываются различные техники параллельного рендеринга. Большинство из них [6, 10–12, 15–33] разработаны для серверного применения с использованием нескольких однородных GPU, соединенных между собой через высокоскоростные аппаратные интерфейсы [4, 5, 7, 8, 34–36], что позволяет эффективно распределять нагрузку и синхронизировать данные. Однако эти подходы не применимы к популярным пользовательским конфигурациям, таким как ноутбуки или настольные ПК, в которых комбинируется дискретная (dGPU) и интегрированная (iGPU) видеокарта.

Между серверными и пользовательскими системами существуют принципиальные различия:

1. Неоднородность производительности, выражаемая в существенной разнице между более мощной дискретной видеокартой и значительно менее мощной интегрированной, что делает все традиционные схемы и методы балансировки нагрузки неэффективными;
2. Ограниченная пропускная способность, связанная с тем, что вместо высокоскоростных шин по типу NVLink и XGMI, соединение между dGPU и iGPU осуществляется через PCIe. Из-за этого мы имеем значительно меньшую пропускную способность и большую задержку;
3. Отсутствие различных специализированных аппаратных средств синхронизации и общения между устройствами из-за ограничений на уровне драйверов и операционной системы.

Все эти ограничения заставляют взглянуть на гибридные пользовательские системы под другим углом с целью разработки алгоритмов, которые бы нивелировали все описанные ограничения. Для этого мы решили рассмотреть возможность реализации параллельного алгоритма, абстрагируясь от разделения расчетов по данным, а сконцентрировавшись на разделении вычислений между адаптерами на основе выполняемых ими задач.

#### В. Анализ рендер-графов и выбор стратегии

Для того чтобы определить наиболее подходящую задачу для распределения между графическими адаптерами, мы провели анализ рендер-графов современных игр и игровых движков [37–41] с использованием инструментов профилирования, таких как NVIDIA Nsight [42, 43].

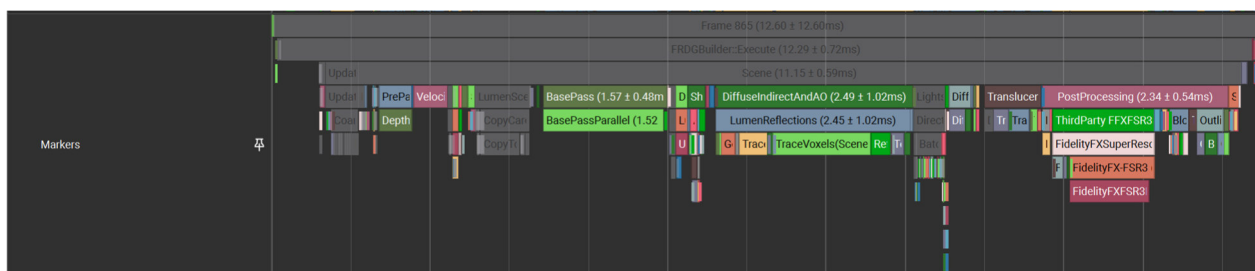


Рисунок 1. Nvidia Nsight профайлер

Исходя из проведенного анализа приложений и игр, мы выявили, что этап локального затенения (Ambient Occlusion, AO) – идеальный кандидат для переноса расчетов на дополнительную видеокарту.

Расчет реалистичного глобального освещения является сложной задачей в области компьютерной графики. Эта техника моделирует количество фонового освещения, достигшего точки на поверхности видимого объекта в зависимости от ее доступности [44], т.е. доли направлений, с которых свет может достичь этой точки, помогает пользователю лучше воспринять геометрию сцены. Такая аппроксимация не является точной с точки зрения физики, однако вычисления проводятся сравнительно быстро и дают убедительный результат на практике, что и привело к широкому распространению этого подхода [45].

В большинстве реализаций для реального времени применяют локальное затенение в экранном пространстве [46], которое использует буфер глубины и дополнительные данные в качестве исходных. Подобное решение снижает вычислительную сложность и увеличивает производительность, однако вычисления осуществляются на одном графическом адаптере, что накладывает ограничение на другие этапы в конвейере отрисовки, поскольку существует максимальный предел во времени построения кадра для систем реального времени.



Рисунок 2. Пример влияния окружающего затенения на поверхность

Результаты расчета АО представляют собой текстуру, размер которой заранее известен и пропорционален размеру экрана, что значительно меньше объемов данных, необходимых для отрисовки сцены. АО может быть рассчитано с небольшой задержкой относительно основного конвейера отрисовки, выполняемого на основной видеокарте, что позволяет смягчить проблемы, связанные с задержкой коммуникаций между видеоадаптерами.

Популярные подходы к параллельному рендерингу, такие как SFR, распределяют пространственные области кадра между GPU, что требует сложной синхронизации и приводит к избыточным вычислениям [15, 16]. На основе проведенного анализа мы разработали алгоритм, который эффективно использует комбинацию dGPU и iGPU в популярных пользовательских системах. В отличие от существующих решений, которые стремятся распределить всю нагрузку рендеринга между адаптерами, наш подход фокусируется на выносе конкретной, вычислительно емкой задачи – расчета Ambient Occlusion – на дополнительную видеокарту.

### Выбор техники АО

Алгоритм построения локального затенения довольно прост: для каждого пикселя, отображаемого пользователю, рассчитывается коэффициент затенения (occlusion factor) на основе значений из буфера глубины.

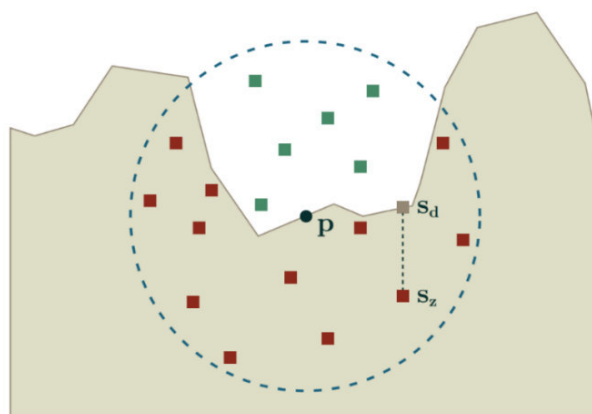


Рисунок 3. АО на основе буфера глубины



Вычисленный коэффициент затенения используется для уменьшения интенсивности фонового освещения, вплоть до полного его исключения. Для расчета этого коэффициента требуется сбор данных о глубине от множества выборок из сферической области, окружающей рассматриваемый фрагмент, и сравнения полученных значений глубины с глубиной рассматриваемого фрагмента [46, 47]. Число выборок, у которых значение глубины больше, нежели глубина непосредственно текущего фрагмента, определяют коэффициент затенения. Пример можно наблюдать на рисунке 3.

На нем изображено следующее:  $p$  – затеняемая точка,  $s_z$  – координата сэмпла, а  $s_d = \text{depth}(s_x, s_y)$  – значение буфера глубины в соответствующем пикселе, где  $s_x, s_y$  координаты пикселя. Здесь также легко заметить, что каждая красная точка лежит внутри некоторого геометрического объекта, а потому осуществляет вклад в значение коэффициента затенения. Чем больше выборок окажется внутри геометрии окружающих объектов, тем меньше будет остаточная интенсивность фонового затенения в этой области.

При разработке нашего подхода к распределённому рендерингу между двумя видеоадаптерами мы тщательно оценили различные алгоритмы Ambient Occlusion (AO) и выбрали именно Screen Space Ambient Occlusion (SSAO) [46] и Horizon-Based Ambient Occlusion (HBAO) [48–51] для реализации, несмотря на существование более современных вариантов, таких как Ground-Truth AO (GTAO) [52–54] и Ray-Traced AO (RTAO) [55, 56].

Наш выбор обусловлен несколькими причинами. Во-первых, эффективное распределение вычислений между GPU критически зависит от баланса нагрузки и соответствия вычислительных требований возможностям каждого устройства: интегрированные видеокарты, особенно в ноутбуках, имеют значительно меньшую вычислительную мощность по сравнению с дискретными. Во-вторых, для нашего подхода критически важен объём передаваемых и обновляемых каждый кадр данных на каждом устройстве, что накладывает ограничения на количество пересылаемых данных, а HBAO+ и GTAO используют дополнительные структуры данных, которые требуют расчетов и пересылки на дополнительную видеокарту [48, 49]. RTAO также исключается, поскольку для этой техники необходимо обновлять геометрию сцены в случае использования динамических объектов, которые распространены в современных компьютерных играх. В-третьих, учет аппаратной совместимости графических техник. В-четвертых, наш фокус направлен на принципиальную возможность осуществления и эффективности предложенного решения, поэтому выбор более простых и самодостаточных алгоритмов позволит оценить именно накладные расходы предложенного нами способа разделения, а не оценивать сложность самого алгоритма.

Именно поэтому SSAO и HBAO, с их понятной и легко реализуемой математикой, являются идеальными кандидатами для демонстрации разработанного нами механизма разделения вычислений в системах с несколькими видеокартами. Оптимизация и внедрение более сложных и продвинутых алгоритмов является следующим шагом по развитию предложенного нами способа взаимодействия видеокарт.

### Разделение вычислений АО

С появлением DirectX 12 Microsoft представила три режима для взаимодействия с видеокартами в системе. В самом простом случае используется упомянутая выше технология AFR с видеокартами одного производителя – AMD или NVIDIA [5, 7, 8, 57].

Но, как мы выяснили из обзора существующих подходов, данный режим ограничивает варианты использования видеокарт в системе, хоть и уменьшает этим вероятность потенциального возникновения ошибок. В этом режиме большая часть работы для поддержки выполняется на уровне драйвера, а не на уровне DirectX. Однако DirectX 12 обеспечивает более расширенный доступ к аппаратной части системы. Для этого существует режим Explicit Multi-Adapter (EMA) [57]. Для каждого одиночного GPU определяется доступ к памяти, описывается взаимодействие GPU между собой – вся эта поддержка должна быть запрограммирована заранее. В EMA доступны два разных варианта использования: Linked Mode и Unlinked Mode.

Режим Linked Mode можно рассматривать как форму SLI или Crossfire под DirectX 12: аппаратные ресурсы в Linked Mode комбинируются в один большой виртуальный адаптер с общей памятью и

пользователь работает с ним как с единой видеокартой в системе. Однако комбинация оборудования в данном режиме ограничена из-за аппаратных особенностей, например, оба физических видеоадаптера должны быть от производителя NVIDIA.

В режиме Unlinked Mode становится возможно комбинировать видеокарты в том числе и от разных производителей. Также появляется возможность комбинирования дискретного и интегрированного GPU. В этом режиме каждый графический процессор считается самостоятельной аппаратной единицей со своими аппаратными ресурсами. Демонстрацию того, как работает данный режим можно наблюдать на рисунке 4.

### Explicit Multiadapter Cross-adapter memory

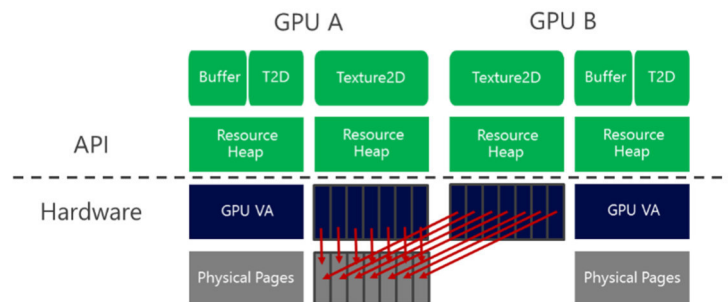


Рисунок 4. Механизм работы разделенных между адаптерами ресурсов [57]

В этом режиме ЕМА позволяет обмениваться данными между видеокартами, причем на глубоком уровне, а не просто готовыми кадрами, что является ключевой особенностью, которую не используют современные AFR алгоритмы [6, 15, 16, 28, 34, 58]. Обмен возможен как частично просчитанными кадрами, так и данными в буферах, что позволяет разрабатывать новые механизмы взаимодействия на нескольких GPU. Но присутствует также дополнительное ограничение в виде использования PCI Express шины, обмен данными через которую намного медленнее, нежели с применением аппаратных интерфейсов [5]. Пример того, как осуществляется взаимодействие между видеокартами, представлен на рисунке 5.

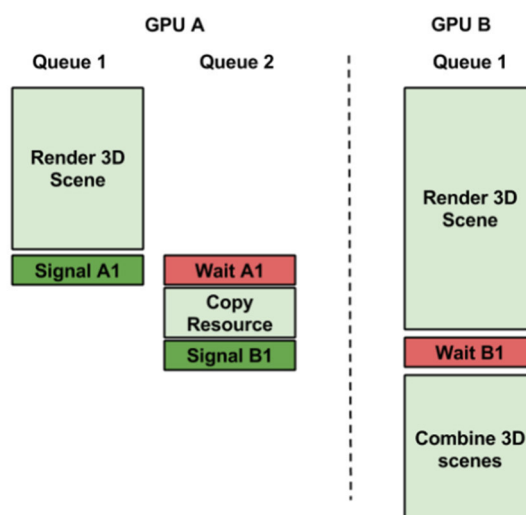


Рисунок 5. Механизм взаимодействия видеокарт в режиме ЕМА

Для реализации гибридного расчета АО мы разработали графический бенчмарк и изменили конвейер рендеринга, а именно разделили расчет локального затенения на две составляющие: подготовка данных для алгоритма и непосредственно сам алгоритм расчета затенения. Подготовкой

данных будет заниматься основная видеокарта в системе, так как эти данные необходимы будут не только АО, но и другим этапам в конвейере. Общую схему иллюстрирует рисунок 6.

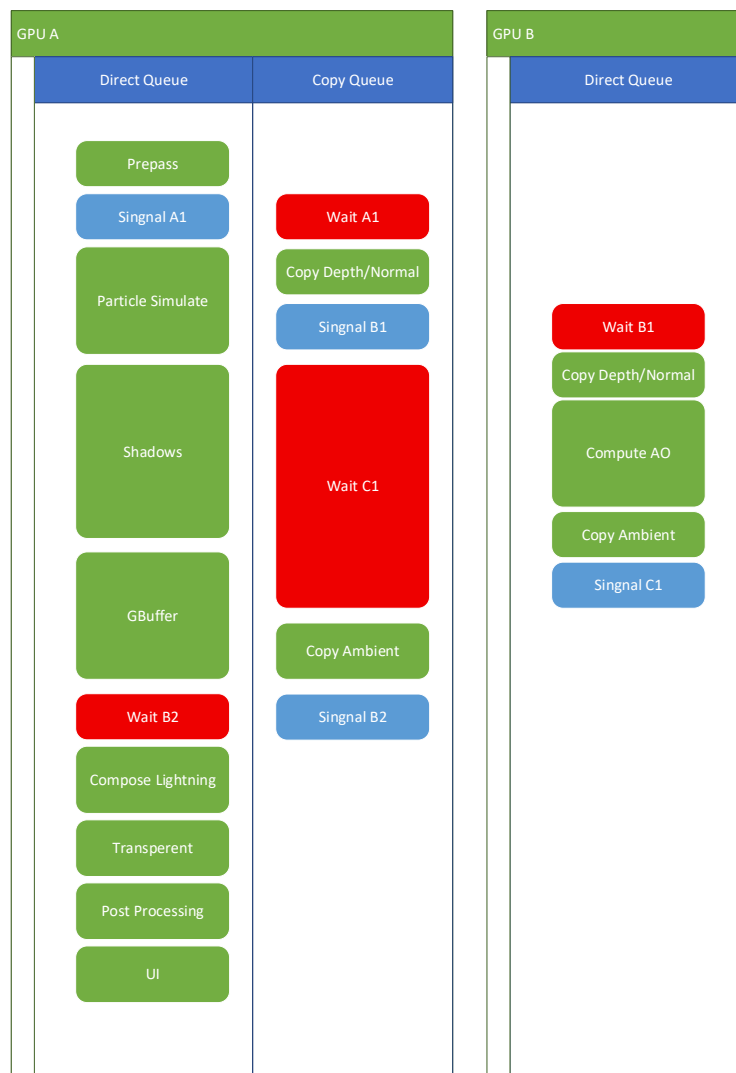


Рисунок 6. Диаграмма выполнения вычислений между видеоадаптерами в рамках одного кадра

После предварительного расчета глубины сцены и нормалей осуществляется копирование полученных данных в общую для устройств память. Перед моментом расчета освещения добавляется копирование рассчитанной текстуры с данными о локальном затенении из общей памяти видеокарт, а затем продолжается вычисление кадра. Параллельно этому дополнительная видеокарта копирует в свою память данные для расчета локального затенения, считает окклюзию и копирует результат обратно в общую память.

### Сравнительный анализ

Нагрузочная сцена состоит из вращающейся камеры, направленной в центр сцены, на которой происходит симуляция 200000 частиц и расположены статические объекты – их суммарное количество вертексов составляет примерно 2 миллиона. На сцене присутствует пять источников освещения, включая направленный источник освещения, эмулирующий солнце. Также каждый кадр пересчитывается карта теней.

Для оценки эффективности разработанного алгоритма мы производили тестирование на следующих системах: Intel Core I7 3770 со связкой Nvidia GTX 660 и AMD RX 470, а также Intel Core i7 9900K с интегрированным видеоядром Intel UHD Graphics 630 в связке с Nvidia GTX 1050Ti. Все тесты производились на системе Windows 10 в разрешении Full HD (1920x1080). Каждый запуск первые 30 секунд происходило прогревание кэшей, после чего в течение 120 секунд каждый кадр собиралась

статистика по времени построения кадра на каждом из адаптеров. На изображениях с 7 по 10 отображено время построения кадра, где по вертикальной оси указываются миллисекунды, рассчитанные как средние для 10 запусков бенчмарка, а по горизонтальной – время.

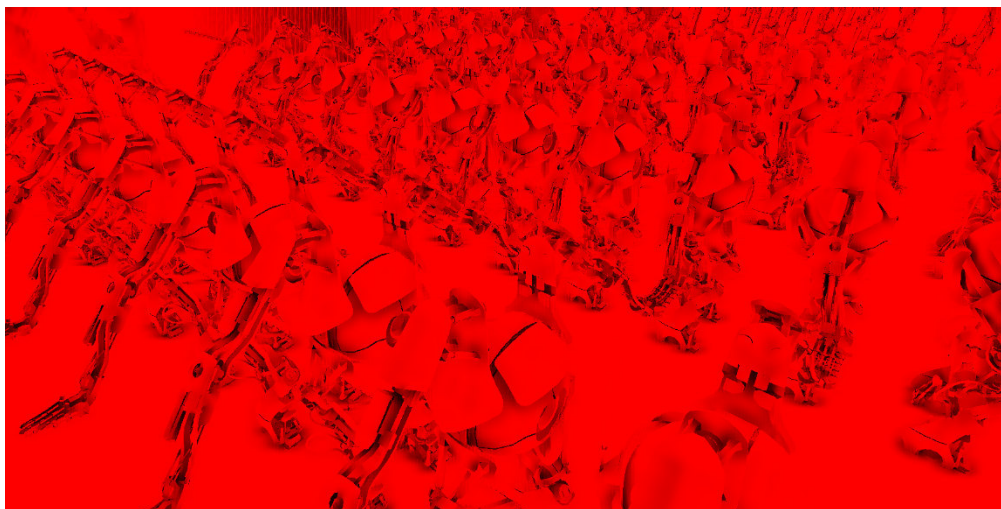


Рисунок 7. Демонстрация рассчитанного АО

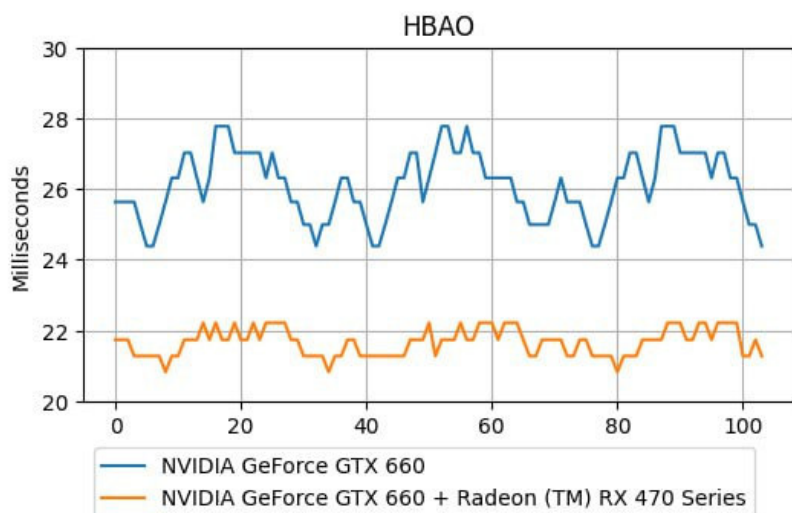


Рисунок 8. Рендер кадра при Native и Shared реализации HBAO на NVIDIA GTX 660 + AMD RX 470



Рисунок 9. Рендер кадра при Native и Shared реализации SSAO на NVIDIA GTX 660 + AMD RX 470

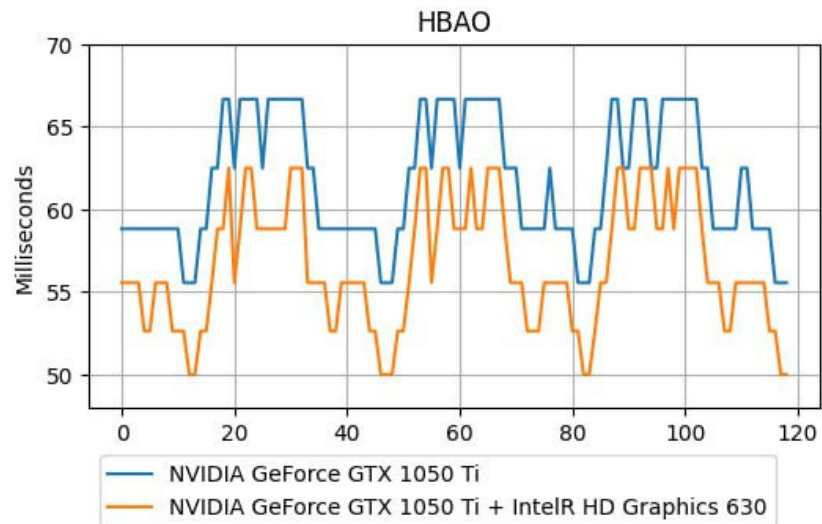


Рисунок 10. Рендер кадра при Native и Shared реализации HBAO на NVIDIA GTX 1050 Ti + Intel HD 630

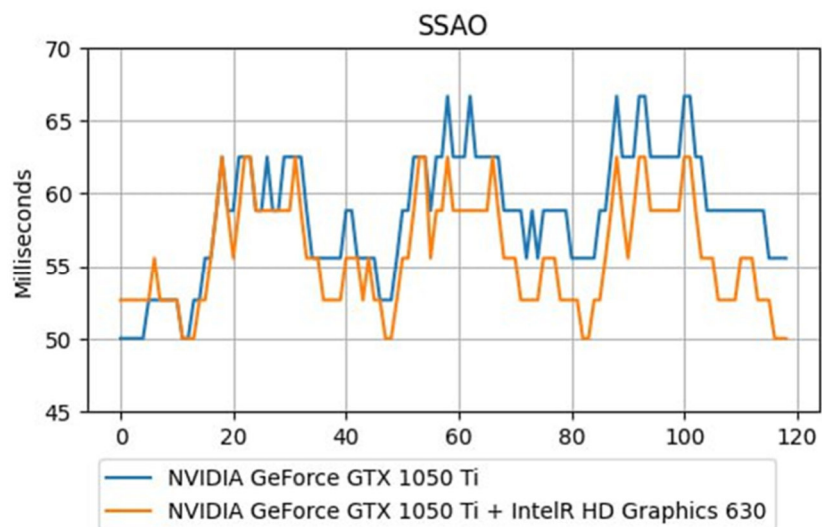


Рисунок 11. Рендер кадра при Native и Shared реализации SSAO на NVIDIA GTX 1050 Ti + Intel HD 630

Из анализа полученных данных видно, что гибридные реализации, вне зависимости от используемых устройств, показывают прирост производительности в разделенном режиме. В среднем копирование между адаптерами в зависимости от конфигурации занимало от 0.3 до 1 миллисекунды, что существенно быстрее, чем расчет локального затенения на основной видеокарте. Для сравнения: на конфигурации Nvidia GTX 1050 Ti и Intel HD Graphics 630 основная видеокарта 1050 Ti рассчитывала карту локального затенения в среднем 3 мс, что в 3 раза медленнее копирования данных между адаптерами. Для оценки загруженности нами были предприняты попытки использования сторонних профилировщиков, таких как Nvidia Nsight и AMD GPU Profiler, но указанные профилировщики не поддерживают работу с приложениями DirectX 12 в режиме EMA с включенным Unlinked Mode. Загруженность определялась по встроенному в Windows профилировщику видеокарт.

Согласно нашим наблюдениям, производительность разработанного нами подхода к разделению зависит от используемой пары видеоадаптеров. Так, при включении разделённого режима загруженность видеокарт в связке Nvidia GTX 1050 Ti и Intel HD Graphics 630 была 83 и 95 % соответственно, согласно встроенному в Windows профилировщику. Однако при использовании связки Nvidia GTX 660 и AMD RX 470 загруженность видеокарт в разделенном режиме была 80 и 60 % соответственно. Разница в загруженности показывает, что эффективность разделения техники зависит не только от выбранного алгоритма, но и от разницы в производительности связки видеокарт, что необходимо учитывать в дальнейших исследованиях.

В ходе дальнейших исследований мы планируем оценить эффективность переноса других популярных графических техник и возможность комбинированного запуска выявленных техник на дополнительном видеоадаптере в гибридных системах.

## Выводы

Наша гибридная реализация окружающего затенения с копированием данных между устройствами позволяет сократить время рендеринга приблизительно на 20 % по сравнению с использованием одной GPU. Гибридный подход может улучшить пользовательский опыт из-за большей частоты кадров, а также позволить использовать усложненные вычислительные техники из-за сэкономленного времени при построении кадра на основной видеокарте.

## Список литературы

1. Luebke D., Humphreys G. How GPUs work // Computer (Long Beach Calif). 2007. Vol. 40, no. 2.
2. NVIDIA 2023 Annual Review. URL: [https://s201.q4cdn.com/141608511/files/doc\\_financials/2023/ar/2023-Annual-Report-1.pdf](https://s201.q4cdn.com/141608511/files/doc_financials/2023/ar/2023-Annual-Report-1.pdf) (accessed: 21.08.2024).
3. Nanite as a Disruptive Technology for the Interactive Visualisation of Cultural Heritage 3D Models: A Case Study / Díaz-Alemán M.D., et al. // Heritage. 2023. Vol. 6, no. 8.
4. Scaling Deep Learning workloads: NVIDIA DGX-1/Pascal and Intel Knights Landing / Gawande N.A., et al. // Future Generation Computer Systems. 2020. Vol. 108.
5. Evaluating Modern GPU Interconnect: PCIe, NVLink, NV-SLI, NVSwitch and GPUDirect / Li A., et al. // IEEE Transactions on Parallel and Distributed Systems. 2020. Vol. 31, no. 1.
6. MCM-GPU: Multi-chip-module GPUs for continued performance scalability / Arunkumar A., et al. // Proceedings - International Symposium on Computer Architecture. 2017. Vol. Part F128643.
7. NVIDIA Corporation. Introduction to SLI Technology. 2017. URL: <https://www.geforce.com/whats-new/guides/introduction-to-sli-technology-guide#2> (accessed: 07.01.2025).
8. Advanced Micro Devices. AMD Crossfire Technology. 2017. URL: <https://www.amd.com/ru/technologies/crossfire> (accessed: 07.01.2025).
9. How to Fix Stuttering of Crossfire? URL: [https://www.reddit.com/r/crossfire/comments/ddlcl9/how\\_to\\_fix\\_stuttering\\_of\\_crossfire/](https://www.reddit.com/r/crossfire/comments/ddlcl9/how_to_fix_stuttering_of_crossfire/) (accessed: 23.03.2024).
10. Eilemann S., Makhinya M., Pajarola R. Equalizer: A scalable parallel rendering framework // IEEE Trans Vis Comput Graph. 2009. Vol. 15, no. 3.
11. Eilemann S., Steiner D., Pajarola R. Equalizer 2.0-convergence of a parallel rendering framework // IEEE Trans Vis Comput Graph. 2020. Vol. 26, no. 2.
12. Chromium: A stream-processing framework for interactive rendering on clusters / Humphreys G., et al. // Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02. 2002.
13. Mueller C. Sort-first rendering architecture for high-performance graphics // Proceedings of the Symposium on Interactive 3D Graphics. 1995.
14. Steam Hardware Stats. URL: <https://store.steampowered.com/hwsurvey> (accessed: 14.12.2024).
15. Ren X., Lis M. CHOPIN: Scalable Graphics Rendering in Multi-GPU Systems via Parallel Image Composition // 2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA). IEEE, 2021. P. 709–722.
16. GPUd: A fast and scalable multi-GPU architecture using cooperative projection and distribution / Kim Y., et al. // Proceedings of the Annual International Symposium on Microarchitecture, MICRO. IEEE Computer Society, 2017. Part F131207. P. 574–586.
17. Eilemann S., Steiner D., Pajarola R. Equalizer 2.0-convergence of a parallel rendering framework // IEEE Trans Vis Comput Graph. IEEE Computer Society, 2020. Vol. 26, no. 2. P. 1292–1307.
18. Molnar S., Eyles J., Poulton J. PixelFlow: high-speed rendering using image composition // Computer Graphics (ACM). 1992. Vol. 26, no. 2.
19. Moll L., Heirich A., Shand M. Sepia: Scalable 3D compositing using PCI Pamette // IEEE Symposium on FPGAs for Custom Computing Machines, Proceedings. 1999.
20. Laine S., Karras T. High-performance software rasterization on GPUs // Proceedings - HPG 2011: ACM SIGGRAPH Symposium on High Performance Graphics. 2011.
21. WireGL: A scalable graphics system for clusters / Humphreys G., et al. // Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001. 2001.
22. Hsu W.M. Segmented ray casting for data parallel volume rendering // Proceedings of the 1993 Parallel Rendering Symposium. 1993.
23. Eldridge M., Igehy H., Hanrahan P. Pomegranate: A fully scalable graphics architecture // Proceedings of the ACM SIGGRAPH Conference on Computer Graphics. 2000.
24. ATTILA: A cycle-level execution-driven simulator for modern GPU architectures / Del Barrio V.M., et al. // ISPASS 2006: IEEE International Symposium on Performance Analysis of Systems and Software, 2006. Vol. 2006.



25. Griffin: Hardware-software support for efficient page migration in multi-GPU systems / Baruah T., et al. // Proceedings - 2020 IEEE International Symposium on High Performance Computer Architecture, HPCA 2020. 2020.
26. Arnau J.M., Parcerisa J.M., Xekalakis P. Eliminating redundant fragment shader executions on a mobile GPU via hardware memoization // Proceedings - International Symposium on Computer Architecture. 2014.
27. Tauray: A Scalable Real-Time Open-Source Path Tracer for Stereo and Light Field Displays / Ikkala J., et al. // Proceedings - SIGGRAPH Asia 2022: Technical Communications. 2022.
28. Granja P., Pereira J. Hybrid-Rendering Techniques in GPU. 2023.
29. GPU Accelerated Path Tracing of Massive Scenes / Jaros M., et al. // ACM Trans Graph. 2021. Vol. 40, no. 2.
30. MDScale: Scalable multi-GPU bonded and short-range molecular dynamics / Barreales G.N., et al. // J Parallel Distrib Comput. 2021. Vol. 157. P. 243–255.
31. MIGPerf: A Comprehensive Benchmark for Deep Learning Training and Inference Workloads on Multi-Instance GPUs / Zhang H., et al. 2023.
32. Balin M.F., Sancak K., Çatalyürek Ü. V. MG-GCN: Scalable Multi-GPU GCN Training Framework. 2021.
33. Dong Y., Peng C. Multi-GPU multi-display rendering of extremely large 3D environments // Vis Comput. 2023. Vol. 39, no. 12. P. 6473–6489.
34. rCUDA: Reducing the number of GPU-based accelerators in high performance clusters / Duato J., et al. // Proceedings of the 2010 International Conference on High Performance Computing and Simulation, HPCS 2010. 2010. P. 224–231.
35. Beyond the socket: NUMA-aware GPUs / Milic U., et al. // Proceedings of the Annual International Symposium on Microarchitecture, MICRO. 2017. Part F131207.
36. Kirk D. NVIDIA CUDA software and GPU parallel computing architecture // International Symposium on Memory Management, ISMM. 2007.
37. O'Donnell Y. FrameGraph: Extensible Rendering Architecture in Frostbite // Gdc 2017. 2017. Vol. 1999, no. December.
38. Electronic Arts. Halcyon: Rapid Innovation using Modern Graphics. 2019. URL: <https://media.contentapi.ea.com/content/dam/ea/seed/presentations/wihlidal2019-rebootdevelopblue-halcyon-rapid-innovation.pdf> (accessed: 23.08.2025).
39. Unity. HDRP Render Graph. URL: <https://docs.unity3d.com/Packages/com.unity.render-pipelines.high-definition@12.0/manual/Custom-Post-Process.html> (accessed: 14.03.2023).
40. Epic Games. Unreal Rednering Pipeline. URL: <https://dev.epicgames.com/community/learning/tutorials/lyJK/unreal-engine-rendering-pipeline-ue5> (accessed: 23.07.2025).
41. Epic Games. A first look at Unreal Engine 5. 2020. URL: <https://www.unrealengine.com/en-US/blog/a-first-look-at-unreal-engine-5> (accessed: 23.08.2024).
42. Installing NVIDIA Nsight into Visual Studio // Accelerating Matlab with GPUs. 2014.
43. Iyer K., Kiel J. GPU debugging and profiling with NVIDIA parallel nsight // Game Development Tools. 2016.
44. Zhukov S., Iones A., Kronin G. An ambient light illumination model. 1998.
45. Landis H. Production-Ready Global Illumination. 2004.
46. The Alchemy screen-space ambient obscurance algorithm / McGuire M., et al. // Proceedings - HPG 2011: ACM SIGGRAPH Symposium on High Performance Graphics. 2011.
47. Bavoil L., Sainz M. Screen space ambient occlusion // NVIDIA developer information: .... 2008. no. September.
48. Bavoil L., Sainz M., Dimitrov R. Image-space horizon-based ambient occlusion // SIGGRAPH'08: ACM SIGGRAPH Talks 2008. 2008.
49. Song Y., Liu F., Xu J. Horizon-based screen-space ambient occlusion using mixture sampling // ACM SIGGRAPH ASIA 2010 Posters, SA'10. 2010.
50. McGuire M., Mara M., Luebke D. Scalable ambient obscurance // High-Performance Graphics 2012, HPG 2012 - ACM SIGGRAPH: Eurographics Symposium Proceedings. 2012.
51. Mara M., McGuire M., Luebke D. Lighting Deep G-Buffers: Single-Pass, Layered Depth Images with Minimum Separation Applied to Indirect Illumination // Graphics.Cs.Williams.Edu. 2013.
52. Jimenez J. Practical Realtime Strategies for Accurate Indirect Occlusion // Siggraph 2016. 2016.
53. Vardis K., Papaioannou G., Gaitatzes A. Multi-view ambient occlusion with importance sampling // Proceedings of the Symposium on Interactive 3D Graphics. 2013.
54. Vermeer J., Scandolo L., Eisemann E. Stochastic-Depth Ambient Occlusion // Proceedings of the ACM on Computer Graphics and Interactive Techniques. 2021. Vol. 4, no. 1.
55. Gautron P. Real-time ray-traced ambient occlusion of complex scenes using spatial hashing // Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks, SIGGRAPH 2020. 2020.
56. Multi-GPU room response simulation with hardware raytracing / Thoman P., et al. // Concurr Comput. 2022. Vol. 34, no. 4.
57. Microsoft Corporation. Multi-Adapter. 2017. URL: <https://devblogs.microsoft.com/directx/directx-12-multiadapter-lighting-up-dormant-silicon-and-making-it-work-for-you/>. (accessed: 07.01.2025).
58. Marchesin S., Mongenet C., Dischler J.-M. Multi-GPU Sort-Last Volume Visualization. 2008. 1–8 p.

## Разработка и анализ метода отражений в экранном пространстве для систем с дискретным и интегрированным GPU

Т. Р. Галеев, М. К. Богданов

Университет ИТМО, Санкт-Петербург, Россия

**Аннотация.** В работе рассмотрены возможности распределения вычислительной нагрузки при рендеринге отражений в экранном пространстве (Screen Space Reflections, SSR) между дискретным и интегрированным GPU. Проведен обзор основных подходов к мульти-GPU рендерингу (Alternate Frame Rendering, Split Frame Rendering, Hybrid Multi-GPU Rendering) и методов построения отражений (геометрические техники, environment-mapping, SSR, трассировка лучей). На основании анализа выбран гибридный подход, при котором дискретная видеокарта отвечает за основной рендеринг сцены, а интегрированная – за вычисление отражений методом SSR. Описана архитектура графического приложения, реализующего предложенный подход, и проведены рекомендации по оптимизации параметров алгоритма для минимизации задержек и повышения производительности.

**Ключевые слова:** мульти-GPU, интегрированный GPU, дискретный GPU, экранные отражения, Screen Space Reflections, cross-adapter ресурсы, синхронизация fence, гибридный рендеринг

## Development and analysis of a screen space reflection method for systems with discrete and integrated GPUs

T. R. Galeev, M. K. Bogdanov

ITMO University, St. Petersburg, Russia

**Abstract.** This work investigates the distribution of computational load for Screen Space Reflections (SSR) between a discrete and an integrated GPU. It provides a survey of multi-GPU rendering techniques (Alternate Frame Rendering, Split Frame Rendering, Hybrid Multi-GPU Rendering) and reflection methods (geometry-based techniques, environment mapping, SSR, real-time ray tracing). Based on the analysis, a hybrid approach is adopted where the discrete GPU performs the main scene rendering, and the integrated GPU computes SSR reflections. The architecture of a graphical application implementing the proposed approach is described, and recommendations for optimizing the algorithm parameters to minimize delays and improve performance are made.

**Keywords:** multi-GPU, integrated GPU, discrete GPU, screen space reflections, SSR, cross-adapter resources, fence synchronization, hybrid rendering

### Введение

Современным видеоиграм предъявляют высокие требования к качеству визуализации. Одним из главных критериев хорошей 3D-графики является наличие отражений окружающих объектов на поверхностях, которое добавляет визуализируемой 3D-сцене реалистичности. Среди различных методов реализации отражений особое место занимает метод отражений в экранном пространстве (Screen Space Reflections, SSR), который позволяет генерировать отражения в реальном времени, основываясь на информации, доступной в буфере изображения, что делает его одним из самых быстрых методов построения отражений. Однако рендеринг сцены с отражениями, даже при помощи SSR, может существенно повлиять на производительность в условиях использования одного графического процессора.

Современные пользовательские ЭВМ часто имеют как дискретные GPU, обеспечивающие высокую вычислительную мощность, так и интегрированные GPU, встроенные в центральный процессор или материнскую плату и обеспечивающие базовый уровень графики. Эти два типа GPU значительно различаются по архитектуре, производительности и объему памяти. В частности, интегрированные GPU могут испытывать трудности с обработкой сложных эффектов, в то время как дискретные GPU предлагают значительно больше возможностей для их реализации. По этой причине вся вычислительная нагрузка во время 3D-рендеринга приходится на дискретную видеокарту, однако в условиях использования устаревшего оборудования добиться качественных изображений и одновременно высокой производительности системы крайне проблематично, в то время как интегрированный GPU практически не используется. Это создает проблему распределения вычислительной нагрузки между дискретным и интегрированным GPU для обеспечения высокой производительности рендера в реальном времени.

Развитие современных графических API, таких как DirectX и Vulkan [1], предоставило новые возможности для одновременного управления несколькими GPU [2], что открывает перспективы повышения производительности графической визуализации в контексте игровых приложений путём синхронной работы интегрированной и дискретной видеокарт.

### Постановка задачи

Целью данной работы является исследование перспективы разделения вычислительной нагрузки рендеринга 3D-сцены между дискретной и интегрированной видеокартами для повышения производительности визуализации в реальном времени, а также разработка графического приложения, использующего дискретную видеокарту для основного рендеринга 3D-сцены и интегрированную видеокарту для рендеринга отражений с помощью метода SSR.

Для достижения данной цели были поставлены следующие задачи:

- 1) исследование подходов Multi-GPU рендеринга для синхронной работы дискретных и интегрированных видеокарт;
- 2) исследование методов построения отражений;
- 3) программная реализация функционала графического приложения для визуализации 3D-сцены с использованием дискретных и интегрированных видеокарт для исследования производительности выбранного подхода;
- 4) сравнительный анализ полученных результатов и оценка применения выбранного подхода в игровых приложениях.

### Теория

#### 1. Методы Multi-GPU рендеринга

Современные алгоритмы взаимодействия между несколькими графическими процессорами позволяют распределять вычислительную нагрузку между несколькими GPU для увеличения производительности графической визуализации [2]. Рассмотрим три ключевых подхода: Alternate Frame Rendering, Split Frame Rendering и Hybrid Multi-GPU Rendering.

Alternate Frame Rendering (AFR) – это метод рендеринга для систем с несколькими графическими процессорами, при котором каждый GPU обрабатывает полный кадр, но выполняет это попеременно [2]. Например, один GPU рендерит нечётные кадры, а другой – чётные [3]. В то время как один GPU завершает рендеринг кадра, другой начинает работу над следующим кадром, что обеспечивает высокий уровень параллелизма (рис. 1) [4].

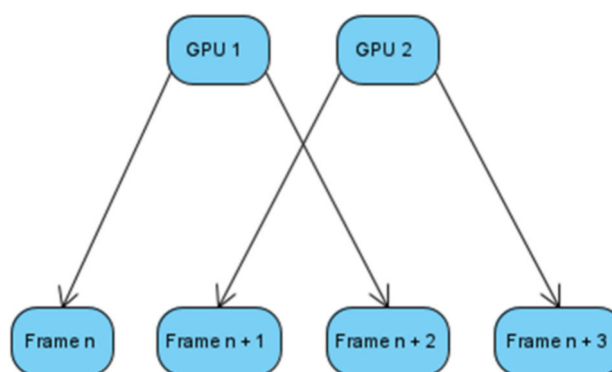


Рисунок 1. Работа алгоритма AFR

Split Frame Rendering (SFR) – это метод рендеринга для систем с несколькими графическими процессорами, при котором один кадр делится на несколько частей, каждая из которых обрабатывается отдельным GPU. В отличие от AFR, SFR распределяет рендеринг внутри одного кадра, а не между кадрами [5]. Ключевая идея SFR заключается в том, чтобы разделить экран на области (например, горизонтальные или вертикальные полосы) и распределить их между GPU [4]. Каждый GPU рендерит свою область параллельно, а затем результаты объединяются для вывода на экран (рис. 2).

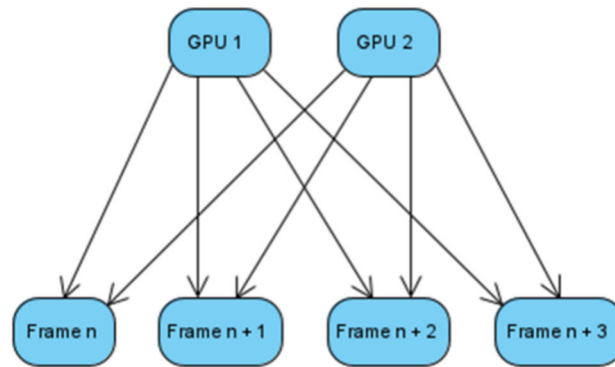


Рисунок 2. Работа алгоритма SFR

Основная идея Hybrid Multi-GPU Rendering заключается в том, чтобы использовать несколько графических процессоров, где каждый GPU выполняет задачи, наиболее подходящие для его архитектуры. Примером может быть сценарий, где один GPU выполняет рендеринг основной геометрии сцены, а другой GPU обрабатывает специфические задачи, такие как расчёт освещения, тени или отражения. Такой подход особенно актуален в системах с различными по мощности GPU, например, дискретными и интегрированными.

Пример данного алгоритма, представленный на рисунке 3, показывает, как можно выиграть время на обработку каждого кадра, разделив задачи рендеринга на две видеокарты, различающиеся по мощности. Основная часть всех графических задач выполняется на основной видеокарте. Она введет расчет различных текстур, таких как Карта нормалей и Карта локального затенения, которые необходимы для построения сцены. В это время дополнительное устройство занимается просчетом карты теней и копирует полученные данные в память ведущей видеокарты для последующего финального построения изображения сцены [6].

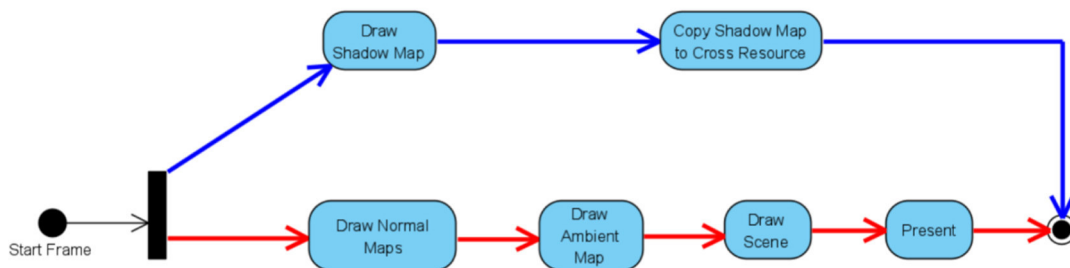


Рисунок 3. Алгоритм Shared Shadow Map

Проанализировав каждый подход Multi-GPU взаимодействия, можно сделать вывод о том, что AFR и SFR не подходят для реализации графического приложения в данной работе, так как не учитывают разницу в мощности между дискретным и интегрированным GPU. В AFR производительность будет ограничена слабой мощностью интегрированной видеокарты. А в SFR интегрированной видеокарте необходимо также пройти по всем ресурсам, что и основной видеокарте, чтобы сформировать свой кусок изображения, что значительно снизит производительность. Hybrid Multi-GPU Rendering является наиболее подходящим методом для реализации, так как учитывает малую мощность интегрированной GPU.

## 2. Техники построения отражений

Основные подходы делятся на три категории: геометрические, основанные на изображениях, и трассировка лучей.

Геометрические методы (Planar Surfaces, Curved Surfaces) используют трансформацию самой геометрии объектов, чтобы создать отражения [7]. Это значит, что создаются виртуальные объекты, которые размещаются так, чтобы имитировать отражение относительно отражающей поверхности. Подходы для плоских и криволинейных поверхностей различаются.

Для создания отражений на плоской поверхности (Planar Surfaces) требуется только одно линейное преобразование, поскольку нормаль к плоскости отражения остаётся неизменной. Для каждого объекта сцены создаётся виртуальная копия, которая отражается относительно плоскости (рис. 4). Это выполняется с помощью матрицы трансформации [7, 8].

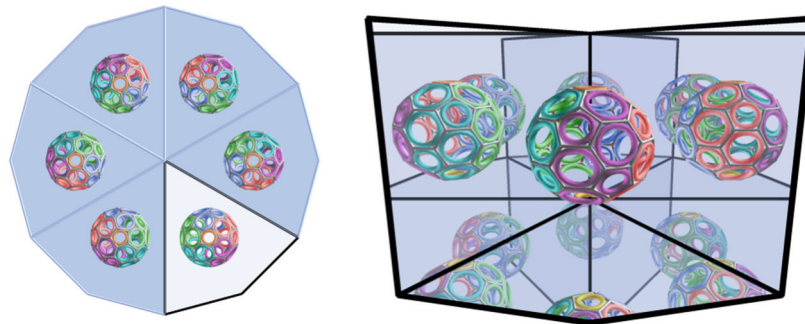


Рисунок 4. Пример построения отражений с помощью Planar Surfaces

Для создания отражений на кривой поверхности (Curved Surfaces) сложность возрастает, поскольку отражения зависят от точки обзора. Для расчета отражения каждого вершинного элемента применяются дополнительные алгоритмы. Для каждой вершины объекта находят пересечение луча зрения с поверхностью отражателя (рис. 5) [7, 9].

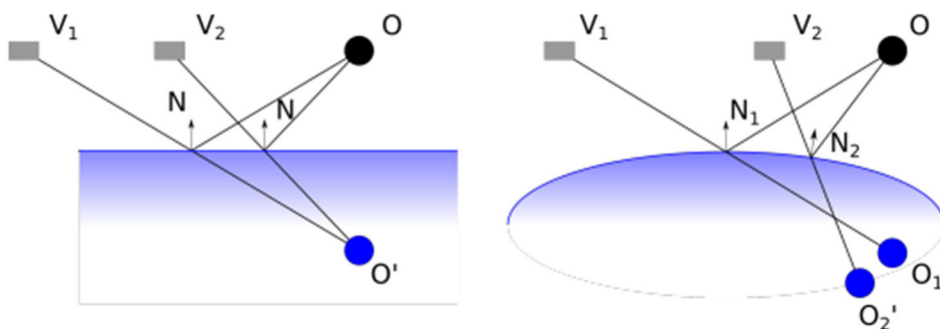


Рисунок 5. Пример построения отражений с помощью Curved Surfaces

Методы, основанные на изображениях (Cube Mapping, Sphere Mapping, SSR), используют текстуры (или уже отрендеренные изображения сцены) для симуляции отражений. Эти техники более просты и эффективны для отображения отражений на криволинейных поверхностях, где точное геометрическое моделирование слишком сложное.

Отражение окружающей среды предполагает, что сцена проецируется на текстуру, которая затем применяется к отражающим объектам. Отражение создаётся с использованием заранее подготовленной текстуры [7]. При построении отражений с помощью кубической карты (Cube Mapping) сцена визуализируется из центра отражающего объекта в шесть направлений (по осям X, Y, Z) для создания шести текстур, формирующих куб. Кубические текстуры используются для определения цвета отражения на основе направления отражённого луча. Этот метод популярен благодаря простоте реализации, а также высокой производительности (рис. 6) [7, 10]. При построении отражений с помощью сферической карты (Sphere Mapping) отражения кодируются в одной текстуре, проецируемой на сферу. Текстура затем преобразуется в плоскую 2D-карту. Преимущество данного метода в том, что требуется только одна текстура, а не шесть, как в кубическом методе. Однако значительная потеря точности на краях текстуры, а также неравномерность выборки пикселей (линейная интерполяция на сфере создаёт артефакты) [7].

Screen-Space Reflections (SSR) – это метод, который работает в пространстве экрана и использует уже отрендеренную сцену для создания отражений [11]. Сначала сцена рендерится в G-Buffer – структуру данных, содержащую глубину каждого пикселя, нормали поверхностей и положение объектов. Для каждого пикселя экрана рассчитывается отражённый луч на основе его нормали. Луч

отслеживается в пространстве экрана, где проверяется его пересечение с данными глубины в G-Buffer. Как только находится пересечение, цвет соответствующего пикселя используется для отображения отражения, как показано на рисунке 7 [12, 13].

Методы трассировки лучей (Real-Time Ray Tracing) являются наиболее точным способом создания отражений, так как они моделируют физическое поведение света (рис. 8). В процессе лучи отслеживаются от наблюдателя через отражающую поверхность к объектам, которые они пересекают [7].

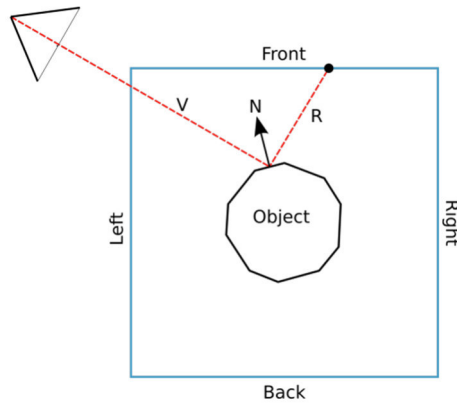


Рисунок 6. Пример построения отражений с помощью Cube Mapping

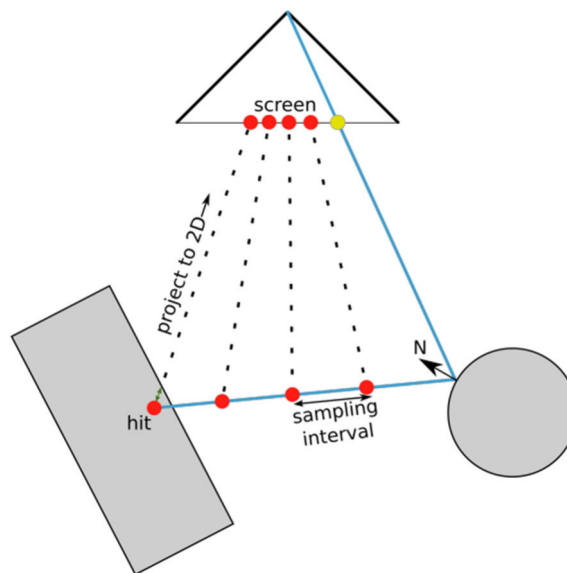


Рисунок 7. Алгоритм построения отражений с помощью SSR

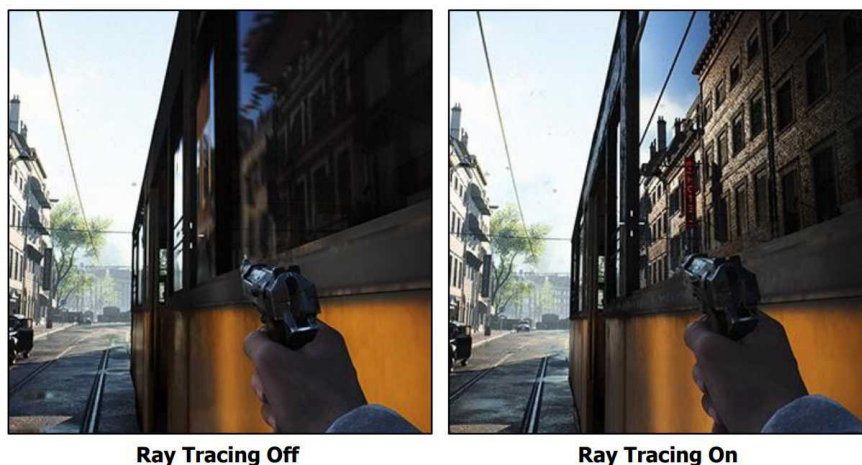


Рисунок 8. Пример работы Ray Tracing



Для каждого пикселя экрана рассчитывается луч, который начинает своё движение от точки обзора и пересекает сцену. Луч проверяется на пересечения с геометрическими объектами в сцене, чтобы определить ближайшую точку контакта. Если поверхность в этой точке зеркальная, рассчитывается направление отражённого луча, который отслеживается дальше. Процесс повторяется для заданного числа отражений, чтобы избежать бесконечных вычислений (рис. 9). Одно из преимуществ трассировки лучей – способность моделировать многократные отражения, т.е. многократное отражение лучей между поверхностями. Это автоматически учитывается при каждом отскоке луча.

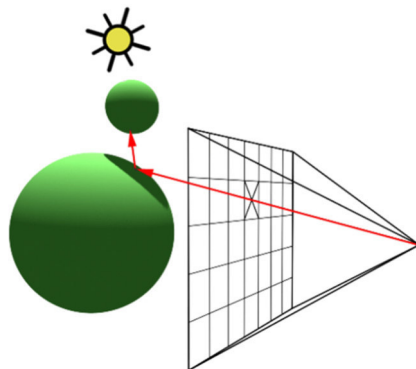


Рисунок 9. Пример построения отражений с помощью Ray Tracing

Проанализировав различные алгоритмы построения отражений с помощью дополнительной видеокарты, можно сделать вывод о том, что наиболее подходящим методом для реализации графического приложения с использованием гибридного Multi-GPU подхода является Screen Space Reflections, так как обеспечивает отражения динамических объектов в реальном времени, а также высокую производительность отрисовки.

Planar Surfaces не подходит, так как, несмотря на высокую производительность, работает только для плоских отражающих поверхностей. Curved Surfaces и Ray Tracing не подходят, так как имеют высокую вычислительную сложность, что не подходит для интегрированной GPU с маленькой вычислительной мощностью. Техники Environment Mapping, такие как Cube Mapping и Sphere Mapping, не подходят, так как отражения строятся для статических объектов.

### Предложенный метод

На основе выводов анализа алгоритмов Multi-GPU рендеринга в данной работе предлагается использование гибридного подхода, учитывающего малую мощность интегрированной GPU. Для начала необходимо разобрать текущий пайплайн графического приложения, представленный на рисунке 10.

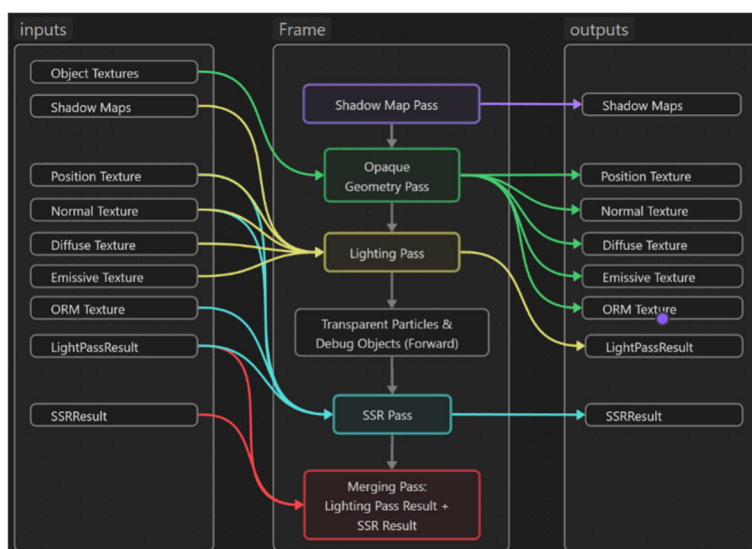


Рисунок 10. Рендер граф Single-GPU графического приложения

На данном графе можно увидеть, что этап вычисления отражений SSR Pass находится в конце графического пайплайна, так как требует G-Buffer текстуры позиций, нормалей и текстуру ORM (Occlusion, Roughness, Metalness), а также результат Lighting Pass в виде текстуры. Затем в Merging Pass две текстуры SSR Pass Result и Lighting Pass Result смешиваются воедино.

Для вычисления отражений на дополнительной видеокарте нужно подготовить и передать на неё необходимые текстуры GBuffer, а также текстуру Light Pass Result. Результат работы SSR Pass должен отправляться на основную видеокарту для дальнейшего смешивания в Merging Pass. Таким образом, рендер граф системы будет выглядеть, как на рисунке 11.

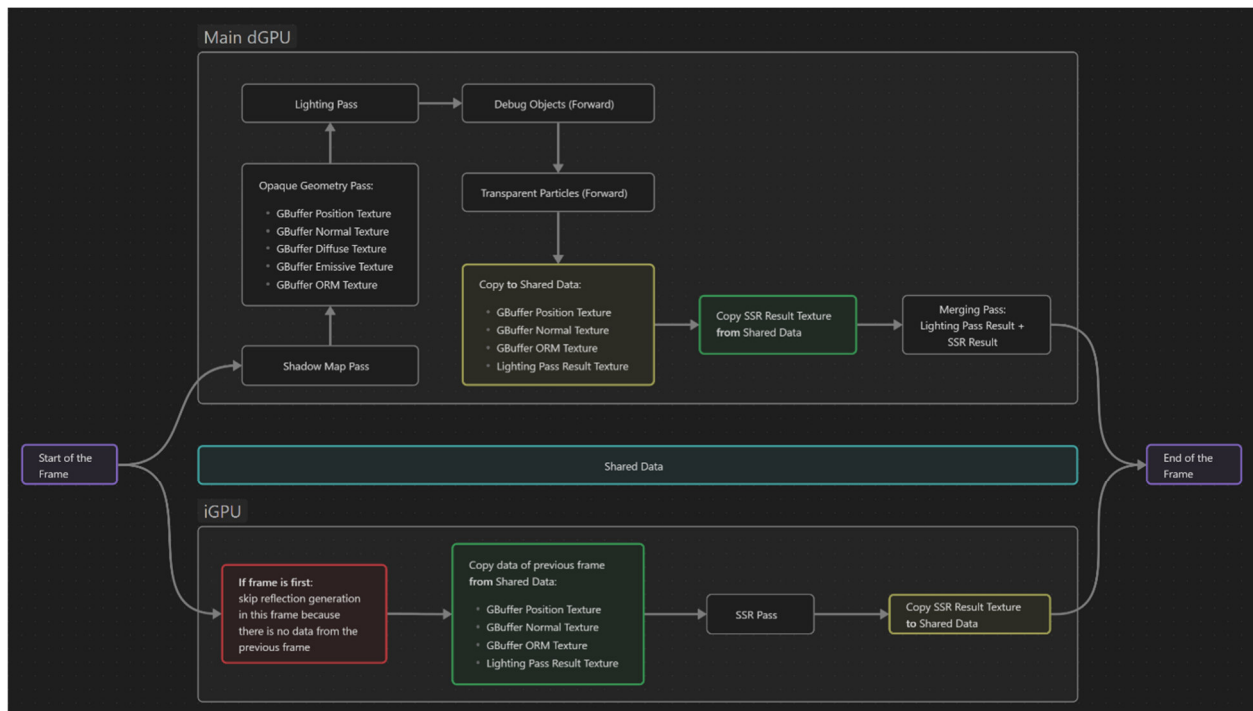


Рисунок 11. Рендер граф Multi-GPU графического приложения

Shared Data является «буферной зоной» для передачи данных между двумя видеокартами. Зеленым и желтым цветами выделены соответственно этапы загрузки ресурсов в Shared Data и выгрузки из неё. Генерация отражений в первом кадре пропускается, так как основная видеокарта ещё не рассчитала данные для G-Buffer и Lighting Pass Result, необходимые для SSR Pass. Когда результаты Geometry Pass и Lighting Pass из предыдущего кадра будут получены интегрированной видеокартой, она выполнит SSR Pass и скопирует результат в Shared Data. После этого в следующем кадре основная видеокарта скопирует SSR Pass Result, посчитанный в предыдущем кадре, из Shared Data.

### Результаты экспериментов

На основе полученного Multi-GPU пайплайна была проведена предварительная реализация графического приложения с использованием всех вышеперечисленных этапов (рис. 12).

Также был проведен предварительный эксперимент сравнения производительности подходов Single-GPU и Hybrid Multi-GPU в условиях сильной загруженности основной видеокарты. Результаты эксперимента представлены на рисунке 13.



Рисунок 12. Предварительная реализация графического приложения с отражениями

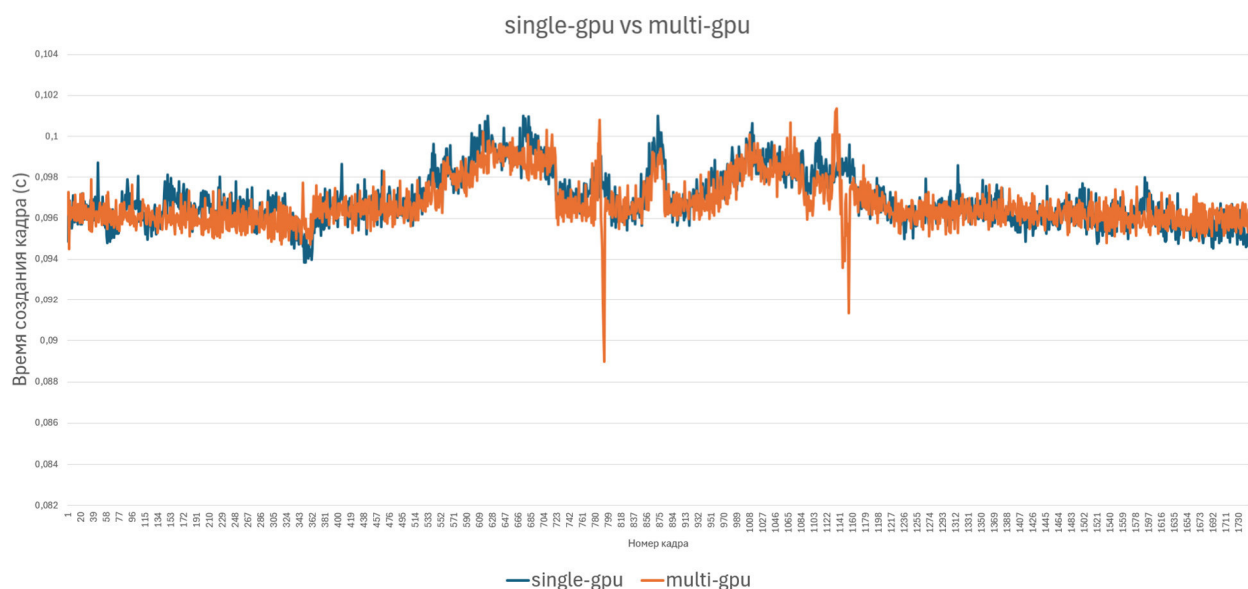


Рисунок 13. Сравнение производительности подходов Single-GPU (RTX 4070 Super) и Hybrid Multi-GPU (AMD iGPU + RTX 4070 Super)

### Обсуждение результатов

Предварительный эксперимент сравнения производительности подходов показал, что подход Hybrid Multi-GPU более производительный, чем Single-GPU. Однако такие результаты были получены в условиях сильной нагрузки на основную видеокарту. Даже при том, что при Hybrid Multi-GPU основная видеокарта не тратит время на вычисление отражений, она всё равно тратит какое-то время на копирование ресурсов в общую память и из неё. Если нагрузка на основную видеокарту слабая, а время на рендер отражений на интегрированной видеокарте и на копирование ресурсов между графическими устройствами превосходит время рендера на основной видеокарте, то целесообразнее использовать Single-GPU рендеринг на основной видеокарте.

Предполагается, что дальнейшее улучшение и оптимизация гибридного подхода multi-GPU взаимодействия минимизирует простой на дискретной и интегрированной видеокартах, а также

повысит производительность системы в условиях равного времени работы обеих видеокарт. Однако при таком подходе отражения 3D-сцены будут всегда отставать на два кадра от самой сцены, что некритично при большом количестве кадров в секунду. Также возможным решением для минимизации простоев является динамическая настройка параметров для SSR, таких как RayStepLength и MaxDistance, от которых зависит время выполнения работы алгоритма.

### Выводы

В ходе данной работы проведен анализ подходов Multi-GPU рендеринга, в результате чего для реализации графического приложения был выбран алгоритм Hybrid Multi-GPU Rendering. Также осуществлен анализ методов построения отражений, в результате чего для реализации отражений был выбран алгоритм SSR.

Результаты проведенного исследования и предварительная реализация показали, что гибридный Multi-GPU подход с переносом вычислений SSR на интегрированный GPU позволяет снизить нагрузку на основную видеокарту и добиться прироста производительности в условиях её высокой загрузки. Результаты экспериментов подтвердили, что при достаточной нагрузке на дискретный GPU гибридная схема превосходит классический Single-GPU рендеринг, тогда как при низкой нагрузке рендеринг на дополнительной видеокарте может снизить производительность. Дальнейшие работы будут направлены на динамическую адаптацию параметров SSR и оптимизацию графического приложения для уменьшения задержек и выравнивания нагрузки между GPU.

### Список литературы

1. Szczerbiński A. Multithreaded game engine architecture. Wrocław, Poland: Wrocław University of Science and Technology, 2021. P. 16.
2. Sjöholm J. Explicit Multi GPU Programming with DirectX 12 // Game Developers Conference. 2016.
3. Aguaviva R., Baker D. Explicit DirectX 12 Multi GPU rendering // Game Developer Conference. 2017.
4. Roca J., Grossman M. Scaling of 3D game engine workloads on modern multi-GPU systems // Proceedings of the Conference on High Performance Graphics 2009. New York, USA: ACM, 2009. P. 37–46.
5. Кожемяко А. История и перспективы multi-GPU на рынке игровой графики. 2023. URL: <https://www.ixbt.com/3dv/multi-gpu-pc-retrospect.html> (дата обращения 29.12.2024).
6. Богданов М. Разработка и анализ алгоритмов мульти-GPU рендеринга с использованием дискретного и интегрированного GPU. Санкт-Петербург: Университет ИТМО, 2021.
7. Clemenz C., Weydemann L. Reflection Techniques in Real-Time Computer Graphics // KoG. 2021. Vol. 25, no. 25. P. 87–95.
8. McReynolds T., Blythe D. Advanced Graphics Programming Using OpenGL. Elsevier, 2005. P. 407–411.
9. Ofek E., Rappoport A. Interactive reflections on curved objects // Proceedings of the 25th annual conference on Computer graphics and interactive techniques. New York, USA: ACM Press, 1998. P. 333–342.
10. Cube Mapping. 2018. URL: <https://research.ncl.ac.uk/game/mastersdegree/graphicsforgames/cubemapping/Tutorial%2013%20-%20Cube%20Mapping.pdf> (дата обращения 29.12.2024).
11. Beug A. Screen Space Reflection Techniques. Regina, Canada: University of Regina, 2020.
12. Roberts J. Screen Space Reflections (SSR). 2023.
13. De Macedo D. V., Serpa Y. R., Rodrigues M. A. F. Fast and Realistic Reflections Using Screen Space and GPU Ray Tracing – A Case Study on Rigid and Deformable Body Simulations // Computers in Entertainment (CIE). New York, USA: ACM, 2018. Vol. 16, no. 4. P. 1–18.
14. Гвоздкова У., Мойсейчук А. Алгоритм рейтрейсинга в современных играх // Информационные технологии и управление: материалы 59-й научной конференции аспирантов, магистрантов и студентов (2023). Минск: БГУИР, 2023. 117 с.
15. Redmon J. RAY TRACING – Middlebury. Middlebury College, USA, 2011.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-192-205

## Метод восстановления текстуры объектов сцены

Д. Д. Жданов, С. И. Куприянов

Университет ИТМО, Санкт-Петербург, Россия

**Аннотация.** В статье представлен метод восстановления текстуры диффузных поверхностей объектов сцены. Для восстановления оптических свойств объектов создается проекция UV-развертки треугольной сетки на текстурную плоскость, для которой формируются карты яркостей, полученные в результате проекции изображений сцены с каждой камеры на текстурную плоскость. Полученная карта яркостей и оригинальные изображения в сочетании с параметрами геометрии и источников света сцены позволяют рассчитать первичное и оценить вторичное освещение сцены. На основе полученных данных рассчитывается начальный цвет поверхности. Далее запускается процесс оптимизации, который рассчитывает на текстурных картах вторичное освещение методом прямой стохастической трассировки лучей и с учетом вычисленного распределения вторичной освещенности выполняет переоценку оптических свойств сцены. Процесс повторяется до тех пор, пока отклонение яркости полученного изображения от яркости исходного изображения не окажется в заданных пределах. Предложенный подход требует заранее определенной геометрии сцены, параметров источников света и камер, а также оригинальных изображений для заданных камер. В работе приводятся примеры восстановления и оценки точности восстановления оптических свойств текстурированных и нетекстурированных объектов сцены.

**Ключевые слова:** оптические свойства, прямое освещение, вторичное освещение, яркость, освещенность, цвет, развертка геометрии, текстурирование, оптимизация.

## Method of texture reconstruction for scene objects

D.D. Zhdanov, S. I. Kupriyanov

ITMO University, Saint-Petersburg, Russia

**Abstract.** The article presents a method for restoring the texture of diffuse surfaces of scene objects. To restore the optical properties of objects, a projection of the uv scan of a triangular grid onto a texture plane is created, for which luminance maps are generated, obtained as a result of the projection of images of the scene from each camera onto the texture plane. The resulting luminance map and original images, combined with the parameters of the geometry and light sources of the scene, allow us to calculate the primary and evaluate the secondary illumination of the scene. Based on the data obtained, the initial surface color is calculated. Next, the optimization process is started, which calculates secondary illumination on texture maps using direct stochastic ray tracing and, taking into account the calculated distribution of secondary illumination, performs a reassessment of the optical properties of the scene. The process is repeated until the luminance deviation of the received image from the luminance of the original image is within the specified limits. The proposed approach requires a predefined geometry of the scene, parameters of light sources and cameras, as well as original images for the specified cameras. The paper provides examples of restoring and evaluating the accuracy of restoring the optical properties of textured and non-textured scene objects.

**Keywords:** optical properties, direct illumination, secondary illumination, luminance, illuminance, color, geometry scanning, texturing, optimization.

### Введение

Задача восстановления оптических свойств сцены находит применение не только в приложениях виртуальной и дополненной реальностей, где физическая корректность не является определяющим аргументом, но и в приложениях виртуального прототипирования, требующих физическую корректность модели оптических свойств. В качестве основного решения, обеспечивающего физическую корректность восстановления оптических свойств, как правило, используется дифференцируемый рендеринг. Дифференцируемый рендеринг представляет собой метод, обеспечивающий математически формализованное вычисление градиентов целевых функций по параметрам трехмерной сцены (в данном случае оптические характеристики материалов). Он устанавливает дифференцируемое соответствие между параметрами сцены и цветовыми значениями результирующего изображения.

Расчет дифференциалов яркости по параметрам сцены позволяет рассчитывать градиенты яркости по параметрам сцены, что дает возможность применить методы оптимизации на основе градиентного спуска в задачах определения свойств сцены по целевым метрикам. Такой подход позволяет не только реконструировать геометрию сцены и характеристики источников света, но и восстановить параметры модели оптических свойств объектов сцены.

Для восстановления параметров модели оптических свойств сцены дифференцируемому рендерингу необходимо иметь ряд дополнительных данных сцены. К этим данным относятся геометрия объектов сцены, изображения сцены, информация о камерах, для которых были получены эти изображения, данные источников света. Необходимо отметить, что все дополнительные данные могут быть восстановлены средствами дифференцируемого рендеринга, но проблема их восстановления не является целью данного исследования.

В основе данной работы лежит идея уменьшения объема данных для последующего применения алгоритмов дифференцируемого рендеринга. То есть восстановленная текстура объекта определяет базовые параметры оптических свойств объекта (коэффициенты яркости по каналам цвета). Если оптические свойства близки к свойствам ламбертовых поверхностей, то предложенный алгоритм корректно восстанавливает их. В противном случае мы имеем первое приближение, которое позволяет уменьшить объем данных для последующего дифференцируемого рендеринга [1] и ускорить процесс определения параметров модели оптических свойств. Уменьшение объема данных достигается за счет преобразования яркости изображения в псевдоосвещенность, полученную в результате ее деления на восстановленную текстуру изображения. В результате получается достаточно равномерное изображение, к которому применяется алгоритм, представленный в работе [1].

### Постановка задачи

В рамках исследования решается проблема определения оптических свойств поверхности при заданных параметрах геометрии сцены, источников света, камер и набора изображений, полученных на данных камерах. Предполагается, что камеры формируют изображения широкого динамического диапазона в единицах реальной яркости, т.е. яркость источников света и оптические свойства сцены физически корректно соотносятся с яркостью полученных изображений. Модель оптических свойств – диффузная с переменным коэффициентом отражения по цветовым каналам. Допускается вариация оптических свойств по поверхности сцены, которая восстанавливается в виде текстуры на поверхностях сцены. Если оптические свойства сцены близки к диффузным, то предложенное решение должно физически корректно их восстановить. В противном случае данное решение позволит уменьшить объем данных и ускорить процесс восстановления оптических свойств средствами дифференцируемого рендеринга.

### Теория

Задача восстановления текстуры разбивается на две подзадачи: во-первых, формирование UV-развертки геометрических объектов сцены на текстурную плоскость; во-вторых, восстановление параметров текстуры, т.е. распределения коэффициентов яркости по каналам цвета для каждой точки текстуры, связанной с геометрическим объектом.

**Формирование UV-развертки.** UV-развертка представляет собой процесс проекции трехмерной поверхности на плоскость для построения текстуры. Это процесс позволяет установить соответствие между вершинами трехмерной модели и координатами текстуры ( $u, v$ ) [2]. Математически UV-развертка может быть описана следующим образом: каждая точка  $p \in M$  (где  $M$  – многообразие, представляющее 3D-объект) отображается в точку  $(u, v) \in R^2$ , сохраняя локальную геометрическую информацию [3].

Параметризация поверхности предполагает нахождение биективного отображения  $\phi : M \rightarrow \Omega \subset R^2$ , минимизирующего искажения. При биективном отображении каждому элементу одного множества соответствует ровно один элемент другого множества, при этом определено обратное отображение, которое обладает тем же свойством. Для треугольной сетки задача сводится к определению таких координат  $(u_i, v_i)$  для каждой вершины  $i$ , чтобы деформация у развертки была минимальной.

Классические алгоритмы UV-развертки включают различные методы.

*Методы на основе граничных условий*

Методы, использующие фиксацию границ на выпуклом многоугольнике, упрощают параметризацию за счет снижения вычислительной сложности [3]. Например, привязка граничных вершин к выпуклому контуру гарантирует однозначность отображения и предотвращает самопересечения [2]. Однако такие методы могут вызывать искажения внутри области из-за неравномерного распределения деформаций, особенно для поверхностей со значительной кривизной [4].



*Глобальные методы: Angle-Based Flattening (ABF)*

Метод ABF минимизирует угловые деформации через решение системы нелинейных уравнений, сохраняя локальные углы между треугольниками [5].

*Многосеточные подходы*

Многосеточные методы адаптируют разрешение, комбинируя грубые и детальные сетки [6]. Например, на уровне низкого разрешения решается задача с аппроксимированной геометрией, после чего результаты уточняются на высокодетализированных уровнях [7]. Это позволяет снизить сложность алгоритмов с  $O(N^3)$  до  $O(N \log N)$ , где  $N$  – число вершин. Такие методы эффективны для сложных поверхностей, таких как органические формы или анатомические структуры [8].

*Современные методы: машинное обучение*

Нейросетевые архитектуры, такие как MeshVAE, обучаются на датасетах 3D-моделей для предсказания UV-координат [9]. Эти модели минимизируют зависимость от ручной настройки, автоматизируя развертку для неструктурированных сеток.

Кроме того, для построения развертки существуют готовые библиотеки, например *xatlas*. Библиотека проста в реализации и применении. Алгоритм работы библиотеки следующий:

1. Сегментация поверхности – разделение сетки на связные области (чарты) с минимизацией метрики угловых искажений. Для этого используются методы спектрального анализа, дополненные условиями сохранения локальной геометрической согласованности;
2. Оптимизация UV-размещения – упаковка чартов в текстурное пространство с применением итеративного алгоритма, снижающего перекрытия и растяжения за счет решения систем уравнений, учитывающих тензорные деформации на проекции.

Основным преимуществом *xatlas* является способность генерировать UV-развертки для высокодетализированных сеток с сохранением топологии, включая модели с нерегулярной триангуляцией и геометрическими аномалиями. Библиотека демонстрирует высокую устойчивость к артефактам, что критично для последующего построения HDR-карт яркостей.

К основным ограничениям можно отнести склонность к фрагментации на моделях с множеством тонких элементов и отсутствие поддержки параллельных вычислений, что ограничивает производительность при работе с большими сетками.

**Формирование текстур.** После построения развертки геометрии извлекается текстура (в данном случае цвет) из набора изображений. При наличии известной геометрической развертки (UV-развертки) объекта задача построения текстурной карты сводится к проецированию информации с множества изображений, снятых под разными углами, на соответствующую параметризованную поверхность. Этот процесс требует решения задач согласования данных, коррекции освещения и устранения артефактов. Классический метод предполагает проецирование точек каждого из изображений на UV-карту с использованием матриц камеры и данных о геометрии. Для текстурных координат  $(u, v)$  значение цвета текстуры определяется как средневзвешенное значение по всем изображениям, которые он видит:

$$T(u, v) = \frac{\sum_{i=1}^N w_i \cdot L_i(x_i, y_i)}{\sum_{i=1}^N w_i}, \quad (1)$$

где  $L_i(x_i, y_i)$  – яркость пикселя в  $i$ -м изображении;  $w_i$  – веса, учитывающие двунаправленную функцию отражения, условия освещения и наблюдения [10].

Кроме того, для устранения шума применяются различные фильтры (например, медианный), а для HDR-текстур используется линейное сочетание экспозиций [11]. Поскольку двунаправленная функция отражения неизвестна, то основная задача сводится к определению истинного значения цвета текстуры, т.е. цвета текстуры для некоторого опорного источника света, как правило, D65. Если предположить, что модель оптических свойств поверхности близка к ламбертовым, то истинное значение цвета текстуры это практически есть двунаправленная функция рассеивания в каждой ее точке. Это приводит нас к возможному решению [1], позволяющему ускорить процесс восстановления оптических свойств сцены средствами дифференцируемого рендеринга для объектов с постоянными свойствами (нормализация яркости изображения на «истинное» значение цвета текстуры позволяет получить однородные значения яркости изображения).

Рассмотрим идею метода восстановления оптических свойств равноярких объектов сцены. Исходя из предположения, что нам известны изображения сцены для заданных параметров камеры, геометрия сцены с наложенной на нее текстурой в виде карты реальных ее яркостей (методы построения UV-развертки на геометрии сцены и проекция карты яркостей изображений на данную развертку представлены выше) и параметры источников света, мы можем представить яркость в виде суммы яркостей прямого и вторичного освещений, зависящих от одной величины  $K_d(p_j, c)$  – коэффициента яркости в каждой точке  $p_j$  поверхности по каналам цвета  $c$ . Поскольку между точками текстуры на поверхности сцены и точками изображения имеется взаимно-однозначное соответствие, то расчет яркости изображения  $L(p_j, c)$  можно выполнять в координатах сцены  $p_j$ , непосредственно связанных с точками текстуры, где  $j$  – индекс точки текстуры.

$$L(p_j, c) = \frac{1}{\pi} K_d(p_j, c) \cdot \left( \sum_{k=1}^{K_L} E_{d_k}(p_j, c) + E_i(p_j, c) \right), \quad (2)$$

где  $E_{d_k}(p_j, c)$  – значение прямой освещенности от  $k$ -го источника света в  $j$ -й точке текстуры (суммирование происходит по всем  $K_L$  источникам света);  $E_i(p_j, c)$  – значение вторичной освещенности в  $j$ -й точке текстуры.

Поскольку геометрия сцены и параметры источников света известны, то вычисление значения  $E_{d_k}(p_j, c)$  не представляет большого труда. Для точечных источников света  $E_{d_k}(p_j, c) = \frac{I_k}{\|p_j - p_k\|^2} \cos \gamma$ , где  $\gamma$  – угол между направлением освещения и нормалью к поверхности;  $p_k$  – координата точечного источника света;  $I_k$  – сила света  $k$ -го источника света на освещаемую точку. Для протяженных источников света  $E_{d_k}(p_j, c)$  может быть вычислена интегрированием по поверхности источника с применением метода выборки точек на поверхности источника света. Для бесконечно удаленных источников света применяются аналогичные выражения с единственной разницей, что величина  $\frac{I_k}{\|p_j - p_k\|^2}$  заменяется на плотность светового потока.

В результате из выражения (2) можно определить оптические свойства поверхности:

$$K_d(p_j, c) = \pi \frac{L(p_j, c)}{\sum_{k=1}^{K_L} E_{d_k}(p_j, c) + E_i(p_j, c)}. \quad (3)$$

Как было отмечено,  $\sum E_{d_k}(p_j, c)$  может быть вычислена,  $L(p_j, c)$  – известная яркость изображения. Единственная неизвестная – это  $E_i(p_j, c)$ , которая некоторым образом может быть оценена. Оценка этой величины осуществляется из общих соображений. Например, эта величина составляет некоторую долю прямого освещения в данной точке. Если прямого освещения нет, то выражение принимает следующий вид:

$$K_d(p_j, c) = \pi \frac{L(p_j, c)}{E_i(p_j, c)}. \quad (4)$$

В данном случае вторичное освещение может рассматриваться как некоторая доля среднего освещения сцены, например половина. Естественно, данные оценки не являются корректными, и первое условие для возможной корректировки этих значений – интегральное отражение поверхности не может превышать 1.

$$0 \leq K_d(p_j, c) \leq 1. \quad (5)$$

Если данное условие нарушается, то вторичное освещение корректируется до уровня (5). Например, если мы хотим скорректировать  $K_d(p_j, c)$  до уровня  $K_{d_0}(p_j, c)$ , то

$$E_i(p_j, c) = \frac{\pi L(p_j, c) - K_{d_0}(p_j, c) \sum_{k=1}^{K_L} E_{d_k}(p_j, c)}{K_{d_0}(p_j, c)}.$$

Как было отмечено, выражения (2) и (3) имеют две зависимые переменные. Это коэффициент отражения и вторичная освещенность. Кроме того, в данных выражениях отсутствует формальное

взаимное влияние точек изображения друг на друга, что позволяет осуществлять поиск коэффициента яркости каждой точки объекта сцены независимо от коэффициентов яркости других точек и решать проблему оптимизации независимо для каждой точки объекта с использованием единого критерия – среднеквадратичного отклонения яркости по всем точкам изображения. Естественный параметр оптимизации – коэффициент яркости в каждой точке поверхности (текстуре), однако для его вычисления необходимо определить освещенность (прямую и вторичную). Прямая освещенность вычисляется либо аналитически, либо с использованием численного интегрирования яркости по поверхности источника света. Для расчета вторичной освещенности был реализован метод прямой трассировки лучей с сохранением результата на текстурной карте (обратная трассировка в общем случае неприменима, поскольку может оценить освещенность только в направлении наблюдения, и даже в случае ламбертовых поверхностей пересчет яркости в освещенность может создавать артефакты на текстурных картах).

Для запуска процесса оптимизации необходимы начальные значения коэффициентов яркости поверхности. Возможно несколько решений, как определить эти коэффициенты. Во-первых, можно установить некоторые константные значения для всех объектов сцены (не очень удачное решение для текстурированных объектов). Во-вторых, можно взять их пропорциональными яркости текстуры (также не очень удачное решение, поскольку яркость сцены может меняться на порядки, что не соответствует вариации реальных коэффициентов) или пересчитать их с учетом прямого освещения (решение не подходит для затененных участков). В-третьих, можно «определить» вторичную освещенность как некоторую долю прямой освещенности (или средней освещенности для затененных объектов сцены) и исходя из нее вычислить коэффициенты яркости в каждой точке объектов сцены (решение неоптимальное, но, как показал опыт, на рассматриваемых сценах наиболее удачное). Важным положительным моментом последнего решения является то, что вторичная освещенность, как правило, не имеет значительной вариации освещенности по поверхностям объектов сцены, и ее расчет может быть выполнен с меньшими вычислительными затратами, например на меньшем разрешении или с использованием фильтрации.

В результате был предложен и реализован общий алгоритм восстановления цвета поверхности в точке. Алгоритм состоит из следующих шагов:

1. Определение исходных параметров. Построение текстурных карт истинной яркости на сцене. Расчет прямого освещения для всех точек текстурных карт. Начальная оценка вторичной освещенности для всех точек текстурных карт (либо некоторая доля прямой освещенности, либо доля средней освещенности, если прямое освещение отсутствует).

2. Вычисление коэффициента яркости для всех точек текстурных карт по формулам (3) и (4). Для проверки и коррекции коэффициентов яркости используется условие (5). На данный момент сцена полностью сформирована и готова к расчету вторичного освещения.

3. Запуск процесса рендеринга. Вычисляется только вторичное освещение, прямое освещение не меняется.

4. Расчет яркости и оценка результатов вычислений. Для каждой из камер яркость вычисляется по выражению (2). Оцениваются разность рассчитанной яркости и исходная яркость в каждой точке текстуры, а также среднеквадратичное отклонение исходной и рассчитанной яркостей по всей текстуре.

5. Если точность достигнута или выполнено заданное число шагов оптимизационного процесса, формируется сигнал прекращения расчета.

6. На основе полученной разности исходной яркости текстуры и вычисленной яркости формируется ошибка оценки вторичной освещенности. На основе этой ошибки для каждой точки текстуры определяется необходимое изменение значения вторичной освещенности, для которой новые значения коэффициентов яркости поверхности  $K_d(p_j, c)$ , вычисленные по выражениям (3) и (4), должны приблизить вычисленную яркость к исходной.

7. Переход на пункт 2.

Общая схема алгоритма представлена на рисунке 1.

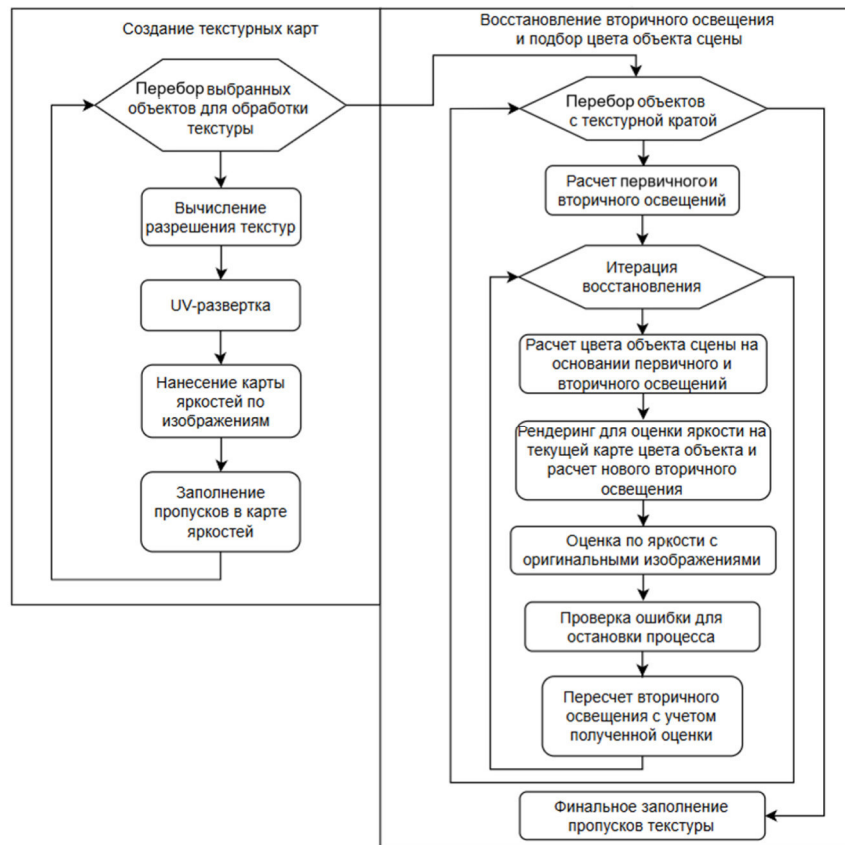


Рисунок 1. Схема алгоритма расчета коэффициентов яркости объектов сцены

Основной частью процесса оптимизации является расчет оценочной функции и выбор шага оптимизации. Оценочная функция для текстурной точки  $p_j$  и компоненты цвета  $c$  считается достаточно просто как разность оригинальной яркости, усредненной по всем видимым изображениям сцены  $L_{orig}(p_j, c)$ , и усредненной яркости, полученной в результате рендеринга для аналогичных изображений сцены  $L_{render}(p_j, c)$ :

$$\Delta L(p_j, c) = L_{orig}(p_j, c) - L_{render}(p_j, c). \quad (6)$$

В предположении ламбертовских свойств поверхности яркость изображения может быть вычислена не только на изображении, но и на текстуре, поскольку яркость вторичного освещения вычисляется как  $\frac{1}{\pi} K_d(p_j, c) E_i(p_j, c)$ , а значение  $E_i(p_j, c)$  вычисляется непосредственно на текстуре методом прямой стохастической трассировки лучей. Яркость прямого освещения  $\sum_{k=1}^{K_L} E_{d_k}(p_j, c)$  также может быть вычислена на текстуре, поскольку для каждой точки текстуры имеется обратное преобразование ее координат в систему координат сцены. Оригинальная яркость изображения сцены также существует в текстурных координатах (это первое действие, переводящее яркость изображения в текстурную карту яркости). В результате оценка производится не на изображении, а на текстуре, и оценочная функция для каждой точки текстуры – независимая величина. То есть нет необходимости формирования единой оценочной функции, алгоритм позволяет работать с большим числом независимых оценочных функций, зависящих только от одной переменной, что на порядки упрощает решение данной проблемы и сводит ее к решению большого числа однопараметрических задач.

Выразим дифференциал оценочной функции от значения вторичной освещенности. Для каждой точки текстуры  $j$  и компоненты цвета  $c$

$$\begin{aligned} F(p_j, c) &= L_{orig}(p_j, c) - L_{render}(p_j, c) = \\ &= L_{orig}(p_j, c) - \frac{K_d(p_j, c)}{\pi} \cdot \left( \sum_{k=1}^{K_L} E_{d_k}(p_j, c) + E_i(p_j, c) \right). \end{aligned} \quad (7)$$

Для каждой компоненты цвета производная по вторичному освещению

$$\frac{\partial F(p_j, c)}{\partial E_i(p_j, c)} = \frac{\partial}{\partial E_i(p_j, c)} \left( L_{orig}(p_j, c) - \frac{K_d(p_j, c)}{\pi} \cdot \left( \sum_{k=1}^{K_L} E_{d_k}(p_j, c) + E_i(p_j, c) \right) \right). \quad (8)$$

Градиент оценочной функции становится равным коэффициенту диффузного отражения точки поверхности для заданной компоненты цвета  $-\frac{K_d(p_j, c)}{\pi}$ . Тогда изменение вторичного освещения становится пропорциональным градиенту вторичного освещения для заданной компоненты цвета:

$$E_{i_{mod}}(p_j, c) = -\alpha \cdot \frac{K_d(p_j, c)}{\pi} \cdot F(p_j, c). \quad (9)$$

Новое значение вторичной яркости в точке  $j$  определяется как

$$E_{i_{new}}(p_j, c) = E_i(p_j, c) - E_{i_{mod}}(p_j, c),$$

где  $\alpha$  – шаг оптимизации, определяющий скорость коррекции. Для некоторого шага  $n$   $\alpha_n = \frac{\alpha_0}{1+n}$ , где  $\alpha_0$  – начальный шаг (в проводимых экспериментах начальный шаг был равен 2).

По разработанной методике были реализованы алгоритмы и выполнены тестовые расчеты по восстановлению параметров оптических свойств поверхностей.

### Результаты экспериментов

Для демонстрации работы алгоритма было создано несколько вариантов сцены cornel box с разными оптическими свойствами поверхностей (однородными и текстурированными). Для поверхностей этой сцены было выполнено восстановление оптических свойств. Первый вариант сцены – это классическая сцена cornel box. На рисунке 2 показаны изображения сцены, для которых производилось восстановление оптических свойств.

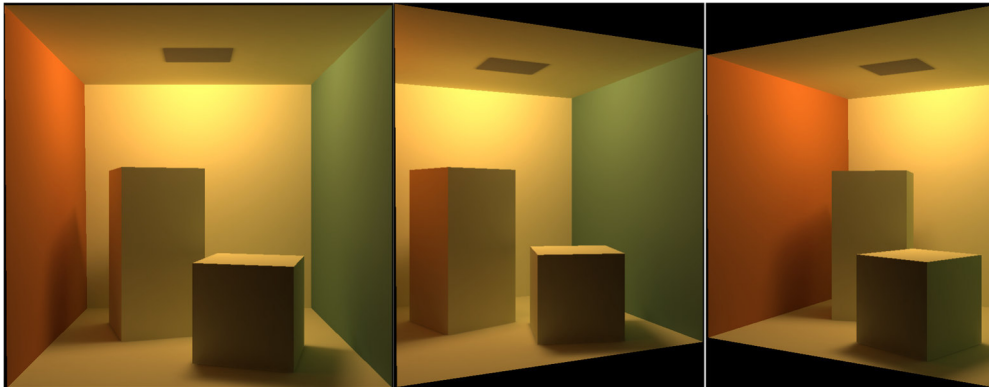


Рисунок 2. Сцена cornel box для трех ракурсов камеры

С использованием библиотеки *hAtlas* была построена UV-развертка геометрических объектов сцены на текстурную плоскость, на которую затем было спроецировано изображение сцены и построена карта яркостей. Пример UV-развертки и карты яркостей для правой зеленой стенки сцены представлен на рисунке 3.

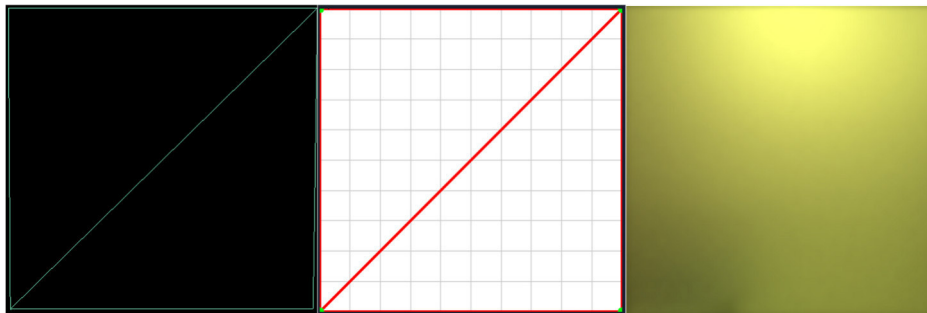


Рисунок 3. Геометрия правой (зеленой) стенки сцены (слева), текстурная развертка (по центру) и исходная карта яркости (справа)

Первые эксперименты с восстановлением оптических свойств были проведены для нетекстурированных поверхностей, однако алгоритм восстановления не был ориентирован на постоянство оптических свойств и для каждой точки поверхности восстановление осуществлялось независимым образом.

В соответствии с приведенным выше алгоритмом был выполнен первый шаг оптимизации, т.е. расчет яркости прямого освещения и оценка коэффициента яркости по компонентам цвета. После выполнения данного шага запустился итерационный процесс восстановления оптических свойств. На рисунке 4 приведены результаты первых четырех шагов процесса восстановления. Результат представлен в виде относительной погрешности восстановления, усредненной по всем компонентам цвета  $\frac{|L_{orig}(p_j,c) - L_{render}(p_j,c)|}{L_{orig}(p_j,c)} * 100$ .

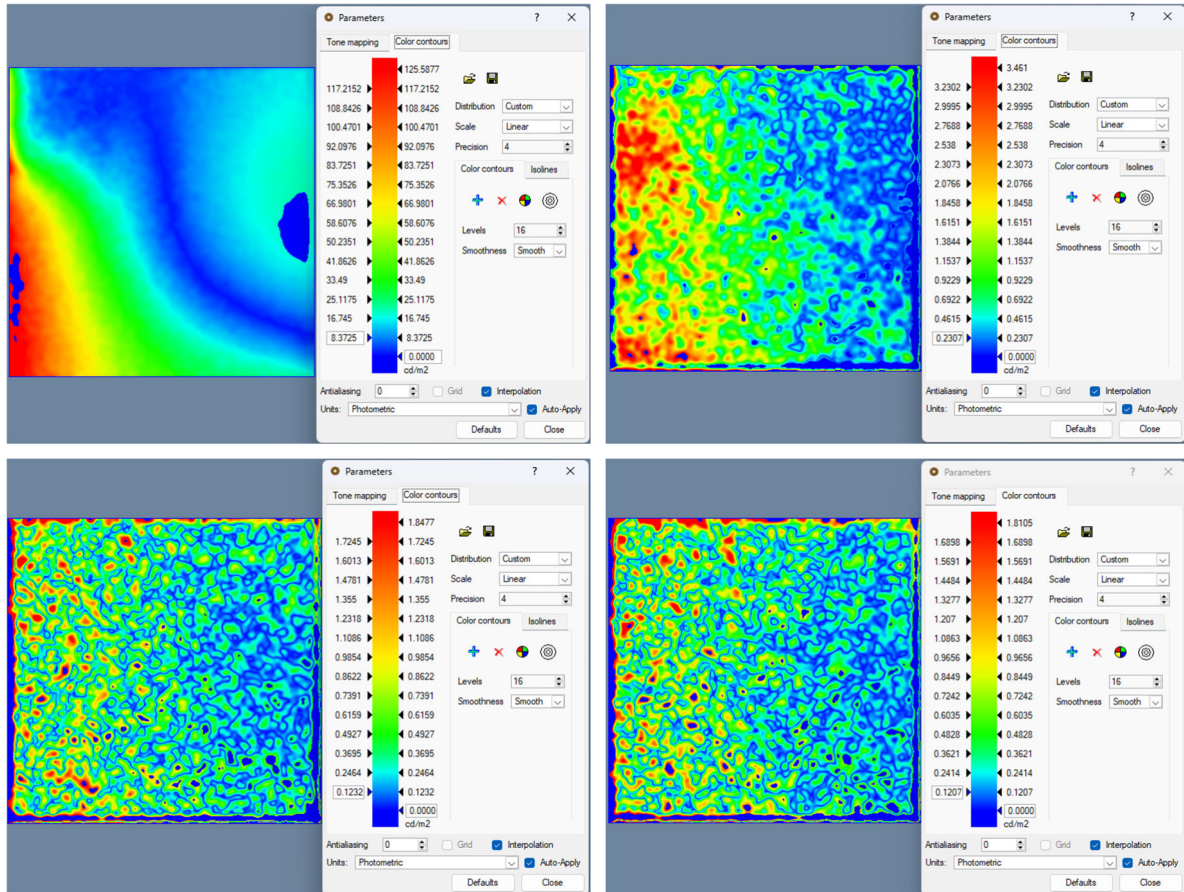


Рисунок 4. Карты нормализованной ошибки: на первом шаге (верхний левый рисунок), на втором шаге (верхний правый рисунок), на третьем шаге (нижний левый рисунок), на четвертом шаге (нижний правый рисунок)

Видно, что за первые четыре шага максимальная ошибка не превышает 2 %, что вполне допустимо и визуально не различимо на изображении. На рисунке 5 показаны оригинальное изображение и изображение после восстановления цвета.

Был проведен анализ соответствия исходного и восстановленного цвета поверхности. Анализ производился по результатам рендеринга в области, показанной на рисунке 5. Результаты анализа представлены в таблице 1.

Кроме того, была проведена оценка скорости сходимости компонент цвета поверхности к истинным значениям. Рисунок 6 демонстрирует скорость сходимости по шагам расчета.

Очевидно, что трех шагов достаточно для корректного восстановления цвета поверхности.

В приведенном примере правая стенка практически не содержит полностью или частично затененных участков. Для проверки работы алгоритма на затененных областях сцены было проведено восстановление свойств поверхности на левой стенке и потолке. Рисунок 7 демонстрирует восстановление цвета поверхности левой стенки.



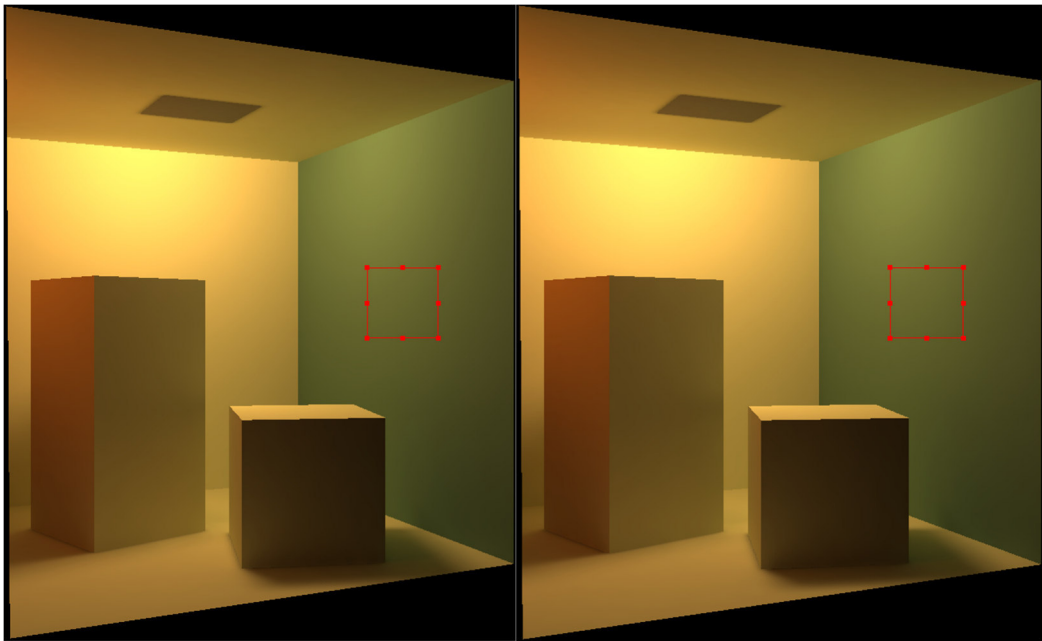


Рисунок 5. Оригинальное изображение (слева), результат рендеринга, полученный после восстановления цвета (справа)

Таблица 1. Сравнение средних значений яркости и цвета для области 90 × 90 точек на оригинальном изображении и изображении, полученном с помощью рендеринга с восстановленными оптическими свойствами зеленой стенки

Изображение	Оригинал	Рендеринг на полученной текстуре
Средняя яркость (cd/m2)	0.392047	0.391723
Среднеквадратичное отклонение яркости (cd/m2)	0.0513821	0.0516017
Цвет RGB (cd/m2)	0.415489, 0.417419, 0.0716505	0.415249, 0.417045, 0.0715795
Цветность xy (CIE)	0.391717, 0.460432	0.391744, 0.460422

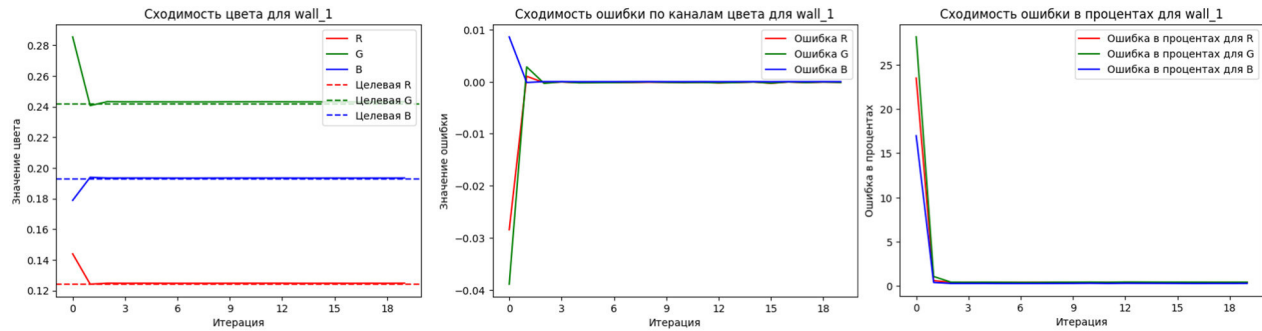


Рисунок 6. График сходимости среднего значения цвета к истинному значению цвета правой зеленой стенки (слева), абсолютная ошибка (по центру) и относительная ошибка (справа)

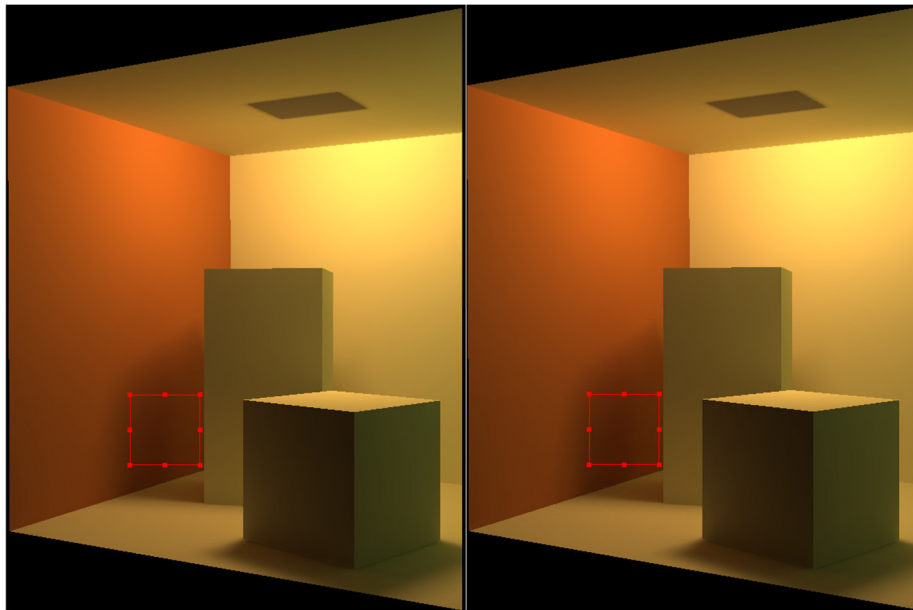


Рисунок 7. Оригинальное изображения (слева), результат рендеринга, полученный после восстановления цвета (справа)

Также был проведен анализ соответствия исходного и восстановленного цвета поверхности. Анализ проводился по результатам рендеринга в области, показанной на рисунке 7. Результаты анализа представлены в таблице 2.

Таблица 2. Сравнение средних значений яркости и цвета для области  $90 \times 90$  точек на оригинальном изображении и изображении, полученном с помощью рендеринга с восстановленными оптическими свойствами красной стенки

Изображение	Оригинал	Рендеринг на полученной текстуре
Средняя яркость (cd/m2)	0.0893196	0.0892520
Среднеквадратичное отклонение яркости (cd/m2)	0.0181009	0.0179845
Цвет RGB (cd/m2)	0.276969, 0.0424213, 0.00120226	0.276751, 0.0423917, 0.001120092
Цветность xy (CIE)	0.562333, 0.38754	0.562329, 0.387545

Рисунок 8 демонстрирует восстановление цвета поверхности на потолке.

Анализ проводился по результатам рендеринга в области, показанной на рисунке 8. Результаты анализа сравнения оригинального и восстановленного изображений представлены в таблице 3.

Аналогичные эксперименты были проведены и для реально текстурированных поверхностей. На рисунке 9 приведен результат восстановления истинного цвета текстуры. Восстановление было выполнено за 4 шага расчета.

Максимальная относительная ошибка не превысила 2.7 %. Поскольку сцена имеет достаточно сложное освещение с точки зрения цвета прямого и вторичного освещений, то истинный цвет текстуры может сильно отличаться от наблюдаемого. Рисунок 10 демонстрирует, как исходный цвет текстуры отличается от наблюдаемого и насколько он близок к восстановленному.

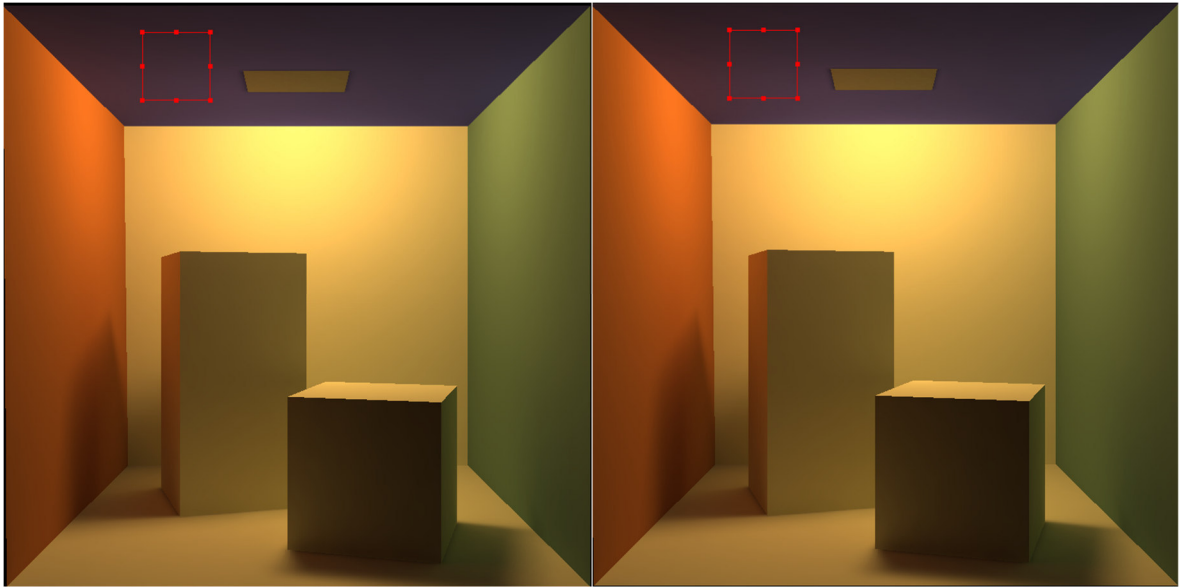


Рисунок 8. Оригинальное изображение (слева), результат рендеринга, полученный после восстановления цвета (справа)

Таблица 3. Сравнение средних значений яркости и цвета для области 90 × 90 точек на оригинальном изображении и изображении, полученном с помощью рендеринга с восстановленными оптическими свойствами потолка

Изображение	Оригинал	Рендеринг на полученной текстуре
Средняя яркость (cd/m2)	0.0781676	0.0781493
Среднеквадратичное отклонение яркости (cd/m2)	0.0214328	0.0215023
Цвет RGB (cd/m2)	0140054, 0.0590389, 0.085381	0.13989, 0.0590521, 0.085481
Цветность xy (CIE)	0.357999, 0.296823	0.357786, 0.296719

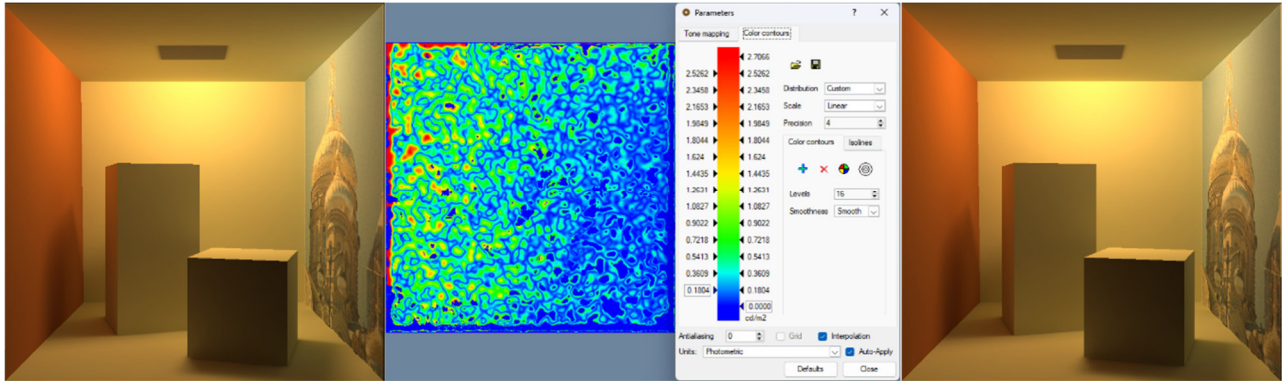


Рисунок 9. Исходное изображение церкви (слева), карта ошибок (по центру) и результат рендеринга с восстановленной текстурой (справа)



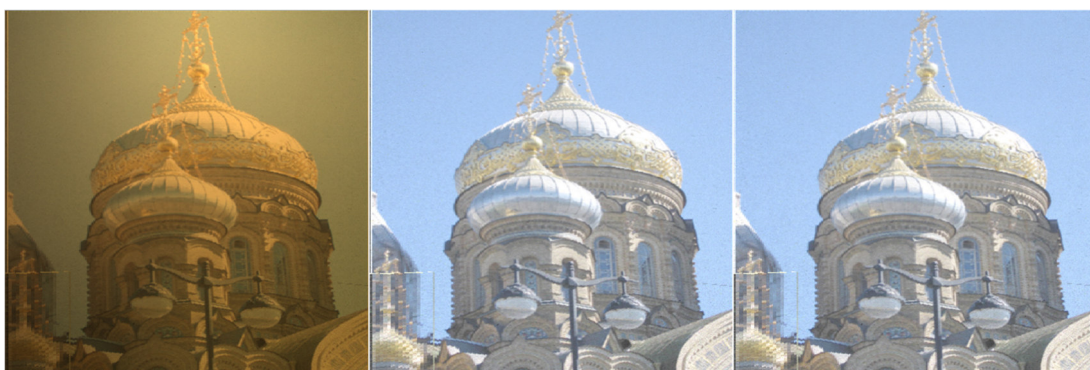


Рисунок 10. Слева оригинальное изображение, по центру расположена оригинальная текстура, а справа – восстановленная

Видно, что даже мелкие детали текстуры (дефект в нижней левой четверти – маленькая белая линия) были восстановлены.

Также в качестве эксперимента было проведено восстановление трех различных текстур, назначенных на стенки и потолок. Рисунок 11 демонстрирует исходную и восстановленную сцены.

Исходные и восстановленные текстуры представлены на рисунке 12.

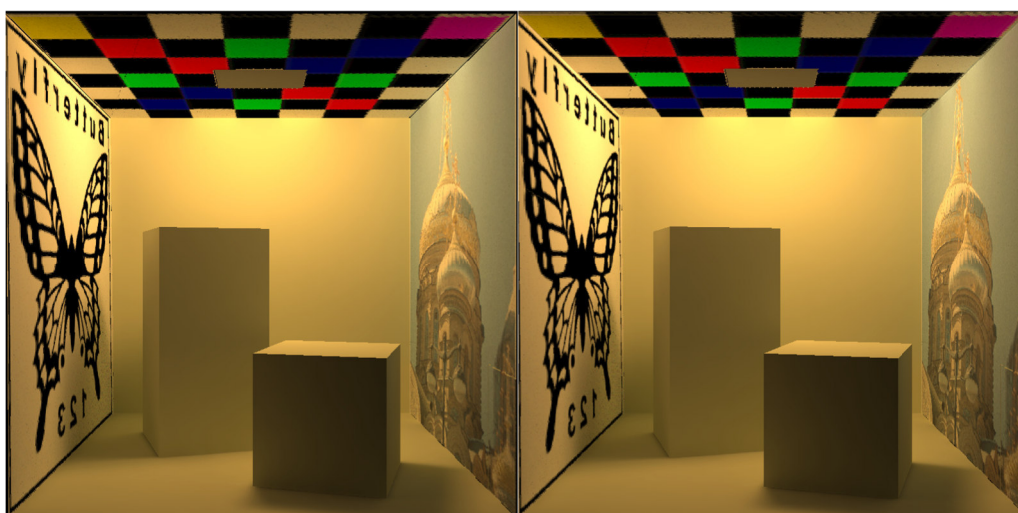


Рисунок 11. Слева оригинальное изображение, справа результат рендеринга для восстановленных текстур

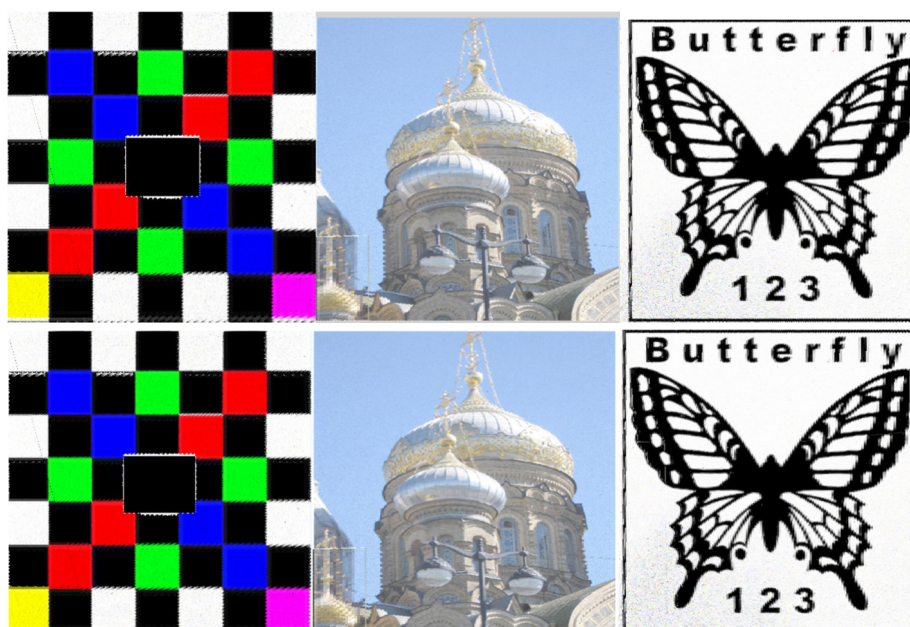


Рисунок 12. Верхний ряд – исходные текстуры, нижний ряд – текстуры после восстановления

### Обсуждение результатов

Приведенные результаты были получены на компьютере AMD RYZEN 9, 5000 series, 32 Gb RAM. Для расчета вторичной освещенности использовалось 100,000,000 световых лучей на одну фазу расчета. Исследование показало, что 300,000,000 лучей позволяют повысить точность восстановления, однако замедление расчета не оправдывает незначительное повышение качества восстановления для данных сцен. Время, затрачиваемое на восстановление текстурного объекта, составляет порядка 120 секунд, в то время как Mitsuba затрачивает порядка 900 секунд на восстановление текстуры для данной сцены [12].

Согласно результатам экспериментов, разработанный метод хорошо работает как с однородными поверхностями, так и с текстурированными. Кроме того, полученная текстура представляет ценность в контексте предложенного алгоритма отбора точек для дифференцируемого рендеринга [1]. Так как точки выбираются исходя из яркости оригинальных изображений, сильная вариация цвета объекта сцены может привести к не совсем равномерному и корректному выбору областей для расчета. Это может происходить, когда у объекта присутствуют одинаково освещенные зоны, но с большим разбросом по цвету – от практически черных областей до пересыщенно белых. Для того чтобы сделать выбор независимым от цвета объекта и тем самым устранить вариацию цвета, можно использовать текстурную карту и «обесцветить» поверхность сцены.

Для примера можно устранить вариацию цвета на стенке с изображением «церковь». Для того, чтобы обесцветить изображение, понадобится текстурная карта цвета объекта и яркость оригинального изображения. Для получения «обесцвеченного» значения в точках текстуры (с малой вариацией цвета) необходимо выполнить нормировку на цвет текстуры:

$$L_{colorless}(p_j, c) = \frac{\sum_{i=1}^N \frac{L_{render}(p_j, c)}{K_d(p_j, c)}}{\sum_{c=1}^C K_d(p_j, c)}, \quad (10)$$

где  $C$  – число компонент цвета. На рисунке 13 показана яркость стенки с текстурой «церковь» и соответствующая «обесцвеченная» яркость.

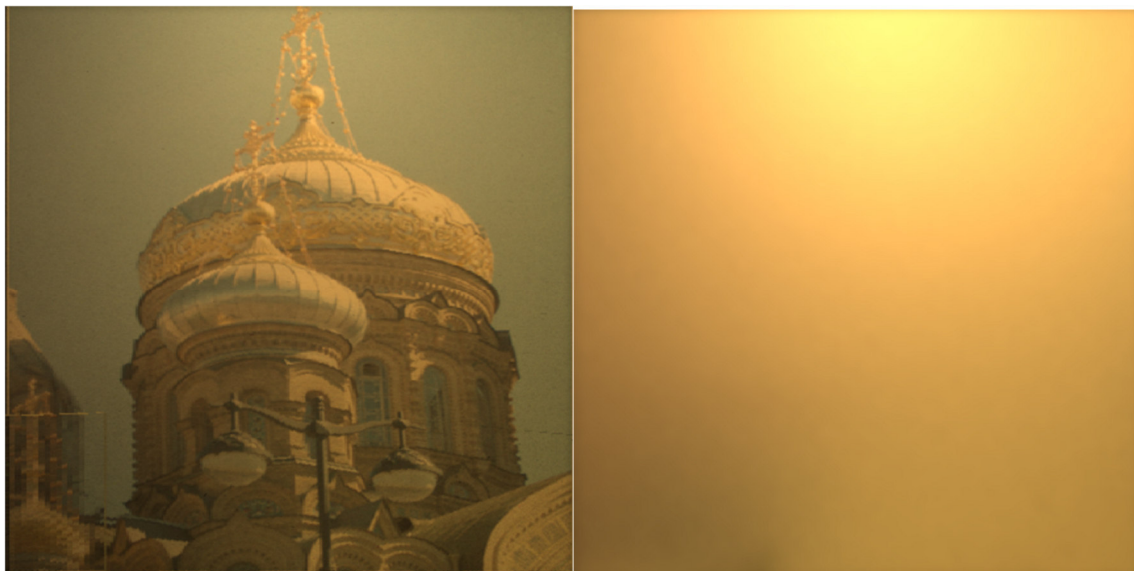


Рис. 13. Слева оригинальная яркость, справа яркость после устранения вариации цвета объекта сцены

Таким образом можно добиться устранения вариации цвета объекта и тем самым отбирать точки по яркости, более корректно используя метод, предложенный в [1].

### Выводы

Предложенный алгоритм позволяет предварительно построить текстурные карты, собрать яркости по оригинальным изображениям и рассчитать прямое и вторичное освещение на развертке геометрии для ускорения процесса восстановления оптических свойств модели сцены средствами дифференцируемого

рендеринга. Кроме того, алгоритм использует эти данные для восстановления цвета равноярких поверхностей сцены.

Проведенные исследования показывают эффективность восстановления цвета поверхности на сценах как с однородными свойствами, так и со сложными цветными текстурами. Данный алгоритм позволяет убрать вариацию цвета объектов сцены и тем самым использовать предложенный ранее метод [1] для выбора наиболее информативных точек по одноцветному изображению сцены. Построенная текстура может быть передана в дифференцируемый рендеринг как вариация цвета, что позволяет снизить размерность задачи дифференцируемого рендеринга и повысить скорость и точность его работы. Кроме того, предложенный алгоритм может использоваться как независимое решение для физически корректного восстановления оптических свойств сцены, содержащих равнояркие объекты.

### Список литературы

1. Куприянов С.И., Жданов Д.Д., Валиев И.В. Предобработка изображений для решения задач дифференцируемого рендеринга // Препринты ИПМ им. М.В. Келдыша. 2025. № 24. С. 1-23
2. Desbrun, M., Meyer, M., & Alliez, P. (2002). Intrinsic Parameterizations of Surface Meshes. *Computer Graphics Forum*, 21(3), 209–218
3. Floater, M.S., & Hormann, K. (2005). Surface Parameterization: a Tutorial and Survey. In: *Advances in Multiresolution for Geometric Modelling* (pp. 157–186). Springer.
4. Lévy, B., Petitjean, S., Ray, N., & Maillot, J. (2002). Least Squares Conformal Maps for Automatic Texture Atlas Generation. *ACM Transactions on Graphics*, 21(3), 362–371.
5. Sheffer, A., & de Sturler, E. (2001). Parameterization of Faceted Surfaces for Meshing Using Angle-Based Flattening. *Engineering with Computers*, 17(3), 326–337.
6. Zayer, R., Lévy, B., & Seidel, H.-P. (2007). Linear Angle Based Parameterization. In: *Proceedings of the Fifth Eurographics Symposium on Geometry Processing* (pp. 135–141).
7. Sorkine, O., Cohen-Or, D., & Toledo, S. (2002). High-Pass Quantization for Mesh Encoding. In: *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (pp. 42–51).
8. Gu, X., Luo, F., Sun, J., & Yau, S.-T. (2016). Convergence of Discrete Conformal Geometry and Computation of Teichmüller Space. *Foundations of Computational Mathematics*, 16(3), 649–685.
9. Tatarchenko, M., Dosovitskiy, A., & Brox, T. (2016). Multi-view 3D Models from Single Images with a Convolutional Network. *arXiv preprint arXiv:1605.03557*.
10. Debevec, P.E., Taylor, C.J., Malik, J. (1996). Modeling and Rendering Architecture from Photographs. *ACM SIGGRAPH*.
11. Matusik, W., et al. (2003). Image-Based 3D Photography. *Eurographics Workshop on Rendering*.
12. Nimier-David M., Vicini D., Zeltner T., Jakob W. (2019). Mitsuba 2: A Retargetable Forward and Inverse Renderer. *Transactions on Graphics*. In: *Proceedings of SIGGRAPH Asia*. 38(6).



УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-206-214

## Алгоритм расчета бликов в оптических системах

Ван Янь, Д. Д. Жданов, И. С. Потемин, А. Д. Жданов

Университет ИТМО, Санкт-Петербург, Россия

**Аннотация.** Оценка влияния рассеянного света, вызванного переотражением между преломляющими поверхностями оптических элементов, является одним из ключевых моментов в процессе проектирования оптических систем, формирующих изображение. Однако существующие методы, используемые для такой оценки, такие как метод Монте-Карло и стохастическая трассировка лучей, страдают от проблем, связанных громоздкостью вычислений из-за необходимости трассировать десятки, а иногда и сотни миллионов лучей для получения качественного изображения бликов на приемнике изображения. В данной статье предлагается алгоритм эффективного расчета изображений бликов, позволяющий за счет использования адаптивной интерполяции на порядки как ускорить расчет изображения бликов заданного порядка, так и оценить энергетику блика.

**Ключевые слова:** расчет оптических систем, рассеянный свет, расчет бликов, адаптивная интерполяция, детерминистическая трассировка лучей, построение изображения бликов.

## Algorithm for calculating ghosts in optical systems

Wang Yan, D. D. Zhdanov, I. S. Potemin, A. D. Zhdanov

ITMO University, Saint-Petersburg, Russia

**Abstract.** Estimation of the influence of scattered light caused by re-reflection between the refractive surfaces of optical elements is one of the key points in the design of the imaging optical systems. However, existing methods used for such an assessment, such as the Monte Carlo method and stochastic ray tracing, suffer from problems associated with the cumbersomeness of calculations due to the need to trace tens and sometimes hundreds of millions of rays to obtain a high-quality ghost image on the image receiver. This paper proposes an algorithm for efficient calculation of ghost images that allows, through the use of adaptive interpolation, to accelerate both the calculation of a flare image of a given order and to estimate the ghost energy by orders of magnitude.

**Keywords:** optical system design, stray light, ghosts simulation, adaptive interpolation, deterministic ray tracing, ghosts image formation.

### Введение

При проектировании оптических систем, формирующих изображение (например, в фотообъективах) одним из важных источников ухудшения качества изображения может быть рассеянный свет, вызванный нежелательными френелевскими отражениями или бликами от поверхностей линз [1]. Разработчикам оптических систем важно понимать влияние рассеянного света на возможности системы и предпринимать соответствующие шаги на ранних этапах проектирования для его контроля [2–3].

В большинстве линзовых оптических систем, предназначенных для формирования изображения, предполагается только один путь света, определяемый преломлением на поверхностях линз. К сожалению, бликовые отражения от преломляющей оптики приводят к появлению множества нежелательных путей света, которые не соответствуют предполагаемому. Процесс проектирования системы с адекватным контролем бликов, по сути, заключается в выявлении путей переотраженного между оптическими поверхностями света и их последующем уменьшении или устранении.

Для расчета бликов необходимо оценить распределение энергии (освещенности) в плоскости изображения, вызванное переотражениями от преломляющих поверхностей оптической системы. Обычно для этой цели осуществляется расчет большого количества лучей методом стохастической трассировки. Такой расчет из-за сравнительно низких значений коэффициентов отражения очень неэффективен по времени, так как количество отраженных лучей, формирующих изображение блика, существенно меньше количества преломленных лучей, формирующих основное изображение наблюдаемого объекта. По этой причине использование методов стохастической трассировки для расчета бликов в большом наборе оптических систем является крайне трудоемким по времени [4]. А именно большие наборы вариантов оптических систем необходимы для обучения нейросетей, предназначенных для решения задачи поиска исходных схем для расчета оптических систем, формирующих изображение [5, 6]. Таким образом, весьма актуальной является проблема создания

алгоритма, обеспечивающего гораздо более эффективный расчет бликов, чем это позволяют сделать алгоритмы, основанные на методах стохастической трассировки лучей.

### Постановка задачи

Как поднять эффективность расчета бликов в оптических системах? Известно, что блики в зависимости от числа отражений бывают разных порядков. Блики второго порядка вызваны двукратным отражением от оптических поверхностей (рис. 1а). Первое отражение от одной из оптических поверхностей системы происходит в направлении от пространства изображений к пространству объектов, второе отражение – в направлении от пространства объектов к пространству изображений. Блики четвертого порядка вызваны четырехкратным отражением (рис. 1б) от оптических поверхностей. Блики шестого и более высоких порядков вызваны соответственно шестикратным, восьмикратным и так далее отражениями в направлении от изображения к объекту и обратно. Очевидно, что блики второго порядка несут в разы больше энергии, чем блики четвертого порядка, а тем более, чем блики более высоких порядков. Поэтому для повышения эффективности расчета бликов предлагается использовать метод детерминистической трассировки лучей при условии заданного порядка блика.

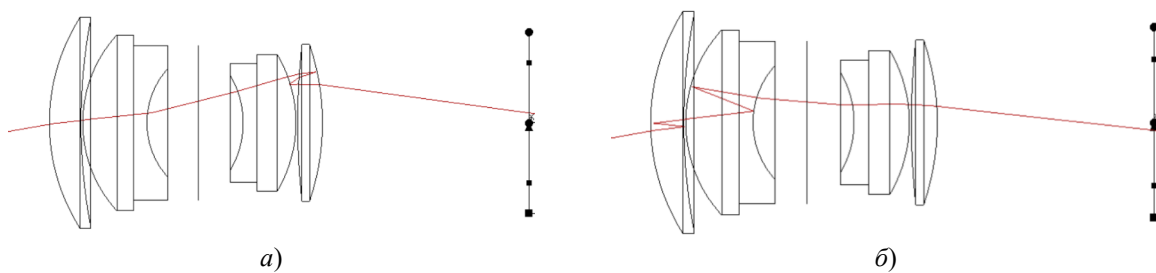


Рисунок 1. Примеры формирования бликов различных порядков

Кроме того, для сокращения числа трассируемых лучей предлагается осуществлять адаптивную интерполяцию как на зрачке оптической системы, так и в плоскости ее изображения. Такой подход позволит существенно сократить временные затраты за счет сокращения числа трассируемых лучей.

Укрупненно предлагаемый алгоритм включает следующие этапы:

1. Генерация сетки лучей на входном зрачке оптической системы.
2. Интерполяция точек на изображении по точкам на входном зрачке.
3. Расчет (накопление) пропускания и оценка размеров блика по среднеквадратичному отклонению в плоскости изображения.
4. Переход на этап 2, если заданное разрешение не достигнуто.
5. Нормировка пропускания и вывод финального изображения блика.

### Теория

#### 1. Генерация сетки лучей на входном зрачке оптической системы

Объект находится на бесконечности и направления всех лучей, идущих от объекта, одинаковы. Начальная точка координат всех лучей находится на входном зрачке объектива. Форма зрачка круглая, радиус зрачка  $R_p$ .

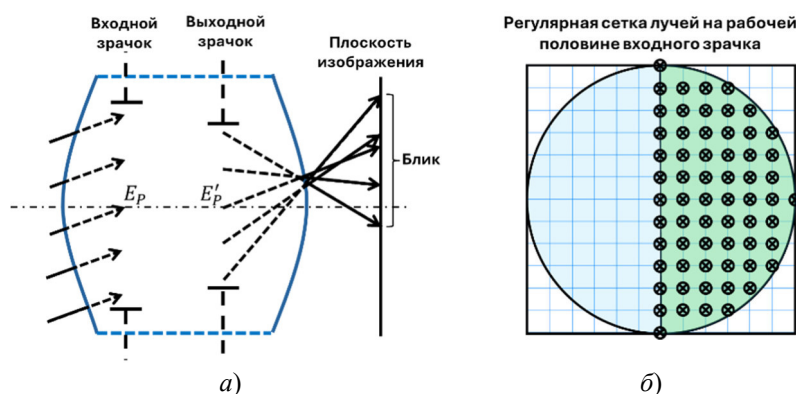


Рисунок 2. Формирование сетки лучей на входном зрачке

Важно отметить, что поскольку оптическая система симметрична относительно меридиональной плоскости, то распределение точек начальных координат лучей достаточно выполнить на половине зрачка, что позволяет вдвое сократить число трассируемых лучей. Вторая половина точек строится зеркально относительно плоскости симметрии.

## 2. Интерполяция точек на изображении по точкам на входном зрачке

После формирования сетки начальных координат лучей на входном зрачке осуществляется трассировка этих лучей заданного направления через оптическую систему до плоскости изображения для предварительно заданного порядка блика. В результате такой трассировки получаем контур реально прошедших лучей на плоскости входного зрачка. Пример контуров области прошедших лучей в плоскости входного зрачка для различных углов поля зрения представлен на рисунке 3. Контур входного зрачка обозначен пунктирной линией.

В пределах полученного контура формируется новая (интерполированная) сетка лучей, по результатам трассировки которых строятся изображения бликов заданного порядка.

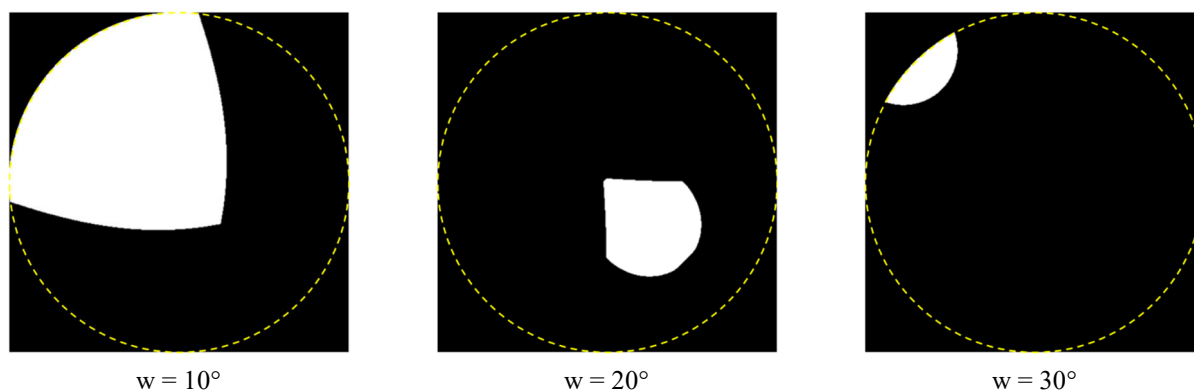


Рисунок 3. Формирование контура прошедших лучей на входном зрачке

Для ускорения расчетов предлагается трассировать только часть лучей из сформированной на входном зрачке сетки, а часть точек на изображении интерполировать по уже имеющимся координатам ограниченного набора лучей. Возможность такой интерполяции определяется по степени близости координат лучей, окружающий интерполируемую точку на изображении. Для этого был модифицирован Adaptive Interpolation Density (AID) алгоритм адаптивной интерполяции, используемый в рендеринге для интерполяции цвета пикселя. В настоящей модификации интерполируются координаты лучей на изображении по четырем точкам на входном зрачке объектива. Интерполяция, осуществляемая с помощью этого алгоритма, позволяет многократно снизить вычислительную нагрузку, связанную с заменой трассировки большого числа лучей простой билинейной интерполяцией. Пример работы алгоритма для прямоугольной сетки на зрачке представлен на рисунке 4.

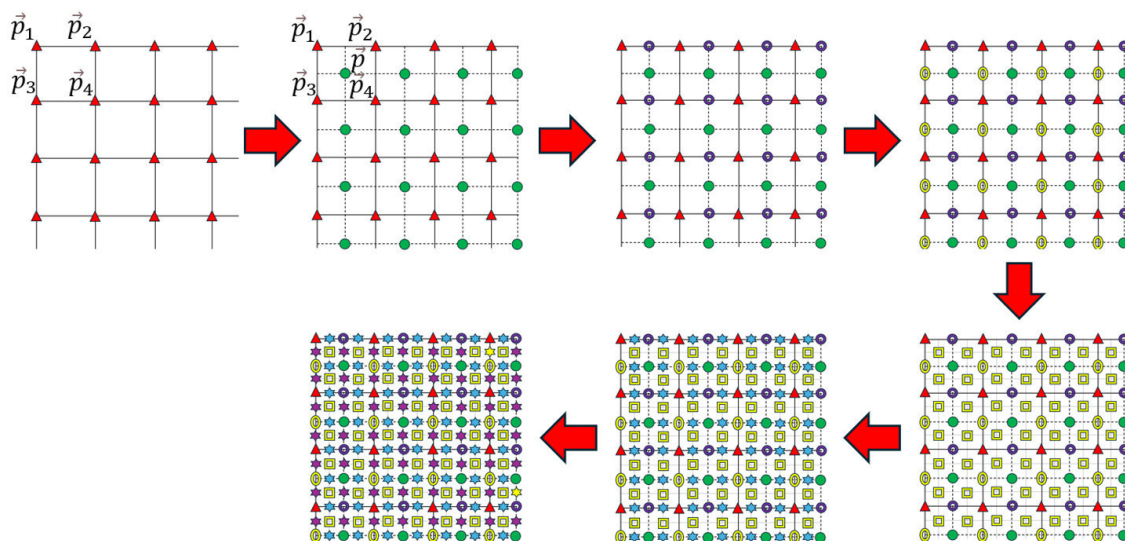


Рисунок 4. Принцип работы AID алгоритма для прямоугольной сетки на входном зрачке

Здесь красным треугольником обозначены точки на входном зрачке, из которых осуществляется трассировка лучей до плоскости изображения без интерполяции. Для остальных лучей координаты в плоскости изображения вычисляются посредством интерполяции между четырьмя точками, сформированными лучами, трассированными из четырех точек на зрачке, окружающих исследуемую точку. Для граничных зон входного зрачка число точек для интерполяции может быть уменьшено естественным образом. Интерполяция для точки  $\vec{p}$  возможна, если четыре соседние точки (или три для граничных зон) на входном зрачке ( $\vec{p}_1, \vec{p}_2, \vec{p}_3, \vec{p}_4$ ) формируют точки в плоскости изображения ( $\vec{i}_1, \vec{i}_2, \vec{i}_3, \vec{i}_4$ ), расположенные на «близком» расстоянии друг от друга, а интерполируемая точка  $\langle \vec{i} \rangle_4$  на плоскости изображения (со средневзвешенными координатами  $\vec{i}_i S_i$ ) не увеличивает среднеквадратичное отклонение (СКО) четырех базовых точек. Критерий «близости», заключающийся в том, что СКО четырех точек не превышает заданную величину  $\sigma_0$  можно представить следующим образом:

$$\langle \vec{i} \rangle_4 = \frac{\vec{i}_1 S_1 + \vec{i}_2 S_2 + \vec{i}_3 S_3 + \vec{i}_4 S_4}{S_1 + S_2 + S_3 + S_4}; \quad (3.1)$$

$$\sigma_4^2 = \frac{1}{S_1 + S_2 + S_3 + S_4} \left( \vec{i}_1^2 S_1 + \vec{i}_2^2 S_2 + \vec{i}_3^2 S_3 + \vec{i}_4^2 S_4 - \langle \vec{i} \rangle_4 (\vec{i}_1 S_1 + \vec{i}_2 S_2 + \vec{i}_3 S_3 + \vec{i}_4 S_4) \right); \quad (3.2)$$

$$\sigma_5^2 = \frac{1}{S_1 + S_2 + S_3 + S_4 + S_5} \left( \vec{i}_1^2 S_1 + \vec{i}_2^2 S_2 + \vec{i}_3^2 S_3 + \vec{i}_4^2 S_4 + \langle \vec{i} \rangle_4^2 S_5 - \frac{(\vec{i}_1 S_1 + \vec{i}_2 S_2 + \vec{i}_3 S_3 + \vec{i}_4 S_4 + \langle \vec{i} \rangle_4 S_5)^2}{S_1 + S_2 + S_3 + S_4 + S_5} \right); \quad (3.3)$$

$$\sigma_4^2 \leq \sigma_0^2 \wedge \sigma_5^2 \leq \sigma_4^2 - \text{условие интерполяции}, \quad (3.4)$$

где  $S_i$  – вес  $\vec{p}_i$ -й точки на зрачке. Вес точки на зрачке определяется площадью области зрачка, к которой привязана данная точка.

В случае выполнения условия интерполяции координаты луча в плоскости изображения формируются по результатам интерполяции и луч маркируется как удачно интерполированный. В дальнейшем такой луч может быть использован в качестве исходных данных для следующей интерполяции. В случае невыполнения условия интерполяции для данной точки на зрачке выполняется трассировка лучей.

В дополнение к адаптивной интерполяции в декартовых координатах был разработан вариант в полярных координатах, представленный на рисунке 5.

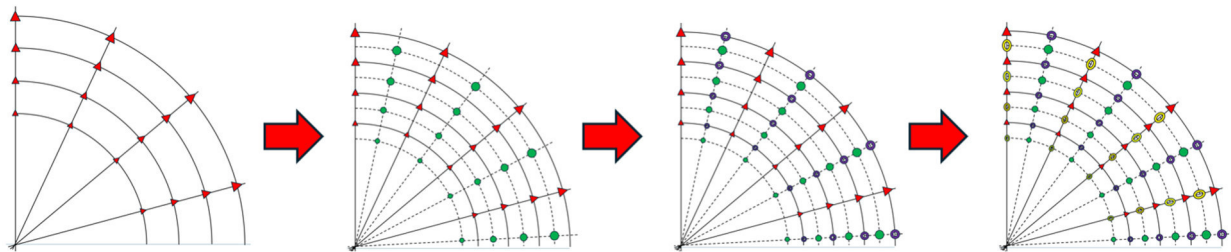


Рисунок 5. Пример формирования радиальной адаптивной сетки на входном зрачке

Адаптация, выполняемая на сетке, позволяет ограничить число трассируемых лучей для вычисления корректной оценки влияния паразитной засветки, вызванной бликом между заданной группой поверхностей, на качество изображения. В большинстве случаев для корректной оценки достаточно выполнить трассировку порядка 1000 лучей, которая позволяет оценить энергетику блика и его концентрацию. Оценка выполняется по параметрам энергетики и концентрации пятна рассеивания.

### 3. Расчет пропускания и оценка размеров блика по СКО в плоскости изображения

Энергетика оценивается по пропусканию оптической системы. Пропускание оптической системы  $\tau$  можно рассматривать как взвешенное отношение числа лучей, прошедших через оптическую систему, к общему числу лучей на входном зрачке. Косинусная зависимость освещенности зрачка от направления падения света в данном случае не рассматривается, поскольку для всех вариантов переотражений между поверхностями эта величина постоянная и может быть добавлена позже в виде масштабного множителя при анализе влияния направления паразитной засветки. С учетом пропускания материалов стекла и отражения его поверхностей для заданной точки поля и длины волны можно записать:

$$\tau = \frac{\sum_{i=0}^N \prod_{j=1}^L \tau_j \prod_{k=1}^M T_k S_i}{\pi R_p^2}, \quad (3.5)$$

где  $R_p$  – радиус зрачка;

$N$  – число сегментов трассы луча;

$L$  – число линз;

$\tau_j$  – коэффициент пропускания линзы на длине трассы луча в стекле  $t_j$  и при заданном пропускании на единичную длину  $\tau_{0j}$  ( $\tau_j = \tau_{0j}^{t_j}$ );

$M$  – число поверхностей линз, граничащих с двумя различными средами;

$T_k$  – коэффициент пропускания на границе стекла, который определяется либо по «зеркальной» двунаправленной функции пропускания (ДФП) покрытия, либо по формуле Френеля при отсутствии покрытия.

Для неполяризованного света

$$T_k = \frac{1}{2} \left( \left| \frac{n_{k1} \cos \theta_{k1} - n_{k2} \cos \theta_{k2}}{n_{k1} \cos \theta_{k1} + n_{k2} \cos \theta_{k2}} \right|^2 + \left| \frac{n_{k1} \cos \theta_{k2} - n_{k2} \cos \theta_{k1}}{n_{k1} \cos \theta_{k2} + n_{k2} \cos \theta_{k1}} \right|^2 \right), \quad (3.6)$$

где  $\theta_{k1}$  и  $\theta_{k2}$  – угол падения и преломления соответственно. Углы вычисляются исходя из соотношения

$$\vec{s}_{k2} n_{k2} = \vec{s}_{k1} n_{k1} + \gamma \vec{N}_k,$$

где  $n_{k1}$  – показатель преломления стекла со стороны падения света;

$n_{k2}$  – показатель преломления стекла со стороны, противоположной падению света;

$\vec{N}_k$  – единичный вектор нормали в точке падения луча на поверхность;

$\vec{s}_{k1}$  и  $\vec{s}_{k2}$  – единичные векторы падающего и преломленного луча;

$\gamma$  – нормировочный коэффициент, обеспечивающий «единичность» всех векторов.

В случае интерполяции пропускание вычисляется аналогично (3.1), причем пропускание для точки на зрачке может также являться частью критерия возможной интерполяции по аналогии с (3.4).

$$\langle \tau \rangle_4 = \frac{\tau_1 S_1 + \tau_2 S_2 + \tau_3 S_3 + \tau_4 S_4}{S_1 + S_2 + S_3 + S_4}, \quad (3.7)$$

Концентрация лучей в изображении блика оценивается по среднеквадратичному отклонению точек на изображении:

$$\sigma^2 = \frac{\sum_{i=0}^N S_i (\tilde{t}_i)^2}{\sum_{i=0}^N S_i} - \sum_{i=0}^N \left( \frac{S_i \tilde{t}_i}{\sum_{i=0}^N S_i} \right)^2, \quad (3.8)$$

где  $N$  – число лучей, попавших на изображение;

$\sigma^2$  вместе с пропусканием дает некоторую среднюю оценку влияния паразитного излучения на качество изображения.

Кроме того, опционально вычисляются значения пропускания системы в радиусе  $\frac{\sigma}{2}$ ,  $\sigma$  и  $2\sigma$ . Поскольку оценка влияния паразитного излучения на качество изображения выполняется при малом числе лучей на входном зрачке, то параметры концентрации лучей и пропускания оптической системы позволяют отсеять большинство комбинаций поверхностей, не формирующих блики, что на порядки ускоряет процесс расчета бликов.

### Результаты экспериментов

Расчеты бликов с использованием предложенного алгоритма были проведены для шестилинзового объектива «Двойной Гаусс» с фокусным расстоянием 100 мм и относительным отверстием 1:3 и полем зрения  $2\omega = 28^\circ$ . Конструктивные параметры объектива представлены в таблице. Оптическая схема объектива с номерами поверхностей приведена на рисунке 6.

Конструктивные параметры объектива

№ поверхности	Радиус кривизны	Толщина	Марка стекла	Световой диаметр
1	54.15325	8.746658	SK2	58.45
2	152.5219	0.5		56.28

Окончание таблицы

№ поверхности	Радиус кривизны	Толщина	Марка стекла	Световой диаметр
3	35.95062	14	SK16	48.59
4	$\infty$	3.776966	F5	42.59
5	22.26992	14.25306		29.83
6	$\infty$	12.42813		20.45
7	-25.68503	3.776966	F5	26.37
8	$\infty$	10.83393	SK16	32.93
9	-36.98022	0.5		37.85
10	196.4173	6.858175	SK16	42.62
11	-67.14755	57.31454		43.29
Изображение	$\infty$			49.14

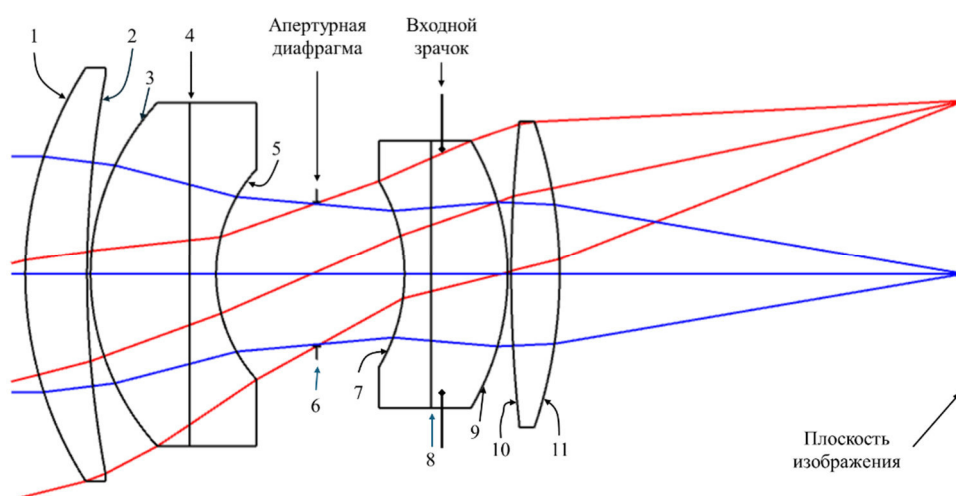


Рисунок 6. Оптическая схема объектива

Входной зрачок диаметром 33.33 мм расположен на расстоянии 58.93976 мм от первой поверхности.

На рисунке 7 представлены результаты расчета бликов с использованием грубой сетки на входном зрачке и с применением интерполяции многократно увеличивающей разрешение сетки. Расчеты производились на компьютере с восьмиядерным процессором AMD Ryzen 9 5000 и 32 Гб оперативной памяти. В данном примере рассчитывался блик второго порядка, возникающий при переотражении между седьмой и пятой поверхностями. Шестая поверхность – апертурная диафрагма. Угол засветки  $\omega = 10^\circ$  попадает в поле зрения объектива. Размер изображения 35x35 мм. Разрешение на изображении 501x501.

Результаты расчета бликов с углами засветки вне поля зрения объектива  $\omega = 20^\circ$  и  $\omega = 30^\circ$  приведены на рисунках 8 и 9 соответственно.

На рисунке 10 представлены результаты сравнения изображений бликов, возникающих при переотражении между 7-й и 5-й поверхностями для угла засветки  $\omega = 10^\circ$  внутри поля зрения. Время расчета изображения, полученного с использованием алгоритма адаптивной интерполяции, составило 1285 миллисекунд, а время расчета с использованием стохастической трассировки – 16 часов. В стохастической трассировке лучей был установлен критерий, который отслеживает только один блик (второго порядка).

На рисунке 11 представлены результаты сравнения изображений бликов, возникающих при переотражении между 11-й и 9-й поверхностями для угла засветки  $\omega = 20^\circ$  вне поля зрения.

Время расчета изображения, полученного с использованием алгоритма адаптивной интерполяции, составило 1007 миллисекунд, а время расчета с использованием стохастической трассировки – 12 часов.



		<p>Параметры расчета</p> <p><math>\omega = 10^\circ</math></p> <p>Сетка на входном зрачке: 32 x 32</p> <p>Трасса: 1 2 3 4 5 6 -7 -6 5 6 7 8 9 10 11</p> <p>Пропускание: <math>\tau = 0.412162</math></p> <p>Центр тяжести: <math>x = 0, y = -13.946711</math></p> <p>СКО от центра тяжести: <math>\sigma_x = 7.329427, \sigma_y = 5.182688</math></p> <p>СКО от центра изображения: <math>\sigma = 15.755357</math></p> <p>Время расчета: 8мс</p>
а)	б)	в)
		<p>Параметры расчета</p> <p><math>\omega = 10^\circ</math></p> <p>Сетка на входном зрачке: 8192 x 8192</p> <p>Трасса: 1 2 3 4 5 6 -7 -6 5 6 7 8 9 10 11</p> <p>Пропускание: <math>\tau = 0.419743</math></p> <p>Центр тяжести: <math>x = 0, y = -13.96032</math></p> <p>СКО от центра тяжести: <math>\sigma_x = 7.522507, \sigma_y = 5.319216</math></p> <p>СКО от центра изображения: <math>\sigma = 15.858076</math></p> <p>Время расчета: 1285мс</p>
г)	д)	е)

Рисунок 7. Формирование блика методом интерполяции: а) исходная (грубая) сетка лучей на входном зрачке; б) изображение блика, построенное по исходной сетке входных лучей; в) условия и результаты расчета блика по грубой сетке; г) интерполированная сетка с увеличенным разрешением; д) изображение блика, построенное по интерполированной сетке; е) условия и результаты расчета блика по интерполированной сетке

		<p>Параметры расчета</p> <p><math>\omega = 20^\circ</math></p> <p>Сетка на входном зрачке: 32 x 32</p> <p>Трасса: 1 2 3 4 5 6 7 8 9 10 -11 -10 9 10 11</p> <p>Пропускание: <math>\tau = 0.125676</math></p> <p>Центр тяжести: <math>x = 0, y = 11.0855</math></p> <p>СКО от центра тяжести: <math>\sigma_x = 7.1966, \sigma_y = 5.0888</math></p> <p>СКО от центра изображения: <math>\sigma = 17.8503</math></p> <p>Время расчета: 8мс</p>
а)	б)	в)
		<p>Параметры расчета</p> <p><math>\omega = 20^\circ</math></p> <p>Сетка на входном зрачке: 8192 x 8192</p> <p>Трасса: 1 2 3 4 5 6 7 8 9 10 -11 -10 9 10 11</p> <p>Пропускание: <math>\tau = 0.124943</math></p> <p>Центр тяжести: <math>x = 0, y = 11.0045</math></p> <p>СКО от центра тяжести: <math>\sigma_x = 7.2516, \sigma_y = 5.1277</math></p> <p>СКО от центра изображения: <math>\sigma = 17.1694</math></p> <p>Время расчета: 1007мс</p>
г)	д)	е)

Рисунок 8. Формирование блика внеполевой засветки  $\omega = 20^\circ$ : а) исходная (грубая) сетка лучей на входном зрачке; б) изображение блика, построенное по исходной сетке входных лучей; в) условия и результаты расчета блика по грубой сетке; г) интерполированная сетка с увеличенным разрешением; д) изображение блика, построенное по интерполированной сетке; е) условия и результаты расчета блика по интерполированной сетке

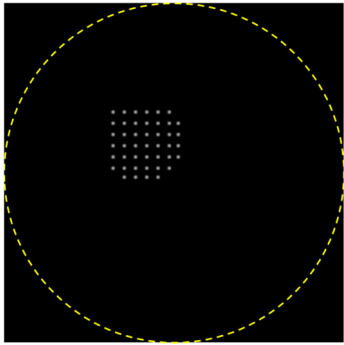
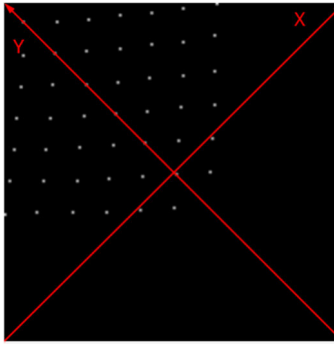
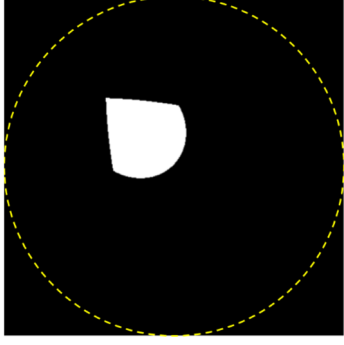
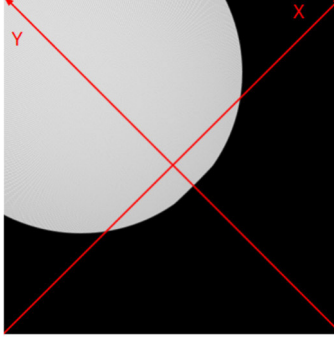
		<p>Параметры расчета</p> <p><math>\omega = 30^\circ</math></p> <p>Сетка на входном зрачке: 32 x 32</p> <p>Трасса: 1 2 3 4 -5 -4 -3 2 3 4 5 6 7 8 9 10 11</p> <p>Пропускание: <math>\tau = 0.062162</math></p> <p>Центр тяжести:</p> <p><math>x = 0, y = 6.406597</math></p> <p>СКО от центра тяжести:</p> <p><math>\sigma_x = 9.217027, \sigma_y = 6.517423</math></p> <p>СКО от центра изображения: <math>\sigma = 12.924495</math></p> <p>Время расчета: 8мс</p>
а)	б)	в)
		<p>Параметры расчета</p> <p><math>\omega = 30^\circ</math></p> <p>Сетка на входном зрачке: 8192 x 8192</p> <p>Трасса: 1 2 3 4 -5 -4 -3 2 3 4 5 6 7 8 9 10 11</p> <p>Пропускание: <math>\tau = 0.062445</math></p> <p>Центр тяжести:</p> <p><math>x = 0, y = 6.378792</math></p> <p>СКО от центра тяжести:</p> <p><math>\sigma_x = 9.501703, \sigma_y = 6.718719</math></p> <p>СКО от центра изображения: <math>\sigma = 13.101922</math></p> <p>Время расчета: 586мс</p>
г)	д)	е)

Рисунок 9. Формирование блика внеполевой засветки  $\omega = 30^\circ$ : а) исходная (грубая) сетка лучей на входном зрачке; б) изображение блика, построенное по исходной сетке входных лучей; в) условия и результаты расчета блика по грубой сетке; г) интерполированная сетка с увеличенным разрешением; д) изображение блика, построенное по интерполированной сетке; е) условия и результаты расчета блика по интерполированной сетке

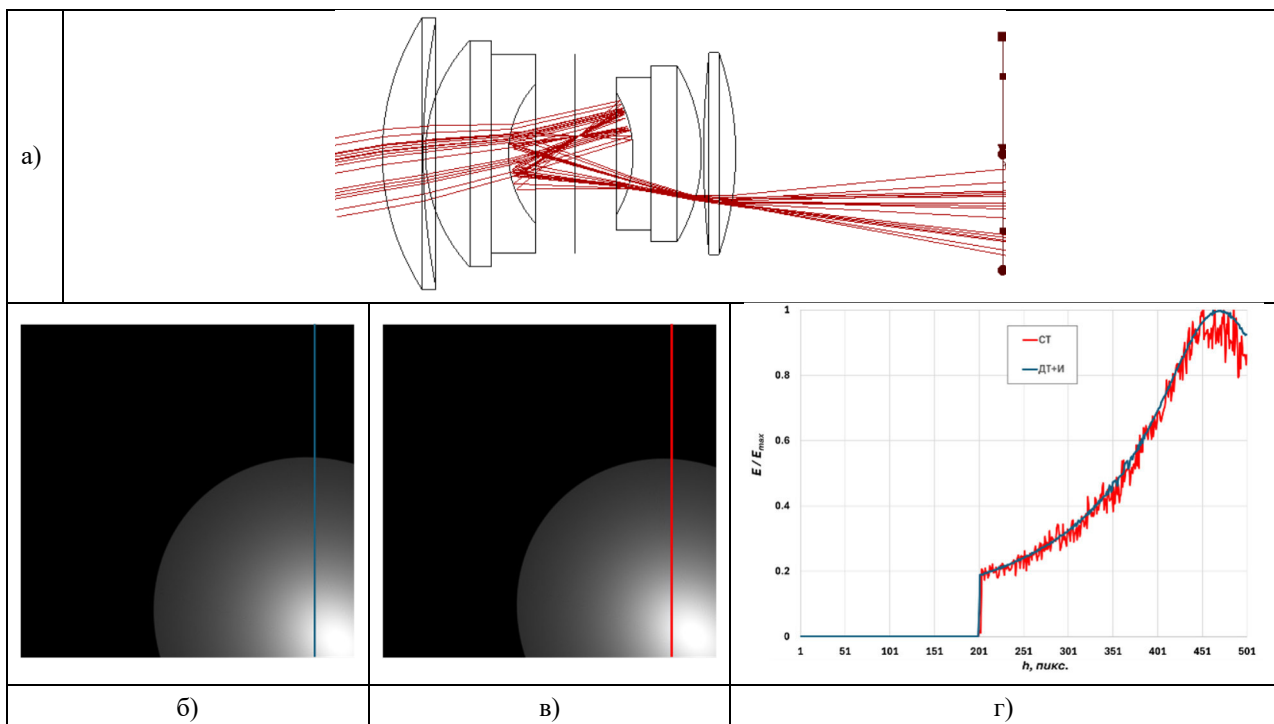


Рисунок 10. Сравнение изображения внутриполевых бликов 2-го порядка для угла засветки  $\omega = 10^\circ$  (а), полученных с использованием алгоритма адаптивной интерполяции (б) и с использованием стохастической трассировки (в), сравнение относительного распределения освещенности в вертикальном сечении (г)

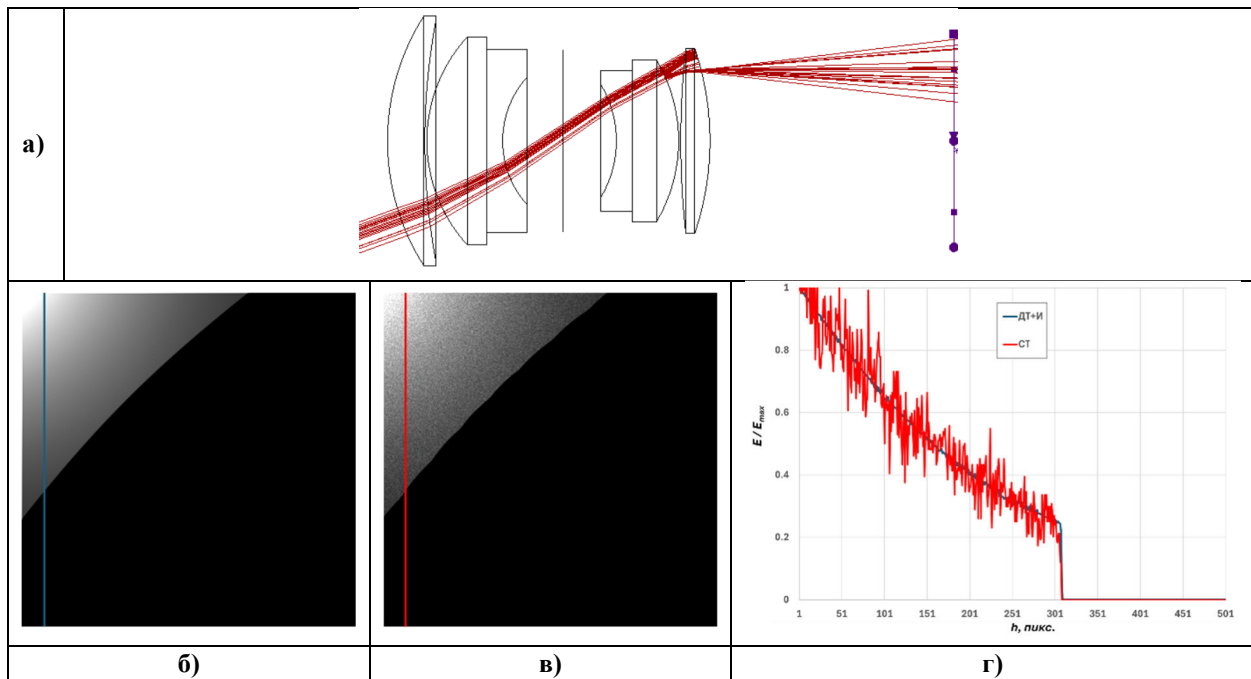


Рисунок 11. Сравнение изображения внутривольных бликов 2-го порядка для угла засветки  $\omega = 20^\circ$  (а), полученных с использованием алгоритма адаптивной интерполяции (б) и с использованием стохастической трассировки (в)

### Заключение

Алгоритм, представленный в данной работе, отличается от большинства существующих решений скоростью расчета и возможностью анализировать отдельные блики, возникающие между двумя и более поверхностями. В отличие от стохастических методов, данное решение требует определения поверхностей, между которыми ожидается появление блика. Несмотря на кажущееся неудобство, данное решение имеет ряд существенных преимуществ перед стохастическими методами, которые с достаточно низкой эффективностью оценивают и смешивают результаты расчета всех возможных вариантов переотражения света в рамках одного вычислительного процесса. Основным преимуществом предлагаемого решения является то, что появляется возможность выделить пары (в ряде случаев четверки) поверхностей, создающих основную паразитную засветку, и оценить их влияние на качество изображения. Кроме того, адаптивный подход к распределению координат начала луча на входном зрачке позволяет значительно ускорить процесс расчета. Высококачественные изображения бликов могут быть получены в сотни (в ряде случаев в тысячи) раз быстрее, чем методами стохастической трассировки лучей. Принимая во внимание адаптивность подхода, анализ блика (т.е. необходимость его расчета между заданными парами поверхностей) может быть выполнен за единицы миллисекунд с точностью оценки его СКО и пропуска до второго знака после запятой. Поэтому данное решение может быть использовано не только как средство анализа блика, но и как часть оптимизационного критерия в процессе автоматического проектирования оптических систем.

### Список литературы

1. Fest, E.C. Stray Light Analysis and Control; SPIE Press: Bellingham, WA, USA, 2013; 212p.
2. Zeng, C.; Xia, G.; Zhong, X.; Li, L.; Qu, Z.; Yang, Q.; Wang, Y. Fast Stray Light Performance Evaluation Based on BSDF and Radiative Transfer Theory. *Sensors* 2023, 23, 9182. <https://doi.org/10.3390/s23229182>
3. Davenport, T. L. R.; Cassarly, W. J.; Crowther, B. Simulation-based computation of the glare spread function in camera systems, Proc. SPIE 12913, Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) V. 12 March 2024, 1291314. <https://doi.org/10.1117/12.3002837>
4. Li, J.; Yang, Y.; Qu, X.; Jiang, C. Stray Light Analysis and Elimination of an Optical System Based on the Structural Optimization Design of an Airborne Camera. *Appl. Sci.* 2022, 12, 1935. <https://doi.org/10.3390/app12041935>
5. Côté, G.; Lalonde, J.; Thibault, S. Deep learning-enabled framework for automatic lens design starting point generation. *Opt. Express* 2021. 29, 3841-3854.
6. Nguyen, D. H.; Bakholidin, A. V. Calculation automation of “cemented doublet with a singlet” and “singlet with a cemented doublet” objectives. *J. Opt. Technol.* 2023. 90, 6-13.

## Исследование методов отслеживания зрачков для систем автономного вождения

Ч. Ван, Д. Д. Жданов, А. Д. Жданов

Университет ИТМО, Санкт-Петербург, Россия

**Аннотация.** В ответ на сложную задачу прогнозирования намерений пешеходов в условиях автономного вождения на городских дорогах в данном исследовании предлагается инновационный метод прогнозирования внимания, основанный на отслеживании движения глаз. Традиционное определение позы всего тела страдает от значительных ошибок прямой видимости ( $\pm 15^\circ$ ) и плохо работает с небольшими, удалёнными целями, например размером глазного яблока всего  $10 \times 10$  пикселей. В отличие от этого, данное решение использует стандартную RGB-камеру и обеспечивает эффективное восприятие благодаря двухрежимной архитектуре. В режиме RAW система напрямую использует Mediapipe FaceMesh для извлечения опорных точек лица и объединяет алгоритм PnP для оценки положения головы. Основным компонентом является усовершенствованная модель оценки взгляда EfficientNet-B0. Для улучшения извлечения периорбитальных признаков реализован облегчённый механизм внимания, а модуль симуляции расстояния предназначен для имитации ухудшения дальних визуальных признаков, что значительно повышает надёжность при низких разрешениях. Модель в режиме RAW обучается на наборе данных MPIIGaze. Экспериментальные результаты показывают, что угловая погрешность в реальном времени для целей на близком расстоянии ( $< 5$  м) составляет менее  $0,1^\circ$ , а алгоритм оценки взгляда позволяет эффективно различать ключевые состояния, такие как «взгляд прямо перед собой», «взгляд вбок» и «голова опущена». По сравнению с моделью лица YOLOv8s, обученной на архитектуре YOLOv8s и наборе данных WIDER FACE, предлагаемый метод обеспечивает повышение скорости примерно на 30 кадров в секунду при обнаружении одной цели и сохраняет преимущество в 3 кадра в секунду при обнаружении нескольких целей ( $\geq 5$  человек). Этот подход представляет собой более перспективную альтернативу традиционным методам определения поз, предлагая недорогую и высокоэффективную поддержку безопасности в реальном времени для принятия решений автономным вождением в сложных городских условиях.

**Ключевые слова:** прогнозирование намерений пешеходов, автономное вождение в городских условиях, отслеживание движения глаз, оценка направления взгляда, низкая угловая погрешность, высокоскоростной вывод, недорогая RGB-камера

## Research on pupil tracking techniques for autonomous driving systems

Z. Wang, D. D. Zhdanov, A. D. Zhdanov

ITMO University, Saint-Petersburg, Russian

**Abstract.** In response to the challenging problem of pedestrian intention prediction in urban road autonomous driving, this study innovatively proposes an attention prediction method based on eye tracking. Traditional full-body pose detection suffers from significant line-of-sight errors ( $\pm 15^\circ$ ) and performs poorly for small, distant targets—for instance, the eyeball may be as small as  $10 \times 10$  pixels. In contrast, this solution employs a standard RGB camera and achieves efficient perception through a dual-mode architecture. In RAW mode, it directly utilizes Mediapipe FaceMesh to extract facial fiducial points and combines the PnP algorithm to estimate head pose. The core component is an improved EfficientNet-B0 gaze estimation model. A lightweight attention mechanism is introduced to enhance periorbital feature extraction, and a distance-simulator module is designed to simulate the degradation of long-distance visual features, thereby significantly improving robustness at low resolutions. The RAW mode model is trained on the MPIIGaze dataset. Experimental results show that the real-time angular error for short-distance ( $< 5$  m) targets is less than  $0.1^\circ$ , and the gaze estimation algorithm can effectively distinguish the key states such as "looking straight ahead", "side looking" and "head down". Compared with the YOLOv8s-face model trained on the YOLOv8s architecture and the WIDER FACE dataset, the proposed method achieves a speed improvement of approximately 30 FPS for single-target detection and maintains a 3 FPS advantage in multi-target detection scenarios ( $\geq 5$  people). This approach represents a more forward-looking alternative to traditional pose detection methods, offering low-cost, high-real-time safety support for autonomous driving decisions in complex urban environments.

**Keywords:** pedestrian intention prediction, urban autonomous driving, eye tracking, gaze estimation, low angular error, high-speed inferenc, low-cost RGB camera

### Introduction

The analysis of human postures represents a key research direction within the current field of artificial intelligence. This area of study originally emerged from the development of robotics. In today's era of rapid artificial intelligence advancement, a fully functional artificial intelligence system should not be confined to server-based operations or voice interaction capabilities. Rather, it should embody a humanoid form and possess autonomous learning abilities. Consequently, human posture detection technology has garnered

increasing attention in recent years. A comprehensive understanding of human posture information is essential for its effective integration into intelligent robotic systems. Moreover, in-depth research on human behavior and postures enables artificial intelligence to more accurately interpret user intentions and even anticipate future human actions. Currently, human action recognition primarily relies on video data, as it offers continuous sequences of motion frames. Artificial intelligence systems detect and classify human subjects within these video sequences, analyzing changes between adjacent frames to identify specific actions. In contrast, posture prediction presents a greater challenge. It requires the system to accurately recognize current postures even when dealing with incomplete time-series data and to forecast subsequent action trends based on this recognition. Therefore, research on posture prediction holds significant practical value and has broad application potential, particularly in the areas of home safety for infants, children, and the elderly. For instance, posture recognition can be employed to issue early warnings for incidents such as falls. Furthermore, this technology significantly supports the development of autonomous driving systems by enhancing vehicles' ability to predict pedestrian behavior.

In existing research, it is already feasible to interpret the command intentions of traffic police by identifying their body postures and further predicting the lane-changing gestures they are about to perform [1]. This capability enables vehicles to avoid congested or under-construction road sections and prevents them from entering incorrect driving routes. Although current traffic police action recognition is primarily limited to a few predefined postures—typically identified by detecting key joint positions and matching them with pre-stored action templates in a database—it is evident that this technology holds significant potential for future development. At present, research and applications of autonomous driving systems are becoming increasingly mature. On highways, Level 3 assisted driving functionality has been largely realized, allowing drivers to allocate minimal attention to perform basic autonomous driving tasks. However, autonomous driving technology still encounters numerous challenges in urban road environments and has yet to achieve substantial breakthroughs. The primary reason lies in the abundance of unpredictable dynamic factors present in urban settings, such as pedestrians and cyclists, whose behaviors often lack regular patterns and are difficult to model accurately using traditional methods. Additionally, compared to highways, urban areas contain a high density of intersections, including alleys and entrances to residential zones. These locations are numerous, complex, and highly variable, making autonomous navigation riskier and more challenging. If posture detection technology can be leveraged to accurately predict the future actions of pedestrians, it would significantly enhance the adaptability of autonomous driving systems in urban environments, thereby enabling broader real-world applications.

Currently, autonomous vehicles generally employ a method based on the extraction of overall target features to detect pedestrians. By utilizing the significant differences in shape, size, and color between pedestrians and other traffic participants, semantic segmentation is performed to identify pedestrians [2]. However, this approach often results in a relatively large region of interest (ROI), which introduces a substantial amount of redundant data. This, in turn, increases computational load and reduces processing efficiency. Therefore, narrowing the ROI has become crucial for improving detection efficiency. This study proposes a novel approach that focuses on feature extraction from the head and facial regions of pedestrians. On one hand, since adults are typically taller than 1.2 meters and children are usually accompanied by adults, the area below 1.2 meters from the ground can be reasonably ignored, allowing detection efforts to be concentrated on regions above this height. The head region not only possesses distinct features that differentiate pedestrians from other objects on the road, making it easier to identify, but also provides valuable information. Specifically, the head point cloud data obtained through radar can offer an initial estimation of the pedestrian's orientation. This enables the system to infer whether the pedestrian is aware of the autonomous vehicle and whether sudden movement is likely, which could affect the vehicle's path planning. This method not only accelerates scene modeling during the perception stage but also enhances the accuracy and safety of the vehicle's path planning and risk prediction to a certain extent.

### **Statement of the problem**

In previous studies, eye-tracking technology has been widely utilized as a critical method for monitoring the driver's state. By accurately capturing key ocular features, it provides quantitative data to support fatigue detection and attention evaluation. Currently, a prevalent approach employs a cascaded detection framework: first, a robust facial detection algorithm (e.g., YuNet) identifies the facial region; subsequently, the eye ROI is localized

within this area; finally, a gaze estimation model (e.g., the MPIIGaze training framework) generates a three-dimensional gaze direction vector. Compared to the conventional HOG (Histogram of Oriented Gradients) combined with SVM (Support Vector Machine) method, this architecture demonstrates significant improvements in real-time performance and adaptability to complex environmental conditions [3]. Particularly in challenging scenarios involving illumination variations or head pose deviations, it maintains reliable detection stability.

The research presented in this paper introduces eye tracking technology into the detection of pedestrians outside the vehicle. In urban traffic scenarios, the ability to assess pedestrians' attention states directly influences the decision-making safety of autonomous vehicles. The system faces three primary challenges. First, long-distance detection is critical for safety: when a vehicle is traveling at 60 km/h on urban roads, it requires a braking distance of 20–30 meters. Therefore, the detection range must be at least 30 meters to ensure sufficient response time. However, at such distances, the eye region may occupy only  $10 \times 10$  pixels, rendering traditional eye keypoint detection ineffective. Second, the system must accurately analyze the direction of gaze, even when the head is turned. Third, it must determine whether the pedestrian is aware of the approaching vehicle. Conventional approaches rely on full-body pose estimation methods, such as OpenPose, but these techniques can result in gaze direction errors as high as  $\pm 15^\circ$ . To address these challenges, this paper integrates the gaze estimation model MPIIGaze and transfers its robust gaze decoupling capability into the in-vehicle external perception system.

### Object Detection in Autonomous Driving

The key to autonomous driving technology lies in enabling vehicles to accurately perceive their surrounding environment and make autonomous decisions. Within this system, the object detection algorithm serves as the core component of the perception module, directly influencing the vehicle's understanding of its surroundings and its ability to respond effectively. In recent years, with continuous advancements in deep learning and multi-modal sensing technologies, object detection methods in autonomous driving have evolved from initial single-image recognition approaches to integrated architectures that fuse data from multiple sensors. This evolution has enabled systems to efficiently identify and accurately locate objects such as vehicles, pedestrians, and obstacles in complex and dynamic traffic environments. However, challenges such as variations in lighting conditions, adverse weather, object occlusion, and the detection of small-sized objects remain significant technical hurdles. Currently, commonly used object detection methods in autonomous driving include the You Only Look Once (YOLO) algorithm [4], two-stage detection models (e.g., Fast R-CNN) [5], Transformer-based architectures [6], multi-sensor fusion models [7], and end-to-end (E2E) learning models [8].

The environmental perception capability of autonomous driving systems largely relies on the collaborative integration of multi-modal sensors and the input of high-quality data. Currently, mainstream autonomous driving vehicles are typically equipped with cameras, light detection and ranging (LiDAR) systems, millimeter-wave radars, and infrared sensors, forming a perception architecture that captures a wide range of physical characteristics. Specifically, cameras capture detailed texture and color information, LiDAR generates high-precision three-dimensional spatial point cloud data, millimeter-wave radars maintain stable performance under adverse weather conditions, and infrared sensors enhance visual perception in low-light environments. This multi-sensor fusion framework not only improves the system's adaptability to complex and dynamic environments but also enhances the redundancy of perceptual data. However, it also introduces significant challenges in the spatio-temporal alignment and fusion of heterogeneous data sources.

Table 1. Characteristics of Common Sensors in Autonomous Driving Systems

Type	Data	Advantageous scenarios	Main limitations	Sampling rate
Visible light camera	RGB array	Texture recognition, semantic understanding	Light-sensitive and weak night vision	30-60 FPS
LiDAR	3D point clouds	Precise distance measurement, 3D reconstruction	Rain and snow scattering, high cost	10-20 Hz
Millimeter-wave radar	Distance-Doppler matrix	Penetrate rain and fog, speed measurement	Low resolution, no height information	5-20 Hz
Infra-red sensor	Thermal radiation intensity map	Nighttime target, biological detection	Colorless texture, affected by heat sources	30-60 FPS



In terms of data representation, object detection algorithms must address the challenge of mapping between two-dimensional image space and three-dimensional real space. The point cloud data acquired by LiDAR is typically represented in the form of three-dimensional coordinates:

$$P = \{p_i \in \mathbb{R}^3 \mid i = 1, \dots, N\}.$$

Among these points, each point  $p_i$  is represented by its spatial coordinates  $(x, y, z)$ , as well as the corresponding reflectance intensity  $I$ . The camera data is then projected onto the image plane based on the pinhole camera model:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R \quad T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

The key to multimodal fusion is establishing an effective spatio-temporal alignment mechanism across different sensors. For instance, calibration enables the projection of LiDAR point clouds onto the image plane, generating depth-enhanced features.

### 1. Current Object Detection Algorithm

In recent years, the perception module of autonomous driving systems has undergone rapid development, evolving into a multi-paradigm framework that integrates visual, LiDAR, and multi-sensor fusion techniques. Current mainstream object perception algorithms can be broadly divided into four categories.

Vision-driven 2D/3D detection has become the most widely adopted approach due to the maturity of deep learning in computer vision. The YOLO series, as a representative, leverages end-to-end convolutional neural networks to achieve real-time inference with high accuracy. Its enhanced variants, such as YOLOv8n-CSS with attention mechanisms and multi-scale pooling, significantly improve the recognition of small objects in complex urban environments [9]. In parallel, stereo vision extends monocular detection by estimating object depth through binocular disparity maps, thereby enhancing detection performance under nighttime or low-light conditions [10].

Point cloud-based 3D object detection plays a central role in leveraging LiDAR data. Voxelization-based methods, exemplified by PointPillars, convert sparse point clouds into pseudo-images for efficient processing with 2D CNNs. Alternatively, point-based methods such as PointNet++ directly learn hierarchical local features from raw point sets, enabling more precise geometric representation. Despite their accuracy, these methods remain sensitive to weather-induced noise, necessitating domain adaptation and point cloud denoising strategies to maintain robustness [11].

Multi-sensor fusion frameworks are increasingly recognized as indispensable for achieving reliable perception in diverse real-world scenarios. By integrating complementary sensing modalities—cameras for texture, LiDAR for geometry, millimeter-wave radar for robustness in adverse weather, and infrared sensors for night vision—fusion strategies at the data, feature, or decision level provide redundancy and adaptability. Pre-fusion approaches, such as MV3D, combine depth maps with RGB features before convolutional processing, while post-fusion methods integrate independently processed detection outputs via optimization algorithms like Non-Maximum Suppression.

Bird's-Eye View (BEV) perception represents a novel paradigm that unifies heterogeneous sensor inputs into a top-down spatial representation [12]. BEV facilitates global consistency in target localization and scene understanding, thereby supporting advanced planning and decision-making modules [13]. Industrial implementations such as Tesla's HydraNet and XPeng's NGP 3.0 exemplify the state-of-the-art, combining image features, LiDAR point clouds, and high-definition maps to construct dynamically updated BEV grids with strong cross-view association.

In summary, the mainstream perception algorithms for autonomous driving are converging toward hybrid frameworks that balance the efficiency of visual methods, the precision of LiDAR-based 3D detection, and the robustness of multi-modal fusion. The integration of BEV-based representations further enhances spatial awareness, paving the way for end-to-end perception–planning pipelines in next-generation autonomous driving systems.

### 2. RAW Model

The RAW model begins with the acquisition of camera frames and proceeds through a series of processing stages, including multi-person facial key point and iris point extraction, eye-axis-oriented cropping and

normalization, a lightweight batch forward pass through the gaze estimation network, direction determination via the fusion of geometric and learning-based signals, stabilization using hysteresis and multi-frame consistency, and finally, visualization via discretized small arrows and log recording—thereby forming a fully end-to-end real-time system. First, MediaPipe FaceMesh (with `refine_landmarks` enabled) is employed to extract 468 facial landmarks and four high-precision iris coordinates per eye for each frame. For any detected face, let the pixel coordinates of the outer and inner corners of the right eye be denoted as  $P_{out}^r$  and  $P_{in}^r$ , respectively, and those of the inner and outer corners of the left eye as  $P_{in}^l$  and  $P_{out}^l$ , respectively. Based on these points, the "horizontal axis" vectors for both eyes are then constructed and normalized:

$$u_r = \frac{P_{in}^r - P_{out}^r}{\|P_{in}^r - P_{out}^r\|}, \quad u_l = \frac{P_{out}^l - P_{in}^l}{\|P_{out}^l - P_{in}^l\|}.$$

To ensure that the subsequent network processes eyes with a consistent orientation, we construct an affine transformation matrix  $M$  at the midpoint  $c = \frac{1}{2}(P_1 + P_2)$  between the eyes, based on the rotation angle  $\theta = \tan^{-1}2(\Delta y, \Delta x)$  of the line connecting the eye corners. An "orientation-aware cropping" is then applied to a square local window using the function `warpAffine(·;M)`. The window's side length is determined proportionally to the inter-pupillary distance and constrained within a reasonable range. After aligning both eye regions, each local image is converted to grayscale, resized to  $224 \times 224$ , and linearly normalized using  $(x - 0.5)/0.5$ . These images are then stacked across all samples and both eyes into a batch tensor of shape  $[N, 1, 224, 224]$ . We adopt the 'channels\_last' memory format and utilize CUDA's 'autocast' for half-precision computation. A single forward pass is executed under 'torch.inference\_mode()' to minimize CPU/GPU data transfer overhead and scheduling latency. This preprocessing pipeline ensures that the network consistently receives eye patches with aligned geometric poses and uniform scales, thereby significantly reducing domain shift caused by minor variations in head pose.

The gaze regression network adopts EfficientNet-B0 as the backbone, with the first convolutional layer modified to accept single-channel grayscale input. At a mid-level feature layer with 24 channels in the backbone, a  $1 \times 1$  gating attention mechanism is introduced. This mechanism first compresses the channel dimension using a  $1 \times 1$  convolution followed by a ReLU activation, and then applies another  $1 \times 1$  convolution followed by a Sigmoid activation to generate channel-wise weights. These weights are then applied via element-wise multiplication with the original feature maps, effectively emphasizing channels that contain stronger eye anatomical cues. Following this, a  $1 \times 1$  "distance simulation" convolution is employed to reduce the feature dimensionality to 256 channels. After applying adaptive pooling to resize the features to  $4 \times 4$  and flattening the output, two fully connected layers are used to regress the 3D gaze vector. Finally, L2 normalization is applied to the output to ensure a consistent vector magnitude, representing a directional signal. To align the "looking straight ahead" direction across individuals in the vector space, the system implements a "center calibration" procedure. Upon triggering the calibration, the system computes the average gaze vector from a short sequence of recent frames. This vector is then rotated to align with the forward direction of the camera using Rodrigues' rotation formula. The resulting vector is transformed into screen coordinates, where the x-axis increases to the right and the y-axis increases upward. The resulting offset is stored as the individual's zero-point reference. For each subsequent frame, the network output is first aligned to this reference before computing the final gaze angle and direction.

The vertical direction is entirely determined by the pitch angle derived from the network. Let the aligned unit vector be denoted as  $s = (s_x, s_y, s_z)$ , and the pitch angle is represented by

$$pitch = \tan^{-1}2(s_y, s_z) \cdot \frac{180}{\pi}.$$

Subtract  $pitch_0$  obtained during center calibration from the current measurement to obtain  $\Delta pitch$ . To suppress critical jitter, the system defines distinct entry and exit thresholds for the Up and Down directions, thereby creating a hysteresis loop. A certain direction is entered only when the entry threshold is reached, and to return to a non-target state, the signal must fall back below the tighter exit threshold. This mechanism effectively prevents rapid oscillations such as "up/center/up/center" caused by micro-vibrations and environmental noise in real-world conditions.

For the left - right direction, instead of directly using the yaw angle of the network, the geometric "iris offset difference method" is adopted. Inside each eye, taking the line connecting the corners of the eye as

the axis, the centroid of four iris points is calculated and projected onto this axis. For the right eye, the positive direction is from the outer corner to the inner corner, and for the left eye, the positive direction is from the inner corner to the outer corner. Thus, two dimensionless offset quantities "relative to the mid - point of the eye corners" are obtained, which respectively describe the "inward" and "outward" displacements. In order to symmetrically fuse the information of both eyes, the system defines the horizontal difference scalar:

$$d_h = \frac{1}{2}(c_r - c_l).$$

Among them,  $c_r$  denotes the offset of the right eye with "inward as positive", and  $c_l$  denotes the offset of the left eye with "outward as positive". Therefore,  $d_h > 0$  indicates a rightward gaze, while  $d_h < 0$  indicates a leftward gaze. Considering the potential interference caused by pixel noise and eyelid occlusion,  $dh$  is not used directly. Instead, it is passed through a first-order IIR low-pass filter to achieve signal smoothing. The filter is updated only when the eye aspect ratio (EAR) exceeds a predefined threshold, which indicates that the eyes are fully open. The filtering process follows the equation:

$$\widetilde{d}_h[k] = (1 - \alpha)\widetilde{d}_h[k - 1] + \alpha d_h[k].$$

Among them,  $\alpha$  denotes the smoothing coefficient. The EAR (Eye Aspect Ratio) measures the degree of eye opening and closing by calculating the ratio of "the average vertical distance between the upper and lower eyelids to the horizontal distance between the outer eye corners". This approach helps prevent misjudgment in cases where facial keypoints are detected but the iris is temporarily occluded. The system defines the "direct gaze central zone" as the region where  $|\Delta pitch|$  is within a predefined pitch bandwidth and  $|\widetilde{d}_h|$  is within a specified horizontal bandwidth. If the gaze falls within this central zone, the current frame's  $\widetilde{d}_h$  value is added to the "direct gaze sample pool". The standard deviation  $\sigma$  of this sample pool is then used to determine an adaptive threshold: entry thresholds for left and right directions are set based on  $K\sigma$ , while the exit threshold is defined as a fixed proportion of the entry threshold. This mechanism enables automatic adjustment of sensitivity to match individual noise levels, balancing responsiveness ("easy to trigger") with stability ("not overly sensitive").

When the up-down and left-right channels each provide candidate directions, the system does not simply follow the principle of "whoever triggers first prevails". Instead, it calculates the intensity of each relative to its respective entry threshold and selects the stronger side as the "principal axis". When the two sides are similar, the system preferentially retains the principal axis from the previous frame to reduce the back-and-forth flickering of the "diagonal state". A candidate direction must remain consistent across several consecutive frames before it is confirmed as the final label. This multi-frame confirmation mechanism is particularly effective in mitigating false transitions caused by glare, slight head movements, or brief keypoint jitter. Considering that most front-facing cameras display images in a mirrored manner, the system applies mirroring compensation at the left-right label level to align the on-screen arrow direction with the user's visual intuition. During the visualization stage, the continuous three-dimensional gaze vector is no longer used to render the "tracking arrow", as this would amplify and display the system's internal noise through arrow oscillation, potentially causing misleading visual illusions. The current approach directly maps the final labels into small discrete arrows on the screen plane: Left is represented by a short, stable left-pointing arrow, Right by a right-pointing arrow, Up by an upward arrow, and Down by a downward arrow. The arrow length is adjusted slightly based on the "intensity," albeit within a very narrow range. For the Direct mode, no arrow is displayed; instead, only the eye key points and a minimalist HUD are shown. This maintains the intuitive perception of "appearing to point at a location" while minimizing psychological discomfort caused by visual jitter. In the upper-left corner of the interface, only the current model mode, the two-second sliding window average FPS, and the number of detected faces are displayed. All other intermediate variables are logged into a CSV file, facilitating subsequent comparisons of robustness, angular error, and efficiency across different parameter configurations.

To maintain time consistency in a multi-person scenario, the system maintains a set of lightweight trajectories. Each trajectory uses the "eye center" from the previous frame as an anchor point. In the new frame, this anchor point is matched to the newly detected "eye center set" using the nearest neighbor method. If the matching distance exceeds a predefined threshold, a new trajectory is initiated. For existing trajectories that fail to find a match, the "missing count" is incremented, and if it exceeds a specified upper limit, the trajectory is discarded. Each trajectory independently stores the zero offset, threshold, adaptive statistics, and internal

filter state for the corresponding individual. As a result, even when multiple individuals alternate in and out of the frame, the calibration and threshold parameters remain free from mutual interference or "contamination." Additionally, the code includes a "demo mode" that utilizes ONNX face detection based on YOLOv8s, which runs on the GPU when the CUDAXecutionProvider is available. However, the main processing pipeline continues to focus on the iris and eyelid geometry provided by FaceMesh, as these are essential for the "iris offset difference method". In terms of performance, the main bottlenecks occur in FaceMesh on the CPU side and in image preprocessing. To mitigate this, we batch all individuals' eye patches and feed them into the model collectively. On the GPU side, we enable channels\_last memory layout, half-precision computation, and cuDNN algorithm optimization, while limiting the number of OpenCV threads to allocate computational resources more efficiently toward forward passes. In the background, the system continuously accumulates per-frame processing times to estimate FPS. It records various metrics—including average FPS, angular error between the alignment vector and the forward axis, a proxy for robustness and false detection rate, directional binary accuracy, and the ratio of the minimum eye area to the total frame area—into a CSV file. Combined with several interactive components such as center calibration, mirror compensation, and mode switching, the entire pipeline is capable of not only sensitively responding to changes in gaze direction but also stably outputting discrete, application-friendly semantic outputs such as "Up/Down/Left/Right" in real, noisy, and multi-person front-camera environments. In conclusion, the RAW model offers several advantages over the currently prevalent object detection models.

Table 2. The main innovations of the RAW model

Model design	RAW	Others
Input and Key Points	Directly employed MediaPipe FaceMesh, which consists of 468 facial landmarks and 4 iris landmarks, for real-time processing of the video stream.	Make more use of pre-annotated datasets; key points/irises are often used in offline annotation or weak supervision.
Eye pretreatment	"Crop along the eye axis", scaled equally to 224×224 grayscale patches.	Fixed-frame cropping or normalization to the eye/head coordinate system.
Model and Input	EfficientNet-B0 single-channel + 1×1 gated attention + lightweight regression head; output unitized 3D vector.	Multi-channel inputs such as ResNet/EfficientNet/Hourglass; sometimes combined with head pose input.
Adaptive Thresholding	$d_h$ standard deviation was counted in the "direct-looking center band", and the left and right thresholds were adapted to the center bandwidth.	Relying on fixed thresholds or data-driven confidence thresholds.
Anti-blinking / Anti-obscuring	The EAR threshold was used to suppress $d_h$ update. Keep the last stable state.	Time-domain filtering or robust loss.

### Experimental results

During the evaluation phase, the system is implemented in two modes: the RAW mode and the YOLO mode. In the RAW mode, FaceMesh is directly applied to the full image to estimate the gaze vector. In contrast, the YOLO mode utilizes YOLOv8s-face to detect the pedestrian's facial region before gaze estimation. From the perspective of autonomous driving applications, this pupil tracking approach is anticipated to deliver substantial performance enhancements. First, in terms of computational efficiency, focusing on the eye region significantly reduces the size of the ROI, thereby lowering the computational overhead associated with feature extraction. On the experimental platform based on the NVIDIA RTX 4070 Ti chip, the system achieves stable processing speeds of over 60 FPS for single-target detection, which is critical for real-time perception in in-vehicle systems. According to NVIDIA's architectural documentation, the INT8 throughput of the Tensor Core is approximately eight times that of FP32. Based on this specification, the equivalent computing power of the NVIDIA RTX 4070 Ti can be estimated at approximately 360 TOPS.

Table 3. Comparison of chip computing power

Chip	INT8 TOPS	Power (W)	TOPS/W
RTX 4070 Ti	360	285	1.26
NVIDIA Orin	254	60	4.23
Tesla FSD (HW3)	144	72	2.00
Mobileye EyeQ Ultra	176	100	1.76

After conversion based on computing power ratio, taking the widely adopted NVIDIA Orin chip as an example, the single target detection speed of this algorithm when deployed on an automotive-grade chip is also greater than 40 FPS.

The RAW model demonstrated clear advantages in speed and stability when handling a single target. In RAW mode, the average frame rate reached 64 FPS, as shown in Figure 1. Since the maximum frame rate of the camera used in the experiment is limited to 30 FPS, the 64 FPS value listed here represents the frame rate achieved without accounting for camera blockage.

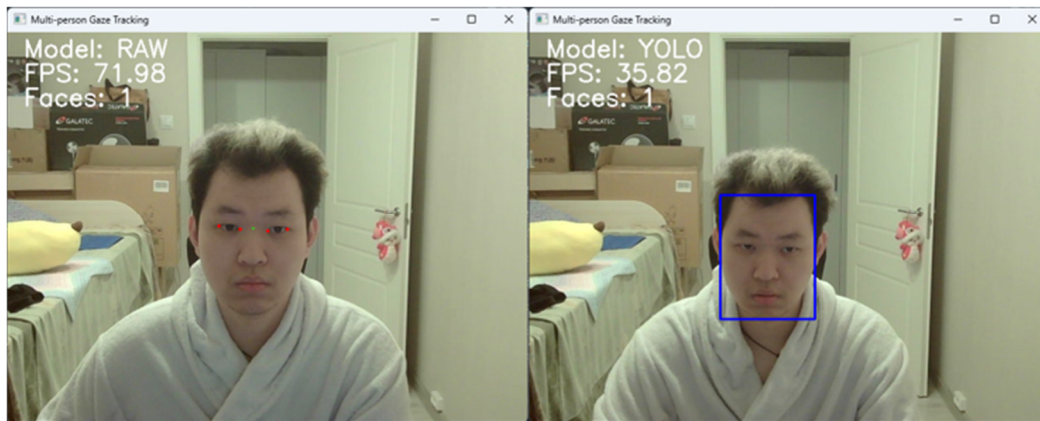


Figure 1. RAW model and YOLOv8s-face

In the YOLO model family, the YOLOv8s-face model trained on the WIDER FACE dataset demonstrates the best performance. To further enhance the model's robustness during training, we incorporated data augmentation techniques such as random rotation, motion blur, and adjustments to brightness and contrast. As a result, the model achieved an average frame rate of 32 FPS in the single-object detection experiment. As illustrated in Figure 2, the frame rate comparison chart visually highlights the substantial differences in real-time performance between the two approaches.

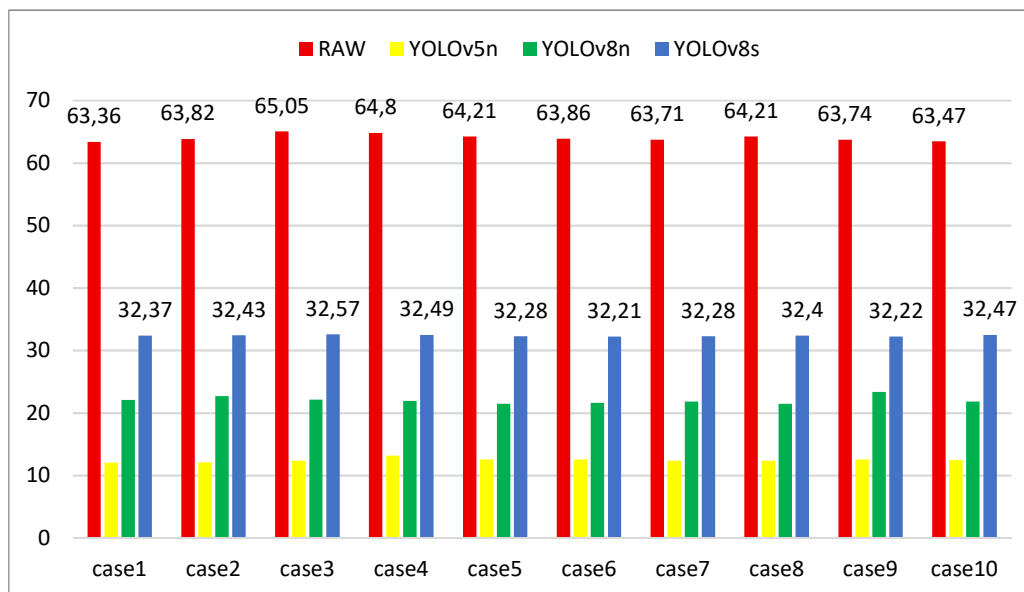


Figure 2. Comparison with other models (single object)

This difference suggests that in single-person or low-density population scenarios, pure eye tracking can substantially reduce computational demands while still fulfilling the real-time processing requirements of autonomous driving systems. Conversely, in environments involving multiple individuals or complex backgrounds, the performance of the RAW mode degrades more significantly due to differences in computational methodology. In multi-target real-time detection scenarios, the RAW mode relies on Mediapipe FaceMesh, which operates through a two-stage process: first, the lightweight BlazeFace model detects and localizes facial regions across the entire frame. Subsequently, for each detected target, 468 facial keypoints are independently regressed, followed by cropping of the eye region, which is then fed into the gaze estimation network. In single-person cases, this procedure exhibits nearly constant time complexity. However, as the number of targets increases, the operations involved in keypoint regression and ROI cropping accumulate linearly, leading to an overall time complexity that grows proportionally with the number of detected individuals:

$$T_{RAW}(N) = T_{det} + N \cdot (T_{mesh} + T_{gaze}).$$

Among them,  $T_{det}$  represents the global face detection time consumption, while  $T_{mesh}$  and  $T_{gaze}$  denote the time required for single-object keypoint regression and gaze prediction, respectively. When  $N$  is relatively large, the linear term becomes dominant, leading to a significant decrease in frame rate. Additionally, FaceMesh incorporates an internal temporal smoothing mechanism to stabilize the keypoint trajectory, utilizing a forward rolling window to fuse features across multiple frames. Consequently, during sudden increases in frame volume or in the presence of multiple objects, cache read/write operations and weighted updates further contribute to the computational load and introduce additional latency.

In contrast, the YOLO model employs a single-stage dense prediction architecture, in which the entire detection process involves only a single global feature extraction and multi-scale prediction. Consequently, its time complexity is approximately constant and can be roughly expressed as

$$T_{YOLO}(N) \approx T_{backbone} + T_{neck} + T_{head} + T_{NMS}.$$

Among them, the target number  $N$  introduces only a slight linear overhead during the NMS stage, which is typically much lower than the dominant computational cost associated with the feature extraction phase. Consequently, even when detecting dozens of targets in dense scenes, the frame rate of YOLO remains largely stable.

### Discussion of results

The experimental results validate the aforementioned analysis. In a single-target scenario, utilizing `frame_time` to measure the data prior to `imshow/waitKey` (without accounting for potential camera blockage), the frame rate in RAW mode can achieve approximately 60 FPS. The angular error remains stable at around  $0.05^\circ$ , and due to temporal smoothing, the key-point stability is relatively high. The algorithm effectively controls the error range through the fusion and normalization of binocular gaze vectors. This enables reliable discrimination of key attention states such as "looking ahead", "looking sideways", or "looking down", as shown in Figure 3. Both RAW and YOLO modes are extremely robust in the single-player case, with no keypoint loss.

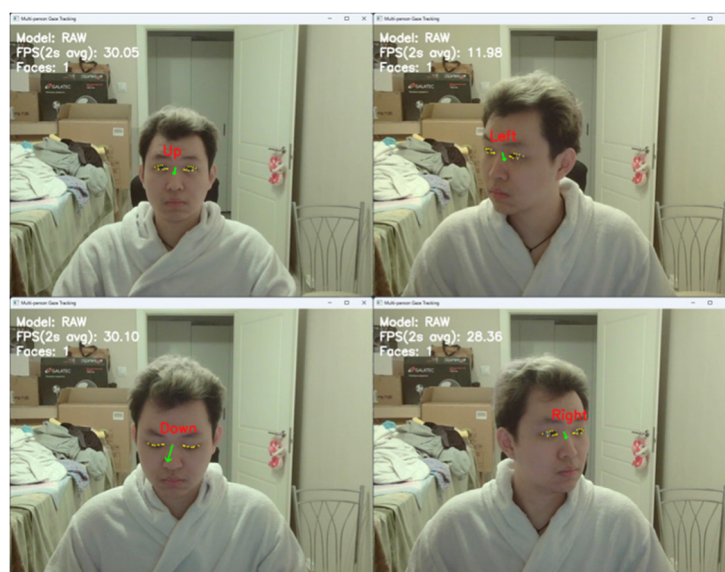


Figure 3. Key gaze direction recognition



In contrast, in the multi-target scenario ( $\geq 5$  people), the frame rate of the RAW mode decreases to approximately 23 FPS, primarily due to the increased computational load from per-target key-point regression and the rolling window mechanism, but an accurate determination of the line-of-sight direction is still achieved, as illustrated in Figure 4. In multi-object detection, the YOLO mode does not introduce keypoint detection to obtain robustness of 1 for the number of detected faces/actual number of faces. In RAW mode, Boolean indications are adopted to assess whether the system is working properly. If FaceMesh has detected at least one face in a frame, it will be 1, otherwise it will be 0, some keypoint information will be lost when the angle is large. But the combined availability coverage coefficient will be no less than 0.98.

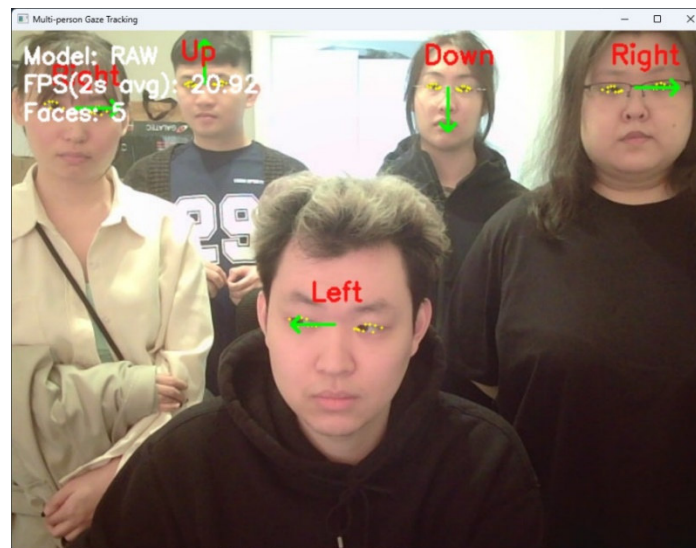


Figure 4. Gaze direction in multi-object scenariosGaze direction recognition in multi-object scenarios

Under the same conditions, the YOLO mode (YOLOv8s-face) exhibits a frame rate that remains relatively stable at around 30 FPS. The reduction from the original 32 FPS is minimal, and the accuracy in detecting the number of people remains largely unchanged, as illustrated in Figure 5.

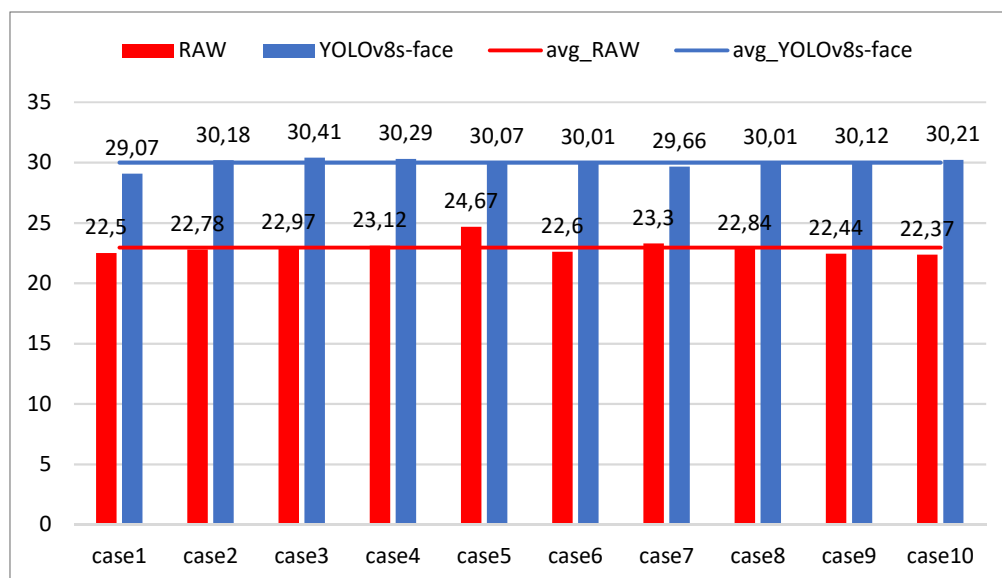


Figure 5. Comparison with YOLOv8s-face (multi-object)

Although the performance of the RAW mode degrades significantly in multi-target detection, it still outperforms the YOLO mode by maintaining a frame rate that is approximately 2 FPS higher. This clearly demonstrates the objective advantages of the RAW mode over the YOLO mode. Moreover, compared with YOLO, which only detects the faces of pedestrians for early warning, the RAW mode calculates the line of sight direction based on full-face recognition. As a direct manifestation of a pedestrian's attention, the line of sight

direction can predict potential intentions 2 to 3 seconds before a jaywalking behavior occurs. This provides a greater lead time than methods relying solely on human key points or movement trajectories, thereby enhancing risk prediction capabilities. In addition, in multi-target scenarios, the system can aggregate the gazing directions of multiple pedestrians to generate an "attention heat map," which assists the decision-making module in identifying the focal points of crowd attention. This feature is particularly valuable in complex environments such as traffic light intersections. More importantly, this method operates using only standard RGB cameras, without requiring additional infrared light sources or expensive dedicated hardware. It offers low cost and high scalability, making it suitable for large-scale deployment across various autonomous driving platforms, including passenger cars and urban buses.

In terms of accuracy, the average angular error of the model on the MPIIGaze validation set is approximately  $13^\circ$  to  $14^\circ$ , which is comparable to that reported in public studies such as Gaze360 and ETH-XGaze [14], [15]. In actual experimental settings, under ideal conditions including sufficient distance, good lighting, and a simple surrounding environment, the average angular error was found to be extremely small—only about  $0.05^\circ$ . However, even in less controlled, real-world scenarios such as vehicle-mounted applications. Such capability is crucial for autonomous vehicles to anticipate pedestrian intentions and make timely decisions regarding deceleration or avoidance.

Meanwhile, the issue mentioned previously—that the pedestrian's eye area may be smaller than  $10 \times 10$  pixels at long distances—was also confirmed in the experiment. As shown in Figure 6, a standard 1080P computer camera was used, under which the eye region in the image was already significantly small. Specifically, the recorded eye area measured only  $22 \times 1$  pixels when the subject was 3 meters away from the camera. Although the current performance is far from achieving the 30-meter recognition distance mentioned at the beginning of the paper, this study still offers novel insights into vehicle and pedestrian target detection systems. Future work should include experiments using high-resolution cameras and further algorithmic improvements to continuously optimize the model, with the ultimate goal of attaining performance that meets the requirements of in-vehicle systems.



Figure 6. The size of the eye area at the maximum distance-3 meters

It is worth noting that the mPA (mean Pixel Accuracy/mean Point Accuracy) metric, commonly utilized in facial keypoint detection or segmentation tasks, is not included in the primary comparison. This exclusion is due to the fact that the central objective of this study is 3D gaze direction prediction. Furthermore, the keypoint detection component employs the pre-trained Mediapipe FaceMesh model without any retraining or fine-tuning, and the accuracy of its detected keypoints has been rigorously validated in the original paper [16]. Additionally, there is no strong linear relationship between keypoint localization accuracy and the final gaze angle error. Hence, incorporating mPA into the evaluation would not provide an accurate reflection of the system's actual performance in gaze estimation. Consequently, this paper emphasizes the angle error as a direct indicator of gaze prediction quality, and complements it with metrics such as frame rate and robustness to offer a comprehensive assessment of the system's practicality and stability. Future work may consider integrating metrics like mPA or NME (Normalized Mean Error) to enable a more detailed analysis of error sources within the system.

## Summary

In response to the challenges of pedestrian intention prediction in urban road autonomous driving, this study innovatively proposes an attention prediction method based on pupil tracking. An efficient perception framework is achieved through a dual-mode processing architecture: in RAW mode, Mediapipe FaceMesh is directly utilized to extract full-image facial key points and estimate head pose; in YOLO mode, YOLOv8s-face is employed to pre-detect pedestrian facial regions. The core of this method is an improved EfficientNet-B0 gaze estimation model. A lightweight attention mechanism is integrated to enhance feature responses around the eyes, while a distance-simulator module is introduced to simulate the degradation of long-distance visual features, thereby significantly improving robustness under low-resolution conditions. Experimental results demonstrate that in short-distance scenarios (<5 meters), RAW mode achieves an ultra-low angular error of  $0.05^\circ$  and operates at a real-time frame rate of 60 FPS for a single target, effectively distinguishing critical states such as "looking straight," "looking sideways," and "bowing the head." In multi-target scenarios involving five or more pedestrians, the system still maintains a frame rate of 30 FPS, which is 10 FPS faster than YOLO mode. The proposed method requires only a standard RGB camera, eliminating the need for infrared equipment. When deployed on the NVIDIA Orin automotive chip, the system achieves a processing speed of over 40 FPS and enables the prediction of pedestrian intentions to cross the road 200–300 milliseconds in advance, offering a low-cost, high-real-time solution for autonomous driving decision-making in complex urban environments. Future work will focus on improving detection performance for small targets (e.g., eye regions smaller than  $10 \times 10$  pixels) at long distances (>30 meters) and exploring synergistic enhancements between high-resolution cameras and algorithmic improvements.

## Reference

1. Ma, Nan, et al. "A survey of human action recognition and posture prediction." *Tsinghua Science and Technology* 27.6 (2022): 973-1001.
2. Zhang, Bing, et al. "Using artificial neural networks for human body posture prediction." *International Journal of Industrial Ergonomics* 40.4 (2010): 414-424.
3. Li, Yunyang, et al. "Eye-gaze tracking system by haar cascade classifier." *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2016.
4. Cheng, Siqiang, Lingshan Chen, and Kun Yang. "DGSS-YOLOv8s: A Real-Time Model for Small and Complex Object Detection in Autonomous Vehicles." *Algorithms* 18.6 (2025): 358.
5. Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2016): 1137-1149.
6. Beal, Josh, et al. "Toward transformer-based object detection." *arXiv preprint arXiv:2012.09958* (2020).
7. Senel, Numan, et al. "Multi-sensor data fusion for real-time multi-object tracking." *Processes* 11.2 (2023): 501.
7. Wang, Ao, et al. "Yolov10: Real-time end-to-end object detection." *Advances in Neural Information Processing Systems* 37 (2024): 107984-108011.
8. Behley, Jens, et al. "Semantickitti: A dataset for semantic scene understanding of lidar sequences." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.
9. Roy, Payel, et al. "Adaptive thresholding: A comparative study." *2014 International conference on control, Instrumentation, communication and Computational Technologies (ICCICCT)*. IEEE, 2014.
10. Lang, Alex H., et al. "Pointpillars: Fast encoders for object detection from point clouds." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
11. Li, Hongyang, et al. "Delving into the devils of bird's-eye-view perception: A review, evaluation and recipe." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.4 (2023): 2151-2170.
12. Zhou, Hao, et al. "Review of learning-based longitudinal motion planning for autonomous vehicles: research gaps between self-driving and traffic congestion." *Transportation research record* 2676.1 (2022): 324-341.
13. Zhang, Xucong, et al. "Appearance-based gaze estimation in the wild." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
14. Kellnhofer, Petr, et al. "Gaze360: Physically unconstrained gaze estimation in the wild." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.
15. Zhang, Xucong, et al. "Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation." *European conference on computer vision*. Cham: Springer International Publishing, 2020.
16. Zhang, Xucong, et al. "Mpiigaze: Real-world dataset and deep appearance-based gaze estimation." *IEEE transactions on pattern analysis and machine intelligence* 41.1 (2017): 162-175.

## Объемно-ориентированная визуализация каустик и динамических преломляющих объектов со сложными свойствами материала в реальном времени

С. И. Вяткин, Б. С. Долговесов

Институт автоматизации и электрометрии СО РАН, Новосибирск, Россия

**Аннотация.** Преломление и рассеяние света при прохождении через различные материалы остается сложной задачей, особенно когда целью является рендеринг в реальном времени динамичных сцен, в которых освещение, материалы и геометрия могут меняться. При визуализации сцен с глобальным освещением важно точно рассчитать освещение, чтобы получить детальные каустические характеристики. И в то же время достаточно быстро, чтобы можно было изменять освещение, геометрию и материалы в режиме реального времени. Наиболее интересные явления зависят от непрямого освещения, поэтому необходимо решить глобальную проблему освещения. Предлагается объемно-ориентированный подход к визуализации каустик и динамических преломляющих объектов со сложными свойствами материала в реальном масштабе времени. В конвейере рендеринга преобразуются поверхности в объемные данные, с отслеживанием кривых траекторий фотонов, преломляющихся в объеме, с отображением произвольных видов результирующего распределения яркости. В результате предложенный метод позволяет визуализировать каустики, поглощение и рассеяние света при выполнении физических симуляций и при манипулировании поверхностями в режиме реального времени.

**Ключевые слова:** объемно-ориентированная визуализация, вокселизация, воксел, воксельное восьмеричное дерево, глобальное освещение, преломление, каустика

## Volume-oriented visualization of caustics and dynamic refractive objects with complex material properties in real time

S.I. Vyatkin, B.S. Dolgovessov

Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences,  
Novosibirsk, Russia

**Abstract.** Refraction and scattering of light as it passes through different materials remains a challenging task, especially when the goal is to render dynamic scenes in real time, in which lighting, materials, and geometry can change. When visualizing scenes with global illumination, it is important to calculate accurately the lighting in order to obtain detailed caustic characteristics. In addition, at the same time, it is fast enough to be able to change lighting, geometry, and materials in real time. The most interesting phenomena depend on indirect lighting, so it is necessary to solve the global lighting problem. A volume-oriented approach to visualization of caustics and dynamic refractive objects with complex material properties in real time is proposed. In the rendering pipeline, surfaces are transformed into volumetric data, tracking the curves of the trajectories of photons refracted in the volume, displaying arbitrary views of the resulting brightness distribution. As a result, the proposed method makes it possible to visualize caustics, absorption and scattering of light when performing physical simulations and manipulating surfaces in real time.

**Keywords:** volume-oriented visualization, voxelization, voxel, voxel octal tree, global illumination, refraction, caustics

### Введение

Реалистичные эффекты освещения важны для восприятия сложных пространственных структур [1-4]. Однако они также значительно и более затратные в вычислительном отношении, чем обычные модели локального освещения. Это еще более актуально для рендеринга объемных данных с улучшенными эффектами освещения ввиду отсутствия дискретных однородных объектов, так как вклад освещенности должен оцениваться и распространяться в каждой точке пространства. Одним из аспектов, которым часто пренебрегают, является преломление света. Эффекты преломления обычно игнорируются, и все лучи рассматриваются как прямые. Из-за высоких вычислительных затрат на физический рендеринг интерактивные методы обычно используют упрощенные оптические модели. Физически корректная визуализация скалярных полей выполняется с помощью объемных фотонных карт [5]. Однако этот метод не достигает даже интерактивной скорости вычислений. Большинство работ по фотонным картам посвящены рендерингу, при котором каждый пиксель в сцене вычисляется заранее перед отображением конечного результата, где качество и точность результата ценятся выше интерактивности. Однако в [6] описывается отображение фотонов, реализованное на графическом процессоре. Методы преломления

света в пространстве изображения реалистично отображают рефракцию, но многие из этих методов не отображают каустику [7, 8], или могут обрабатывать поверхностную каустику, но не объемную каустику [9]. Кроме того, в этих методах любое изменение освещения, материалов или геометрии сцены требует повторного расчета распределения яркости, что занимает несколько секунд. В работе [10] представлен физически корректный рендеринг в реальном времени эффектов однократного рассеяния, но без отображения преломления. Метод [11] позволяет отображать однократное рассеяние и преломление, однако он не учитывает поглощение. Кроме того, как и все методы обработки в пространстве изображения, он не учитывает преломления за пределами поля зрения, что делает каустику зависимой от обзора. В статье [12] представлена основанная на экране техника рендеринга объемных каустик, подход обеспечивает высокую производительность для сцен, в которых каустики влияют на небольшое количество пикселей. Этот подход также поддерживает неоднородные среды, но требует очень низкой частоты дискретизации для интерактивности. Что касается объемной визуализации, то в [13] был предложен фильтр на основе сплайнов для высококачественной реконструкции градиентов, с целью улучшения внешнего вида эффектов преломления. В [14] описано восьмеричное дерево для ускорения вычислений. Но эти методы также работают с отдельными выборками на поверхностях. В нашей работе [15] представлены как преломления, так и объемные каустики, которые визуализируются в интерактивном режиме. Используется прямой рендеринг функционально заданных объектов.

В данной работе мы представляем метод, позволяющий визуализировать эффекты, связанные с преломлением, поглощением и анизотропным одиночным рассеянием в динамических сценах в реальном масштабе времени. Предлагается конвейер рендеринга, который состоит из нескольких этапов, каждый из которых реализован на графическом процессоре. На вход конвейера поступает поверхностное описание сцены, которое вокселизуется, а также набор точечных и направленных источников света, а на выходе получаем физически корректное визуализированное изображение.

### Постановка задачи

Поставлена задача разработки конвейера рендеринга таким образом, чтобы он начинался с вокселизации "на лету", чтобы можно было поддерживать объемное представление даже при изменении входной геометрии. Поскольку воксельное представление позволит в полной мере использовать параллелизм графического процессора на каждом из последующих этапов конвейера, необходимо учесть непрерывное изменение показателя преломления по всему объему. Для этого фотоны и лучи просмотра должны следовать по изогнутым траекториям, определяемым уравнением луча геометрической оптики. Благодаря воксельному представлению необходимо построить восьмеричное дерево для ускорения алгоритма. То есть восьмеричное дерево нужно использовать для выбора переменного размера шага при распространении фотонов по их криволинейным траекториям. Реализация трассировки фотонов на базе графического процессора должна делить работу на этап перемещения фотонов, который вычисляет новое положение, направление и излучение каждого фотона после одного шага, и этап хранения фотонов, который накапливает излучение в объеме. При рендеринге преломляющих материалов необходимо иметь возможность изменять не только параметры просмотра в реальном времени, но и изменять параметры освещения и материала, сохраняя при этом скорость обновлений. Так, например, пользователи системы должны иметь возможность свободно изменять расположение и цвета источников света при просмотре эффектов каустики, поглощения, одиночного рассеяния и теней. Также конвейер должен поддерживать динамическое обновление свойств материала, а также освещение. Необходимо предусмотреть интеграцию в систему алгоритмов моделирования поверхностей, сохраняя при этом эффекты преломления. Например, пользователь должен иметь возможность в реальном времени деформировать преломляющий объект и наблюдать за влиянием деформации на освещение сцены.

Целью представленной работы является создание объемно-ориентированного метода визуализации каустик и динамических преломляющих объектов со сложными свойствами материала в реальном масштабе времени.

## Теория

### 1. Функциональные объекты

Для компактного хранения базы данных используются функциональные объекты [16]. Функциональные объекты описываются с помощью квадрик и функций отклонения от квадрики:

$$q'(x, y, z) = q(x, y, z) + \sum_{i=1}^N d_i(x, y, z), \quad (1)$$

где  $q'(x, y, z)$  – функциональная поверхность;  $q(x, y, z)$  – квадрика;  $i=1 \dots N$  – количество функций отклонения;  $d_i(x, y, z)$  – функция отклонения,

$$d_i(x, y, z) = \begin{cases} q_i^3(x, y, z), & \text{if } q_i(x, y, z) \geq 0 \\ 0, & \text{if } q_i(x, y, z) < 0 \end{cases} \quad (2)$$

где  $q_i(x, y, z)$  – возмущающая функция второго порядка.

Для формирования моделей сложных объектов используются теоретико-множественные операции. Перед визуализацией в режиме реального времени функциональные объекты триангулируются [17]. Далее алгоритм работает в интерактивном режиме.

### 2. Вокселизация 3D-объектов

С целью плавного изменения значений показателя преломления в объеме необходимо при вокселизации присваивать дробные значения покрытия тем вокселям, через которые проходит поверхность. Для этого сначала выполняется дискретизация объема с субпиксельным разрешением. Это означает, что преобразуется сетка треугольников в объемную текстуру, которая в каждом измерении в четыре раза больше, чем на выходе. Однако при уменьшении выборки результирующей текстуры стоимость считывания выборок текстуры для каждого из вокселей на выходе является очень высокой. Поэтому используется понижающая дискретизация только для вокселей вблизи поверхности. На рисунке 1 показаны основные этапы вычисления вокселизации.

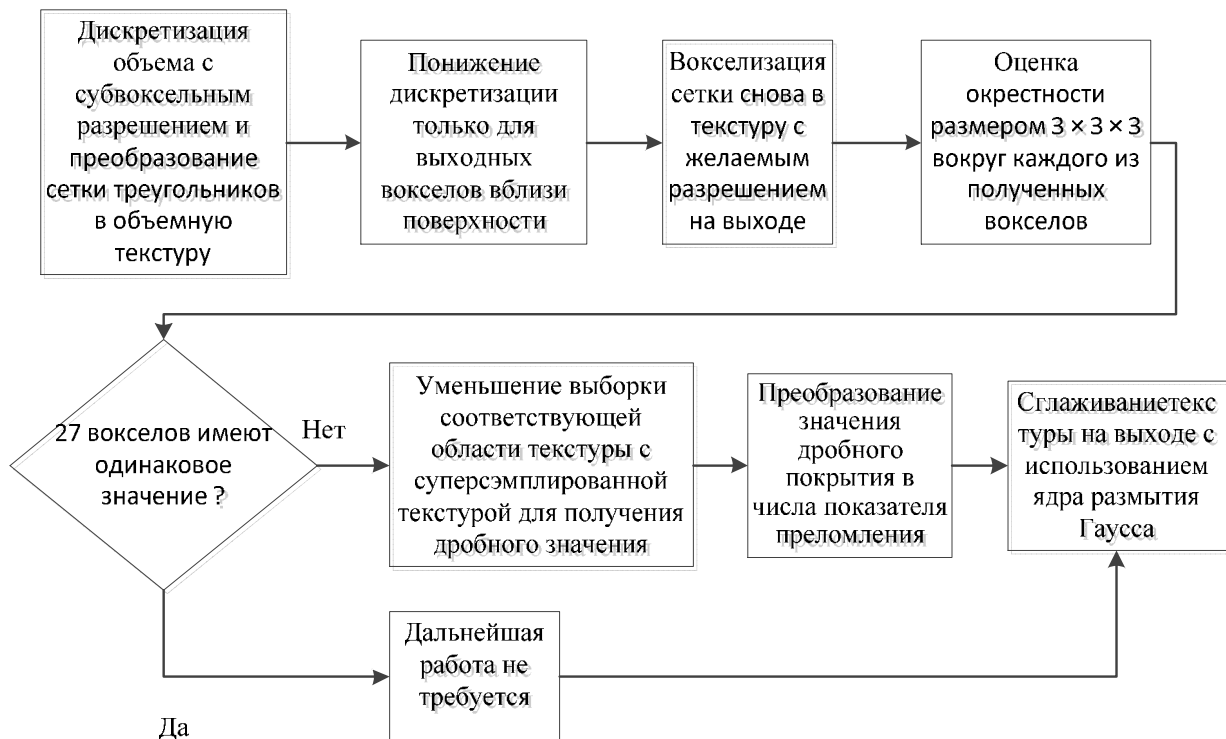


Рисунок 1. Структура вычислительного процесса вокселизации объекта

Вокселизируется сетка треугольников снова в текстуру с желаемым разрешением на выходе, затем проверяется окрестность размером  $3 \times 3 \times 3$  вокруг каждого из полученных вокселей. Если значения 27 вокселей отличаются в этой окрестности, то уменьшаем выборку соответствующей области



текстуры с суперсэмплированной текстурой для получения дробного значения. Далее преобразуются значения дробного покрытия в числа показателя преломления и сглаживается текстура на выходе с использованием ядра размытия Гаусса. Используется супервыборка  $4 \times 4 \times 4$  и ядро  $9 \times 9 \times 9$ . Суперсэмплирование эффективно повышает точность нормалей к поверхности, представленных градиентом показателя преломления, в то время как размытие размывает граничную область на более широкий диапазон вокселей.

Были определены разрешение сверхдискретизации и ядро размытия по Гауссу на основе экспериментальных изображений. Было отмечено, что изображения улучшаются по мере увеличения суперсэмплирования, независимо от того, какое ядро размытия используется. Как только было выбрано суперсэмплирование  $4 \times 4 \times 4$  (самое большое, которое сочли целесообразным), увидели небольшое улучшение, когда размер ядра blur увеличился с 5 до 7–9. Однако существенных изменений не произошло, когда размер ядра увеличился до 11. Поэтому были использованы суперсэмплинг  $4 \times 4 \times 4$  и ядро  $9 \times 9 \times 9$ .

Далее строится восьмеричное дерево (рис. 2). Построение представления в виде восьмеричного дерева аналогично построению mip-map пирамиды двумерного аналога. Пирамида строится из трехмерных массивов, в которых записаны минимальные и максимальные показатели преломления, присутствующие в каждой объемной области. Затем анализируется пирамида от самого грубого уровня до самого точного, записывая номер текущего уровня в восьмеричное дерево всякий раз, когда диапазон значений показателя преломления находится в пределах входного допуска.

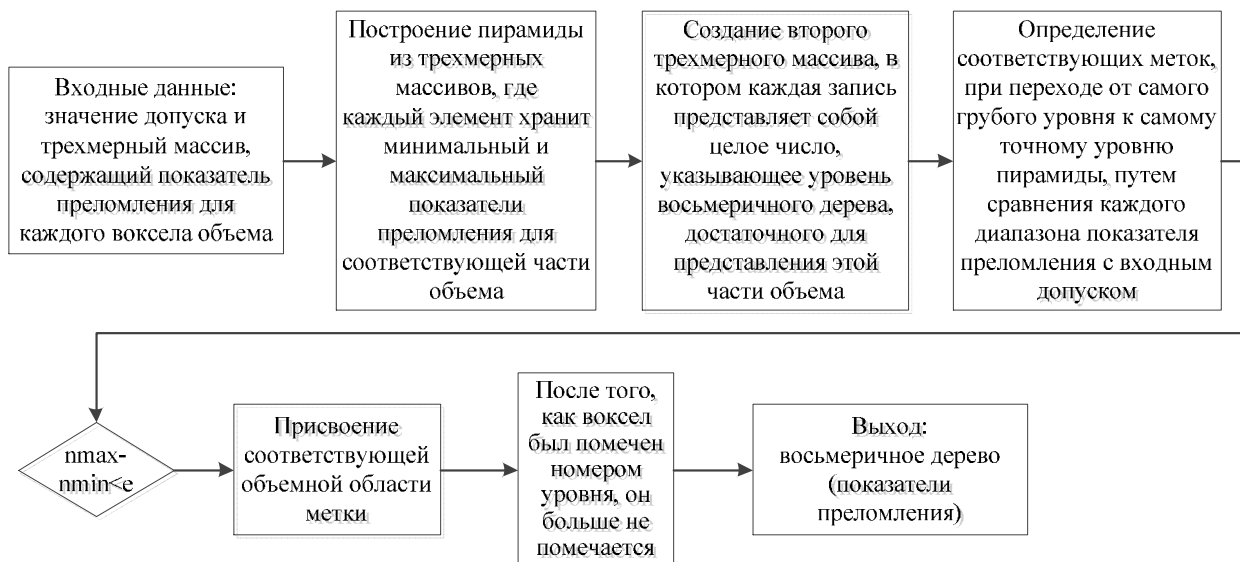


Рисунок 2. Алгоритм построения восьмеричного дерева

На вход (см. рис. 2) поступает кубический объем элементов показателя преломления  $n$  размером  $2^M$  в каждом измерении. Начинаем с построения пирамиды из  $M$  трехмерных массивов, где каждый элемент хранит минимальный и максимальный показатели преломления  $n_{\min}$  и  $n_{\max}$  для соответствующей части объема. Далее создается второй трехмерный массив размером  $2^M$  в каждом измерении, где каждая запись представляет собой целое число, указывающее уровень восьмеричного дерева, достаточный для представления этой части объема. Используется ноль в качестве метки для самого точного уровня дерева и  $m$  в качестве метки для самого грубого (одновоксельного) уровня. Затем определяем соответствующие метки, переходя от самого грубого уровня к самому точному уровню пирамиды, сравнивая каждый диапазон показателя преломления с входным допуском. При просмотре уровня пирамиды  $m$ , если встретится запись, удовлетворяющая условию  $n_{\max} - n_{\min} < \varepsilon$ , происходит присваивание соответствующей объемной области метки  $m$ . После того как воксел был помечен номером уровня, он больше не помечается. Построение восьмеричного дерева происходит таким образом, чтобы оба показателя преломления  $n$  и коэффициент затухания  $\beta$  были

приблизительно постоянны для каждого конечного узла, поскольку на алгоритм отслеживания фотонов влияют оба этих свойства материала.

### 3. Алгоритм отслеживания фотонов

Сначала настраиваем графический конвейер таким образом, чтобы положение камеры совпадало с источником света и было ориентировано на интересующий объем. Затем визуализируем грани куба, ограничивающие объем. Далее отображаем текстуру с использованием шейдера, который записывает альфа-значение равное единице вместе с позицией в 3D для каждого пикселя, представляющего обращенную вперед поверхность. Пустые пиксели имеют нулевые альфа-значения. Затем преобразуем текстуру в список точечных примитивов, представляющих фотоны, используя геометрический шейдер. Каждый пиксель с ненулевым значением альфа генерирует один фотон, положение фотона определяется по значению пикселя. Направление фотона определяется по координатам текстуры и положению источника света. Яркость фотона определяется характеристиками излучения света.

Уравнение луча геометрической оптики [18]

$$\frac{d}{ds} \left( n \frac{d\vec{x}}{ds} \right) = \nabla n, \quad (3)$$

где  $\vec{x}(s)$  – криволинейный световой путь,  $n$  – показатель преломления.

Запишем уравнение (3) в виде системы дифференциальных уравнений первого порядка:

$$\frac{d\vec{x}}{ds} = \frac{\vec{v}}{n}, \quad (4)$$

$$\frac{d\vec{v}}{ds} = \nabla n, \quad (5)$$

где  $\vec{v} = n \frac{d\vec{x}}{ds}$ .

Кусочно-линейные аппроксимации кривых выглядят следующим образом:

$$\vec{x}_{i+1} = \vec{x}_i + \frac{\Delta s}{n} \vec{v}_i, \quad (6)$$

$$\vec{v}_{i+1} = \vec{v}_i + \Delta s \nabla n, \quad (7)$$

где  $\Delta s$  – размер шага.

По мере движения фотонов вдоль этих кривых меняется размер шага  $\Delta s$ , для этого используется структура данных восьмеричного дерева. Входные данные состоят из значений показателя преломления восьмеричного дерева, трехмерного массива коэффициентов ослабления RGB и списка фотонов. Каждый фотон имеет начальное положение  $\vec{x}_0$ , направление  $\vec{v}_0$  и значение яркости RGB  $\tilde{L}_0$ . Результатом является трехмерный массив распределений яркости RGB, описывающий освещенность, которая поступает в каждый воксел. Каждая итерация отслеживания фотонов продвигает фотоны на один шаг вперед в соответствии с уравнениями 6 и 7. Для каждого фотона используется восьмеричное дерево для определения наибольшего размера шага  $\Delta s_{octree}$ , чтобы удерживать фотон в области с приблизительно постоянным показателем преломления. То есть положение фотона используется для определения уровня окружающего его конечного узла восьмеричного дерева. А размер шага  $\Delta s_{octree}$  вычисляется для движения фотона к границе узла восьмеричного дерева. Затем выбираем  $\Delta s$ , который будет больше  $\Delta s_{octree}$ , и минимальный размер шага  $\Delta s_{min}$ , указанный пользователем. Минимальный размер шага обычно равен ширине одного или двух вокселей и может регулироваться пользователем в зависимости от точности и производительности.

В каждом вокселе аппроксимируется распределение входящего излучения с сохранением только средневзвешенного значения направления приходящих фотонов и значения излучения для RGB-длин волн. При прохождении фотоном пути от  $\vec{x}_i$  до  $\vec{x}_{i+1}$  должно происходить излучение в каждом из вокселей, через которые он проходит. Поэтому, когда движется фотон на один шаг вперед,

генерируются две вершины в буфере вершин для записи старого и нового значений положения, направления и яркости фотона. После продвижения всех фотонов на один шаг вперед обрабатывается буфер вершин в виде списка линейных сегментов. Сегменты растриваются в выходной массив распределения яркости. В графическом конвейере используется интерполяция положения, направления и яркости фотона между двумя конечными точками каждого сегмента линии. Применяется пиксельный шейдер, в котором добавляется излучение фотона к распределению, сохраненному в каждом вокселе. Взвешивается направление движения фотона по сумме значений его RGB-излучения, прежде чем добавляется к направлению, сохраненному в каждом вокселе. После выделения всех фотонов и сохранения их вклада в распределение излучения исключаем каждый фотон, объем или излучение которого упало ниже порогового значения, и затем повторяется весь процесс. Продолжаются эти итерации до тех пор, пока количество активных фотонов не составит лишь малую долю от первоначального количества фотонов. Далее происходит сглаживание объемного распределения излучения с использованием приближения гауссова ядра  $3 \times 3 \times 3$  для уменьшения шума. Алгоритм отслеживания фотонов показан на рисунке 3.

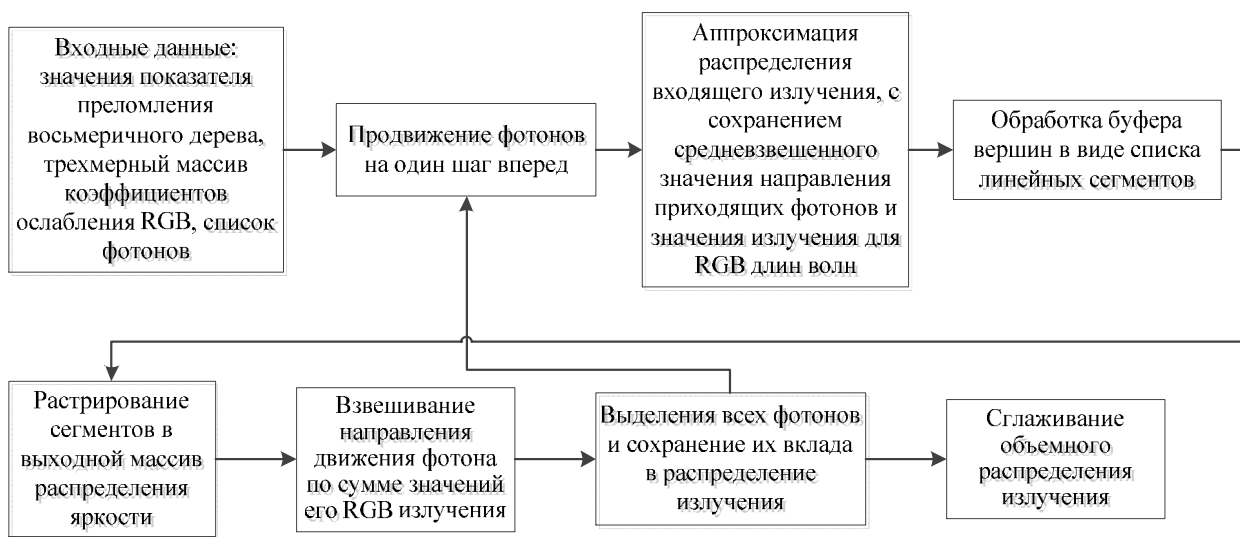


Рисунок 3. Алгоритм отслеживания фотонов

Отметим, что внешний вид каустик чувствителен к размеру шага перемещения фотонов. Размер шага, в свою очередь, зависит от допуска, используемого в алгоритме построения октодерева. Для некоторых объемных данных сложно выбрать допуск, который обеспечивает точную визуализацию с разумным количеством шагов перемещения фотонов. В частности, иногда приемлемое качество рендеринга требует такого низкого допуска, что размеры шагов слишком малы для достижения частоты кадров в реальном времени. Небольшое увеличение допуска может привести к объединению множества узлов октодерева, что приводит к значительно большему размеру шага и значительному снижению качества рендеринга. Поэтому создаем узлы октодерева с промежуточными размерами шага. В каждом вокселе хранится максимальный размер шага в дополнение к номеру уровня окружающего узла восьмеричного дерева конечного уровня. Выбирается размер шага между заданным пользователем минимальным размером шага и максимальным размером шага узла восьмеричного дерева. Выбираем максимальный размер шага для каждого вокселя на этапе построения октодерева. Когда вокселу присваивается номер уровня октодерева, если изменение показателя преломления меньше чем  $\varepsilon = 0,005$ , устанавливается ограничение на максимальный размер шага вокселя равное бесконечности. Если изменения показателя преломления больше чем  $\varepsilon_1 = 0,005$ , но меньше второго заданного пользователем допуска  $\varepsilon_2 = 0,02$ , то устанавливается значение  $\Delta s_{\max}$  вокселя на конечный размер шага, выбранный пользователем. Эта схема гарантирует, что в узлах с практически постоянным показателем преломления продвигаются фотоны до самой границы узла, в то время как в узлах с некоторым изменением показателя преломления ограничивается размер шага.

#### 4. Визуализация

Отслеживание лучей происходит в обратном направлении траектории, по которым световые лучи достигают камеры. Процесс начинается с инициализации начала и направления отдельного луча для каждого пикселя выходного изображения. Если луч пересекает треугольную сетку (после триангуляции функционально заданного объекта), то происходит шаг за шагом продвижение вдоль луча, пока он не выйдет из объема. Используются уравнения (6) и (7) для определения траекторий, по которым проходят лучи, точно так же, как это делается для фотонов. Однако при трассировке лучей обзора используется фиксированный размер шага, равный ширине одного вокселя, а не переменный размер шага. На каждом шаге по лучу обзора просматриваются значение и направление излучения, сохраненные в соответствующем вокселе. Затем оценивается функция фазы рассеяния для определения части излучения, рассеиваемой в направлении камеры. Для этого умножаем полученный результат на локальный коэффициент рассеяния и общее значение затухания вдоль луча из-за поглощения и рассеяния. Это произведение дает вклад в яркость одного вокселя, которое затем добавляется к общей яркости текущего луча. Как только луч выходит за пределы объема, включается фон для завершения вычисления цвета выходного пикселя. На рисунке 4 показана общая схема конвейера вычислений.



Рисунок 4. Общая схема конвейера

#### Результаты тестирования

Тестирование производилось на процессорах Intel Core i7 3820 и GPU NVIDIA GeForce GTX 1080 Ti. Для тестирования использовалось несколько сцен, пять из них показаны на рисунках 5–9. Разрешение изображения для всех примеров рендеринга составляет 1024 x 768. Все сцены полностью динамичные.

Можно отметить, что по мере того, как показатель преломления объекта увеличивается по сравнению с единицей, волновые фронты пересекаются раньше, и артефакты распространения становятся все более заметными. С другой стороны, отслеживание фотонов дает приемлемые результаты независимо от показателя преломления. Мы можем повторно высвечивать сцены, подобные изображенному на рисунке 5, примерно со скоростью восемьдесят кадров в секунду. На рисунке 6 показан эффект, как изменение освещения влияет на визуализацию преломляющего объекта, при этом видны каустические изменения. Пользователь в интерактивном режиме может деформировать преломляющий объект и наблюдать за влиянием деформации на освещение сцены (рис. 7 и 8). Была применена динамическая структура для обработки больших деформаций модели, в частности для обнаружения самопересечения для явного контроля топологии, создания гладких переходов между двумя объединёнными частями поверхности и т.д.

Сочетание вокселизации и рендеринга требует использования современных графических процессоров. Центральный процессор используется для выполнения деформации. В результате можно интегрировать в систему множество алгоритмов моделирования поверхностей, сохраняя при этом эффекты преломления. Метод позволяет генерировать сцены, содержащие неоднородные преломляющие объекты. На рисунке 9 показана модель аморфного полимера (с фрагментами внутри) с низкой степенью кристалличности и низкой интенсивностью рассеяния.

Производительность метода зависит от нескольких параметров, включая количество треугольников во входной сетке, количество занятых вокселей и количество исходных фотонов. Производительность метода показана в таблице.



Рисунок 5. Объект с показателем преломления 1,1



Рисунок 6. Каустика с изменением положения источника света в реальном времени



Рисунок 7. Объемная и поверхностная каустика, создаваемая преломляющим объектом, с деформацией поверхности в реальном времени (рис. 8)



Рисунок 8. Деформация поверхности в реальном времени



Рисунок 9. Слабоанизотропный материал

### Результаты тестирования

Сцена	Рис. 5	Рис. 6	Рис. 7	Рис. 9
Кол-во функционально заданных объектов	1	3	8	9
Кол-во треугольников	53K	55K	100K	118K
Разрешение объема	256 <sup>3</sup>	256 <sup>3</sup>	256 <sup>3</sup>	256 <sup>3</sup>
Занятые воксели	90 <sup>3</sup>	80 <sup>3</sup>	77 <sup>3</sup>	103 <sup>3</sup>
Исходные фотоны	1024 <sup>2</sup>	1024 <sup>2</sup>	1024 <sup>2</sup>	1024 <sup>2</sup>
Количество шагов трассировки фотонов, млн	2,7	2,0	1,7	3,81
Количество фотонов, прошедших трассировку по объему	36K	41K	35K	143K
Генерация градиента после вокселизации, построение октодеревя, объединение освещенности после отслеживания фотонов, время просмотра и пр., мс	3,79	6,25	9,38	9,75
Вокселизация, мс	2,65	3,38	5,9	6,45
Отслеживание фотонов, мс	10,8	9,6	8,52	13,8
Общее время, мс	17,24	19,23	23,8	30
Кол-во кадров в секунду	58	52	42	33



Примечания:

- 1) время вокселизации зависит от количества треугольников и занятых вокселей (см. таблицу);
- 2) время отслеживания фотонов зависит от количества занятых вокселей и исходных фотонов;
- 3) количество шагов трассировки фотонов зависит от количества занятых вокселей и исходных фотонов;
- 4) разрешение объема, количество треугольников, занятых вокселей и исходных фотонов можно менять.

Метод поддерживает динамическое обновление свойств материала, а также освещение. Материалы, находящиеся в объеме, характеризуются тремя свойствами: показателем преломления, коэффициентом затухания и коэффициентом рассеяния. Изменения показателя преломления определяют, как лучи света изгибаются при прохождении через объем. Коэффициент рассеяния влияет на то, сколько света распространяется по объему от источников света к камере. Коэффициент рассеяния определяет, насколько сильно излучение внутри каждого вокселя влияет на конечное изображение.

Поскольку работы [1] и [4] посвящены сложным зеркальным поверхностным отражениям с каустиками, а не объемам, рассмотрим наиболее близкие методы в сравнении с предлагаемым подходом.

Для визуализации скалярных полей с преломляющими средами использовались такие методы, как объемное фотонное отображение [5]. Однако эти методы являются непомерно дорогими для интерактивных приложений, поскольку количество фотонов, необходимое для того, чтобы результат казался непрерывным, слишком велико.

Метод «эikonальной визуализации», предложенный в [8], наиболее близок к визуализации всех эффектов преломления и участвующих в нем сред в режиме реального времени. Несмотря на то что визуализация с помощью этого метода позволяет создавать новые изображения в интерактивном режиме, любое изменение освещения, материалов или геометрии сцены требует повторного расчета распределения яркости, что занимает несколько секунд.

Наш метод позволяет более точно воспроизвести размер и форму каустик, чем метод [8], поскольку в нашем методе накапливается вклад излучения всех входящих фотонов в вокселе, в то время как при визуализации в методе [8] сохраняется вклад только участка волнового фронта с наибольшей энергией, который проходит через воксел. В результате при рендеринге в методе [8] не улавливаются каустические волны, которые должны генерироваться несколькими волновыми фронтами низкой интенсивности, сходящимися с разных направлений.

Также в нашем методе можно беспрепятственно распространять фотоны по объему, когда они проходят через каустику, в то время как изображения в методе [8] содержат искаженные четырехугольники волнового фронта. Это происходит, когда четыре угла участка проходят через каустические сингулярности в разное время. Некоторые из этих дефектных участков обнаруживаются и устраняются, однако это приводит к искусственной потере яркости, в то время как те, которые остаются, вносят числовую ошибку, видимую в виде шумных, неровных краев каустики. По мере того как показатель преломления объекта увеличивается по сравнению с единицей, волновые фронты пересекаются раньше, и артефакты распространения становятся все более заметными.

Кроме того, наш подход намного быстрее, чем рендеринг [8], и позволяет создавать изображения после изменения параметров освещения или материала. Можно также повторно высвечивать сцены в реальном времени, в то время как визуализация в [8] занимает порядка семи секунд всего для одного кадра. Высокая производительность заключается в использовании адаптивных размеров шагов при отслеживании фотонов. Делая меньше шагов по объему, генерируется меньше примитивов для рендеринга и используется скорость графического процессора для растеризации сегментов линий вместо точек.

Скорость нашего подхода открывает возможности для многих новых применений. Если предыдущие методы рендеринга преломляющих материалов позволяли изменять только параметры просмотра в интерактивном режиме, то теперь можно также изменять параметры освещения и материала, сохраняя при этом интерактивность обновлений. Пользователи нашей системы могут свободно изменять расположение и цвета источников света при просмотре эффектов каустики, поглощения, одиночного рассеяния и теней. Предлагаемый метод можно использовать для физического моделирования реальных процессов и анимации воображаемых процессов, например для

моделирования динамики твердого тела и жидкости, а также турбулентного перемешивания неоднородных материалов.

Весь объём обрабатываемых данных распределён между CPU и GPU. На CPU данные обрабатываются с использованием технологии для многопоточного программирования OpenMP. На GPU данные обрабатываются с использованием технологии NVIDIA CUDA.

Весь объём обрабатываемых данных делится на две неравные части: зону, обрабатываемую CPU, и зону, обрабатываемую GPU.

Граница между зоной CPU и зоной GPU в случае статической балансировки является фиксированной. В случае динамической балансировки вычислительная нагрузка изменяется в процессе выполнения программы.

Для оценки необходимости внесения корректировки в положение границы зон, на каждом  $n$ -м шаге итерации запускается процесс определения общей эффективности вычислений при текущем распределении вычислительной нагрузки. В результате этого процесса принимается решение о сохранении текущего положения границы зон или о сдвиге его на определённую величину в ту или иную сторону. В случае сдвига организовывается перераспределение данных, которые в результате сдвига начинают обрабатываться другим вычислителем.

Поскольку программа гибридная и наряду с вычислениями на CPU используются вычисления и на GPU, то в случае, когда GPU обрабатывает выделенные ему данные медленнее, чем CPU – выделенные ему, программа будет простаивать в точках синхронизации в ожидании окончания работы GPU. Вследствие этого загрузка программой CPU будет меньше 100 %. Получение программой данных о том, насколько она потребляет выделенный ей ресурс CPU, делается с помощью системного вызова.

Допускается, чтобы CPU простаивал в ожидании GPU не более 5 %. Если значение загрузки CPU больше нормального, значит, уменьшается его зона и, соответственно, увеличивается зона GPU. Если наоборот, меньше нормального, то увеличивается зона CPU.

Однако скачки в виде постоянного то уменьшения, то увеличения размеров зон неприемлемы. Для этого нормальной загрузкой CPU считается не конкретное значение, а диапазон от 85 до 95 %.

## **Заключение**

Разработан объёмно-ориентированный метод визуализации преломляющих объектов в режиме реального времени. В методе используется воксельное восьмеричное дерево для определения наибольшего размера шага при отслеживании фотонов. Также применение воксельной сетки и конвейера растеризации графического процессора для хранения данных о влиянии излучения фотонов во время прохождения процесса освещения позволяет избежать значительных затрат на поиск близлежащих фотонов на каждом этапе просмотра.

Представлена методика, позволяющая воспроизводить эффекты преломления, однократного рассеяния и поглощения даже при изменении освещения, свойств материала, геометрии и параметров обзора. Это позволяет одновременно изменять освещение, материалы и геометрию в реальном времени. Конвейер рендеринга использует графический процессор для преобразования поверхностных сеток в объёмные данные, отслеживания фотонов в сцене и рендеринга изображения результирующего распределения яркости. Центральный процессор применяется для таких приложений, как деформация поверхности и моделирование физических процессов. Метод позволяет учитывать непрерывное изменение показателя преломления по всему объёму.

Имеется несколько направлений для будущей работы. К таким работам относятся реалистичное отображение тумана, дыма и пара, физическое моделирование таких явлений, как жидкости и миражи, где эффекты преломления имеют большое значение. Для этого необходимо будет учесть многократное рассеяние во время адаптивного процесса отслеживания фотонов. Имеется также интерес к работе по включению карт окружения и других сложных источники света в конвейер рендеринга.

## **Источник финансирования**

Работа выполнена при поддержке Министерства науки и высшего образования в рамках выполнения работ по Государственному заданию № 124041700102-4 в ИАиЭ СО РАН.

**Список литературы**

1. Looking Through the Glass: Neural Surface Reconstruction Against High Specular Reflections / J. Qiu [and others] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023. Vancouver, BC, Canada. P. 1–20.
2. Loube G., Zeltner T., Holzschuch N. Slope-space integrals for specular next event estimation // ACM Transactions on Graphics. 2020. Vol. 39, no. 6. P. 1-13.
3. Glossy Probe Reprojection for Interactive Global Illumination / S. Rodriguez [and others] // December 2020 ACM Transactions on Graphics. 2020. Vol. 39, no 6. P. 1-16.
4. Neural Point Catacaustics for Novel-View Synthesis of Reflections / G. Kopanas [and others] // ACM Transactions on Graphics. 2022. Vol. 41, no. 6. P. 1-15.
5. Non-linear Volume Photon Mapping / D. Gutierrez [and others] // Proceedings of the Eurographics Symposium on Rendering. 2005. P. 291–300
6. Kern R., Brull F., Grosch T. Accelerating Photon Mapping for Hardware-based Ray Tracing // Journal of Computer Graphics Techniques. 2023. Vol. 12, no. 1. P. 1-28.
7. Davis S. T., Wyman C. Interactive refractions with total internal reflection // Proceedings of Graphics Interface. 2007. P. 185–190.
8. Eikonal Fields for Refractive Novel-View Synthesis / M. Bermana [and others] // ACM SIGGRAPH Conference Proceedings. 2022. P. 1 – 9.
9. Wyman C., Davis S.T. Interactive image-space techniques for approximating caustics // Proceedings of the Symposium on Interactive 3D Graphics and Games. 2006. P. 153–160.
10. A Practical Analytic Single Scattering Model for Real Time Rendering / B. Sun [and others] // ACM Transactions on Graphics. 2005. Vol. 24, no. 3. P. 1040–1049.
11. Kruger J., Bugrer K., Westerman R. Interactive Screen-Space Accurate Photon Tracing on GPUs // Rendering Techniques. 2006. P. 319–329.
12. Li S., Mueller K. Spline-based gradient filters for high-quality refraction computations in discrete datasets // Proceedings of EuroVis. 2005. P. 215–222.
13. Li S., Mueller K. Accelerated, high-quality refraction computations for volume graphics // Proceedings of Volume Graphics. 2005. P. 73–81.
14. Interactive volume caustics in single-scattering media/ W. Hu [and others] // I3D '10: Proceedings of the ACM SIGGRAPH symposium on Interactive 3D Graphics and Games. 2010. P. 109 – 117.
15. Вяткин С. И., Долговесов Б. С. Интерактивное вычисление преломления света и каустик с применением графического процессора // Программирование. 2024. № 1. С. 100-112.
16. Вяткин С. И., Долговесов Б. С. Метод сжатия геометрических данных с применением функций возмущения // Автометрия. 2018. Т. 54, № 4. С. 18- 25.
17. Vyatkin S. I. Polygonization method for functionally defined objects // International Journal of Automation, Control and Intelligent Systems. 2015. Vol. 1, no. 1. P. 1–8.
18. Principles of Optics / M. Born [and others]. (7-th expended edition). Cambridge University Press, 1999. 99 p.

## Расчет показателя дискомфорта с помощью технологий компьютерной графики

Д. Н. Макаров, Д. К. Слюсарев, В. П. Будак

Национальный исследовательский университет «МЭИ», Москва, Россия

**Аннотация.** Статья посвящена усовершенствованию расчета объединенного показателя дискомфорта (UGR) для оценки качества освещения. Авторы критикуют ограничения современных систем проектирования (DIALux, Relux), не учитывающих вторичные зеркальные блики, и предлагают метод на базе технологий компьютерной графики (3ds Max) и математического моделирования (Matlab). Ключевое нововведение – алгоритм кластеризации источников бликов, заменяющий трудоемкий попиксельный расчет. Это повышает точность UGR на 10 %, лучше отражая физиологию зрительного восприятия. Результаты показывают, что учет коэффициентов отражения (особенно при  $\rho > 0.8$ ) существенно снижает дискомфорт. Метод актуален для проектировщиков и архитекторов, позволяя прогнозировать UGR на этапе моделирования.

**Ключевые слова:** объединенный показатель дискомфорта (UGR), зрительный дискомфорт, зеркальные отражения, компьютерная графика, 3D-визуализация, кластеризация источников бликов, коэффициент отражения, трассировка лучей, программное обеспечение светотехнического проектирования, физиология зрительного восприятия

## Calculation of the discomfort index using computer graphics technologies

D. N. Makarov, D. K. Slyusarev, V. P. Budak

National Research University «Moscow Power Engineering Institute», Moscow, Russia

**Abstract.** The article addresses the improvement of the Unified Glare Rating (UGR) calculation for lighting quality assessment. The authors critique limitations of current design tools (DIALux, Relux), which ignore secondary specular glares, and propose a method leveraging computer graphics (3ds Max) and mathematical modeling (Matlab). The core innovation is a clustering algorithm for glare sources, replacing pixel-level calculations. This enhances UGR accuracy by 10%, better aligning with human visual perception. Results demonstrate that accounting for surface reflection coefficients (notably at  $\rho > 0.8$ ) significantly reduces discomfort. Practical recommendations for material selection are provided. The approach enables UGR prediction at the design stage, offering architects and lighting designers a tool to optimize visual comfort while accounting for complex reflections.

**Keywords:** Unified Glare Rating (UGR), visual discomfort, specular reflections, computer graphics, 3D visualization, glare source clustering, reflection coefficient, ray tracing, lighting design software, visual perception physiology

### Введение

Качественные показатели освещения – ключевые светотехнические характеристики осветительных установок. Они, в отличие от количественных показателей, позволяют оценить её визуальное восприятие и понять, насколько комфортно будет людям и возможно ли выполнение зрительных задач [1].

Глаз человека реагирует именно на яркость, а не на освещенность поверхности. Поэтому критерии качества освещения определяются яркостью освещаемых объектов и окружения сцены. В современной светотехнике становится актуальным разработка методов моделирования и оценки пространственно-углового распределения яркости (ПУРЯ). Одним из инструментов служит создание фотореалистичных изображений на основе метода адаптивной трассировки лучей.

Среди большого количества качественных показателей освещения особое место занимает объединенный показатель дискомфорта (UGR). Он характеризует качество освещения и закреплён в нормативных документах. Этот показатель позволяет определить, насколько комфортно человеку будет находиться в помещении с проектируемой осветительной установкой.

Важно подчеркнуть, что широко используемые в светотехническом проектировании программы моделирования DIALux (Германия) и Relux (Швейцария) построены на основе метода конечных элементов и решают не уравнение глобального освещения относительно яркости, а уравнение излучательности в диффузном приближении. При этом вторичные блики не учитываются.

Важным шагом вперед в светотехническом проектировании стало появление DIALux evo (Германия), расчетное ядро которой основано на методе фотонных карт. Но метод расчета UGR в нем не изменился. В результате при проектировании освещения дается лишь приблизительная оценка комфортности для человека.

Сегодня в существующих системах автоматизированного проектирования, таких как DIALux и Relux, в UGR входят только сами светильники, для которых не нужно рассчитывать яркость. В связи с этим расчета UGR оказывается недостаточно для полноценной оценки дискомфортных ощущений, вызываемой сценой освещения.

Поэтому в настоящее время активно разрабатываются новые методы оценки качества освещения. Однако развитие сложных методик, основанных на построении новых шкал психовизуальных ощущений, требует больших исследований и будет внедрено в светотехническое программное обеспечение (ПО) еще нескоро, в то время как развитие архитектурных приемов, решений и отделочных материалов рабочих и жилых пространств требует особого внимания к оценке влияния зеркальных бликов на комфортность пребывания в них людей уже сейчас.

В настоящей статье предлагается использовать не реализованный в современных системах проектирования потенциал расчета UGR с помощью определения блеских источников по распределению яркости в программах для 3D-визуализации на примере 3Ds Max.

### Постановка задачи

Основной целью работы является создание программы для более точного расчета показателя дискомфорта UGR по изображению распределения яркости и сравнение результатов работы с расчетом в программе DIALux evo.

В работе предполагается выполнение следующих задач:

- 1) сравнение светотехнических расчетов в DIALux evo и 3Ds Max;
- 2) разработка программы для попиксельного расчета UGR в Matlab;
- 3) проверка метода расчета UGR на реальной ОУ по фотометрическим характеристикам;
- 4) разработка методики уточненного расчета показателя дискомфорта UGR.

### Теория

Что такое зрительный комфорт? Одна из точек зрения состоит в том, что зрительный комфорт – это просто отсутствие зрительного дискомфорта [2]. Зрительный дискомфорт определяется в [3] как неприятные ощущения, характеризующиеся чувством неудобства и напряженности, возникающие в основном в результате наличия в освещаемом пространстве резких различий яркости или неудачного выбора спектра излучения источника света. Также принято разделять зрительный дискомфорт, относящийся к психологическому восприятию, и нарушение зрительных функций или ослепление.

Основные механизмы, которые приводят к восприятию дискомфортной ослепленности, до сих пор мало изучены. Важность оценки ослепленности подчеркивается Международной комиссией по освещению (CIE) в документе под названием «Исследовательская дорожная карта для здорового внутреннего освещения», опубликованном в 2016 году.

Наличие яркости, значительно превышающей среднюю для поля зрения, вызывает дискомфорт и называется бликами. Блескость может проявляться в двух формах: ослепляющая блескость (disability glare) и дискомфортная блескость (discomfort glare) [2].

Ослепляющая блескость – это форма, оказывающая негативное влияние на зрительную работоспособность. Блескость, вызывающая ощущение дискомфорта, называется дискомфортной. Общий негативный эффект влияния ослепляющей засветки на зрительную работоспособность можно рассчитать, поскольку Холладей (1927) на основе тестов видимости вывел формулу, предсказывающую эквивалентную вуалирующую яркость. Её значение зависит от освещённости на глазу ( $E_{eye}$ ) и угла ( $\theta$ ) между направлением взгляда и направлением падения света от источника засветки [2].

С распространением светодиодных технологий особую актуальность приобрела проблема блескости и связанного с ней зрительного дискомфорта. Одним из способов её решения стало внедрение объединенного показателя дискомфорта (UGR) среди множества качественных характеристик освещения.

В настоящее время данный показатель, регламентированный стандартом [6], служит основным критерием оценки качества неспециальных осветительных установок и реализован в большинстве современных системах автоматизированного проектирования осветительных установок.

Несмотря на достаточную эффективность методики расчета обобщенного показателя дискомфорта, в профессиональном сообществе продолжают дискуссии относительно того, насколько он подходит для оценки систем на основе полупроводниковых источников света [7].

Согласно [9], UGR – это критерий оценки дискомфортной блёскости, вызывающей неприятные ощущения при неравномерном распределении яркостей в поле зрения.

Разработанный Международной комиссией по освещению (МКО) метод UGR (Unified Glare Rating), представленный в публикациях МКО 1995 и МКО 2010, получил мировое признание и включен во многие международные и национальные нормативные документы [7].

Методика предназначена для количественной оценки дискомфортной блескости, создаваемой светильниками внутреннего освещения, и базируется на расчетной формуле, представленной уравнением (1). CIE объединила лучшие элементы различных методов оценки бликов, использовавшихся на тот момент, в единую формулу. Благодаря этому «методу консенсуса» результат получил название «CIE Unified Glare Rating (UGR)» [7].

$$UGR = 8lg \left[ \frac{0,25}{L_a} \sum_{i=1}^N \frac{L_i^2 \omega_i}{p_i^2} \right], \quad (1)$$

где  $L_i$  – яркость блеского источника в направлении наблюдателя, кд/м<sup>2</sup>;

$\omega_i$  – угловой размер блеского источника, стер;

$p_i$  – индекс позиции блеского источника относительно линии зрения по Гуту;

$L_a$  – средняя яркость поля адаптации наблюдателя, кд/м<sup>2</sup>. Рассчитывается через вертикальную непрямую освещенность на уровне глаз наблюдателя по формуле  $L_b = E_{\text{непрям}}/\pi$ .

Суммирование происходит по всем источникам освещения в поле зрения.

Современный подход к оценке дискомфорта от освещения основан на расчете объединенного показателя дискомфорта UGR, который позволяет прогнозировать ослепляющий эффект осветительных установок на психофизической шкале дискомфорта.

Шкала UGR основана на устаревшей Британской системе индекса ослепления, успешно применявшейся почти 30 лет [2].

Описание зрительных ощущений по шкале UGR приведено в таблице.

**Шкала ощущений UGR**

UGR	Ощущение	Где применяется
$\leq 10$	Блеск отсутствует или незаметен	Высочайшие требования к визуальному комфорту и концентрации внимания: операционные в больницах, прецизионные лаборатории (электроника, часовое дело)
13	Блеск едва заметен	Высококочувствительные задачи (хирургические операции, читальные залы библиотек)
16	Заметный, но не раздражающий блеск	Стандартные офисные помещения, учебные классы. Требуется для большинства рабочих мест с постоянным пребыванием людей
19	Заметный блеск, начинает вызывать дискомфорт	Помещения, где зрительная работа не является основной или постоянной: промышленные цеха (рабочие зоны без постоянной работы с мелкими деталями), столовые, спортивные залы, некоторые зоны розничной торговли
22	Ощутимый дискомфорт	Технические помещения, гаражи, некоторые виды производств, подсобные помещения
25	Сильный дискомфорт	Производственные цеха, склады (где блеск не критичен)
28	Невыносимо	Не рекомендуется для любых рабочих зон

Чем ниже значение UGR, тем меньше дискомфортного блеска и комфортнее освещение. Чем выше UGR, тем сильнее ощущение дискомфорта, напряжения и усталости глаз.

- UGR <10: блики практически незаметны.

- UGR = 28: блики вызывают невыносимый дискомфорт.

Минимально заметная разница в уровне дискомфортного ослепления составляет 1 единицу UGR, а разница в 3 пункта – явно ощутимому.



Меньшая величина показателя UGR соответствует более комфортному освещению. Производители светотехнического оборудования приводят значения UGR для стандартных конфигураций светильников в типовых помещениях, что позволяет проводить сравнительный анализ различных моделей.

Несмотря на широкое применение, метод UGR имеет ряд общепризнанных ограничений:

1. Применимость уравнения (1) – корректна только для малоугловых источников с равномерной яркостью и не учитывает протяженные неоднородные блики.
2. Угловые размеры источников – метод действителен в диапазоне 0,0003–0,1 стерadian; за его пределами требуются альтернативные подходы.
3. Условия наблюдения – расчет предполагает расположение наблюдателя у стены прямоугольного помещения с фиксированным направлением взгляда, что не всегда соответствует реальным условиям (например, работа за столом в центре комнаты).

Для очень малых источников света (телесный угол  $< 0.0003$  ср) дискомфорт от бликов зависит от силы света, а не от яркости. В таких случаях система UGR не применяется (CIE 1995). Для них CIE (2002) предлагает формулу на основе силы света.

Для крупных источников (телесный угол  $> 0.1$  ср, например,  $1 \text{ м}^2$  с расстояния 3 м) система не проверена и её использование не рекомендуется [9].

Тем не менее, несмотря на эти ограничения, метод UGR остается практичным решением, доказавшим свою эффективность в большинстве случаев. Для источников света, выходящих за рамки его применимости, в публикации МКО (2002) предложены альтернативные методики оценки [7]. Однако остаются нерешенные вопросы, связанные с методикой UGR.

### Результаты экспериментов

Так как основной целью нашего исследования является разработка методики расчета UGR, которая будет учитывать зеркальные отражения, а также используемое программное обеспечение должно быть доступно и просто в использовании для любого проектировщика, то этим целям удовлетворяет программа 3ds Max от компании Autodesk, которая считается признанным лидером в своей области.

На рисунках 1 и 2 представлены визуализации расчетов освещения разными методами, на которых можно наблюдать значительные отличия в деталях визуализации. При расчете трассировкой лучей на стенах и на полу наблюдаются многократные переотражения. Причем нужно отметить, что весомый вклад в блёскость и визуальный дискомфорт вносят не только первые переотражения, но и вторые. В небольшом стандартном интерьере при наличии зеркальной компоненты отражения поверхностей больше 0,7 яркость первичных отражений может достигать 20 % от яркости источника света, а яркость вторичных переотражений – до 10 % от яркости источника. В то время как в DIALux evo (см. рис. 2) все эти вторичные источники в расчет UGR просто не войдут.

Также на примере шара на рисунке 3 наблюдаем, что при внесении в сцену одного небольшого предмета с поверхностью с зеркальным характером отражения можно получить дополнительные источники блескости как на самом объекте, так и в отражении на полу, что гарантированно повлечет за собой увеличение UGR. Это подчеркивает важность учета зеркальным отражений в сложных сценах освещения.

Для того чтобы посчитать вклад зеркальных бликов в показатель дискомфорта, проведем расчет в 3ds Max и сравним с расчетом для диффузных отражений в DIALux evo. Построим модель помещения с теми же размерами и ies профилями светильников, но вместо диффузного коэффициента отражения устанавливаем коэффициент направленного зеркального отражения на те же уровни, что и DIALux evo.

Программа 3ds Max очень точно рассчитывает распределение яркости и проводит визуализацию, но не предоставляет численные данные о распределении яркости напрямую пользователю. Для того чтобы получить точные численные значения распределения яркости, полученные при рендере освещения сцены, при экспорте данных из 3Ds Max будем использовать технологию High Dynamic Range Imaging (HDRI) – технология, предназначенная для захвата, обработки, хранения и визуализации изображений и видео с динамическим диапазоном, значительно превышающим возможности стандартных форматов (LDR/SDR).

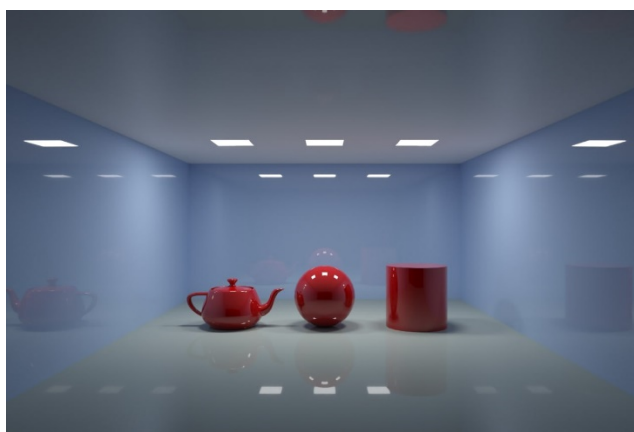


Рисунок 1. Визуализация методом трассировки лучей в 3ds Max

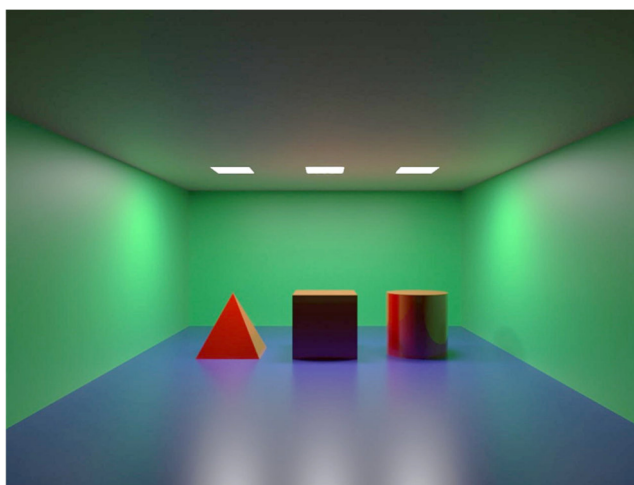


Рисунок 2. Визуализация методом фотонных карт в DIALux evo

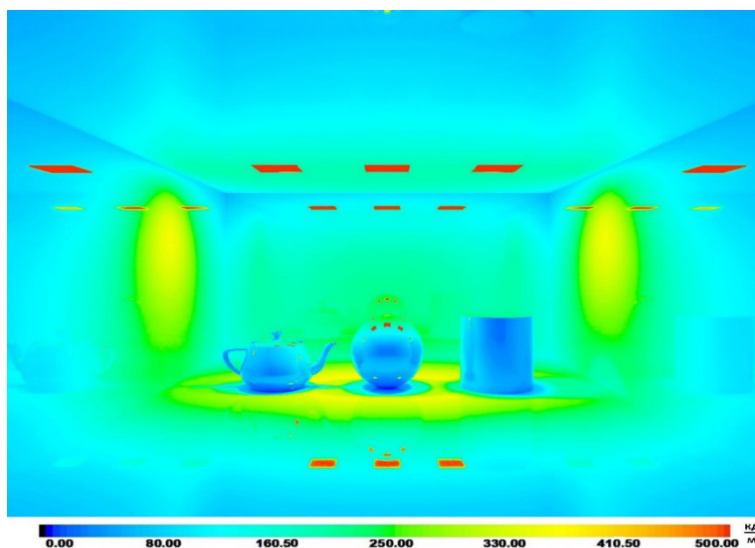


Рисунок 3. Распределение яркости в 3ds Max в псевдоцветах

Для проведения попиксельного расчета по распределениям яркости в качестве среды программирования была выбрана программа Matlab, так как она обладает высокой эффективностью и скоростью расчетов с матричным представлением данных. Именно такими данными и являются HDRI матрицы распределения яркости с большим количеством элементов – пикселей изображения. Также в Matlab есть специальный Image Processing Toolbox – большой набор инструментов для работы с изображениями.

Для чтения EXR файлов в Matlab существует функция `exrread`, однако она доступна в Image Processing Toolbox только начиная с MATLAB R2020b. Так как в исследовании разрабатывается методика, доступная для всех проектировщиков, то она должна работать не только на последней версии ПО. Если версия Matlab устарела, то можно использовать функцию `hdrread` и считывать каждый `hdr` файл, предварительно сохранив их в 3Ds Max по отдельности.

Для расчета UGR необходимо определить индекс позиции Гута для каждого пикселя. Согласно [10], индекс позиции  $p$  рассчитывается по специальной таблице, однако количество значений в таблице недостаточно для работы с изображением, поэтому нужно прибегнуть к интерполяции. Однако данный подход дает очень грубые результаты для большого количества пикселей в изображении. Поэтому было решено обратиться к опыту корейских исследователей. В статье [11] предлагают вычислять индекс позиции по следующей формуле:

$$p = e^{(35.2 - 0.31889 \cdot \alpha - 1.22 \cdot e^{\frac{-2\alpha}{9}}) \cdot 0.001 \cdot \beta + (21 + 0.267 \cdot \alpha - 0.0029663 \cdot \alpha^2) \cdot 10^{-5} \cdot \beta^2}, \quad (2)$$

где  $\alpha$  – угол между линией зрения и линией, соединяющей наблюдателя с точкой под источником, в градусах;

$\beta$  – угол между линией зрения и линией от наблюдателя к источнику.

Определение углов  $\alpha$  и  $\beta$  в пространстве сцены представлено на рисунке 4. Условно обозначены источник  $S$ , перпендикуляр к горизонту  $H$  и наблюдатель значком глаза.

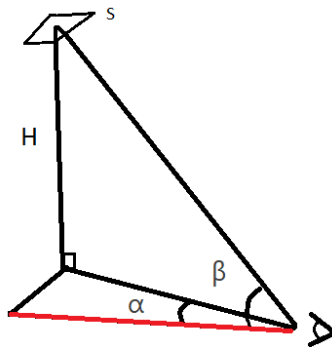


Рисунок 4. Определение углов  $\alpha$  и  $\beta$  для расчета индекса позиции  $p$

Полученная карта значений  $p$  может быть масштабирована для различных угловых размеров изображения. Отметим, что для поля зрения, соответствующего полному человеческому зрению ( $200^\circ \times 120^\circ$ ), распределение значений  $p$  приближается к табличным значениям, что подтверждает справедливость данной формулы.

В результате была получена карта значений  $p$ , представленная на рисунке 5.

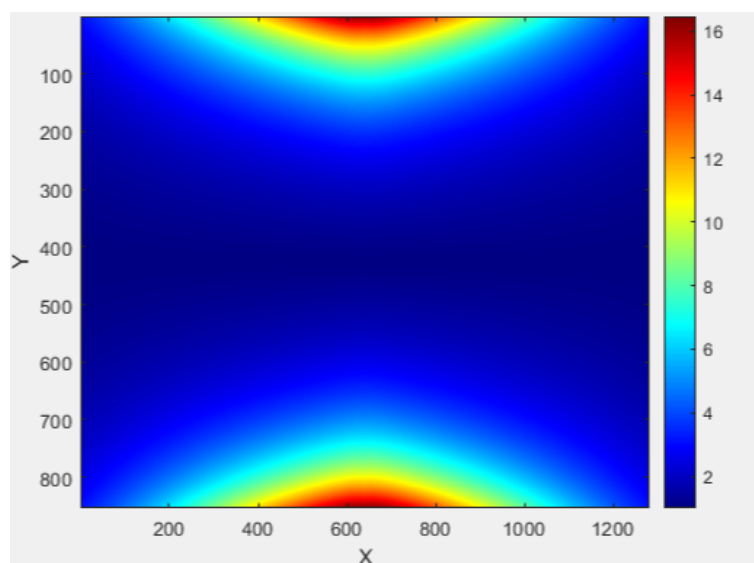


Рисунок 5. Карта значений  $p$  для поля зрения  $200^\circ \times 120^\circ$

По данному распределению наглядно понятен физический смысл индекса позиции. Чем ближе источник к линии зрения, тем меньше значение  $p$  в формуле UGR и тем больше его вклад в дискомфорт.

После вычисления всех необходимых параметров для расчета UGR были рассчитаны вклады всех пикселей в дискомфорт и произведено суммирование по стандартной формуле UGR (1).

Для того чтобы проверить работу методики попиксельного расчета, необходимо сначала определиться, какие пиксели будут являться источниками дискомфорта. Согласно МКО UGR рассчитывается по стандартной формуле только для источников направленного света, то есть формула применима лишь для первичных источников направленного света и вторичных источников, вызванных зеркальными отражениями.

Как же тогда выделить пиксели, которые войдут в расчет UGR? В работе [4] В.С. Желтов предлагает выделить пиксели с яркостью выше некоторого порога. Или выделить пиксели с резким перепадом яркости вокруг. Но этого может быть недостаточно, потому что по полученной визуализации и фотографии уже невозможно определить характер отражения светового пятна на поверхности. Вместо зеркального блестящего источника в расчет могут попасть пиксели с диффузно отраженным светом большой яркости. Помочь в определении природы источника может определение градиента яркости на границах пятна. Но это трудно реализуемая задача, не гарантирующая высокой точности результатов определения. Поэтому мы предлагаем обратиться к продвинутым технологиям 3D-визуализации в КГ [5].

Пакет Corona render дает возможность разделения излучения на каналы в зависимости от пути, по которому свет добрался до камеры.

В 3ds Max (Corona Renderer) каналы рендера Sources, Direct, Reflect и Indirect распределяют свет следующим образом:

1. Sources (свет от источников) – это свет, который напрямую попадает в камеру от заданных первичных фотометрических источников света в сцене, взаимодействуя с поверхностями.

2. Direct (прямой свет) – это свет, который напрямую попадает на объекты из источников освещения (лампы, солнце и т.д.) и отражается диффузно.

3. Reflect (отраженный свет) – это свет, отраженный зеркально (гладкие поверхности, блики). Для полностью диффузных материалов (без зеркальности) этот канал должен быть пустым.

4. Indirect (непрямой, рассеянный свет) – это свет, который отражается рассеянно от других поверхностей при вторичных и многократных отражениях. Для полностью диффузной сцены это основной канал.

В 3D-рендеринге термины Direct, Reflect и Indirect связаны не с тем, что видит камера, а с путем света от источника до камеры.

Для согласования методики расчета с стандартом UGR и для приближения модели к реальной физиологии зрения был предложен новый подход: используя алгоритм кластеризации, выделить области с наиболее резкими перепадами яркости на границе, поскольку глаз реагирует именно на контраст и перепады яркости.

Теперь будем рассматривать каждую яркую область как единый источник с усредненной яркостью и общим телесным углом. Благодаря такому подходу сохраним главное преимущество попиксельного расчета – высокую детализацию при учете блестящих источников.

В качестве алгоритма кластеризации используем функцию в Matlab. Для ее работы нам необходимо определить пороговый контраст источника с фоном, при котором будут детектироваться яркие области.

В случае непрерывного ПУРЯ в сцене освещения контраст является отношением градиента распределения яркости по полю зрения к средней яркости по полю. Обобщенный контраст в точке сцены можно определить с помощью следующего уравнения (3):

$$K(x, y) = \frac{|grad(L(x, y)p(x, y))|}{\bar{L}}, \quad (3)$$

где  $x, y$  – координаты точки на проекции сцены;

$L$  – яркость данной точки в направлении наблюдения;

$\bar{L}$  – средняя яркость по полю зрения;

$p(x, y)$  – функция, учитывающая различный вклад в реакцию в зависимости от расположения в поле зрения.

В уравнении (3)  $\rho$  имеет тот же физический смысл, что и индекс положения в формуле UGR. По данной формуле можно сделать вывод, что теперь изменение яркости по фону будет влиять на градиент яркости. Для того чтобы ограничить порог изменений яркостей, которые не будут играть существенную роль в восприятии качества освещения, в статье [4] исследователи предлагают использование определенного порога, ниже которого контрасты не следует учитывать.

Предполагается, что пороговый контраст будет определяться поставленной визуальной задачей.

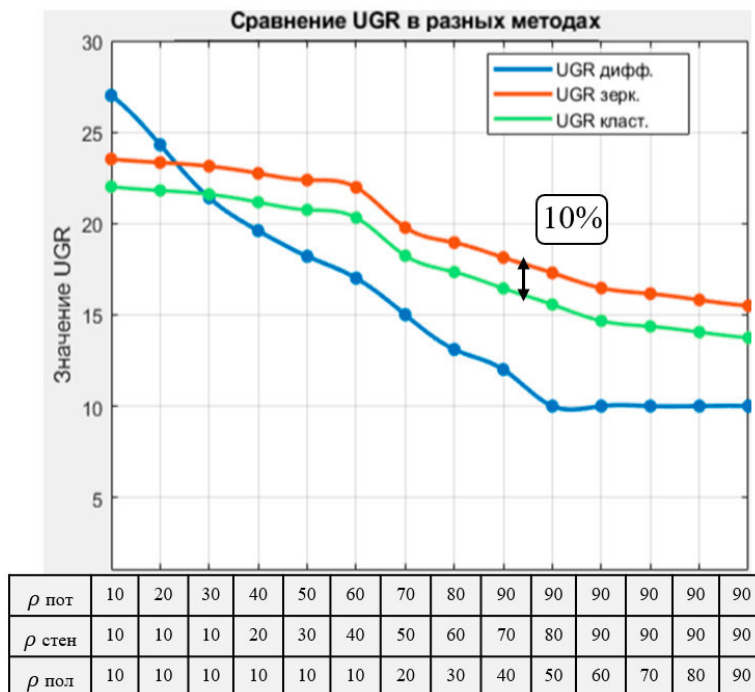


Рисунок 6. Результаты расчетов UGR методами кластеризации и попиксельного расчета

Значения, полученные методом кластеризации, оказались меньше, чем при попиксельном расчете в среднем на 1 единицу по шкале UGR. Средняя разница в процентном соотношении составила примерно 10 % (рис. 6). Это связано с отличиями в расчетных формулах, описанных выше, а также с тем, что мы намеренно не учитывали области с малыми контрастами на границе, которые не вызовут дискомфорт. Также мы избавились от таких нежелательных особенностей работы рендера, как возможность появления шумов и паразитных засветок при низких коэффициентах отражения поверхностей в сцене.

Ожидается, что дальнейшее увеличение яркостей в сцене еще больше увеличит разницу между кластерным и попиксельным расчетом, потому что вклад ложных блеских пикселей при попиксельном расчете будет расти, тогда как точность кластерного расчета не изменится.

Метод кластеризации позволил уточнить значения UGR. Несмотря на то, что процентная разница в результатах между методами не критична для стандартного проектировании освещения, главным преимуществом данной методики является большее приближение данной модели к реальным механизмам зрительного восприятия. Также нужно отметить, что этот алгоритм затрачивает меньше вычислительных ресурсов, поскольку расчет UGR проводится для небольшого количества источников в отличие от принудительного расчета для огромного количества пикселей в попиксельном методе.

С одной стороны, по зависимости для диффузного отражения видна тенденция на уменьшение UGR с ростом коэффициента отражения. Но при появлении зеркальной составляющей отражения в сцене появляются вторичные источники, вносящие вклад в повышение UGR. Эффект двух противоположных тенденции складываются, поэтому в результате получается зависимость, близкая к прямой. Однако снижение UGR за счет малых контрастов при высоких коэффициентах отражения превышает возможную компенсацию от появления вторичных источников, что приводит к итоговой тенденции спада. Соответственно при внесении дополнительных первичных источников света в сцену кривая пройдет еще выше и будет еще больше выпрямляться.

Благодаря данному методу расчета значения показателя UGR были уточнены. В результате исследования можно дать следующие рекомендации архитекторам. Даже при полном отсутствии зеркального отражения при  $\rho < 0,2$  UGR превышает 25, что является недопустимым для большинства типов помещений.

В этом случае, чтобы преодолеть значимый для многих нормативов барьер в  $UGR=20$ , следует использовать материалы с немного большей составляющей диффузного отражения, например покрытых темной матовой краской, с напылением специальных темных порошков и покрытий.

Для приемлемой зоны  $0,3 < \rho < 0,8$  зеркальные блики могут повысить показатель дискомфорта не более чем на один порог ощущений 3-4 единицы. Выбор отделки имеет широкий диапазон допустимых материалов и зависит от типа конкретного помещения и архитектурной концепции.

При высоких коэффициентах отражения  $\rho > 0,8$  зеркальная компонента отражения уже не вносит ощутимого влияния на ощущения. Дискомфортная блескость становится минимальной при высоких коэффициентах отражения. Однако необходимо помнить про возникающую блескость насыщения и ограничивать предельную яркость в поле зрения до  $1000-2000 \text{ кд/м}^2$  в зависимости от типа помещения.

Для подтверждения расчетов UGR по предлагаемой нами методике был проведен натурный эксперимент в реальном помещении учебной аудитории на кафедре светотехники НИУ «МЭИ».

В помещении были установлены два встраиваемых потолочных светильника  $600 \times 600 \text{ мм}$  с матовым рассеивателем и световым потоком 3800 лм каждый. Использовались диммируемые светильники, чтобы провести измерения при разных световых потоках. Светильники абсолютно новые, не эксплуатируемые ранее. Соответственно они имеют неискаженные заводские параметры.

Таким образом, эксперимент в реальном помещении подтвердил теоретические ожидания о росте UGR при наличии в сцене поверхностей с низким коэффициентом отражения. Метод кластеризации позволил исключить шумы, однако значения все еще остаются выше, чем для 3D-модели. Поэтому можно сделать вывод, что при упрощенном моделировании или при моделировании сложных 3М-сцен нужно учитывать возможное занижение оценки UGR по сравнению с реальной осветительной установкой до 10 %.

В результате исследования была сформирована комплексная методика для расчета UGR с помощью современных технологий КГ, ее можно представить в виде блок-схемы на рисунке 7.



Рисунок 7. Методика расчета, представленная в виде алгоритма

Первый этап – стандартное моделирование. Создается 3D-модель помещения, осуществляется подбор светильников.

Второй этап – использование САПР для получения визуализации с точным яркостным распределением.

Третий этап – программа в Matlab. Расчет UGR методом кластеризации.

Благодаря проектно-ориентированному подходу удалось обеспечить удобное практическое применение методики и ее доступность для проектировщиков ОУ.



## Заключение

В работе были рассмотрены исследования в области оценки дискомфорта освещения, выявлены проблемы и недостатки существующих методов расчета показателя дискомфорта, которые требуют дополнительных исследований. Последние исследования, проводимые в этой области, реализуемы только в случае реальных объектов с уже установленным световым оборудованием. Поэтому одной из задач настоящих исследований было создания метода с возможностью оценки показателя дискомфорта еще на стадии проектирования ОУ.

Анализ расчетов светотехнических параметров в программах DIALux 4.13 и DIALux evo подтвердил, что программы используют стандартный метод расчета UGR, в котором невозможен учет зеркальных бликов от поверхностей.

Также проведен более подробный анализ зависимостей UGR от коэффициентов отражения и даны соответствующие рекомендации проектировщикам и архитекторам для уменьшения показателя зрительного дискомфорта.

Увеличение коэффициентов отражения поверхностей однозначно уменьшает ощущение дискомфорта в помещении, однако необходимо ограничивать значения яркости в пределах определенного допустимого уровня, рассчитываемого для различных областей поля зрения в зависимости от соотношения яркости источников и яркости фоновой адаптации.

Разработана новая методика попиксельного расчета UGR с использованием программ КГ. Технология HDRi, доступная в различных программах для качественной 3D-визуализации, позволила получить точное распределение яркости в сцене освещения с учетом зеркальных отражений. А дополнительный функционал рендера Corona в 3Ds Max сделал возможным разделение распределения яркости в сцене на слои в зависимости от характера прямого и отраженного света. Это позволило выделить только источники направленного света и учесть их в дальнейшем расчете UGR.

Разработана программа в Matlab для попиксельного расчета UGR, суммирующая вклад блеских пикселей в общий UGR для всей сцены. В программе введены дополнительные уточнения, связанные с особенностями физиологии зрения. Широкое поле видения, соответствующее человеческому, корректировка формулы UGR для малых источников для изображений с большим разрешением.

Также предложен улучшенный алгоритм расчета с кластеризацией, который лучше согласовывался с механизмом восприятия блескости. Так как глаз реагирует на яркое пятно целиком, поэтому теперь программа выделяет точные области пикселей, соответствующие блеским источникам, а не единичные пиксели. Результаты расчетов UGR в простой сцене освещения показали уменьшение значений UGR на 10 % по сравнению с попиксельным расчетом. В более сложных сценах освещения с наличием зеркальных бликов и более высокими уровнями яркости ожидается еще большее расхождение значений. Таким образом, удалось избежать вклада ложных пикселей в дискомфорт и повысить точность результатов расчета.

В результате работы была сформирована методика уточненного расчета UGR с помощью программ компьютерной графики, позволяющая работать как с распределениями яркости реальных объектов, так и с моделью ОУ еще на стадии проектирования.

Однако проблема учета неравномерного распределения яркости у блеских источников при расчете UGR требует дальнейших исследований. Уточнение критериев кластеризации также нуждается в дополнительном изучении. Перспективой дальнейшего развития в данной области является подключение к методике нейросетей для прогнозирования значения UGR.

## Список литературы

1. Справочная книга по светотехнике: учебное пособие для обучающихся по образовательным программам высшего образования уровня бакалавриат и магистратура по направлению «Электроника и микроэлектроника» / Редакция журн. «Светотехника»; общ. ред. Ю.Б. Айзенберг, Г.В. Боос; науч. ред. С.Г. Ашурков, Л.П. Варфоломеев. 4-е изд., полностью перераб. и доп. Москва: [б.и.], 2019. 892 с.
2. Boyce P.R. Human Factors in Lighting 3rd Edition / CRC Press. 2014.
3. Мешков В.В., Матвеев А.Б. Основы светотехники. Ч.2: учебное пособие для вузов по специальности «Светотехника и источники света». 2-е изд., перераб. М.: Энергоатомиздат 1989. 431 с.
4. Расчет UGR на основе пространственно-углового распределения яркости / В.С. Желтов, В.Д. Чембаев // Автоматизация и моделирование в проектировании и управлении. 2019. № 4(06)

5. Будак В.П., Макаров Д.Н. Компьютерная графика с приложением в светодизайн: учебник для высших учебных заведений. Москва: Редакция журн. «Светотехника», 2022. 264 с.
6. ГОСТ 33392-2015. Метод определения показателя дискомфорта при искусственном освещении помещений.
7. Дискомфорт, обусловленный блескостью светильников с источниками света неравномерной яркости / Питер Торнс // Современная светотехника. 2020. № 5.
8. Wout van Bommel. Interior Lighting. Fundamentals, Technology and Application // Nuenen, The Netherlands. Springer Nature Switzerland. 2019.
9. СП 52.13330.2016. Естественное и искусственное освещение.
10. ГОСТ 55710-2013. Освещение рабочих мест внутри зданий. Нормы и методы контроля.
11. Kim W., Kim. J. T. A Formula of the Position Index of a Glare Source in the Visual Field// Indoor and Built Environment. 2010. T. 20, вып. 1. P. 47-53.

УДК 535.361.21

DOI: 10.25686/978-5-8158-2474-4-2025-250-260

## Модель отражения света от реальных поверхностей: матричный подход с учетом анизотропии рассеяния

В. П. Буда́к, Д. Н. Мака́ров, П. А. Смирнов

Национальный исследовательский университет «МЭИ», Москва, Россия

**Аннотация.** В работе предлагается наиболее полная аналитическая модель отражения света от реальных поверхностей. С учетом глубины проникновения света в реальные объекты часто в несколько длин волн рассматривается отражение от плоского слоя, ограниченного сверху случайно-неровной френелевской поверхностью и ламбертовским дном. Отличительной особенностью предлагаемой модели является выделение в решении анизотропной части, которая приближенно, но аналитически учитывает все особенности точного решения. Выделение особенностей в решении позволяет дискретизировать уравнение переноса излучения, а все описание переноса излучения приобретает матричную форму. Для учета френелевского отражения выделяются зона полного внутреннего отражения и зона закона Снеллиуса, что позволяет описывать перенос через границу в матричной форме. Это дает возможность рассчитывать отражение произвольной стопкой слоев на основе матрично-операторного метода, однако операция звездного произведения, лежащая в его основе, была модифицирована с учетом выделения анизотропной части решения. Полученная модель отличается высокой скоростью счета при реализации и не имеет ограничений на степень анизотропии рассеяния в объеме вещества.

**Ключевые слова:** коэффициент яркости поверхности, объемное рассеяние, поверхностное отражение, случайно-неровная поверхность

## Light reflection model for real surfaces: matrix approach with anisotropic scattering

V. P. Budak, D. N. Makarov, P. A. Smirnov

National Research University "MPEI"

**Abstract.** This paper presents a comprehensive analytical model for light reflection from real surfaces, accounting for multi-wavelength penetration depths. The model treats reflection from a flat layer bounded by a randomly rough Fresnel top surface and a Lambertian base. Its key innovation is the analytical isolation of an anisotropic component, approximating all features of the exact solution while enabling discretization of the radiative transfer equation into a matrix form. Fresnel reflection is handled by separating total internal reflection and Snell's law zones, with boundary conditions expressed as matrices. The model extends to multi-layer systems via a modified matrix operator method (star product adapted for anisotropy). Computational efficiency is high, with no constraints on scattering anisotropy.

**Keywords:** surface radiance factor, volumetric scattering, surface reflection, randomly rough surface.

### Введение

В компьютерной графике модели отражения поверхности (BRDF – Bidirectional Reflectance Distribution Function) играют ключевую роль в создании реалистичных изображений. В этой области имеются многочисленные исследования. Даже краткий обзор существующих моделей отражения представляет собой самостоятельную объемную тему анализа. Поэтому ограничимся ссылками на две фундаментальные монографии энциклопедического содержания, в которых приведена классификация основных моделей отражения [1, 2]. Анализ таких моделей показывает, что среди них преобладают модели, основанные на отражении на случайно-неровной поверхности (СНП). Если применимость такого рода моделей для металлических поверхностей не вызывает сомнения, то отражение от частично-прозрачных объектов подобными моделями описать невозможно. Наиболее полная модель отражения реальными объектами была предложена в работе [3], где отражение включало в себя как поверхностное отражение, так и объемное рассеяние подповерхностной средой. В [3] убедительно показано, что объяснить отражение света только френелевским отражением от поверхности невозможно.

Развитие полноценной модели отражения по [3] сильно сдерживалось отсутствием эффективных методов решения уравнения переноса излучения (УПИ) в рассеивающей среде, по сути, до конца XX века [4]. В настоящее время появились две модели, реализующие модель [3]. В [5] для решения УПИ используется метод дискретных ординат (МДО), что делает УПИ дискретным в матричной форме. Учет отражения от границ проводится на основе матрично-операторного метода (МОМ – ниже

в статье мы остановимся подробнее). В [6] этот подход далее развивается на произвольные плоскостойкие среды.

Сложность вычислений по матричным выражениям определяется размерами матриц, которые определяются анизотропией индикатрисы рассеяния – количество полиномов Лежандра в ее разложении. Подавляющее большинство сред имеет взвешенные частицы существенно больше длины волны [7], что в соответствии с теорией Г. Ми приводит к сотням членов в разложении индикатрисы по полиномам Лежандра. В настоящей работе предлагается эффективный метод расчета отражения слоя с учетом объемного рассеяния для произвольной степени анизотропии, что открывает возможность создания модели отражения реальными объектами.

### Постановка задачи

Рассмотрим краевую задачу УПИ для однородного плоского слоя, освещаемого сверху плоским мононаправленным источником в направлении  $\hat{\mathbf{i}}_0$  и ограниченного сверху плоской френелевской границей, а снизу ламбертовским дном [4]:

$$\begin{cases} \mu \frac{\partial L}{\partial \tau} + L(\tau, \mu, \varphi) = \frac{\Lambda}{4\pi} \iint x(\hat{\mathbf{i}}', \hat{\mathbf{i}}) L(\tau, \mu', \varphi') d\hat{\mathbf{i}}', \\ L(\tau, \hat{\mathbf{i}}) \Big|_{\tau=0, \mu>0} = \rho_F(\hat{\mathbf{i}}, \hat{\mathbf{i}}') L(\tau, \hat{\mathbf{i}}'), \quad L(\tau, \mu, \varphi) \Big|_{\tau=\tau_0, \mu<0} = \frac{\rho_L}{\pi} \int_{(\mu>0)} L(\tau_0, \hat{\mathbf{i}}') \mu d\hat{\mathbf{i}}', \end{cases} \quad (1)$$

где  $L(\tau, \hat{\mathbf{i}})$  – яркость светового поля в точке на оптической глубине  $\tau$  по направлению  $\hat{\mathbf{i}}$ ;  $\Lambda$  – альбеда однократного рассеяния;  $\tau_0$  – оптическая толщина слоя;  $x(\hat{\mathbf{i}}', \hat{\mathbf{i}})$  – индикатриса рассеяния. Здесь декартова система координат  $OXYZ$ , причем  $OZ$  направлена вниз. Значком “^” отмечаются единичные векторы,  $\mu = (\hat{\mathbf{i}}, \hat{\mathbf{z}})$ ;  $\varphi$  – азимут в плоскости  $OXY$  относительно оси  $OX$ ;  $\rho_L$  – коэффициент отражения от диффузного дна (Lambert);  $\rho_F(\hat{\mathbf{i}}, \hat{\mathbf{i}}')$  – коэффициент яркости френелевской (Fresnel) границы [4]:

$$\rho_F(v) = \frac{1}{2} \left\{ \frac{nv - \sqrt{1 - n^2(1 - v^2)}}{nv + \sqrt{1 - n^2(1 - v^2)}} + \frac{v - n\sqrt{1 - n^2(1 - v^2)}}{v + n\sqrt{1 - n^2(1 - v^2)}} \right\}, \quad (2)$$

где  $n$  – показатель преломления вещества плоского слоя;  $v$  – косинус угла отражения.

Важнейшим вопросом является применимость модели плоского слоя к отражению реальными объектами. С точки зрения описания переноса излучения играет роль не геометрия объектов, а геометрия области взаимодействия света со средой под поверхностью границы. Глубина проникновения света в глубину большинства объектов в реальности составляет несколько длин волн, когда кривизной границы можно пренебречь. Это аналогично подходу в описании переноса излучения в атмосфере при оптическом дистанционном зондировании – приближение независимого пикселя (IPA – Independent Pixel Approximation) [8].

Для решения задачи (1) искомую функцию  $L(\tau, \hat{\mathbf{i}})$  необходимо дискретизировать. Вследствие лучевой природы УПИ пространственно-угловое распределение яркости  $L(\tau, \hat{\mathbf{i}})$  неизбежно содержит разрывы и особенности, возникающие из-за точечных источников, резких границ сред или особенностей индикатрисы рассеяния [9]. Сингулярности невозможно точно представить в дискретной форме. Наилучшим решением здесь является представление решения в виде суммы анизотропной части (АЧ)  $L_a(\tau, \hat{\mathbf{i}})$ , которая пусть приближенно, но аналитически описывает всю анизотропную часть решения, включая все особенности и разрывы, и регулярной части (РЧ)  $\tilde{L}(\tau, \hat{\mathbf{i}})$ , которая является гладкой функцией, допускает дискретизацию и находится численно:

$$L(\tau, \hat{\mathbf{i}}) = L_a(\tau, \hat{\mathbf{i}}) + \tilde{L}(\tau, \hat{\mathbf{i}}). \quad (3)$$

Более того, в  $L_a(\tau, \hat{\mathbf{i}})$  можно включить и анизотропию, связанную с угловой анизотропией индикатрисы рассеяния. В работе [9] проведено детальное исследование особенностей решения УПИ и показано, что решение краевой задачи (1) содержит в угловом распределении особенность  $\delta(\hat{\mathbf{i}} - \hat{\mathbf{i}}_0)$ .

Идея такого представления была высказана в [10] для численного вычисления интегралов функций, содержащих особенности и разрывы. Представление (3) модифицирует УПИ в (1):

$$\mu \frac{\partial \tilde{L}}{\partial \tau} + \tilde{L}(\tau, \mu, \varphi) = \frac{\Lambda}{4\pi} \iint x(\hat{\mathbf{l}}', \hat{\mathbf{l}}) \tilde{L}(\tau, \mu', \varphi') d\hat{\mathbf{l}}' + F(\tau, \mu, \varphi), \quad (4)$$

где функция источников  $F(\tau, \mu, \varphi)$  связана с невязкой АЧ от точного решения УПИ.

На решении уравнения (4) с учетом краевых условий из (1) будет построена модель отражения. Идея метода нами впервые высказана в [11]. На ее основе предложена модель отражения в [12], однако не удавалось строго учесть френелевское отражение на границах. Построение полной модели отражения, строго учитывающей объемное рассеяние и отражение на границах, является предметом настоящей работы.

### Матричная теория переноса

Запишем азимутальную зависимость  $\tilde{L}(\tau, \hat{\mathbf{l}})$  в виде ряда Фурье:

$$\tilde{L}(\tau, \mu, \varphi) = \sum_{m=-M}^{m=M} C^m(\tau, \mu) e^{im\varphi} = \sum_{m=0}^{m=M} (2 - \delta_{m0}) C^m(\tau, \mu) \cos m\varphi, \quad (5)$$

а индикатрису рассеяния выразим в виде ряда по полиномам Лежандра:

$$x(v) = \sum_{k=0}^K \frac{2k+1}{4\pi} x_k P_k(v). \quad (6)$$

Подстановка (5) и (6) в (4) приводит к уравнению

$$\mu \frac{dC^m(\tau, \mu)}{d\tau} = -C^m(\tau, \mu) + \frac{\Lambda}{2} \sum_{k=m}^{N_s} (2k+1) x_k Q_k^m(\mu) \int_{-1}^1 Q_k^m(\mu') C^m(\tau, \mu') d\mu' + \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} f_k^m(\tau) Q_k^m(\mu), \quad (7)$$

где  $f_k$  – коэффициенты разложения  $F(\tau, \mu, \varphi)$  в азимутальный ряд Фурье.

Поскольку РЧ  $\tilde{L}(\tau, \hat{\mathbf{l}})$  является гладкой функцией, то интеграл в (7) можно представить по методу гауссовых квадратур

$$\int_{-1}^1 Q_k^m(\mu') C^m(\tau, \mu') d\mu' \approx \frac{1}{2} \sum_{j=1}^{N/2} w_j C_j^-(\tau) Q_k^m(\mu_j^-) + \frac{1}{2} \sum_{j=1}^{N/2} w_j C_j^+(\tau) Q_k^m(\mu_j^+), \quad (8)$$

где  $\mu_j^\pm$  – узлы гауссовых квадратур порядка  $N$  по схеме, при которой разделяются потоки вниз  $\mu_j^+ > 0$  и вверх  $\mu_j^- < 0$ ;  $C_j^\pm(\tau)$ ,  $w_j$  – соответствующие им веса и искомые дискретные ординаты коэффициентов  $C^m(\tau, \mu')$ . Индекс  $m$  из-за очевидности во всех дальнейших выражениях опускаем.

Если подставить (8) в (4) с учетом (5) и (6), то УПИ примет вид конечной системы обыкновенных дифференциальных уравнений. Если все переменные в (7) представить в виде векторов и матриц, то систему обыкновенных дифференциальных уравнений можно записать в виде матричного уравнения

$$\frac{d}{d\tau} \begin{bmatrix} \bar{C}_+ \\ \bar{C}_- \end{bmatrix} = -\bar{M}^{-1} (\bar{I} - \bar{X} \bar{W}) \begin{bmatrix} \bar{C}_+ \\ \bar{C}_- \end{bmatrix} + \begin{bmatrix} \bar{F}_+ \\ \bar{F}_- \end{bmatrix}, \quad (9)$$

где  $\bar{C}_\pm = [C_j^\pm]$ ,  $\bar{W} = \text{diag}(w_i)$ ,  $\bar{M} = \text{diag}(\mu_i^+, \mu_i^-)$ ,  $\bar{F}_\pm = [f_j^\pm]$ ,

$$\bar{X} = \frac{\Lambda}{4} \begin{bmatrix} \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^+) Q_k^m(\mu_j^+) & \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^+) Q_k^m(\mu_j^-) \\ \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^-) Q_k^m(\mu_j^+) & \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^-) Q_k^m(\mu_j^-) \end{bmatrix}.$$

Тем самым описание процесса переноса приобретает дискретную, матричную форму [13]. Система уравнений (9) представляет двухточечную краевую задачу, решение которой можно представить в матричной форме:

$$-\begin{bmatrix} \bar{C}_+(0) \\ \bar{C}_-(0) \end{bmatrix} + e^{\bar{B}\tau_0} \begin{bmatrix} \bar{C}_+(\tau_0) \\ \bar{C}_-(\tau_0) \end{bmatrix} = \int_0^{\tau_0} e^{\bar{B}t} \begin{bmatrix} \bar{F}_+(t) \\ \bar{F}_-(t) \end{bmatrix} dt, \quad (10)$$

где матричная экспонента  $e^{\bar{B}\tau} = \bar{U} e^{\bar{\Gamma}\tau} \bar{U}^{-1}$ ,  $\bar{\Gamma} = \text{diag}(\bar{\Gamma}_-, \bar{\Gamma}_+)$ ,  $\bar{\Gamma}_\pm = \text{diag}(\gamma_i)$ ,  $e^{\bar{\Gamma}\tau} = \text{diag}(e^{\bar{\Gamma}_-\tau}, e^{\bar{\Gamma}_+\tau})$ ,  $\gamma_i$ ,  $\bar{U}$  – собственные значения и матрица собственных векторов матрицы  $\bar{B} \equiv \bar{M}^{-1}(\bar{I} - \bar{X}\bar{W})$ .

Сложностью решения (10) является наличие у матричной экспоненты  $e^{\bar{B}\tau_0}$  положительных и отрицательных показателей, отчего с увеличением оптической глубины  $t_0$  обусловленность матрицы быстро падает. Для устранения этого применяется масштабное преобразование [14] – обе части уравнения умножаются на матрицу:

$$\bar{S} = \begin{bmatrix} 0 & e^{-\bar{F}_+ \tau_0} \\ \bar{I} & 0 \end{bmatrix}, \quad (11)$$

что приводит (10) к виду

$$\begin{bmatrix} -e^{-\bar{F}_+ \tau_0} \bar{u}_{22} & \bar{u}_{21} \\ -\bar{u}_{12} & e^{-\bar{F}_- \tau_0} \bar{u}_{11} \end{bmatrix} \begin{bmatrix} \bar{C}_-(0) \\ \bar{C}_+(\tau_0) \end{bmatrix} = \begin{bmatrix} \bar{J}_+ \\ \bar{J}_- \end{bmatrix} + \begin{bmatrix} e^{-\bar{F}_+ \tau_0} \bar{u}_{21} & -\bar{u}_{22} \\ \bar{u}_{11} & -e^{-\bar{F}_- \tau_0} \bar{u}_{12} \end{bmatrix} \begin{bmatrix} \bar{C}_+(0) \\ \bar{C}_-(\tau_0) \end{bmatrix}, \quad (12)$$

где  $\bar{U} \equiv \begin{bmatrix} \bar{u}_{11} & \bar{u}_{12} \\ \bar{u}_{21} & \bar{u}_{22} \end{bmatrix}$ ,  $\begin{bmatrix} \bar{J}_+ \\ \bar{J}_- \end{bmatrix} = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} \bar{S} \bar{I}_k(\tau_0) \bar{U}^{-1} \begin{bmatrix} \bar{q}_+ \\ \bar{q}_- \end{bmatrix}$ ,  $\bar{I}_k(\tau_0) \equiv \int_0^{\tau_0} e^{\bar{F}_t} f_k^m(t) dt$ ,  $\bar{q}_{\pm} = [Q_k^m(\mu_i^{\pm})]$ .

Уравнение (12) связывает падающие на слой потоки  $\bar{C}_+(0)$ ,  $\bar{C}_-(\tau_0)$  с выходящими из него  $\bar{C}_-(0)$ ,  $\bar{C}_+(\tau_0)$ . Разрешим (12) относительно выходящих потоков  $\bar{C}_-(0)$ ,  $\bar{C}_+(\tau_0)$ :

$$\begin{bmatrix} \bar{C}_-(0) \\ \bar{C}_+(\tau_0) \end{bmatrix} = \begin{bmatrix} \bar{\Phi}_- \\ \bar{\Phi}_+ \end{bmatrix} + \begin{bmatrix} \bar{R} & \bar{T} \\ \bar{T} & \bar{R} \end{bmatrix} \begin{bmatrix} \bar{C}_+(0) \\ \bar{C}_-(\tau_0) \end{bmatrix}, \quad (13)$$

где  $\begin{bmatrix} \bar{\Phi}_- \\ \bar{\Phi}_+ \end{bmatrix} \equiv \begin{bmatrix} -\bar{u}_{11} & e^{-\bar{F}_- \tau_0} \bar{u}_{12} \\ -e^{-\bar{F}_+ \tau_0} \bar{u}_{21} & \bar{u}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \bar{F}_- \\ \bar{F}_+ \end{bmatrix}$ ,  $\begin{bmatrix} \bar{R} & \bar{T} \\ \bar{T} & \bar{R} \end{bmatrix} \equiv \begin{bmatrix} -\bar{u}_{11} & e^{-\bar{F}_- \tau_0} \bar{u}_{12} \\ -e^{-\bar{F}_+ \tau_0} \bar{u}_{21} & \bar{u}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \bar{u}_{12} & -e^{-\bar{F}_- \tau_0} \bar{u}_{11} \\ e^{-\bar{F}_+ \tau_0} \bar{u}_{22} & -\bar{u}_{21} \end{bmatrix}$ .

Матричное уравнение выражает выходящие потоки из слоя  $\bar{C}_-(0)$ ,  $\bar{C}_+(\tau_0)$  через падающие на него  $\bar{C}_+(0)$ ,  $\bar{C}_-(\tau_0)$ , что позволяет его определить как матричное выражение углового спектра коэффициента яркости. Отметим наличие в уравнении источников  $\bar{\Phi}_{\pm}$ , связанных с представлением решения в форме (3) и невязки АЧ  $L_a(\tau, \hat{\mathbf{l}})$  от точного решения УПИ.

### Матрично-операторный метод

Полученная система в матричном виде (12) и (13) является системой из  $N$  уравнений с  $2N$  неизвестными  $\bar{C}_-(0)$ ,  $\bar{C}_+(\tau_0)$ , недостающие  $N$  условий дают граничные условия. Однако эти матричные соотношения открывают возможность для широкого обобщения. Если слой является неоднородным по глубине OZ, то его можно представить в дискретной форме в виде стопки плоских слоёв. Решения для каждого из слоев будут иметь форму матричного соотношения (12) или (13).

Рассмотрим случай двух смежных слоев:

$$\begin{bmatrix} \bar{C}_-^1 \\ \bar{C}_+^1 \end{bmatrix} = \begin{bmatrix} \bar{\Phi}_-^1 \\ \bar{\Phi}_+^1 \end{bmatrix} + \begin{bmatrix} \bar{R}_{1-} & \bar{T}_{1-} \\ \bar{T}_{1+} & \bar{R}_{1+} \end{bmatrix} \begin{bmatrix} \bar{C}_-^1 \\ \bar{C}_+^1 \end{bmatrix}, \quad \begin{bmatrix} \bar{C}_-^2 \\ \bar{C}_+^2 \end{bmatrix} = \begin{bmatrix} \bar{\Phi}_-^2 \\ \bar{\Phi}_+^2 \end{bmatrix} + \begin{bmatrix} \bar{R}_{2-} & \bar{T}_{2-} \\ \bar{T}_{2+} & \bar{R}_{2+} \end{bmatrix} \begin{bmatrix} \bar{C}_-^2 \\ \bar{C}_+^2 \end{bmatrix}, \quad (14)$$

где нижний индекс определяет принадлежность верхнему 1 или нижнему слою 2. Вертикальными стрелками обозначено падающее сверху или снизу на слой излучение.

Отметим, что поскольку слои соприкасаются друг с другом, то

$$\bar{C}_+^1 = \bar{C}_-^2 \equiv \bar{C}_{\downarrow}, \quad \bar{C}_-^1 = \bar{C}_+^2 \equiv \bar{C}_{\uparrow}, \quad (15)$$

и, разрешая систему (14) относительно отраженного  $\bar{C}_-^1$  и прошедшего  $\bar{C}_+^2$  излучений относительно падающего на систему излучения  $\bar{C}_{\downarrow}^1$  и  $\bar{C}_{\uparrow}^2$ , получим

$$\begin{bmatrix} \bar{C}_-^1 \\ \bar{C}_+^2 \end{bmatrix} = \begin{bmatrix} \bar{\Phi}_-^1 + \bar{T}_{1-} \bar{\alpha}_1 (\bar{R}_{2-} \bar{\Phi}_+^1 + \bar{\Phi}_-^2) \\ \bar{T}_{2+} \bar{\alpha}_2 (\bar{\Phi}_+^1 + \bar{R}_{1+} \bar{\Phi}_-^2) + \bar{\Phi}_+^2 \end{bmatrix} + \begin{bmatrix} \bar{R}_{1-} + \bar{T}_{1-} \bar{\alpha}_1 \bar{R}_{2-} & \bar{T}_{1-} \\ \bar{T}_{2+} \bar{\alpha}_2 \bar{T}_{1+} & \bar{R}_{2+} + \bar{T}_{2+} \bar{\alpha}_2 \bar{R}_{1+} \end{bmatrix} \begin{bmatrix} \bar{C}_{\downarrow}^1 \\ \bar{C}_{\uparrow}^2 \end{bmatrix}, \quad (16)$$

где  $\bar{\alpha}_1 = (\bar{I} - \bar{R}_{2-} \bar{R}_{1+})^{-1}$ ,  $\bar{\alpha}_2 = (\bar{I} - \bar{R}_{1+} \bar{R}_{2-})^{-1}$ .

Данный подход в теории переноса получил название матрично-операторного метода (МООМ) [15] и позволяет рассчитывать перенос чрез любую стопу плоских слоев: как рассеивающих, так и отражающих/пропускающих. Наше выражение отличается от соответствующего в [15] тем, что мы представили решение в виде суммы АЧ и РЧ (3), что привело к появлению в (16) первого вектора, связанного с источниками. Уравнение (16) совпадает с [15], если в нем положить  $\bar{\Phi}_{\pm}^{1,2} \equiv 0$ . Аналогичный метод решения был введен в [16] при исследовании решений уравнения Риккати, описывающего



распространение радиоволн в слоистых средах. Для анализа слоистых сред в [16] было введено операторное звёздное произведение (star product), позволяющее комбинировать матрицы рассеяния отдельных слоёв. В наших обозначениях звёздное произведение можно записать в следующем виде:

$$\begin{bmatrix} \tilde{R}_{1-} & \tilde{T}_{1-} \\ \tilde{T}_{1+} & \tilde{R}_{1+} \end{bmatrix} \otimes \begin{bmatrix} \tilde{R}_{2-} & \tilde{T}_{2-} \\ \tilde{T}_{2+} & \tilde{R}_{2+} \end{bmatrix} \equiv \begin{bmatrix} \tilde{R}_{1-} + \tilde{T}_{1-} \tilde{\alpha}_1 \tilde{R}_{2-} & \tilde{T}_{1-} \tilde{\alpha}_1 \tilde{T}_{2-} \\ \tilde{T}_{2+} \tilde{\alpha}_2 \tilde{T}_{1+} & \tilde{R}_{2+} + \tilde{T}_{2+} \tilde{\alpha}_2 \tilde{R}_{1+} \end{bmatrix}. \quad (17)$$

Матрица, описывающая отражение и пропускание слоя, получила название матрицы рассеяния

$$\tilde{S} \equiv \begin{bmatrix} \tilde{R}_- & \tilde{T}_- \\ \tilde{T}_+ & \tilde{R}_+ \end{bmatrix}. \quad (18)$$

Тогда матрицу рассеяния системы из  $m$  слоёв можно записать в виде последовательности звёздных произведений

$$\tilde{S}_\Sigma = \tilde{S}_1 \otimes \tilde{S}_1 \dots \otimes \tilde{S}_m, \quad (19)$$

что является математической записью MOM и позволяет перейти к расчету стратифицированных слоев.

В этом смысле наш подход является обобщением MOM и звёздного произведения на случай выделения в решении АЧ или, в более общем случае, наличия источников в слоях.

### Граничные условия

Для нас существенно, что отражения на границах раздела можно включить в MOM. Рассмотрим нижнюю ламбертовскую поверхность

$$L(\tau_0, \hat{\mathbf{l}}) \Big|_{\mu < 0} = \frac{\rho_L}{\pi} \int_{(\mu > 0)} L(\tau_0, \hat{\mathbf{l}}') \mu d\hat{\mathbf{l}}', \quad (20)$$

что с учетом представления яркости через сумму анизотропной и регулярной частей примет вид

$$L(\tau_0, \hat{\mathbf{l}}) \Big|_{\mu < 0} = \frac{\rho}{\pi} \int_{(\mu > 0)} \left( \tilde{L}_a(\tau, \mu, \varphi) + e^{-\tau/\mu_0} \delta(\hat{\mathbf{l}} - \hat{\mathbf{l}}_0) + \tilde{L}(\tau, \mu, \varphi) \right) \mu d\hat{\mathbf{l}}'. \quad (21)$$

С переходом в дискретные ординаты получим

$$m = 0: \tilde{C}_-(\tau_0) = \tilde{F}_- + \tilde{R}_L \tilde{C}_+(\tau_0), \quad (22)$$

где матрица ламбертовского отражения  $\tilde{R}_L$  состоит из  $N/2$  одинаковых строк  $2\rho\{\mu_j^+ w_j\}$ .

Собирая все вместе, представим перенос излучения через ламбертовскую границу матрицей переноса:

$$\begin{bmatrix} \tilde{C}_-^r \\ \tilde{C}_+^r \end{bmatrix} = \begin{bmatrix} \tilde{F}_- \\ \tilde{0} \end{bmatrix} + \begin{bmatrix} \tilde{R}_L & \tilde{0} \\ \tilde{I} & \tilde{0} \end{bmatrix} \begin{bmatrix} \tilde{C}_+^i \\ \tilde{C}_-^i \end{bmatrix}, \quad (23)$$

где индекс  $i$  относится к падающему на ламбертовскую поверхность излучению, а  $r$  – отраженному.

Существенно сложнее обстоит дело с френелевским отражением на верхней границе. На френелевской границе происходит преломление, и направления ординат (лучей потоков) меняются на границе по закону Снеллиуса

$$n_o \sin \theta_o = n_a \sin \theta_a, \quad (24)$$

где  $n_o$ ,  $n_a$  – показатели преломления среды, для определенности плоский слой будем называть океан, а среда сверху границы – атмосфера. Все, что относится к атмосфере, будем снабжать индексом  $a$  и положим ее  $n_a = 1$ , а вторая среда – океан, индекс  $o$ .

Очевидно, в соответствии с нашими допущениями  $n_o > n_a$ . Косинусы лучей с осью  $OZ$  в обеих средах в соответствии с (24) будут связаны между собой

$$\mu_a = \sqrt{1 - n_o^2 (1 - \mu_o^2)}. \quad (25)$$

Из (25) видно, что для  $\mu_o < \mu_t \equiv \sqrt{1 - 1/n_o^2}$  в океане возникает область полного внутреннего отражения (ПВО), когда лучи не выходят из океана, а идеально отражаются опять в океан (рис. 1). При этом образуются три зоны направлений лучей: две области ПВО I и III, и в центре область II преломления лучей на границе атмосфера – океан.

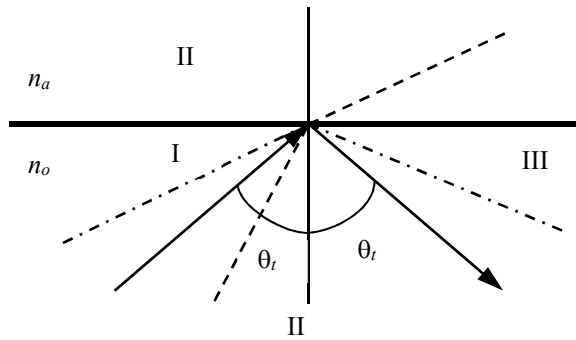


Рисунок 1. Области направлений лучей на границе раздела: лучи в области ПВО показаны штрих-пунктиром, а лучи в области преломления – пунктиром

Рассмотрим интеграл, входящий в оператор рассеяния (8) в соответствии с выделенными на рисунке 1 зонами углов:

$$\int_{-1}^1 Q_k^m(\mu') C^m(\tau, \mu') d\mu' = \int_{-1}^{-\mu_t} Q_k^m(\mu') C^m(\tau, \mu') d\mu' + \int_{-\mu_t}^{\mu_t} Q_k^m(\mu') C^m(\tau, \mu') d\mu' + \int_{\mu_t}^1 Q_k^m(\mu') C^m(\tau, \mu') d\mu'. \quad (26)$$

Первый и последний интегралы связаны с областью преломления, а второй относится к области ПВО. Для второго интеграла выполним преобразование

$$\int_{-\mu_t}^{\mu_t} Q_k^m(\mu') C^m(\tau, \mu') d\mu' = \mu_t \int_{-1}^1 Q_k^m(v) C^m(\tau, v) dv, \quad \mu' = \mu_t v. \quad (27)$$

Для первого и последнего интегралов в (26) сделаем преобразование переменной интегрирования к  $\mu_a$  по выражению (25)

$$\mu_o = \sqrt{1 - \frac{1 - \mu_a^2}{n_o^2}}, \quad d\mu_o = \frac{\mu_a d\mu_a}{\sqrt{n_o^2 - (1 - \mu_a^2)}} \quad (28)$$

что преобразует интегралы к виду

$$\int_{-1}^{-\mu_t} Q_k^m(\mu_o) C^m(\tau, \mu_o) d\mu_o = \int_{-1}^0 Q_k^m(\mu_o) C^m(\tau, \mu_o) \frac{\mu_a d\mu_a}{\sqrt{n_o^2 - (1 - \mu_a^2)}}, \quad \mu_o = -\sqrt{1 - \frac{1 - \mu_a^2}{n_o^2}}, \quad (29)$$

$$\int_{\mu_t}^1 Q_k^m(\mu_2) C^m(\tau, \mu_2) d\mu_2 = \int_0^1 Q_k^m(\mu_o) C^m(\tau, \mu_o) \frac{\mu_a d\mu_a}{\sqrt{n_o^2 - (1 - \mu_a^2)}}, \quad \mu_o = \sqrt{1 - \frac{1 - \mu_a^2}{n_o^2}}. \quad (30)$$

Последние два интеграла (29) и (30) можно объединить, что позволяет для интеграла (26) записать следующее выражение:

$$\int_{-1}^1 Q_k^m(\mu') C^m(\tau, \mu') d\mu' = \int_{-1}^1 Q_k^m(\mu_o) C^m(\tau, \mu_o) \frac{\mu_a d\mu_a}{\sqrt{n_o^2 - (1 - \mu_a^2)}} + \mu_t \int_{-1}^1 Q_k^m(v) C^m(\tau, v) dv. \quad (31)$$

Соответственно, интеграл рассеяния будет представлен двумя интегралами (31), каждый из которых независимо дискретизируем методом гауссовых квадратур с различными выражениями из-за разницы в подынтегральных выражениях. Пусть значения  $\{\mu_{ai}^{\pm}\}$  есть узлы квадратур для интеграла в атмосфере, упорядоченные по возрастанию модуля косинуса угла визирования. Атмосферные и океанские ординаты  $\bar{C}_+^o$ ,  $\bar{C}_-^o$  соответствуют друг другу по номерам, а углы связаны выражением (28). Второй интеграл в (31) позволяет применить двойную гауссову квадратуру с  $N_t$  узлами и впоследствии перейти в этой зоне к двум потоком ординат  $\bar{C}_+^t$ ,  $\bar{C}_-^t$ , которые на границе связаны идеальным зеркальным отражением.

Определим следующие матрицы:

$$\bar{X}_o = \frac{\Lambda}{4} \left[ \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^{\pm}) Q_k^m(\mu_j^{\pm}) \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^{\pm}) Q_k^m(\mu_j^{\pm}) \right]; \quad (32)$$

$$\tilde{X}_i = \frac{\Lambda}{4} \mu_i \left[ \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^+) Q_k^m(\mu_i^+) - \sum_{k=m}^K (2k+1) x_k Q_k^m(\mu_i^+) Q_k^m(\mu_i^-) \right]; \quad (33)$$

$$\tilde{W}_o = \text{diag}(w_i^o), \quad \tilde{W}_i = \text{diag}(w_i), \quad \tilde{M} = \begin{bmatrix} \mu_i^+ & 0 \\ 0 & \mu_i^- \end{bmatrix}, \quad \mu_i^+ = [\mu_o^+ \quad \mu_i^+], \quad \mu_i^- = [\mu_i^- \quad \mu_o^-], \quad (34)$$

где  $\{\mu_j^\pm\}$ ,  $j \in [1, N/2]$ ,  $N_a = N_o \equiv N$ ;  $\{\mu_l^\pm\}$ ,  $l \in [1, N_t/2]$ . Отметим, что  $i \in [1, (N + N_t)/2]$ , т.е. обе матрицы  $\tilde{X}$  прямоугольные. Определим новую матрицу (в формате MATLAB):

$$\tilde{X} = [\tilde{X}_i^- \tilde{W}_i^-, \tilde{X}_o^- \tilde{W}_o^-; \tilde{X}_o^+ \tilde{W}_o^+, \tilde{X}_i^+ \tilde{W}_i^+], \quad (35)$$

и введем вектор-столбцы

$$\tilde{C}_+ = \begin{bmatrix} \tilde{C}_o^+ \\ \tilde{C}_i^+ \end{bmatrix}, \quad \tilde{C}_- = \begin{bmatrix} \tilde{C}_i^- \\ \tilde{C}_o^- \end{bmatrix} \quad (36)$$

и аналогичные им для источников  $\tilde{F}$ .

В этом случае матричное УПИ полностью сохранит вид (9), а решение (12). Для введенных векторов граничные условия на френелевской границе в зоне ПВО можно записать в следующем виде:

$$\tilde{C}_i^+(0) = \tilde{C}_i^-(0), \quad (37)$$

для зоны Френеля

$$\tilde{C}_o^+(0) = \tilde{\rho}_F \tilde{C}_o^-(0), \quad (38)$$

где  $\tilde{\rho}_F = \text{diag}(\rho_F(\mu_i))$ .

Объединяя выражения (37) и (38), граничное условие на верхней френелевской границе можно представить в виде

$$\tilde{C}^+(0) = \tilde{\rho}_o \tilde{C}^-(0), \quad \tilde{C}^\pm = \begin{bmatrix} \tilde{C}_i^\pm \\ \tilde{C}_o^\pm \end{bmatrix}, \quad \tilde{\rho}_o = \begin{bmatrix} \tilde{I} & 0 \\ 0 & \tilde{\rho}_F \end{bmatrix}. \quad (39)$$

Тем самым с разделением лучей в среде на преломленные и область ПВО, с учетом закона Снеллиуса, френелевское отражение также удалось свести к матрице рассеяния. Соответственно, описание переноса излучения в стопке рассеивающих и отражающих слоев сведено к совокупности обобщенных звездных произведений.

### Анизотропная часть решения

Эффективность алгоритмов работы с матрицами – сложение, произведение, нахождение обратной матрицы – определяется размером матрицы, которая в нашем решении зависит от близости углового распределения РЧ к изотропному, что, в свою очередь, зависит от того, насколько близко АЧ к точному. Мы предлагаем в качестве АЧ выбрать малоугловую модификацию метода сферических гармоник (МСГ) [17, 18]. МСГ является наиболее общей формой малоуглового приближения [18], описывает все особенности точного решения в произвольной геометрии. Для случая плоского мононаправленного пучка АЧ в МСГ допускается, что азимутальная асимметрия пренебрежимо мала и решение представляется в виде разложения по полиномам Лежандра относительно направления падения пучка:

$$L_a(\tau, \hat{\mathbf{l}}) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi} l_k(\tau) P_k(\hat{\mathbf{l}} \cdot \hat{\mathbf{l}}_0), \quad (40)$$

что в соответствии с методом сферических гармоник приводит к бесконечной системе обыкновенных дифференциальных уравнений

$$\frac{\mu_0}{2k+1} \frac{d}{d\tau} [(k+1)l_{k+1}(\tau) + kl_{k-1}(\tau)] + (1 - \Lambda x_k) l_k(\tau) = 0, \quad \mu_0 = (\hat{\mathbf{l}}, \hat{\mathbf{z}}). \quad (41)$$

Обычно систему (41) решают, делая ее конечной, положив  $l_k(\tau) = 0$  для  $\forall k$ , большего некоторого  $N$ . Однако очевидно, что такой подход сглаживает анизотропию и устраняет все особенности решения.

Угловое распределение АЧ, включая особенности, сосредоточено в направлении источника  $\hat{\mathbf{l}}_0$ , что в соответствии со свойством масштабирования связи спектра и оригинала [19] позволяет предположить, что коэффициенты  $l_k(\tau)$  медленно монотонно убывают от  $k$ . Это позволяет ввести

непрерывную функцию  $l(k, t)$ , которая в целочисленных точках  $l_k(\tau) = l(k, \tau)$ . Поскольку  $l(k, t)$  медленно монотонно убывает, то ее можно разложить в ряд Тейлора, сохраняя два первых члена:

$$l(k \pm 1, \tau) \approx l(k, \tau) \pm \frac{\partial l(k, \tau)}{\partial k}. \quad (42)$$

Подстановка (42) в (41) приводит к уравнению в частных производных первого порядка

$$\mu_0 \frac{\partial}{\partial \tau} l_k(\tau) + \frac{1}{2k+1} \frac{\partial^2 l_k(\tau)}{\partial \tau \partial k} = -(1 - \Lambda x_k) l_k(\tau). \quad (43)$$

С учетом медленной монотонной зависимости  $l_k(\tau)$  от  $k$  вторым членом в (43) можно пренебречь, что приводит к обыкновенному дифференциальному уравнению, имеющему аналитическое решение

$$l_k(\tau) = \exp(-(1 - \Lambda x_k)\tau/\mu_0). \quad (44)$$

Рисунок 2 наглядно иллюстрирует гипотезу (42) – сравнение результатов численных расчетов  $l_k(\tau)$  по методу сферических гармоник с решением (44) для оптической глубины  $t=2$  среды с  $L=0.8$  и индикатрисой Хенби-Гринштейна  $g=0.97$ .

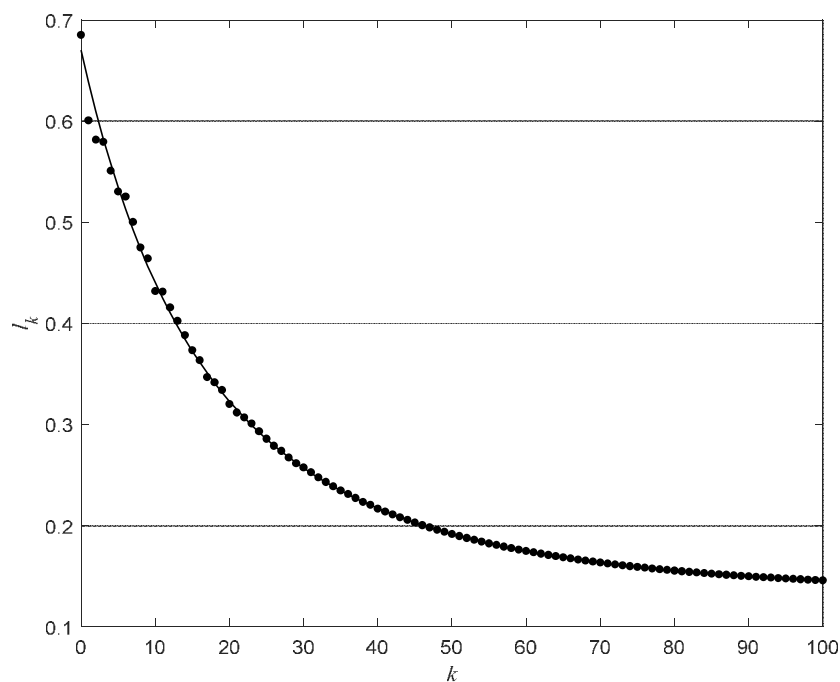


Рисунок 2. Сравнение коэффициентов разложения углового распределения яркости по полиномам Лежандра  $l_k(\tau)$ , вычисленных численно (точки) и в МСГ (сплошная линия)

В работе [20] исследованы свойства МСГ. Показано, что представление (42) позволяет получить в рамках МСГ решение в произвольной геометрии, а МСГ является наиболее общей формой малоугловых приближений.

Однако во все выражение в нашей модели входит не сама АЧ, а функция источников

$$F(\tau, \mu, \varphi) = -\mu \frac{dL_a(\tau, \mu, \varphi)}{d\tau} - L_a(\tau, \mu, \varphi) + \frac{\Lambda}{4\pi} \iint x(\hat{\mathbf{l}}, \hat{\mathbf{l}}') L_a(\tau, \mu', \varphi') d\hat{\mathbf{l}}'. \quad (45)$$

Подставляя в (45) ряд (40) с учетом решения (44) и переходя в дискретное представление – разложение в ряд Фурье по азимуту и дискретным ординатам, получим

$$\bar{F}_m(\tau) = (\bar{M}/\mu_0 - \bar{\mathbf{l}}) \sum_{k=m}^K \frac{2k+1}{4\pi} d_k q_k \tilde{Z}_k(\tau_1 + \tau) \bar{Q}_k^m + \Lambda e^{-(\tau_1 + \tau)/\mu_0} (\bar{\mathbf{l}} - \bar{M}/\mu_0) \sum_{k=m}^K \frac{2k+1}{4\pi} x_k q_k \bar{Q}_k^m, \quad (46)$$

где  $q_k \equiv Q_k^m(\mu_0)$ ,  $\bar{Q}_k^m = [Q_k^m(\mu_i)]$ ,  $\tilde{Z}_k(\tau) = e^{-(1-\Lambda x_k)\tau/\mu_0} - e^{-\tau/\mu_0}$ .

Достоинством (46) является аналитическая простота – ряд по полиному Лежандра, что упрощает расчет всех выражений модели отражения. Количество членов ряда  $K$  определяется индикатрисой рассеяния, но поскольку суммирование ряда однократное, то любая степень анизотропии не усложняет расчет.

Для оценки точности и эффективности сформулированной модели отражения был проведен расчет сравнения угловых распределений яркости отраженного и прошедшего излучений для слоя без границ, разбитого на случайное количество слоев произвольной толщи с оптическими параметрами исходного слоя так, что их суммарная оптическая толщина не изменялась. Сравнение проводилось с расчетами аналогичных распределений для однородного слоя той же оптической толщины по известным и принятым в теории переноса программам [21]. При любом количестве слоев расчеты совпадали с машинной точностью. Отметим, что время счета программы по описанному в настоящей работе методу линейно зависело от числа слоев разбиения.

Для проверки учета границ рассмотрим прозрачную пластину с ламбертовским дном и френелевской границей сверху. Если подставить соответствующие значения матриц отражения и пропускания из раздела «Граничные условия» в выражения (17), (19), то в результате несложных преобразований получим

$$\vec{C}^+(0_+) = \vec{\tau}_F \vec{R}_L \vec{F}_- + \vec{\tau}_F (\vec{I} - \vec{\rho}_F \vec{R}_L)^{-1} \vec{\rho}_F \vec{F}_-, \quad (47)$$

что полностью соответствует физике процесса переноса излучения в такой пластине – первый член соответствует отражению от ламбертовского дна, а второй – многократным переотражениям в пластине от границ.

Оценка точности метода сравнением с отражением реальных поверхностей вызывает существенные проблемы из-за отсутствия оптических параметров среды отражения. Подбор параметров среды дает точность сравнения для угловых распределений около 20 %.

### Обсуждение результатов

В настоящей работе сформулирована комплексная модель отражения и пропускания плоскими поверхностями. Ключевое отличие от существующих аналогов – явное выделение анизотропной составляющей решения на основе малоугловой модификации метода сферических гармоник (МСГ). Это обеспечивает:

- 1) высокую эффективность – время расчёта полного углового распределения яркости отражённого и прошедшего излучения для плоского слоя среды с отражением на границах не превышает сотых долей секунды на стандартном ноутбуке;
- 2) отсутствие ограничений на степень анизотропии рассеяния (например, для частиц произвольного размера). Влияние формы частиц на распределение яркости несущественно, значительно проявляется в поляризации излучения [7], что представляется дальнейшим развитием модели [18].

Нерассмотренными остались шероховатости границ. Здесь необходимо выделить [22] два случая: малых шероховатостей (без затенения друг друга) и больших шероховатостей. Второй случай не соответствует предложенной модели и требует отдельного исследования. В случае малых шероховатостей они представляются набором микрограней, с функцией распределения  $W(\hat{n})$  нормалей  $\hat{n}$ . В этом случае наша модель полностью сохраняет свою справедливость, но формулы Френеля (2) умножаются на  $W(\hat{n})$ .

Хотя модель обеспечивает высокую точность, её использование в компьютерной графике требует предварительной калибровки. Для упрощения предлагаются:

- 1) измерение отражательных характеристик реальных материалов;
- 2) подбор параметров модели по экспериментальным данным;
- 3) аппроксимация результатов упрощёнными BRDF-моделями для интеграции в рендереры.

### Заключение

Сущность предложенной модели основана на представлении решения на АЧ и РЧ. Правильное определение АЧ существенно повышает эффективность модели, но приводит к появлению функции источников от невязки АЧ до точного. Наилучшим способом выделения АЧ является МСГ. Расчет многослойной модели возможен на основе MOM, однако требует модификации операции звездного произведения.

Учет френелевского отражения на границе возможен в рамках MOM при отдельном выделении в отраженном излучении зоны ПВО. В этом случае сохраняется схема MOM. В ламбертовском отражении появляется функция источников, связанная с отражением АЧ.

Математическую модель отражения можно считать полностью завершённой. В развитие можно включить поляризацию, которая приводит к векторному УПИ, однако в соответствии с [18] вся схема сохраняется, но вектора и матрицы увеличиваются в 4 раза в соответствии с параметрами Стокса.

Разработанная модель, в случае ее реализации, имеет потенциальную возможность обеспечить эффективный и точный расчёт отражения и пропускания излучения, включая сложные случаи анизотропного рассеяния. Её ключевые преимущества:

- высокая скорость вычислений – с учетом ограниченности сравнений программы по точности все же можно предположить, что метод обеспечит высокую скорость вычислений (доли секунды для полного углового распределения);
- отсутствие принципиальных ограничений на степень анизотропии;
- гибкость (возможность расширения на векторный случай и многослойные системы).

Для внедрения в приложения (например, фотореалистичный рендеринг) требуется дальнейшая работа по параметризации модели на основе экспериментальных данных и, как следствие, разработка BRDF-аппроксимаций.

Таким образом, модель представляет собой законченный теоретический инструмент с широкими возможностями для практического применения.

### Список литературы

1. Computer Graphics: Principles and Practice / Hughes J.F., van Dam A., McGuire M., et al. Reading, MA: Addison-Wesley, 2013. 1263 p.
2. Pharr M., Jakob W., Humphreys G. Physically Based Rendering: from theory to implementation. Cambridge, MA: Elsevier Inc., 2017. 1233 p.
3. Pokrowski G.I. Zur Theorie der diffusen Lichtreflexioni // Zeit. f. Physik. 1924. B.30. S.66-72.
4. Thomas G.E., Stamnes K., Stamnes J.J. Radiative Transfer in the Atmosphere and Ocean. Cambridge: University Printing House, 2017. 512 p.
5. Stam J. An illumination model for a skin layer bounded by rough surfaces // Rendering Techniques. Proc. Eurographics Workshop, London, United Kingdom, June 25–27, 2001. P.39–52.
6. A comprehensive framework for rendering layered materials / Jakob W., d'Eon E., Jakob O., Marschner S. // ACM Transactions on Graphics. 2014. Vol. 33, no. 4. Article 118.
7. Дейрменджан Д. Рассеяние электромагнитного излучения сферическими полидисперсными частицами. М.: Мир, 1971. 165 с.
8. Harshvardhan Independent Pixel and Monte Carlo Estimates of Stratocumulus Albedo / Cahalan R.F., Ridgway W., Wiscombe W.J., Gollmer S. // Jour. Atm. Sci. 1994. Vol. 51, no. 24. P. 3776-3790.
9. Гермогенова Т.А. Локальные свойства решений уравнения переноса. М.: Наука, 1986. 272 с.
10. Канторович Л.В. О приближенном вычислении некоторых типов определенных интегралов и других применениях метода выделения особенностей // Математический сборник. 1934. Т. 41, вып. 2. С. 235–245.
11. Budak V.P., Klyuykov D.A., Korkin S.V. Complete matrix solution of radiative transfer equation for pile of horizontally homogeneous slabs // Journ. Quan. Spec. Rad. Trans. 2011. Vol. 112. P.1141–1148.
12. Modelling reflection from real surfaces / Basov A., Boos G., Budak V., Grimailo A. // CEUR Workshop Proceedings. 2021. Vol. 3027. P. 800–808.
13. Analysis of the discrete theory of radiative transfer in the coupled “ocean–atmosphere” system: current status, problems, and development prospects / Afanas'ev V.P., Basov A.Yu., Budak V.P., Efremenko D.S., and Kokhanovsky A.A. // J. Mar. Sci. Eng. 2020. Vol. 8, 202.
14. Shettle E.P., Green A.E.S. Multiple scattering calculation of the middle ultraviolet reaching the ground. Appl. Opt. 1974. Vol. 13, no. 7. P. 1567-1581.
15. Plass G.N., Kattawar G.W., Catchings F.E. Matrix operator theory of radiative transfer // Appl. Opt., 1973. Vol. 12. P. 314-329.
16. Redheffer R.M. The relation of transmission-line theory to scattering and transfer // J. Math, and Phys. 1962. Vol. 41, no. 1. P. 1–41.
17. Будак В.П. Малоугловая модификация метода сферических гармоник для расчёта светового поля бесконечно широкого пучка в мутных средах // Сб. науч. трудов. № 106. М.: Моск. энерг. ин-т, 1986. С. 20-25.
18. Budak V.P., Korkin S.V. On the solution of a vectorial radiative transfer equation in an arbitrary three-dimensional turbid medium with anisotropic scattering // Journ. Quan. Spec. Rad. Trans. 2008. Vol. 109. P. 220–234.
19. Титчмарш Э.Ч. Введение в теорию интегралов Фурье. М.-Л.: ОГИЗ, 1948. 418 с.



- 
20. Budak V.P., Klyuykov D.A., Korkin S.V. Convergence acceleration of radiative transfer equation solution at strongly anisotropic scattering // *Light Scattering Reviews 5: Single Light Scattering and Radiative Transfer* / ed. A.A. Kokhanovsky. Springer Praxis Books, 2010. P.147-204.
  21. Benchmark results in vector atmospheric radiative transfer / Kokhanovsky A.A., et al. // *Journ. Quan. Spec. Rad. Trans.* 2010. Vol. 111. P. 1931–1946.
  22. Басс Ф.Г., Фукс И.М. Рассеяние волн на статистически неровной поверхности. М.: Наука, 1972. 424 с.

## Эффективная генерация иерархических уровней детализации: параллельная обработка, инкрементальные обновления, полигональные упрощения на основе видимости

В. Н. Шуткин<sup>1</sup>, Н. К. Морозкин<sup>1</sup>, В. А. Семенов<sup>1,2</sup>, О. А. Тарлапан<sup>1,2,3</sup>

<sup>1</sup> Институт системного программирования им. В.П. Иванникова РАН, Москва, Россия

<sup>2</sup> Московский физико-технический институт, Долгопрудный, Россия

<sup>3</sup> Московский государственный университет имени М.В. Ломоносова,  
Москва, Россия

**Аннотация.** В работе рассматривается задача генерации иерархических уровней детализации (HLOD), широко применяемых в приложениях компьютерной графики для рендеринга больших сцен. Теория и практика уровней детализации HLOD являются хорошо проработанными темами исследований, однако вопросам их эффективного вычисления и оперативного обновления не уделялось должного внимания. Вместе с тем данные требования являются принципиальными для цифровых двойников сложных промышленных объектов и масштабных инфраструктурных программ. Целью работы являлись анализ и оценка подходов к эффективному вычислению уровней детализации HLOD в рамках стратегии агломеративной кластеризации. Для этого разработаны и исследованы специальные техники крупноблочного распараллеливания вычислений, инкрементального обновления уровней детализации и полигональные упрощения с удалением невидимых элементов. Проведенные вычислительные эксперименты с реальными проектными моделями гражданского строительства подтверждают эффективность техник оптимизации, обеспечивающих увеличение производительности генерации HLOD в среднем на порядок.

**Ключевые слова:** HLOD, 3D Tiles, кластеризация, параллельные вычисления, полигональные упрощения, инкрементальное обновление, удаление невидимой геометрии, интерактивный рендеринг

## Effective generation of hierarchical levels of detail: parallel processing, incremental updates, visibility-based polygonal simplifications

V. N. Shutkin<sup>1</sup>, N. K. Morozkin<sup>1</sup>, V. A. Semenov<sup>1,2</sup>, O. A. Tarlapan<sup>1,2,3</sup>

<sup>1</sup> Ivannikov Institute for System Programming RAS, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Dolgoprudny, Russia

<sup>3</sup> Lomonosov Moscow State University, Moscow, Russia

**Abstract.** The paper considers the problem of generating hierarchical levels of detail (HLOD), widely used in computer graphics applications for rendering large scenes. The theory and practice of levels of detail are well-developed research topics, but the issues of their efficient computation and incremental updating have not received due attention. At the same time, these requirements are fundamental for digital twins of complex industrial objects and large-scale infrastructure programs. The aim of the work was to analyse and evaluate approaches to accelerated computation of HLOD within the framework of the agglomerative clustering strategy. For this purpose, optimization techniques for large-grained parallelization of computations, incremental updating of levels of detail and polygonal simplifications with the removal of invisible elements were developed and studied. The conducted computational experiments with real design models of civil construction confirm the effectiveness of the techniques that provide a total increase in the performance of HLOD generation by an order of magnitude on average.

**Keywords:** HLOD, 3D Tiles, clustering, parallel computing, polygonal simplification, incremental update, occlusion culling, interactive rendering

### Введение

В настоящее время иерархические уровни детализации (hierarchical levels of detail, HLOD) являются одним из наиболее распространенных и эффективных подходов к рендерингу больших трехмерных сцен. Теория и практика методов детализации сцен являются хорошо проработанными темами исследований, которым посвящено большое число работ. В частности, установлено, что иерархические уровни детализации обеспечивают хорошую масштабируемость приложений по отношению к сложности сцен и допускают интерактивный рендеринг больших сцен во внешней памяти. Вместе с тем широкое применение цифровых двойников сложных промышленных объектов и масштабных инфраструктурных программ диктует новые требования не только к производительности рендеринга, но и к оперативности подготовки уровней детализации, необходимых для этого. Вопросам

эффективного вычисления и оперативного обновления уровней детализации в условиях перманентно изменяемых проектных данных не уделялось должного внимания, а они оказываются критичными для широкого класса приложений. Например, подготовка уровней детализации HLOD для цифрового двойника города с детальными информационными моделями зданий, сооружений и инфраструктуры может занимать сутки процессорного времени, что является неприемлемым для контроля и мониторинга оперативной ситуации.

Целью работы являлись анализ и оценка подходов к эффективному вычислению уровней детализации HLOD в рамках стратегии агломеративной кластеризации. Для этого разработаны и исследованы специальные техники крупноблочного распараллеливания вычислений, инкрементального обновления уровней детализации и полигональных упрощений с удалением невидимых элементов. Проведенные вычислительные эксперименты с реальными проектными моделями гражданского строительства подтверждают эффективность техник оптимизации, обеспечивающих увеличение производительности генерации HLOD в среднем на порядок. Заметим, что результаты работы распространяются и на иерархические динамические уровни детализации (hierarchical dynamic levels of detail, HDLOD), в основе которых лежит та же агломеративная стратегия [1–5]. В ходе работы были также выявлены дополнительные оптимизационные возможности, которые отмечены в соответствующих разделах статьи и которые предстоит исследовать в дальнейшем.

Применение альтернативных представлений геометрических объектов при отображении сцен имеет давнюю и богатую историю. Идея применения упрощённых представлений для дальних объектов сцены с целью снижения нагрузки на аппаратные средства и увеличения производительности рендеринга была высказана ещё в 1976 году [6]. Окончательно она оформилась в метод и связанный с ним термин «уровни детализации» (level of detail, LOD) со стремительным развитием персональных компьютеров и средств компьютерной графики в 1990-е годы. В это же время идея распространяется на случай подготовки нескольких уровней детализации для каждого индивидуального объекта сцены и выбора подходящего уровня в зависимости от расстояния от точки обзора до объекта [7,8]. Однако данный метод порождает ограничения, обусловленные дополнительным анализом расположения объектов при рендеринге сцены и хранением избыточных представлений объектов в видеопамяти.

Чтобы преодолеть данные ограничения и оперировать не большим числом индивидуальных объектов, а их группами, в 2001 году были разработаны иерархические уровни детализации HLOD [9]. В иерархии каждый узел представляет собой упрощённое представление для группы объектов. Как правило, листовые узлы соответствуют точному представлению группы, а по мере продвижения к корню иерархии и объединения объектов все в большие группы точность представления монотонно уменьшается, так что корень соответствует упрощённому представлению всей сцены. Иерархические уровни детализации расходуют большой объем памяти и поэтому критичными для них становятся вопросы организации хранения во внешней памяти и обеспечения быстрого доступа для обеспечения интерактивности приложений и плавной навигации по сцене [10,11]. Возможности предварительной загрузки уровней детализации [12], оптимизации схемы хранения на внешнем диске [13], минимизации трафика между центральным и графическим процессорами [14], а также кэширования наиболее используемых уровней детализации непосредственно в видеопамяти [15] хорошо проработаны и успешно применяются на практике. Необходимо отметить, что методы HLOD плохо подходят для динамических сцен из-за необходимости пересчета уровней детализации при каждом изменении модельного времени. Предварительная подготовка уровней для тысяч и миллионов временных интервалов, на которых происходят изменения сцены, не представляется возможной из-за требуемых огромных ресурсов, а пересчет уровней во время исполнения графического интерактивного приложения трудно реализуем из-за высокой вычислительной сложности.

Ряд работ посвящён уровням детализации для динамической геометрии, под которой подразумевается деформация полигональных моделей. Деформируемые модели с пространственно-временными уровнями детализации рассматриваются в [16]. В [17] представлена концепция «прогрессивной сетки» с учетом временного фактора. Рендеринг деформируемых изоповерхностей в приложениях моделирования динамики жидкости обсуждается в [18]. Непрерывные и иерархические уровни детализации для толп персонажей со скелетной анимацией рассматриваются в [19] и [20] соответственно.

В 2020-е годы иерархические уровни детализации получили дальнейшее развитие в серии работ, адресованных проблеме рендеринга больших динамических сцен. С этой целью был выделен класс детерминированных псевходинамических сцен с доминированием дискретных событий. Данный класс находит важные приложения, связанные, в частности, с цифровыми двойниками сложных промышленных проектов и масштабных инфраструктурных программ [21, 22]. В подобных приложениях динамика сцены определяется заранее подготовленным календарно-сетевым графиком проектных работ, в соответствии с которым элементы конструкций и оборудования устанавливаются в predetermined положение или удаляются. В сцене могут присутствовать объекты с иным характером динамики, например техника или персонал, перемещаемые по строительной площадке, однако доля таких объектов относительно невелика. Для больших псевходинамических сцен был разработан метод иерархических динамических уровней детализации во внешней памяти HDLOD с поддержкой консервативного (гарантирующего заданную пространственную и временную точность) и интерактивного (стремящегося обеспечить максимально возможную частоту генерации изображений при приемлемом уровне реализма) режимов отображения [1–3]. Также был предложен многовариантный пакетный метод, обеспечивающий одновременный рендеринг нескольких динамических сценариев [4].

Методы упрощения геометрических моделей, прежде всего полигональных сеток как наиболее распространенного вида объектов в приложениях компьютерной графики, играют ключевую роль в методах детализации LOD, HLOD и HDLOD. Полигональные упрощения представляют и самостоятельный интерес. В монографии [23] и тематическом обзоре [24] систематически изложены базовые идеи и принципы, лежащие в их основе, а именно: стягивание рёбер [25], удаление вершин [26] и кластеризация вершин [7,27]. Иногда особенности сцены, например, обусловленные фотограмметрической реконструкцией здания или целого города, позволяют выполнить дополнительные упрощения материалов и текстур и тем самым оптимизировать представление уровней детализации для последующего рендеринга [28–31].

В последние годы теория и практика уровней детализации получили дальнейшее развитие. Как результат иерархические уровни детализации были распространены на нейронные поля излучения NeRF [32], гауссианы 3DGS [33], облака точек [34], воксельные и гибридные представления [35, 36]. В работе [37] предложено гибридное представление полигональная сетка-NeRF.

В последние годы графические интерфейсы (API) претерпевают значительное развитие, благодаря чему становится возможным выбирать необходимое геометрическое представление объекта непосредственно на графическом процессоре. Определенную известность получили Unreal Engine Nanite [38] и подобные ей технологии [39]. Они обеспечивают рендеринг больших полигональных сеток, как правило, представляющих собой топологически замкнутые двумерные поверхности, на основе подготовленных деревьев кластеров. Кластеризация сеток осуществляется сверху-вниз с последующим упрощением фрагментов и исключением разрывов на границах. Обход загруженного в видеопамять дерева кластеров и рендеринг отобранных кластеров осуществляются непосредственно на графическом процессоре, причем загрузка данных кластеров происходит напрямую с диска в видеопамять. В статье [40] предлагается схема управления памятью графического процессора при рендеринге больших полигональных сеток. Статья [41] адресована выборочному уточнению большой сетки при трассировке лучей.

В настоящей работе обсуждается задача рендеринга больших сцен в более общей постановке. Предполагается, что сцена представлена набором полигональных моделей в виде так называемого «полигонального супа» без каких-либо дополнительных требований к топологии, а применяемые методы иерархических уровней детализации HLOD носят универсальный характер, не требуют аппаратной поддержки специальных функций и могут быть применены, в частности, для визуализации сложных сцен в web-браузерах на мобильных устройствах.

### Постановка задачи

Задачами работы являлись разработка и экспериментальное исследование техник оптимизации вычисления уровней детализации HLOD. Поскольку техники разработаны применительно к

конкретной стратегии агломеративной кластеризации объектов [5] со своими особенностями представления уровней детализации и правилами их формирования, приведем ее краткое описание.

Пусть сцена  $S$  определена в трёхмерном евклидовом пространстве  $E^3$  и представлена как набор объектов  $s(g_s, l_s) \in S$ . Каждый объект имеет фиксированное геометрическое представление  $g_s$ , а также уникальный идентификатор  $l_s$ . В рамках данной работы будем считать, что  $g_s$  задаётся в виде полигональной сетки треугольников, а  $l_s$  может быть задан любым способом: строка, номер, GUID или UUID.

Будем рассматривать генерацию HLOD как построение дерева кластеров. Каждый кластер  $c(g_c, b_c, \varepsilon_c)$  имеет фиксированное геометрическое представление  $g_c$ , ограничивающий объём  $b_c$  и геометрическую погрешность  $\varepsilon_c$ . Для кластеров определено отношение агломерации  $<$ . Запись  $c' < c$  означает, что кластер  $c'$  является ребёнком кластера  $c$  в дереве, а также то, что кластер  $c'$  вошёл в кластер  $c$  во время кластеризации. Пусть у кластера  $c$  имеется  $n$  детей  $\forall i = 1, \dots, n \ c_i < c$ . Это означает, что геометрическое представление  $g_c$  получено путём объединения и упрощения представлений всех дочерних кластеров, а именно  $g_c = \bigcup_{i=1, \dots, n} g_{c_i}$ , после чего  $g_c$  упрощено с геометрической погрешностью  $\varepsilon_c$ . Листовые кластеры получаются путём объединения объектов, а кластеры последующих уровней – путём объединения кластеров (и, возможно, объектов). Для листового кластера его геометрическое представление получается путём объединения представлений вошедших в него объектов без упрощения. Если в листовой кластер  $c$  вошли  $n$  объектов  $s_1, \dots, s_n$ , то  $g_c = \bigcup_{i=1, \dots, n} g_{s_i}$ . Идентификатор объекта  $l_s$  сохраняется для каждого треугольника в геометрии кластера. Ограничивающий объём кластера всегда охватывает ограничивающие объёмы дочерних кластеров:  $\forall c' < c \ b_{c'} \subseteq b_c$ . Геометрическая погрешность дочернего кластера не может превышать геометрическую погрешность родительского кластера:  $\forall c' < c \ \varepsilon_{c'} \leq \varepsilon_c$ . Листовые кластеры содержат точную (неупрощённую) геометрию:  $\forall c: \nexists c' < c \ \varepsilon_c = 0$ . Пример дерева кластеров представлен на рис. 1.

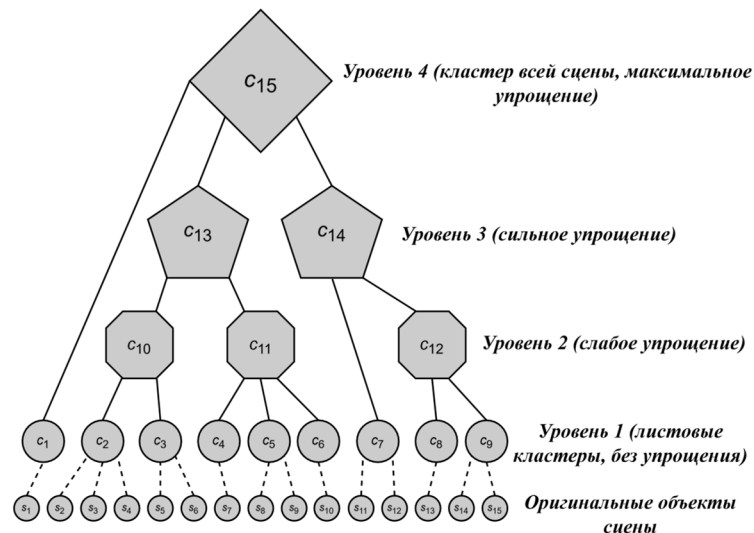


Рисунок 1. Пример дерева кластеров

Опишем общую стратегию построения иерархии уровней детализации HLOD. На первом этапе строится дерево кластеров в результате агломеративной кластеризации объектов снизу вверх. Процедура кластеризации стартует с оригинальных объектов сцены и затем распространяется на сформированные листовые и промежуточные кластеры вплоть до достижения единого корневого кластера всей сцены. На втором этапе вычисляются геометрические представления кластеров путем объединения представлений дочерних кластеров и их последующего упрощения. Данный этап является самым вычислительно затратным.

## Теория

Перечислим основные подходы к эффективному вычислению уровней детализации HLOD на основе описанной выше стратегии агломеративной кластеризации.

Прежде всего, это ускорение полигонального упрощения, которое должно обеспечить меньшие вычислительные затраты на генерацию кластеров приемлемого качества. Причем качество следует оценивать в многокритериальной постановке, которая бы учитывала геометрическую погрешность упрощенной сетки, сохранность ее визуальных особенностей, а также сложность представления, влияющую на производительность рендеринга. К сожалению, на практике не удастся контролировать все критерии из-за особенностей методов упрощения, поэтому представляется более перспективным выбор одного из известных методов, (1) обладающего невысокой вычислительной сложностью, (2) гарантирующего соблюдение заданной геометрической погрешности и (3) допускающего высокую степень упрощения (4) при сохранении визуальных особенностей полигональной сетки.

Важным подходом к оптимизации вычисления уровней детализации HLOD, безусловно, является распараллеливание. Видятся две основные схемы распараллеливания: так называемая мелкозернистая (fine-grained) на уровне подготовки упрощенного полигонального представления для одного обрабатываемого кластера и крупноблочная (coarse-grained) на уровне подготовки представлений для целых групп кластеров. Обе схемы имеют свои достоинства и недостатки.

Мелкозернистая схема предполагает применение методов полигонального упрощения, допускающих эффективное распараллеливание. Это налагает жесткие ограничения на выбор самих методов, а также требует эффективных процедур декомпозиции исходной сетки и объединения результатов упрощения, которые обычно требуют значительных вычислительных ресурсов и выполняются преимущественно последовательным образом. Как результат эффект от распараллеливания вычислений для частей сетки нивелируется затратами на их пред- и постобработку.

Более перспективной представляется крупноблочная схема, поскольку ее реализация не зависит от особенностей применяемых методов полигонального упрощения и не ограничивает в их выборе, диктуемом иными перечисленными выше факторами. Данная схема распараллеливания применима к группам кластеров, относящихся к одному уровню детализации, поскольку их геометрические представления зависят только представлений дочерних кластеров. Данная схема не требует какой-либо пред- и постобработки геометрических представлений и реализуется на доступных мультитядерных и мультипроцессорных архитектурах. В настоящей работе рассматривается одна из возможных реализаций данной схемы.

Наконец, третий обсуждаемый подход состоит в инкрементальном обновлении уровней детализации HLOD. Данная возможность важна для графических приложений поддержки коллективной работы, в ходе которой участниками перманентно вносятся изменения в общие проектные данные. Инкрементальный пересчет уровней детализации целесообразен для цифровых двойников сложных промышленных объектов, параметры которых актуализируются и уточняются на протяжении всего их жизненного цикла. Применение подхода мотивировано для приложений САПР, в которых сложные 3D-модели подлежат постоянным редакторским правкам и нуждаются в оперативном отображении результатов. Если изменения носят локальный характер и распространяются на относительно небольшое число кластеров, инкрементальное обновление имеет хорошую перспективу.

Опишем более подробно техники оптимизации, реализующие перечисленные подходы к ускорению вычисления уровней детализации HLOD.

### *1. Полигональные упрощения*

Методы полигонального упрощения различаются не только указанными выше базовыми принципами (стягивание ребер, удаление вершин и кластеризация), но и параметризацией, способами организации вычислений (итерационные, регулярные), условиями прерывания [23]. Подобные различия приводят к ситуации, когда методы имеют разную вычислительную сложность и генерируют сетки разного качества. В рамках проводимого исследования не ставилась цель комплексного сравнения методов с учетом сложности многокритериального сравнения результатов упрощения.

Вместо этого было принято решение об использовании популярного метода стягивания рёбер с квадратичной метрикой погрешности [25]. В сочетании с процедурой удаления невидимых граней данный метод удовлетворяет всем перечисленным выше требованиям. Метод реализует итеративное стягивание рёбер полигональной сетки, пока не будет достигнуто целевое число треугольников. Для



минимизации геометрической погрешности рёбра обрабатываются в определённом порядке: в первую очередь стягиваются те ребра, которые вносят наименьшие изменения в сетку. Для этого каждому ребру приписывается стоимость и рёбра сортируются по возрастанию стоимости. В данном случае используется квадратичная метрика, оценивающая кривизну поверхности. Стягивание ребра приводит к вырождению опирающихся на него треугольников и, следовательно, к сокращению числа треугольников в сетке.

Для оптимизации базового метода было предложено дополнительно использовать процедуру удаления невидимых граней. Действительно, в сценах гражданских зданий и промышленных сооружений обычно большая часть элементов невидима снаружи из-за наличия стен, перекрытий, коробов инженерного оборудования. Другим фактором учета невидимых граней является сама организация стратегии агломеративной кластеризации. Поскольку при объединении полигональных представлений кластеров значительная часть граней может быть перекрыта другой и впоследствии удалена, процесс объединения кластеров предшествует их упрощению.

Предлагаемая техника удаления невидимых граней в уровнях детализации HLOD налагает дополнительное ограничение на их использование в ходе рендеринга – камера всегда должна находиться снаружи отображаемых кластеров. Данное требование естественным образом реализуется в случае консервативного рендеринга сцены, поскольку отображение ближайших объектов предполагает использование нижних листовых кластеров с точной и полной геометрией. В случае интерактивного рендеринга необходимые кластеры могут быть еще не подгружены и тогда используются доступные верхние кластеры с упрощенным представлением. Когда камера оказывается внутри них, наблюдатель может не увидеть важные удаленные артефакты. Данное ограничение не является столь критичным, если навигация по сцене осуществляется плавно и приложение успевает подгружать необходимые кластеры. Ограничение оказывается принципиальным, если приложение предусматривает функции пространственного анализа сцены с использованием отсекающих плоскостей и объектных фильтров. Удаленные грани и объекты не могут быть реконструированы в представлении HLOD и наблюдатель, возможно, упустит что-то важное в деталях отображаемой сцены.

Указанное ограничение компенсируется важными преимуществами. Во-первых, сокращается время вычисления всей иерархии HLOD за счет удаления значительной части внутренних граней в полигональных представлениях кластеров. Во-вторых, при обеспечении того же внешнего визуального качества кластеров уменьшается их сложность и возрастает скорость рендеринга. Меньшее число полигонов в представлении кластеров снижает требования и к объемам основной памяти и видеопамати, требуемым для их размещения, что уменьшает число обменов и также способствует повышению производительности рендеринга.

Удаление внутренних граней может быть организовано несколькими способами:

- применением техники, аналогичной затенению фоновому освещению (англ. ambient occlusion) [42]. Из каждой грани кластера выпускаются лучи, и если хотя бы один луч достигает «неба» (не встретил препятствий на своем пути), то эта грань считается видимой снаружи. В оригинальной технике фоновому затенению для оценки степени освещённости определяется доля лучей, достигших неба. В нашем случае достаточно провести один непрерывный луч, чтобы сделать вывод о видимости грани;
- построением ограничивающей сферы вокруг кластера и бросанием лучей из точек сферы (распределенных равномерно или каким-либо другим способом) к ее центру с целью обнаружения внешних граней. Грани, не идентифицированные как внешние, удаляются из представления кластера;
- пакетным рендерингом представления кластера с разных ракурсов и применением идентификаторов граней в качестве цветовых индексов. Грани, идентификаторы которых попали хотя бы в один фрейм буфер, считаются внешними, а остальные – внутренними и подлежат удалению.

Отметим, что идентификация и удаление внутренних граней кластеров требуют дополнительных затрат и могут, в конечном счете, нивелировать выигрыш в скорости подготовки уровней детализации HLOD от упрощения кластеров. Поэтому алгоритмические и реализационные детали техник удаления невидимых граней исключительно важны для оценки суммарного эффекта. Прежде всего, достоверность определения внутренних граней, напрямую связанная с числом построенных и пересекаемых лучей или количеством ракурсов для пакетного рендеринга, может радикально влиять

на соотношение оптимизационных факторов. При этом сам выбор базовой техники не играет принципиальной роли. Как показали результаты профилирования программ, реализующих данные техники, основные затраты приходится не на массовые пересечения лучей или пакетный рендеринг параллельными средствами GPU, а на формирование единого массива идентификаторов внешних и внутренних объектов на CPU. В силу указанных факторов целесообразность применения данной оптимизации нуждается в экспериментальном подтверждении.

## 2. Распараллеливание

После того как кластеризация объектов сцены завершена и дерево HLOD построено, начинается процесс формирования геометрических представлений кластеров с использованием описанных выше методов полигонального упрощения и удаления невидимых граней. Из-за объема данных геометрические представления хранятся во внешней памяти и загружаются в оперативную лишь по мере необходимости. Дерево кластеров целиком размещается в оперативной памяти, поскольку содержит лишь информацию об идентификаторах кластеров, их ограничивающих объемах и связях родитель-ребёнок.

Напомним, что кластеры формируются снизу-вверх в соответствии с уровнями в дереве. Сначала обрабатываются кластеры первого нижнего уровня, в результате чего формируются точные и полные геометрические представления из оригинальных моделей объектов сцены. Геометрические представления кластеров второго уровня формируются из представлений кластеров первого уровня. Геометрические представления кластеров третьего уровня формируются из представлений кластеров второго и, возможно, первого уровня (рис. 1). Процесс продолжается до тех пор, пока не будет сформировано представление самого верхнего и единственного корневого кластера, соответствующего всей сцене.

Описанный порядок формирования кластеров позволяет применить схему крупноблочного распараллеливания вычислений. Обработка кластеров одного уровня может осуществляться в нескольких параллельных потоках, поскольку между ними нет зависимостей по данным, а они зависят только от ранее подготовленных представлений дочерних кластеров нижних уровней.

При обработке каждого кластера выполняются следующие действия:

- полигональные сетки дочерних кластеров загружаются в оперативную память и добавляются в кэш, если не находились там ранее;
- полигональные сетки объединяются в единое представление новой сетки;
- к новой сетке применяется процедура упрощения с соответствующими целевыми параметрами сложности и визуального качества;
- полученная полигональная сетка кластера добавляется в кэш и сохраняется во внешней памяти;
- сетки дочерних кластеров удаляются из кэша, проверяется свободный объем кэша и при необходимости применяется политика вытеснения для размещения новых сеток и обработки следующих кластеров.

Описанная схема распараллеливания допускает разные способы реализации. Во-первых, можно отказаться от реализации кэша и подгружать сетки дочерних кластеров из внешней памяти непосредственно перед обработкой родительского и высвобождать сразу после ее завершения. Это обеспечит неблокируемый доступ к сеткам, но может привести к необходимости их повторного чтения с внешнего диска. Во-вторых, планирование заданий на обработку кластеров и их распределение по вычислительным узлам (ядрам или процессорам) может быть осуществлено разными способами. Наиболее простой состоит в предварительном назначении обрабатываемых кластеров одного уровня отдельным вычислительным узлам. Поскольку число листовых кластеров обычно много больше, чем вычислительных узлов, степень параллелизма при обработке кластеров нижних уровней высока, однако падает по мере продвижения к верхним уровням и корню дерева. Другим фактором, препятствующим эффективному распараллеливанию, является неоднородность кластеров, выражающаяся в разной сложности полигональных сеток и приводящая к разному времени их обработки. В зависимости от метода кластеризации степень неоднородности может существенно различаться, и для обеспечения максимальной загрузки вычислительных узлов целесообразно переназначать их на новые задания по мере завершения предыдущих, используя пул тредов. Обычно

программные средства распараллеливания вычислений позволяют реализовать оба варианта относительно просто [43].

### 3. Инкрементальные обновления

Как отмечено выше, существует довольно широкий класс графических интерактивных приложений, оперирующих большими 3D-моделями и нуждающихся в быстрых средствах пересчета уровней детализации HLOD для их отображения в оперативном режиме. Обсудим возможности и алгоритмические детали применения техники инкрементального обновления для этих целей.

Эффективность применения инкрементальных обновлений во многом зависит от характера изменений в сцене. В случае локальных изменений, затрагивающих небольшое число объектов сцены и не меняющих их пространственное положение, габариты и ограничивающие объемы существенно, появляется возможность обновления небольшого числа существующих кластеров без пересчета всей HLOD-иерархии. Прежде всего, это листовые кластеры с изменившимися объектами, а также их родители вплоть до корня дерева при условии, что их представления действительно чувствительны к исходным изменениям. Если это не так, то процесс распространения изменений вверх по дереву и пересчет родительских кластеров не имеет смысла.

В случае глобальных изменений, затрагивающих относительно большое число объектов сцены и существенно меняющих ограничивающие объемы кластеров и плотность заполнения, применение прежней HLOD-иерархии становится неэффективным. При нарушенной пространственной кластеризации отсечение конусом видимости не приведет к желаемому результату и заведомо лишние объекты будут включены в процесс рендеринга. При деградации иерархии ее следует пересчитать заново, проведя повторную кластеризацию объектов и формируя геометрические представления кластеров. К сожалению, трудно выработать критерии для принятия решения о пересчете иерархии, который, с одной стороны, необходим для последующего производительного рендеринга сцены и обеспечения интерактивности графического приложения, а с другой стороны, потребует значительных вычислительных затрат и, скорее всего, прервет работу пользователей.

В нашей работе [5] предлагается подход к пересчету HLOD-иерархии, основанный на оценке ее качества непосредственно в процессе инкрементального обновления. В качестве критериев предлагается использовать показатели плотности кластеров (количество объектов в единице ограничивающего объема кластера) и скорости упрощения кластеров (количество удаленных полигонов или треугольников при единичном увеличении погрешности). Сама иерархия строится при ограничениях на объем кластера, число дочерних кластеров и сложность геометрического представления. Решение о пересчете принимается, если не удастся обновить кластеры иерархии с требуемым качеством при соблюдении указанных ограничений.

Приведем краткое описание ранее предложенной техники инкрементального обновления HLOD-иерархии, используемой в проведенных вычислительных экспериментах. Инкрементальное обновление реализуется в виде процесса обработки и распространения изменений следующих типов: добавление новых объектов, удаление объектов и изменение геометрии или положения существующих объектов.

При добавлении объектов в сцену анализируется возможность их включения в существующие листовые кластеры и распространения изменений вверх по иерархии с соблюдением ограничений построения иерархии. Если это оказывается невозможным, изменения считаются глобальными и выносятся вердикт о пересчете иерархии.

При удалении объектов оценивается качество листовых кластеров с учетом возможного уменьшения ограничивающих объемов и сокращения полигонов в них. Вырожденные кластеры с нарушенными критериями качества удаляются, а приписанные им объекты перераспределяются по соседним кластерам. Если подобные преобразования не удастся реализовать с перечисленными выше требованиями, процесс прерывается и принимается решение о полном пересчете иерархии.

Изменения геометрии или положения удобно интерпретировать как удаление соответствующего объекта и его последующее добавление с новой геометрией и новым расположением в сцене. Поэтому обработка подобных изменений включает все вышеописанные действия. В случаях, когда изменения затрагивают только геометрическое представление объекта, а его положение и габариты не меняются, обработка сводится к анализу сложности обновленного кластера и переоценке скорости упрощения.

Процесс распространения изменений реализуется с помощью меток обновления, которые позволяют контролировать согласованность изменений для всех кластеров дерева.

### Результаты экспериментов

Для каждого из описанных выше трех способов оптимизации была проведена серия вычислительных экспериментов, которая позволила количественно оценить эффект как результат сокращения процессорного времени генерации иерархических уровней детализации HLOD.

Все эксперименты проводились на компьютере с типовой конфигурацией Intel(R) Core(TM) i9-9880 CPU @ 3.60GHz (8 ядер, 16 потоков), 32 ГБ ОЗУ, 2 ТБ NVME SSD. В экспериментах использовался единый набор тестовых сцен, реконструированных по реальным цифровым моделям строительных проектов с соответствующими архитектурными, конструктивными и инженерными элементами (рис. 2). Сцена многоэтажного здания содержала 146 461 объект с тесселированной геометрией в виде полигональных сеток и общим числом треугольников 3 426 267. Сцены стадиона и отеля содержали 41 451 и 47 160 объектов, 7 960 512 и 29 131 139 треугольников соответственно.

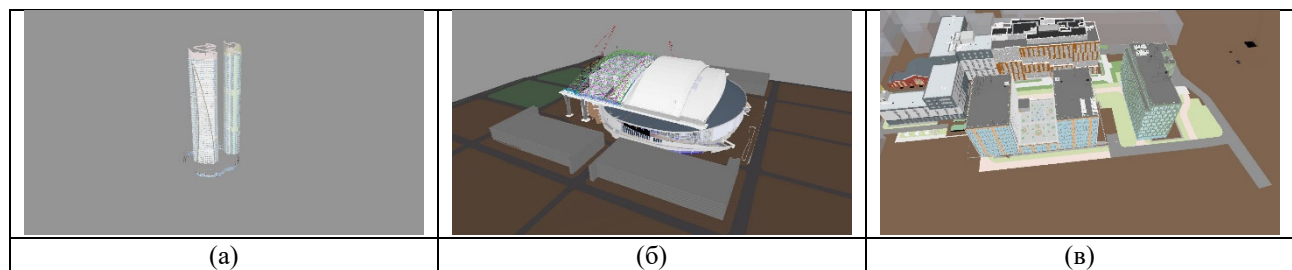


Рисунок 2. Цифровые модели строительных объектов: а) многоэтажное здание; б) стадион; в) отель

Эксперименты были организованы следующим образом: проводилась генерация иерархических уровней детализации HLOD в соответствии с описанной выше стратегией, замерялось время генерации упрощённых представлений кластеров. На первом уровне дерева кластеры формируются из оригинальных представлений объектов сцены, поэтому интерес представляют уровни дерева, начиная со второго, на которых применяются вычислительно затратные упрощения полигональных сеток.

#### 1. Распараллеливание

Первая серия вычислительных экспериментов была связана с применением описанной выше техники распараллеливания. Напомним, что обработка уровней выполняется последовательно. При этом кластеры одного уровня обрабатываются параллельно в несколько потоков, поскольку их представления зависят только от кластеров нижних уровней. После завершения всех потоков на очередном уровне начинается параллельная обработка кластеров следующего уровня. В таблице 1 приведено сравнение времени генерации уровней детализации без распараллеливания и с распараллеливанием для трёх тестовых сцен. В ней же указано количество кластеров на каждом уровне дерева, начиная со второго.

Таблица 1. Эксперименты с распараллеливанием

Модель	Кол-во кластеров на уровнях иерархии	Последовательное выполнение	Параллельное выполнение	Ускорение
Здание	17, 9, 4, 2, 1, 1	7м 57с	3м 2с	2.6х
Стадион	27, 15, 8, 5, 2, 1, 1	36м 28с	19м 56с	1.8х
Отель	127, 64, 32, 16, 8, 4, 2, 1	12ч 51м 38с	3ч 37м 18с	3.6х

#### 2. Удаление внутренних граней

Вторая серия экспериментов предназначалась для оценки техники полигонального упрощения с удалением внутренних граней. Обнаружение внутренних граней было реализовано с использованием описанной выше процедуры бросания лучей. Лучи выпускались из точек, равномерно распределённых по ограничивающей сфере кластера, в направлении ее центра. Для поиска пересечений лучей с гранями полигональной сетки применялась библиотека Embree [44], в которой данные операции эффективно

реализованы с использованием пространственного индексирования на основе иерархий ограничивающих объёмов. Грани, в которые попал хотя бы один луч, помечались как внешние, остальные – как внутренние. После завершения процесса бросания лучей и поиска пересечений все грани, помеченные как внутренние, удалялись. Для каждого кластера поиск и удаление внутренних граней осуществлялись сразу после формирования его полигональной сетки (путём объединения сеток дочерних кластеров) и перед запуском процедуры упрощения полигональной сетки.

Как отмечалось выше, достоверность идентификации внутренних граней описанным способом существенно зависит от числа выпущенных лучей. С увеличением числа лучей достоверность будет монотонно повышаться, однако процесс потребует большее время обработки. При этом будет обнаружено больше внешних граней и меньше внутренних. Поэтому число лучей в проводимых экспериментах варьировалось, начиная от десяти тысяч и заканчивая миллиардом лучей (на каждый кластер). Для каждой тестовой сцены замерялось время генерации всех уровней (начиная со второго) с использованием техники распараллеливания. Также определялись общее количество удалённых граней во всех кластерах, процент от общего числа граней в кластерах без удаления внутренних граней и размер полученных иерархических уровней детализации, сохранённых на диске в формате 3D Tiles [45]. Полученные результаты приведены в таблице 2. В качестве референсной информации также укажем общее число граней в кластерах HLOD-иерархии без применения процедуры удаления внутренних граней: 6 826 213 (здание), 13 853 526 (стадион) и 57 705 028 (отель).

Таблица 2. Эксперименты с удалением внутренних граней

		Без удаления внутренних граней	10 000 лучей	100 000 лучей	1 000 000 лучей	10 000 000 лучей	100 000 000 лучей	1 000 000 000 лучей
З д а н и е	Время генерации	3м 2с	0.58с	2.41с	16.68с	1м 11с	4м 5с	24м 57с
	Кол-во удалённых граней	0	3416063 (50%)	3378397 (49%)	3267249 (48%)	3072979 (45%)	2815698 (41%)	2589898 (38%)
	Размер файлов	969 МБ	482 МБ	508 МБ	554 МБ	613 МБ	654 МБ	714 МБ
С т а д и о н	Время генерации	19м 56с	0.86с	1.56с	13.77с	57с	5м 59с	1ч 3м 22с
	Кол-во удалённых граней	0	7942012 (57%)	7905560 (57%)	7778853 (56%)	7567357 (55%)	7267930 (52%)	6889677 (50%)
	Размер файлов	2.24 ГБ	1.26 ГБ	1.27 ГБ	1.3 ГБ	1.35 ГБ	1.41 ГБ	1.47 ГБ
О т е л ь	Время генерации	3ч 37м 18с	2.72с	3.41с	10.81с	1м 4с	7м 44с	1ч 7м 59с
	Кол-во удалённых граней	0	29103008 (50%)	29044720 (50%)	28876971 (50%)	28480811 (49%)	27715526 (48%)	26732145 (46%)
	Размер файлов	11.4 ГБ	5.78 ГБ	5.81 ГБ	5.86 ГБ	5.98 ГБ	6.18 ГБ	6.42 ГБ

### 3. Инкрементальные обновления

Финальная серия экспериментов предназначалась для оценки техники инкрементального обновления иерархических уровней детализации HLOD при существенно локальных изменениях в сцене. Эксперименты были организованы следующим образом: некоторые объекты сцены помечались как изменившиеся, после чего запускалось обновление уровней детализации. По набору изменившихся объектов определялись листовые кластеры первого уровня, требующие обновления. Затем итеративно определялись кластеры, требующие обновления на следующих уровнях, согласно следующему

принципу: кластер требует обновления, если хотя бы один из его детей изменился. После чего запускался процесс генерации представлений для помеченных кластеров. Заметим, что положение объектов, их габариты и ограничивающие объемы не менялись, поэтому все вычисления выполнялись с прежней иерархией кластеров. Было реализовано два способа моделирования локальных изменений в сцене: путем случайного отбора объектов и путем выбора объектов, компактно расположенных в центре сцены при минимизации их общего ограничивающего объёма. Эксперименты проводились для следующих случаев: когда изменился один случайный объект, когда изменилось 5 % объектов сцены (смежных и случайных) и когда поменялось 10 % объектов сцены (смежных и случайных). Замерялось время генерации кластеров всех уровней, начиная со второго, а также число кластеров второго уровня, которые потребовали обновления. В таблице 3 приведены полученные результаты. В экспериментах полигональные упрощения кластеров проводились параллельно. Техника удаления внутренних граней не применялась, чтобы не исказить эффект от самих инкрементальных обновлений.

Таблица 3. Эксперименты с инкрементальными обновлениями

		Полная генерация	Изменения 1 объекта	5 % смежных изменений	5 % случайных изменений	10 % смежных изменений	10 % случайных изменений
З д а н и е	Время	3м 2с	1м 6с	1м 20с	2м 16с	1м 18с	2м 26с
	Кол-во объектов	146464	1	7323	7323	14646	14646
	Кол-во кластеров уровня 2	17	1	3	12	3	12
С т а д и о н	Время	19м 56с	3м 1с	12м 53с	20м 1с	12м 59с	20м 32с
	Кол-во объектов	41451	1	2072	2072	4145	4145
	Кол-во кластеров уровня 2	27	1	7	24	8	26
О т е л ь	Время	3ч 37м 18с	10м 27с	44м 17с	1ч 36м 1с	1ч 48с	2ч 9м 20с
	Кол-во объектов	47160	1	2358	2358	4716	4716
	Кол-во кластеров уровня 2	127	1	9	39	13	53

### Обсуждение результатов

Обсудим результаты в порядке описания проведенных вычислительных экспериментов. Главное внимание при этом уделим применяемым техникам оптимизации и факторам, влияющим на производительность генерации иерархических уровней детализации HLOD.

#### 1. Распараллеливание

Прежде всего, отметим не столь значительное ускорение вычислительных процедур обработки кластеров в результате их распараллеливания. На компьютерной архитектуре с 8-ядерным процессором ускорение на тестовых сценах варьировалось в диапазоне 1.8–3.6. С одной стороны, даже двух- и трёхкратное ускорение может быть значимым, особенно в тех случаях, когда подготовка HLOD занимает часы процессорного времени (например, тест с отелем в таблице 1) и визуализация является частью непрерывного рабочего процесса. С другой стороны, применяемая схема крупноблочного распараллеливания допускает лучшее масштабирование. Полученные результаты объясняются тем, что для экспериментов намеренно были выбраны сцены среднего размера (50000 – 150000 объектов). Построенные иерархии ограничивались 7-9 уровнями; верхние четыре уровня, обрабатываемые существенно последовательно, имели низкое среднее число кластеров (2, 2.25 и 3.75 для сцен здания, стадиона и отеля соответственно). Это приводило к деградации применяемой схемы



распараллеливания с предварительным планированием заданий из-за низкой загрузки вычислительных узлов на верхних уровнях иерархии. Для больших сцен, соответствующих, например, проектам промышленного строительства и содержащих миллионы объектов, следует ожидать лучшие результаты из-за более глубокой иерархии и эффективной обработки существенно большей части уровней. Более сложный тест со сценой отеля хорошо демонстрирует эту закономерность.

Вместе с тем существуют и другие факторы, препятствующие эффективному распараллеливанию процедур генерации HLOD. В теоретическом разделе среди них обсуждались неоднородность кластеров, обусловленная особенностями оригинальной сцены и ограничениями метода кластеризации, а также повторное чтение полигональных сеток из внешней памяти при обработке каждого следующего уровня иерархии. Организация пула тредов с динамическим назначением заданий на обработку кластеров и кэширование полигональных сеток в оперативной памяти могут обеспечить более полную загрузку вычислительных узлов и достичь большей производительности. Данные возможности нуждаются в дополнительном экспериментальном подтверждении, основанном на конкретных программных реализациях.

## *2. Полигональные упрощения с удалением внутренних граней*

Полигональные упрощения с удалением невидимых граней приводят к более значительному и предсказуемому результату, который выражается как в ускорении времени генерации HLOD, так и в меньшем объеме оперативной памяти и видеопамяти, необходимой для хранения и отображения уровней детализации.

Ускорение достигается за счёт сокращения числа граней в полигональной сетке и более быстрой последующей работы алгоритма упрощения (с меньшим числом итераций, необходимым для достижения целевого числа треугольников, и меньшим размером вспомогательных индексных структур, поиск в которых, как правило, осуществляется за логарифмическое время от числа элементов). Удаление внутренних граней имеет накопительный эффект в рамках применяемой стратегии кластеризации. Поскольку при объединении представлений кластеров часть треугольников перекрывается и удаляется, на каждом следующем уровне иерархии всё меньше треугольников участвует в полигональных упрощениях. Упрощение уровней детализации HLOD при их прежнем визуальном качестве улучшает также и производительность рендеринга.

Конкретные значения ускорения существенно зависят от достоверности идентификации внутренних граней и варьируются в диапазоне от нескольких единиц до трех порядков (см. таблицу 2). Максимальное ускорение в 4 793 раза достигается на тестовой сцене отеля при числе лучей 10 000, однако достоверность определения внутренних граней невелика и, по крайней мере, 4 % граней идентифицируется неверно (оценка получается путем сравнения доли внутренних или внешних граней, обнаруженных при числе лучей 10 000 и 1 000 000 000). Это может привести к нежелательным артефактам, например к дырам во внешних поверхностях или удаленным видимым элементам внутри сцены. При числе лучей 10 000 000 вычислительные затраты на определение пересечений возрастают и частично нивелируют эффект от сокращения числа полигонов. Ускорение на тестовой сцене отеля составляет уже 203. Дальнейшее увеличение числа лучей до 1 000 000 000 приводит к снижению ускорения, а для сцен со зданием и стадионом эффект от применяемой оптимизации и вовсе становится отрицательным.

Результаты экспериментов показывают, что для разных тестовых сцен и разных кластеров одни и те же оценки достоверности достигаются при существенно разном числе лучей. Поэтому на практике их довольно сложно контролировать. Возможная итеративная схема для получения оценки заключается в кратном увеличении числа лучей для каждого обрабатываемого кластера до тех пор, пока изменения в доле внутренних или внешних граней не зафиксируются с заданной погрешностью. В любом случае, компромисс между производительностью генерации уровней детализации HLOD и достоверностью их отображения должен искажаться с учетом требований, предъявляемых к графическому приложению.

## *3. Инкрементальные обновления*

Наконец, проведенные эксперименты с инкрементальными обновлениями HLOD подтверждают перспективность данной техники для приложений, нуждающихся в оперативном пересчете уровней

детализации. Ускорение на тестовых сценах здания, стадиона и отеля с одним изменившимся объектом составило 2.75, 6.6 и 20.8 соответственно (см. таблицу 3).

Ключевым фактором, влияющим на эффективность инкрементальных обновлений, оказывается пространственная локализация изменений, а не просто их число. В самом деле, если изменения затронули объекты одного листового кластера и не привели к его вырождению, то обновление будет осуществлено относительно быстро в результате распространения изменений по родительской ветке вверх по иерархии. И наоборот, если изменившиеся объекты находятся в разных листовых кластерах, то потребуются инкрементальная обработка большей части кластеров, что может оказаться более затратным, чем полный пересчет всей иерархии HLOD заново.

Так, в экспериментах со сценой здания время обновления при 5 % изменившихся объектов и при 10 % не отличается, поскольку изменения затрагивают одинаковое число кластеров (3 в случае смежных изменений и 12 в случае изменения случайных объектов). Более того, на всех сценах 10 % изменившихся смежных объектов затрагивают меньше кластеров, чем 5 % изменившихся случайных объектов, и, как следствие, обновление для 10 % смежных происходит быстрее, чем для 5 % случайных.

Необходимо также учитывать фактор параллельности в проведенных экспериментах, который несколько завуалирует эффект от инкрементальной обработки кластеров. Если изменился, к примеру, один листовый кластер, это потребует обработать его и всех его родителей (по пути от листа до корня). При этом на каждом уровне будет обрабатываться только один кластер, поэтому параллельность не будет задействована. Если изменилось несколько кластеров на одном уровне дерева, включится фактор параллельности.

В целом, инкрементальные обновления необходимы в приложениях, в которых большие сцены подвергаются перманентным изменениям и изменения локализованы в пространстве.

## Выводы

В данной работе анализируются проблемы эффективной генерации иерархических уровней детализации HLOD и их оперативного обновления в условиях перманентно изменяемых проектных данных. Данные требования характерны, в частности, для цифровых двойников сложных промышленных объектов и масштабных инфраструктурных программ.

Для рассмотренной стратегии вычисления уровней детализации HLOD на основе агломеративной кластеризации предложены и исследованы техники крупноблочного распараллеливания вычислений, инкрементального обновления уровней детализации и полигональные упрощения с удалением невидимых элементов. Проведены серии вычислительных экспериментов с моделями гражданского строительства, которые подтверждают перспективность данных оптимизаций, обеспечивающих увеличение производительности генерации HLOD в среднем на порядок. Показана зависимость полученных результатов, прежде всего производительности подготовки уровней детализации, от множества факторов, таких как сложность сцены, характер изменений, степень достоверности отображения.

Выявлены возможности для дальнейших оптимизаций, касающиеся, в частности, динамического планирования заданий при параллельной обработке кластеров и итеративного оценивания достоверности отображения внешних и внутренних граней. Данные возможности планируется исследовать в ходе дальнейшей работы.

## Список литературы

1. Расширение метода иерархических уровней детализации для динамических сцен с детерминированным характером событий / Семёнов В.А., Шуткин В.Н., Золотов В.А., Морозов С.В. // Труды конференции ГрафиКон-2019. С. 37-41. DOI: 10.30987/graphicon-2019-1-37-41.
2. Visualization of Large Scenes with Deterministic Dynamics / Semenov V.A., Shutkin V.N., Zolotov V.A., Morozov S.V., Gonakhchyan V.I. // Programming and Computer Software. 2020. 46(3). P. 223-232.
3. Semenov V., Shutkin V., Zolotov V. Conservative Out-of-Core Rendering of Large Dynamic Scenes Using HDLODs // Proceedings of the 31st International Conference on Computer Graphics and Vision (GraphiCon 2021), 2021, Nizhny Novgorod, Russia, 2021. P. 105-115. DOI:10.20948/graphicon-2021-3027-105-115.

4. Alternative HDLOD method for large scenes with multiple dynamic behaviors / Shutkin V., Semenov V., Zolotov V., Morozkin N. // *Proceedings of 17th International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing*, Porto, Portugal 16 – 18 July 2023. P. 141-148.
5. Оптимизации генерации иерархических уровней детализации для масштабных полигональных сцен / Шуткин В.Н., Морозкин Н.К., Семенов В.А., Тарлапан О.А. // *Труды ИСП РАН*. 2025. Т. 37, вып. 3. С. 311-324. DOI: 10.15514/ISPRAS-2025-37(3)-22.
6. Clark J. H. Hierarchical geometric models for visible surface algorithms // *Communications of the ACM*. 1976. Vol. 19, no. 10. P. 547-554.
7. Rossignac J., Borrel P. Multi-resolution 3D approximations for rendering complex scenes // *Modeling in computer graphics: methods and applications*. Berlin, Heidelberg: Springer, 1993. P. 455-465.
8. Heckbert P., Garland M. Multiresolution modeling for fast rendering // *Graphics Interface*. Canadian Information Processing Society, 1994. P. 43-43.
9. Erikson C., Manocha D., Baxter III W. V. HLODs for faster display of large static and dynamic environments // *Proceedings of the 2001 symposium on Interactive 3D graphics*. 2001. P. 111-120.
10. Varadhan G., Manocha D. Out-of-core rendering of massive geometric environments // *IEEE Visualization*, 2002. VIS 2002. IEEE, 2002. P. 69-76.
11. Lakhia A. Efficient interactive rendering of detailed models with hierarchical levels of detail // *Proceedings. 2nd International Symposium on 3D Data Processing, Visualization and Transmission*, 2004. 3DPVT 2004. IEEE, 2004. P. 275-282.
12. Zhou Z., Chen K., Zhang J. Efficient 3-D scene prefetching from learning user access patterns // *IEEE Transactions on Multimedia*. 2015. Vol. 17, no. 7. P. 1081-1095.
13. Optimally redundant, seek-time minimizing data layout for interactive rendering / Chen J. et al. // *The Visual Computer*. 2017. Vol. 33. P. 139-149.
14. Peng C., Cao Y. A GPU-based approach for massive model rendering with frame-to-frame coherence // *Computer Graphics Forum*. Oxford, UK : Blackwell Publishing Ltd, 2012. Vol. 31, no. 2pt2. P. 393-402.
15. Peng C., Cao Y. Parallel LOD for CAD model rendering with effective GPU memory usage // *Computer-Aided Design and Applications*. 2016. Vol. 13, no. 2. P. 173-183.
16. Shamir A., Pascucci V. Temporal and spatial level of details for dynamic meshes // *Proceedings of the ACM symposium on Virtual reality software and technology*. 2001. P. 77-84
17. Kircher S., Garland M. Progressive multiresolution meshes for deforming surfaces // *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*. 2005. P. 191-200.
18. Ponchio F., Hormann K. Interactive rendering of dynamic geometry // *IEEE Transactions on Visualization and Computer Graphics*. 2008. Vol. 14, no. 4. P. 914-925.
19. Ramos F., Ripolles O., Chover M. Continuous level of detail for large scale rendering of 3d animated polygonal models // *International Conference on Articulated Motion and Deformable Objects*. Berlin, Heidelberg : Springer, 2012. P. 194-203.
20. Toledo L., De Gyves O., Rudomín I. Hierarchical level of detail for varied animated crowds // *The Visual Computer*. 2014. Vol. 30. P. 949-961.
21. 4D modeling of large industrial projects using spatio-temporal decomposition / V.A. Semenov, K.A. Kazakov, S.V. Morozov, et al. // *eWork and eBusiness in Architecture, Engineering and Construction* / eds. K. Menzel and R. Scherer. CRC Press, Taylor & Francis Group, London, UK, 2010. P. 89-95.
22. Visual Planning and Scheduling of Industrial Projects with Spatial Factors / V.A. Semenov, A.S. Anichkin, S.V. Morozov, et al. // *Proceedings of 20th ISPE International Conference on Concurrent Engineering* / eds. C. Bil, J. Mo, J. Stjepandic. IOS Press, Melbourne, Australia, 2013. P. 343-352, ISBN 978-1-61499-301-8.
23. Level of detail for 3D graphics / Luebke D., et al. Elsevier, 2002. ISBN 9780123991812.
24. Гонахчян В.И. Обзор методов упрощения полигональных моделей на графическом процессоре // *Труды Института системного программирования РАН*. 2014. Т. 26, вып. 2. С. 159-174.
25. Garland M., Heckbert P. S. Surface simplification using quadric error metrics // *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. 1997. P. 209-216.
26. Schroeder W. J., Zarge J. A., Lorensen W. E. Decimation of triangle meshes // *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*. 1992. P. 65-70.
27. Low K. L., Tan T. S. Model simplification using vertex-clustering // *Proceedings of the 1997 symposium on Interactive 3D graphics*. 1997. P. 75-ff.
28. Zhao T., Jiang J., Guo X. A Novel Quadratic Error Metric Mesh Simplification Algorithm For 3D Building Models Based On 'Local-Vertex' Texture Features // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2022. Vol. 48. P. 109-115.

29. Salinas D., Lafarge F., Alliez P. Structure-aware mesh decimation // Computer Graphics Forum, 2015. Vol. 34, no. 6. P. 211-227.
30. Li M., Nan L. Feature-preserving 3D mesh simplification for urban buildings // ISPRS Journal of Photogrammetry and Remote Sensing. 2021. Vol. 173. P. 135-150.
31. A Novel LOD Rendering Method with Multi-level Structure Keeping Mesh Simplification and Fast Texture Alignment for Realistic 3D Models / Ge Y., et al. // IEEE Transactions on Geoscience and Remote Sensing. 2024.
32. InfNeRF: Towards Infinite Scale NeRF Rendering with  $O(\log n)$  Space Complexity / Liang J., et al. // SIGGRAPH Asia 2024 Conference Papers. 2024. P. 1-11.
33. A hierarchical 3d gaussian representation for real-time rendering of very large datasets / Kerbl B., et al. // ACM Transactions on Graphics (TOG). 2024. Vol. 43, no. 4. P. 1-15.
34. Pečnik S., Žalik B. Real-time visualization using GPU-accelerated filtering of lidar data // World Acad. Sci., Eng. Technol., Int. J. Comput., Elect., Automat., Control Inf. Eng. 2014. Vol. 8, no. 12. P. 2108-2112.
35. Efficient Scene Appearance Aggregation for Level-of-Detail Rendering / Zhou Y., et al. // arXiv preprint arXiv:2409.03761. 2024.
36. Loubet G., Neyret F. Hybrid mesh-volume LoDs for all-scale pre-filtering of complex 3D assets // Computer Graphics Forum. 2017. Vol. 36, no. 2. P. 431-442.
37. Edavamadathil Sivaram V., Li T. M., Ramamoorthi R. Neural geometry fields for meshes // ACM SIGGRAPH 2024 Conference Papers. 2024. P. 1-11.
38. Презентация «Nanite: A Deep Dive» с конференции Siggraph 2021. URL: [https://advances.realtimerendering.com/s2021/Karis\\_Nanite\\_SIGGRAPH\\_Advances\\_2021\\_final.pdf](https://advances.realtimerendering.com/s2021/Karis_Nanite_SIGGRAPH_Advances_2021_final.pdf) (дата обращения: 27.08.2025).
39. Continuous level of detail mesh library. URL: [https://github.com/nvpro-samples/nv\\_cluster\\_lod\\_builder](https://github.com/nvpro-samples/nv_cluster_lod_builder) (дата обращения: 27.08.2025).
40. Zhang H., Cao L., Peng C. UltraMeshRenderer: Efficient Structure and Management of GPU Out-of-core Memory for Real-time Rendering of Gigantic 3D Meshes // ACM Transactions on Graphics (TOG). 2025. Vol. 44, no. 4. P. 1-19.
41. Haydel J., Yuksel C., Seiler L. Locally-adaptive level-of-detail for hardware-accelerated ray tracing // ACM Transactions on Graphics (TOG). 2023. Vol. 42, no. 6. P. 1-15.
42. Laine S., Karras T. Two methods for fast ray-cast ambient occlusion // EGSR'10: Proceedings of the 21st Eurographics conference on rendering. P.1325-1333. DOI: 10.1111/j.1467-8659.2010.01728.
43. Michael Klemm et al. OpenMP in a Modern World: From Multi-device Support to Meta Programming // Proceedings of 18th International Workshop on OpenMP, IWOMP 2022. Chattanooga, TN, USA, September 27–30, 2022.
44. Intel® Embree open-source high-performance ray tracing library. URL: <https://www.embree.org/> (дата обращения: 14.07.2025).
45. 3D Tiles Standards Specification. URL: <https://www.ogc.org/standards/3dtiles/> (дата обращения: 14.07.2025).

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-276-285

## Расчетное ядро программы «Свет в ночи»: методы и сравнение с коммерческими аналогами

В. П. Будаков<sup>1,2</sup>, А. В. Гримайло<sup>2</sup>, В. С. Желтов<sup>2</sup><sup>1</sup>НИУ «МЭИ», Москва, Россия<sup>2</sup>Центр инновационных разработок МСК БЛ ГРУПП, Москва, Россия

**Аннотация.** В работе рассмотрена точность расчетного ядра САПР «Свет в ночи», предназначенного для моделирования освещения как в помещениях, так и в наружных сценах. В ядре реализованы два метода расчета многократных отражений: фотонные карты и перспективный метод локальных оценок. Прямое освещение рассчитывается по закону обратных квадратов. Точность моделирования проверялась на двух задачах с известным аналитическим решением – задаче Соболева и расчете средней освещенности методом коэффициента использования. Результаты сопоставлялись с расчетами в DIALux и Lightscape. Показано, что при высоких коэффициентах отражения «Свет в ночи» демонстрирует значительно меньшие отклонения по сравнению с другими программами. В качестве направления развития авторы выделяют внедрение метода двойных локальных оценок, позволяющего напрямую вычислять яркость в произвольной точке сцены и направлении, а также разработку более точной модели отражения и применение методов машинного обучения.

**Ключевые слова:** глобальное освещение, многократные отражения, фотонные карты, локальные оценки.

## Computational Core of «Light in Night» Program: Methods and Comparison with Commercial Solutions

V. P. Budakov<sup>1,2</sup>, A. V. Grimailo<sup>2</sup>, V. S. Zheltov<sup>2</sup><sup>1</sup>NRU «MPEI», Moscow, Russia<sup>2</sup>Innovation Development Centre ILIC BL GROUP, Moscow, Russia

**Abstract.** This work examines the accuracy of the computational core of the CAD system Light-in-Night, designed for modelling lighting in both indoor and outdoor scenes. The core implements two methods for calculating multiple reflections: photon mapping and the progressive local illumination approach. Direct illumination is computed according to the inverse-square law. The accuracy of the modelling was verified using two test cases with known analytical solutions – Sobolev’s problem and the calculation of average illuminance via the utilisation factor method. The results were compared with calculations performed in DIALux and Lightscape. It is demonstrated that for high reflectance coefficients, Light-in-Night exhibits significantly lower deviations compared to other software. As a direction for future development, the authors highlight the implementation of the double local estimation method, enabling direct computation of luminance at an arbitrary point and direction within the scene, alongside the development of a more accurate reflection model and the application of machine learning techniques.

**Keywords:** global illumination, multiple reflections, photon maps, local estimations.

### Введение

Программа автоматизированного проектирования «Свет в ночи» [1] сочетает развитый пользовательский интерфейс с гибкими средствами подготовки сцены: предусмотрена как интеграция с внешними редакторами (3ds max, Blender и др.), так и возможность создания сложной геометрии непосредственно во встроенном редакторе. Однако интерфейсные возможности служат лишь подготовительным этапом. Ключевое назначение системы – физически корректное моделирование глобального освещения, включая многократные отражения в сложных архитектурных и интерьерных сценах. В отличие от распространенных графических движков и визуализаторов, ориентированных преимущественно на реалистичную картинку, «Свет в ночи» разрабатывается как инженерный инструмент, обеспечивающий получение количественно точных результатов, пригодных для практического проектирования систем освещения.

Система автоматизированного проектирования «Свет в ночи» реализует физически корректное моделирование уравнения глобального освещения (УГО) [2]:

$$L(\mathbf{r}, \hat{\mathbf{i}}) = L_0(\mathbf{r}, \hat{\mathbf{i}}) + \frac{1}{\pi} \int L(\mathbf{r}, \hat{\mathbf{i}}') \sigma(\mathbf{r}, \hat{\mathbf{i}}, \hat{\mathbf{i}}') |(\hat{\mathbf{N}}, \hat{\mathbf{i}}')| d\hat{\mathbf{i}}', \quad (1)$$

где  $L(\mathbf{r}, \hat{\mathbf{i}})$  – яркость светового поля в точке  $\mathbf{r}$  по направлению  $\hat{\mathbf{i}}$ ;

$\sigma(\mathbf{r}; \hat{\mathbf{l}}, \hat{\mathbf{l}}')$  – двунаправленная функция отражения или пропускания (BRDF или BTDF);

$L_0(\mathbf{r}, \hat{\mathbf{l}})$  – яркость самосвечения;

$\hat{\mathbf{N}}$  – нормаль к элементу поверхности сцены в точке  $\mathbf{r}$ .

Как известно, данное уравнение не имеет аналитического решения, за исключением отдельных частных случаев в диффузном приближении, таких как фотометрическая сфера и задача Соболев [3]. Поэтому на практике применяются приближенные численные методы. Наиболее известные из них включают:

- метод конечных элементов (radiosity), реализованный в *Lightscape*, *DIALux* (версии 4.13), *Relux* [4];
  - метод фотонных карт (photon maps) [5], лежащий в основе *DIALux evo*;
- а также устаревшие инженерные подходы, такие как метод коэффициента использования.

Фотонные карты (ФК) представляют собой, по существу, один из вариантов прямого метода Монте-Карло, основанного преимущественно на умозрительных заключениях. Такой подход страдает от двух фундаментальных недостатков:

- 1) медленной сходимости;
- 2) усреднения результата по ячейке сетки, что ограничивает точность оценки в конкретной точке.

Эти особенности обуславливают постоянный интерес к совершенствованию методов Монте-Карло, и за последние десятилетия предложено множество модификаций, направленных на повышение их эффективности [6].

Одной из наиболее перспективных и строго обоснованных альтернатив, по мнению авторов, являются локальные оценки (ЛО) Монте-Карло, введенные в оригинальной работе [7]. В отличие от классических схем, они позволяют строить алгоритм, основанный на математически строгом решении УГО и обеспечивают:

- получение оценки яркости (а также любой интегральной характеристики светового поля) в конкретной точке сцены;
- учет вклада каждого первичного луча в итоговую оценку.

Это особенно важно для решения задач, требующих высокой точности и физической интерпретируемости результата.

Следует отметить ряд особенностей, характерных для большинства современных систем светотехнического проектирования.

Во-первых, хотя уравнение глобального освещения формулируется для яркости, нормировка этой величины практически применяется лишь в ограниченном числе прикладных сценариев, например в задачах дорожного или архитектурного освещения, где вклад многократных отражений невелик и может быть проигнорирован. Напротив, в задачах, где такие отражения критичны, нормируются уже интегральные величины, такие как освещенность.

Во-вторых, в большинстве существующих решений используется диффузная модель отражения. В случае radiosity это оправдано самой природой метода, поскольку он базируется на уравнении излучательности, полученном в диффузном приближении. Однако в системах вроде *DIALux evo* использование упрощенной модели, вероятно, связано с общепринятыми инженерными традициями, а не с ограничениями метода.

Заметим также, что уравнение (1) допускает выделение прямой компоненты освещения, которая может быть рассчитана независимо.

В рамках расчетного ядра «Свет в ночи» реализована гибридная схема:

- прямое освещение рассчитывается на основе закона обратных квадратов;
- учет многократных отражений может выполняться одним из двух методов:
  - фотонные карты с финальным сбором;
  - локальные оценки.

При этом у пользователя имеется возможность выбирать метод в зависимости от требований к точности и ресурсоемкости задачи.

Дальнейшие разделы настоящей работы посвящены описанию математических и алгоритмических решений, реализованных в расчетном ядре программы «Свет в ночи», а также анализу результатов,



полученных с его помощью, в сравнении с существующими коммерческими средствами моделирования освещения.

## Расчетное ядро

### 1. Расчет прямого освещения

Для расчета прямого освещения от источников в ядре применяется классическая модель, основанная на законе обратных квадратов:

$$E(\mathbf{r}) = \frac{I(\hat{\mathbf{l}})}{l^2} |(\hat{\mathbf{N}}, \hat{\mathbf{l}})| \quad (2)$$

где  $E(\mathbf{r})$  – освещенность в точке  $\mathbf{r}$ ;

$I(\hat{\mathbf{l}})$  – сила света источника в направлении  $\hat{\mathbf{l}}$  расчетной точки  $\mathbf{r}$ ;

$|(\hat{\mathbf{N}}, \hat{\mathbf{l}})|$  – косинус угла между нормалью к поверхности в точке  $\mathbf{r}$  и направлением  $\hat{\mathbf{l}}$ .

Поскольку формула (2) предполагает точечную природу источника, в расчетном ядре реализована проверка ее применимости: если расстояние до источника оказывается меньше пятикратного размера самого источника, он автоматически разбивается на совокупность элементарных излучателей. Такое приближение позволяет корректно учитывать пространственную протяженность источника, в том числе при визуализации мягких теней.

### 2. Фотонные карты и финальный сбор

Одним из компонентов расчетного ядра является реализация метода фотонных карт – варианта прямого метода Монте-Карло для численного решения уравнения глобального освещения. В рамках этого подхода производится эмиссия фотонов от источников света с использованием выборки по значимости (importance sampling) [8], основанной на распределении силы света. Далее каждый фотон последовательно распространяется по сцене, отражаясь, преломляясь или поглощаясь в точках взаимодействия с поверхностями – выбор дальнейшего события осуществляется по методу русской рулетки (Russian roulette). Для отражения также применяется выборка по значимости.

В текущей версии ядра реализована диффузная модель отражения, однако архитектура алгоритма позволяет использовать произвольные BRDF, что делает подход потенциально расширяемым на более сложные материалы и сценарии.

Результаты трассировки фотонов сохраняются в структуре данных на основе kd-дерева. Для оценки освещенности  $E(\mathbf{r})$  или яркости  $L(\mathbf{r}, \hat{\mathbf{l}})$  в произвольной точке сцены используется накопленная энергия фотонов, попавших в заданную окрестность этой точки:

$$E(\mathbf{r}) \approx \sum_{i=1}^N \frac{\Delta\Phi_i(\mathbf{r}, \hat{\mathbf{l}}')}{\Delta A}, \quad (3)$$

$$L(\mathbf{r}, \hat{\mathbf{l}}) \approx \sum_{i=0}^N \sigma(\mathbf{r}; \hat{\mathbf{l}}, \hat{\mathbf{l}}') \frac{\Delta\Phi_i(\mathbf{r}, \hat{\mathbf{l}}')}{\Delta A}, \quad (4)$$

где  $\Delta\Phi_i(\mathbf{r}, \hat{\mathbf{l}}')$  – мощность  $i$ -го фотона, попавшего в (сферическую) окрестность исследуемой точки  $\mathbf{r}$ ;

$\Delta A$  – площадь проекции окрестности на плоскость, касательную к поверхности сцены в точке  $\mathbf{r}$ ;

$\sigma(\mathbf{r}; \hat{\mathbf{l}}, \hat{\mathbf{l}}')$  – двунаправленная функция отражения (коэффициент яркости).

Возможны две стратегии отбора фотонов: либо по фиксированному радиусу (что может привести к чрезмерному числу фотонов и замедлению), либо по фиксированному количеству (в этом случае радиус может вырасти и снизить точность в областях с низкой плотностью фотонов).

Такая чувствительность метода фотонных карт к плотности распределения фотонов в сцене делает актуальным применение техники финального сбора (final gathering), которая может быть использована в сочетании с любым методом расчета многократных отражений, основанным на решении уравнения глобального освещения.

Суть финального сбора заключается в дополнительной численной аппроксимации интегрального слагаемого уравнения (1) – синтетической итерации [9] – в заданной точке наблюдения. Так, если требуется определить яркость  $L(\mathbf{r}, \hat{\mathbf{l}})$ , то на заключительном этапе вычисляется определенный интеграл по телесному углу:

$$\int L(\mathbf{r}, \hat{\mathbf{l}}) |(\hat{\mathbf{N}}, \hat{\mathbf{l}})| d\hat{\mathbf{l}}. \quad (5)$$

Реализация этого интегрирования не представляет принципиальных трудностей: в точке  $\mathbf{r}$  строится полусфера, ориентированная по нормали  $\hat{\mathbf{N}}$ , и разбивается на равные телесные элементы, вдоль осей которых испускаются вспомогательные лучи. В местах их пересечения с поверхностью сцены выполняется оценка яркости, основанная на данных фотонной карты.

Характерной особенностью метода в контексте его использования в ядре «Свет в ночи» является включение в фотонную карту как отраженных, так и первичных фотонов. Это позволяет учитывать вклад прямого света в рамках единой схемы и повысить эффективность сбора. Хотя прямой свет в принципе может быть рассчитан отдельно и явно учтен в процессе финального сбора, такой подход требует дополнительных ресурсов и снижает производительность. Использование фотонной карты для включения прямого освещения оказывается в данном случае предпочтительным с точки зрения баланса между точностью и скоростью расчета.

Метод фотонных карт отличается сравнительно высокой вычислительной затратностью. Для достижения удовлетворительной точности требуется генерация большого числа фотонов, что ограничивает его применимость в контексте практических задач, где важна скорость получения результата. Именно это обстоятельство обуславливает актуальность поиска альтернативных и более эффективных стратегий расчета глобального освещения, таких как метод локальных оценок, описанный далее.

### 1. Локальные оценки

В отличие от фотонных карт, метод локальных оценок основан на *строгом* решении уравнения глобального освещения (УГО) методом статистических испытаний. Такой подход позволяет получать оценку яркости непосредственно в заданной точке сцены, что существенно повышает эффективность вычислений при необходимости быстрого получения локальной информации о распределении характеристик светового поля.

Для корректного вывода расчетных формул необходимо несколько изменить исходную запись уравнения (1). Заметим, что выбор точек  $\mathbf{r}_s$  и  $\mathbf{r}'_s$  на поверхности сцены однозначно определяет направление луча  $\hat{\mathbf{l}}$  и, наоборот, направление луча из  $\mathbf{r}'_s$  однозначно определяет точку пересечения  $\mathbf{r}_s$ . Учесть эту взаимосвязь можно, введя под интеграл  $\delta$ -функцию и выполнив формальный переход от интегрирования по поверхности к интегрированию по объему. В результате уравнение (1) принимает вид [10]:

$$L(\mathbf{r}_s, \hat{\mathbf{l}}) = L_0(\mathbf{r}_s, \hat{\mathbf{l}}) + \frac{1}{\pi} \int_{(V)} L(\mathbf{r}'_s, \hat{\mathbf{l}}) \sigma(\mathbf{r}_s, \hat{\mathbf{l}}, \hat{\mathbf{l}}') G_s(\mathbf{r}_s, \mathbf{r}'_s) d^3r'_s, \quad (6)$$

где  $V$  – пространство сцены освещения;

$G_s(\mathbf{r}_s, \mathbf{r}'_s)$  – геометрический фактор яркости на поверхности:

$$G_s(\mathbf{r}_s, \mathbf{r}'_s) \equiv \frac{|\hat{\mathbf{N}}(\mathbf{r}_s) \cdot \hat{\mathbf{l}}'|}{(\mathbf{r}_s - \mathbf{r}'_s)^2} \Theta(\mathbf{r}_s, \mathbf{r}'_s) \delta\left(\hat{\mathbf{l}}' - \frac{\mathbf{r}_s - \mathbf{r}'_s}{|\mathbf{r}_s - \mathbf{r}'_s|}\right). \quad (7)$$

В этой форме УГО связывает пары точек  $\mathbf{r}_s$  и  $\mathbf{r}'_s$  с направлением луча  $\hat{\mathbf{l}}'$ , что позволяет применять методы решения интегральных уравнений, разработанные в рамках теории статистических испытаний. Следуя [10, с. 169–174], уравнение удобно записать в операторной форме:

$$L = L_0 + KL, \quad (8)$$

где  $K$  – интегральный оператор с ядром

$$k(\mathbf{x}' \rightarrow \mathbf{x}) = \frac{1}{\pi} \sigma(\mathbf{r}_s; \hat{\mathbf{l}}, \hat{\mathbf{l}}') G_s(\mathbf{r}_s, \mathbf{r}'_s). \quad (9)$$

Для численного решения методом Монте-Карло вводится цепь Маркова с начальной плотностью  $\pi(\mathbf{x})$  и переходной плотностью  $p(\mathbf{x}' \rightarrow \mathbf{x})$ . В процессе построения траекторий определяются вспомогательные веса:

$$Q_0 = \frac{L_0(\mathbf{x}_0)}{\pi(\mathbf{x}_0)}, \quad Q_i = Q_{i-1} \frac{k(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i)}{p(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i)}. \quad (10)$$

Основная оценка решения в виде функционала

$$I_h = (L, h) = \int L(\mathbf{x})h(\mathbf{x}) d\mathbf{x} \quad (11)$$

получается как математическое ожидание по траекториям цепи:

$$I_h = (L, h) = M\xi = M \sum_{i=0}^T Q_i h(\mathbf{x}_i), \quad (12)$$

где  $M$  – оператор математического ожидания;

$T$  – случайный номер последнего состояния.

Подробное описание практического алгоритма расчета методом локальных оценок приведено в работе [10].

Особенность метода локальных оценок заключается в возможности непосредственно вычислять яркость в интересующей точке сцены. Для этого достаточно положить  $h_{\mathbf{x}}(\mathbf{x}') = k(\mathbf{x}' \rightarrow \mathbf{x})$ , где нижний индекс подчеркивает зависимость функции не только от переменной  $\mathbf{x}'$ , но и от фиксированной точки  $\mathbf{x}$ . В результате получаем локальную оценку:

$$L(\mathbf{x}) = M \left[ \sum_{i=0}^T Q_i k(\mathbf{x}_i \rightarrow \mathbf{x}) \right] + L_0(\mathbf{x}). \quad (13)$$

Заметим, что ядро  $k(\mathbf{x}' \rightarrow \mathbf{x})$  содержит  $\delta$ -функцию. Однако это не препятствует применению метода, поскольку можно показать, что аналогичные  $\delta$ -функции содержатся и в плотностях  $\pi(\mathbf{x})$  и  $p(\mathbf{x}' \rightarrow \mathbf{x})$ .

Возникающие из-за наличия в ядре УГО сингулярности вида  $1/(\mathbf{r}_s - \mathbf{r}'_s)^2$  последствия могут быть устранены путем [12] усреднения искомой величины в окрестности исследуемой точки  $\mathbf{r}_s^* = \{\mathbf{r}_s : |\mathbf{r}_s - \mathbf{r}'_s| < \varepsilon\}$ . При этом усреднение выполняется лишь тогда, когда расстояние между текущей точкой столкновения и исследуемой точкой  $|\mathbf{r}_s - \mathbf{r}'_s|$  оказывается меньше заданного порогового значения. Такой подход позволяет избежать нежелательных артефактов, вызванных сингулярностью, и при этом не нарушает ключевого достоинства метода – возможности вычислять требуемую величину непосредственно в указанных точках сцены.

В отличие от фотонных карт, где требуется поиск ближайших фотонов в специализированной структуре данных, метод локальных оценок не использует подобных процедур. Вместо этого результаты трассировки хранятся в виде списка всех точек соударения лучей. Основным фактором, увеличивающим время счета, становится необходимость проверять видимость расчетной точки относительно всех сохраненных точек соударения лучей со сценой. Тем не менее возможность напрямую получать значение искомой величины в заданных точках без дополнительной аппроксимации делает локальные оценки предпочтительными в задачах, где важна высокая точность локальных расчетов.

### Сравнение с аналогами

Для объективной оценки точности моделирования светораспределения необходимо использовать задачи, для которых существует точное аналитическое решение. В контексте уравнения глобального освещения таких задач крайне мало: это фотометрическая сфера и задача Соболева. Первая ввиду полной симметрии малоинформативна при анализе пространственного распределения освещенности и подходит лишь для проверки корректности реализации базовых физических законов. Тем не менее с рядом оговорок, о которых будет сказано далее, к числу эталонных можно отнести также задачу по расчету средней освещенности поверхности параллелепипеда методом коэффициента использования.

#### 1. Задача Соболева

Задача Соболева традиционно рассматривается в светотехнике и компьютерной графике как один из наиболее строгих тестов корректности учета многократных отражений. В ней моделируется точечный изотропный источник света, расположенный между двумя бесконечными параллельными диффузными плоскостями. Для этой конфигурации существует точное аналитическое решение в виде распределения освещенности на каждой из плоскостей:

$$E_1(r) = \frac{h_1}{(h_1^2 + r^2)^{3/2}} + \rho_2 \int_0^\infty \frac{e^{-h_1 k} \rho_1 k K_1(k) + e^{-h_2 k}}{1 - \rho_1 \rho_2 k^2 K_1^2(k)} K_1(k) J_0(kr) k^2 dk, \quad (14)$$

где  $E_i(r)$  – освещенность  $i$ -й плоскости на расстоянии  $r$  от источника;

$\rho_i$  – коэффициент отражения  $i$ -й плоскости;

$h_i$  – расстояние от  $i$ -й плоскости до источника;

$K_1(k)$  – модифицированная функция Бесселя чисто мнимого аргумента (функция Макдональда первого порядка);

$J_0(k)$  – функция Бесселя нулевого порядка.

Первое слагаемое в этой формуле описывает прямую составляющую освещенности, второе – вклад многократно отраженного света.

Для оценки точности расчетов задача Соболева была смоделирована в программах *DIALux evo* и «Свет в ночи». Результаты были сопоставлены с аналитическим решением, а также с данными программы *Lightscape 3.2*, которая, несмотря на прекращение развития в начале 2000-х годов, до сих пор рассматривается как один из эталонов точности [13]. На рисунке 1 представлены графики распределения освещенности, полученные всеми тремя программами, в сравнении с аналитическим решением; там же указано время счета для каждой программы.

Параметры расчета в программе «Свет в ночи» были выбраны таким образом, чтобы длительность вычислений совпадала с временем выполнения в *DIALux evo*, где отсутствует возможность управлять настройками точности. В случае *Lightscape*, реализующей иной метод (radiosity), параметры выбирались с ориентацией на достижение сопоставимой точности с результатами «Свет в ночи». Для количественной оценки точности моделирования в задаче Соболева были рассчитаны три статистических показателя: средняя относительная погрешность, среднее квадратическое отклонение  $\sigma$  и коэффициент корреляции Пирсона  $r$ . Расчеты выполнялись по пяти контрольным точкам, показанными на графиках рисунка 1. Итоговые значения перечисленных показателей для всех рассматриваемых программ сведены в таблицу 1.

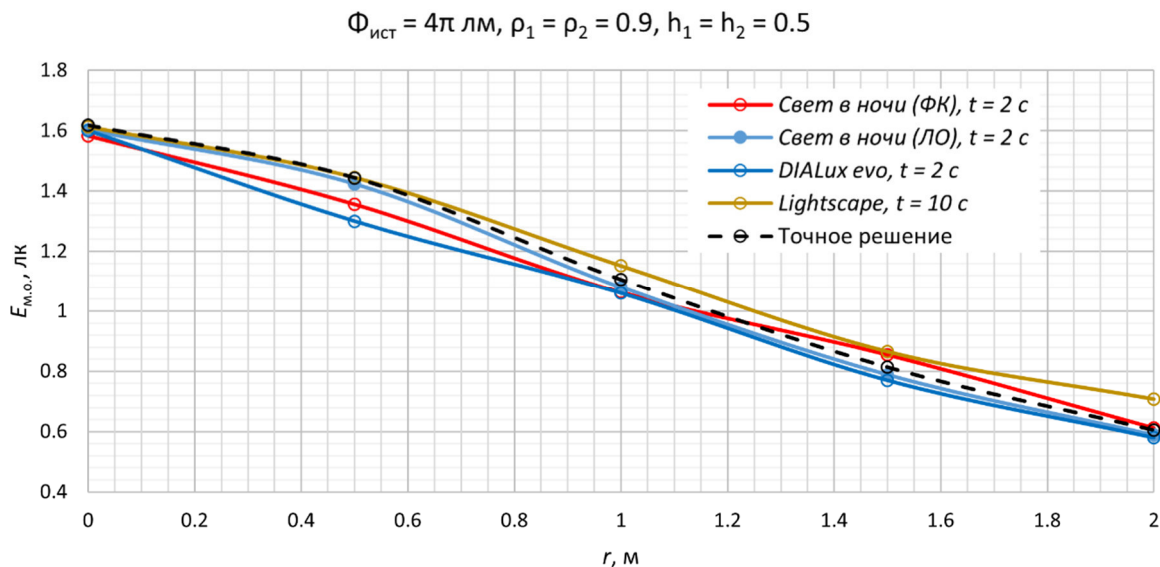


Рисунок 1. Распределение освещенности в задаче Соболева

Таблица 1. Сравнение точности решения задачи Соболева в «Свет в ночи», *DIALux* и *Lightscape*

Программа	Свет в ночи (ФК)	Свет в ночи (ЛО)	<i>DIALux</i>	<i>Lightscape</i>
$\bar{\delta}$ , %	3.64	2.14	4.97	5.55
$\sigma$ , лк	0.0496	0.0214	0.0717	0.0552
$r$	0.9963	0.9999	0.9931	0.9995

Как видно из приведенных данных, все три программы продемонстрировали сопоставимый уровень точности при решении задачи Соболева: коэффициент корреляции Пирсона во всех случаях близок к единице, а средняя относительная погрешность не превышает 6 %. Тем не менее *Lightscape* потребовалось примерно в пять раз больше времени на расчеты по сравнению с *DIALux evo* и «Свет в ночи». Кроме того, при одинаковых временных затратах «Свет в ночи» показала несколько лучшие значения статистических показателей, чем *DIALux evo*, что делает ее более предпочтительной в практическом применении для подобных задач.

## 2. Аналитическое решение УГО для параллелепипеда

До широкого распространения специализированного программного обеспечения, учитывающего многократные отражения, расчет освещенности в помещениях выполнялся с использованием упрощенных инженерных методов. Наиболее известным из них был метод коэффициента использования, в рамках которого освещенность определялась через коэффициент  $u$ , равный отношению светового потока, падающего на расчетную поверхность, к полному потоку осветительного прибора. При этом помещение представлялось в виде прямоугольного параллелепипеда с диффузными стенами, потолком и расчетной поверхностью, которая в частном случае совпадала с полом. Такая постановка задачи имеет точное аналитическое решение, приведенное, в частности, в [14]. Следовательно, расчет средней освещенности на одной из граней параллелепипеда (как правило, на расчетной поверхности) может рассматриваться как еще один пример аналитического решения уравнения глобального освещения. Для наглядности ниже кратко изложены основные этапы этого решения.

Световой поток на каждой из поверхностей параллелепипеда складывается из прямой компоненты, поступающей от светильников, и отраженной, приходящей от остальных поверхностей. Математически это выражается как

$$\Phi_i = \Phi'_i + \sum_{k=1}^6 \Phi_k \rho_k u_{ki}, \quad (15)$$

где  $\Phi'_i$  – прямая составляющая светового потока, падающего на грань  $i$ ;

$\Phi_i$  – световой поток, установившийся на гранях в результате многократных отражений;

$\rho_k$  – коэффициент отражения грани  $k$ ;

$u_{ki}$  – коэффициент использования светового потока, определяющий долю потока грани  $k$ , падающего на грань  $i$ .

Для упрощения расчетов стены можно рассматривать как одну обобщенную отражающую поверхность с коэффициентом отражения  $\rho_c$ . При этом необходимо учитывать потери светового потока, возникающие при переотражении внутри этой поверхности. Эти потери описываются с помощью коэффициента многократных отражений:

$$\gamma_c = \frac{1}{1 - \rho_c(1 - u_{c.п.} - u_{c.р.})}, \quad (16)$$

где  $\rho_c$  – коэффициент отражения стен;

$u_{c.п.}$  – коэффициент использования светового потока стен относительно потолка;

$u_{c.р.}$  – коэффициент использования потока стен относительно расчетной плоскости.

С целью дальнейшего упрощения расчета вогнутую поверхность, образованную потолком и участками стен выше плоскости световых центров осветительных приборов, заменим на эквивалентную плоскую поверхность с аналогичными отражающими свойствами. Ее эффективный коэффициент отражения можно выразить следующим образом:

$$\rho_\phi = \frac{\rho_{ср} u_\phi}{1 - \rho_{ср}(1 - u_\phi)}, \quad (17)$$

где  $\rho_{ср}$  – средневзвешенный по площади коэффициент отражения вогнутой поверхности;

$u_{\phi}$  – коэффициент использования светового потока рассматриваемой поверхности относительно выходного отверстия.

Установившиеся световые потоки на каждой из поверхностей параллелепипеда описываются следующей системой уравнений:

$$\begin{cases} \Phi_{\phi} = \Phi'_{\phi} + \Phi_c \rho_c u_{c.\phi} + \Phi_p \rho_p u_{p.\phi}, \\ \Phi_c = \gamma_c (\Phi'_{\phi} + \Phi_{\phi} \rho_{\phi} u_{\phi.c} + \Phi_p \rho_p u_{p.c}), \\ \Phi_p = \Phi'_{\phi} + \Phi_{\phi} \rho_{\phi} u_{\phi.p} + \Phi_c \rho_c u_{c.p}, \end{cases} \quad (18)$$

где  $\Phi'_{\phi}, \Phi'_c, \Phi'_p$  – прямой световой поток от ОП, падающий соответственно на фиктивную плоскость, стены и расчетную плоскость;

$\Phi_{\phi}, \Phi_c, \Phi_p$  – световые потоки, установившиеся в результате многократных отражений.

Разрешая эту систему уравнений относительно  $\Phi_p$ , получаем выражение для светового потока, установившегося на расчетной плоскости:

$$\Phi_p = A\Phi'_{\phi} + B\Phi'_c + C\Phi'_p, \quad (19)$$

где

$$A = \rho_{\phi} [u_{\phi.p} + \gamma_c \rho_c (1 - u_{\phi.p})^2 \varphi / 2] / D, \quad (20)$$

$$B = [\gamma_c \rho_c (1 - u_{\phi.p}) (1 + \rho_{\phi} u_{\phi.p}) \varphi / 2] / D, \quad (21)$$

$$C = [1 - \gamma_c \rho_{\phi} \rho_c (1 - u_{\phi.p})^2 \varphi / 2] / D, \quad (22)$$

$$D = 1 - [\rho_{\phi} \rho_p u_{\phi.p}^2 \rho_c + \gamma_c (1 - u_{\phi.p})^2 (2\rho_{\phi} \rho_p u_{\phi.p} + \rho_{\phi} + \rho_p) \varphi / 2]. \quad (23)$$

Прямые световые потоки могут быть рассчитаны с помощью различных методик [14]. В данной работе для аналитического расчета применялся метод зональных множителей Джонса и Нейдхарта.

### 3. Средняя освещенность основания параллелепипеда

Рассмотрим прямоугольное помещение размерами 10×10 м и высотой 4 м. Все внутренние поверхности считаются диффузными и обладают одинаковым коэффициентом отражения  $\rho$ . На потолке с равномерным шагом размещено 100 светильников, каждый из которых имеет световой поток 1800 лм и распределение силы света, представленное на рисунке 2.

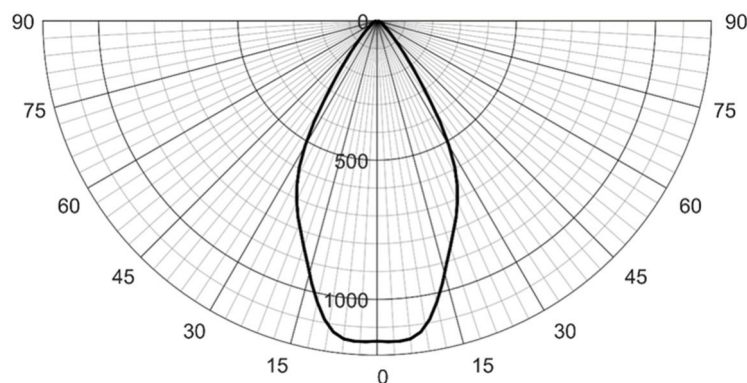


Рисунок 2. Кривая силы света светильника в рассматриваемой сцене

Сцену воспроизвели в программах «Свет в ночи», *DIALux* и *Lightscape* для моделирования многократных отражений и вычисления средней освещенности  $\bar{E}$  пола (основания параллелепипеда) при разных значениях коэффициента отражения поверхностей  $\rho$ . Результаты приведены в таблице 2, где также указано относительное отклонение  $\delta$  от ранее описанного аналитического решения, которое принимается авторами за эталон.

Таблица 2. Расчетные и аналитические значения средней освещенности основания

$\rho$	«Свет в ночи» (ФК)		«Свет в ночи» (ЛО)		<i>DIALux</i>		<i>Lightscape</i>		Аналит. решение
	$\bar{E}$ , лк	$\delta$ , %	$\bar{E}$ , лк	$\delta$ , %	$\bar{E}$ , лк	$\delta$ , %	$\bar{E}$ , лк	$\delta$ , %	



0.9	5777	2.78	5610	0.196	4272	24.0	3870	31.2	5621
0.8	3222	2.61	3141	0.0318	2803	10.7	2707	13.8	3140
0.7	2372	1.98	2331	0.215	2203	5.29	2169	6.75	2326

Окончание таблицы 2

$\rho$	«Свет в ночи» (ФК)		«Свет в ночи» (ЛО)		<i>DIALux</i>		<i>Lightscape</i>		Аналит. решение
	$\bar{E}$ , лк	$\delta$ , %	$\bar{E}$ , лк	$\delta$ , %	$\bar{E}$ , лк	$\delta$ , %	$\bar{E}$ , лк	$\delta$ , %	$\bar{E}$ , лк
0.6	1964	1.71	1945	0.725	1881	2.59	1867	3.31	1931
0.5	1728	1.41	1715	0.646	1688	0.939	1680	1.41	1704
0.4	1580	1.22	1571	0.641	1555	0.384	1555	0.384	1561
0.3	1483	0.953	1474	0.340	1470	0.0681	1470	0.0681	1469
0.2	1419	0.781	1413	0.355	1411	0.213	1413	0.355	1408
0.1	1379	0.657	1378	0.584	1374	0.292	1375	0.365	1370
0.0	1356	0.593	1356	0.593	1353	0.371	1353	0.371	1348

Результаты, представленные в таблице 2, показывают, что при значениях коэффициента отражения  $\rho$  от 0.0 до 0.5 все три программы обеспечивают близкие к аналитическому решению значения средней освещенности. Однако при  $\rho > 0.5$  погрешность в *DIALux* и *Lightscape* существенно возрастает, достигая 24.0 и 31.2 % соответственно при  $\rho = 0.9$ . Особенно ярко это проявляется при анализе только отраженной доли светового потока: в этом случае отклонения составляют 31.7 % для *DIALux* и 41.1 % для *Lightscape*. Хотя точная природа этих расхождений остается неясной, вероятной причиной авторы считают особенности реализации алгоритмов *photon mapping* (в *DIALux*) и *radiosity* (в *Lightscape*). В частности, в последнем случае ошибка в конкретном тесте может быть связана не с принципиальными ограничениями алгоритма, а с ограничениями по памяти и, как следствие, неоптимальной настройкой сетки. На этом фоне программа «Свет в ночи» демонстрирует заметно более точные результаты: отклонения остаются невысокими вплоть до максимальных значений  $\rho$ , что подтверждает корректность и устойчивость вычислений.

### Заключение

В работе проанализировано расчетное ядро САПР «Свет в ночи», в котором реализованы два метода моделирования многократных отражений: фотонные карты и перспективный метод локальных оценок. Прямое освещение рассчитывается по закону обратных квадратов. Для оценки точности применяемых алгоритмов были рассмотрены две тестовые задачи с известными аналитическими решениями: задача Соболева и расчет средней освещенности основания параллелепипеда.

На последнем примере показано, что «Свет в ночи» демонстрирует стабильную и высокую точность расчетов в широком диапазоне коэффициентов отражения – при максимальном отклонении от аналитического решения не более 2.78 %. При тех же условиях *DIALux* и *Lightscape* дают значительно большие отклонения, достигающие 24.0 и 32.1 % соответственно.

В дальнейшем планируется развитие метода двойных локальных оценок, позволяющего напрямую рассчитывать яркость в произвольной точке объема сцены в заданном направлении. Также рассматриваются внедрение более физически обоснованных моделей отражения и интеграция с методами машинного обучения для автоматического подбора параметров моделирования.

Таким образом, САПР «Свет в ночи» представляет собой надежный и точный инструмент для проектирования освещения, сочетающий физическую корректность с практической эффективностью. Будущие улучшения направлены на расширение возможностей ядра и его адаптацию к современным задачам светотехники и компьютерной графики.

### Список литературы

1. Свид. о гос. регистрации программы для ЭВМ № 2023685592 Российская Федерация. LiN.Calculate.light версия 1.0: № 2023685508: заявл. 28.11.2023: опубл. (зарег.): 28.11.2023 / В. П. Будак [и др.]; заявитель ООО «Центр инновационных разработок МСК БЛ Групп». URL: <https://www.l-i-n.ru/>
2. Kajiya J.T. The rendering equation // Proc. 13th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1986. 1986. P. 143–150.
3. Соболев В.В. Точечный источник света между параллельными плоскостями // Доклады Академии наук СССР. 1944. Т. XLII, № 4. С. 175–176.
4. Modeling the Interaction of Light Between Diffuse Surfaces / Goral C.M., et al. // Comput. Graph. 1984. Vol. 18, no. 3. P. 213–222.
5. Jensen H.W. Global Illumination using Photon Maps // In Rendering Techniques'96. Springer-Verlag, 1996. P. 21–30.
6. The current state of the methods for calculating global illumination in tasks of realistic computer graphics / Frolov V., et al. // Proc. Inst. Syst. Program. RAS. 2021. Vol. 33, no. 2. P. 7–48.
7. Kalos M.H. On the Estimation of Flux at a Point by Monte Carlo // Nucl. Sci. Eng. 1963. Vol. 16, no. 1. P. 111–117.
8. Соболев И.М. Численные методы Монте-Карло. М.: Наука, 1973. 312 с.
9. Adams M.L., Larsen E.W. Fast iterative methods for discrete-ordinates particle transport calculations // Prog. Nucl. Energy. 2002. Vol. 40, no. 1. P. 3–159.
10. Budak V.P., Zheltov V.S., Kalakutsky T.K. Local estimations of Monte Carlo method with the object spectral representation in the solution of global illumination // Comput. Res. Model. 2012. Vol. 4, no. 1. P. 75–84.
11. Ермаков С.М., Михайлов Г.А. Курс статистического моделирования. М.: Главная редакция физико-математической литературы изд-ва «Наука», 1976. 320 с.
12. Practical Application of Local Estimations: Singularity Removal in the Solution of the Global Illumination Equation / Budak V.P., et al. // 33rd Int. Conf. Comput. Graph. Vis. 2023. P. 108–115.
13. Макаров Д.Н. Методы компьютерного моделирования осветительных установок: автореф. дис. ... канд. техн. наук: 05.09.07 / Макаров Денис Николаевич. М., 2007. 19 с.
14. Мешков В.В., Епанешников М.М. Осветительные установки: учебное пособие для вузов. М.: Энергия, 1972. 360 с.

## Трассируемые мультиобъекты на основе 3D-гауссианов

П. Ю. Тимохин, М. В. Михайлюк

Научно-исследовательский институт системных исследований  
Национального исследовательского центра «Курчатовский институт», Москва, Россия

**Аннотация.** Важным строительным блоком современного виртуального окружения является мультиобъект – совокупность виртуального объекта-образца и наборов отличительных признаков его экземпляров. Данная работа посвящена направлению визуализации сложных процедурных мультиобъектов с помощью аппаратно-ускоренной трассировки лучей. В частности, рассматривается задача построения трассируемых мультиобъектов, основанных на процедурных примитивах нового типа – 3D-гауссианах. В работе исследуются особенности таких примитивов, предлагаются методы и подходы к построению эффективной модели облака 3D-гауссианов, состоящей из ограничивающих параллелепипедов. Выполнена апробация предложенных решений, включающая сравнительный анализ качества и скорости визуализации отдельного экземпляра образца травы, а также обширного массива, содержащего 4 млн экземпляров. Полученные решения могут быть использованы в системах виртуального окружения, научной визуализации, при разработке визуализаторов нового поколения, основанных на 3D-гауссианах, и др.

**Ключевые слова:** виртуальное окружение, мультиобъектная визуализация, облако точек, 3D-гауссиан, трассировка лучей, GPU

## Traceable multiobjects based on 3D gaussians

P. Y. Timokhin, M. V. Mikhaylyuk

Scientific Research Institute for System Analysis of the National Research Centre  
"Kurchatov Institute", Moscow, Russia

**Abstract.** An essential building block of modern virtual environment is the multiobject, defined as a composition of a sample virtual object and sets of distinctive features of its instances. This work is devoted to the visualization of complex procedural multiobjects by means of hardware-accelerated ray tracing. In particular, the task of constructing traceable multiobjects based on a new type of procedural primitives, 3D Gaussians, is considered. The paper examines the features of such primitives, and suggests methods and approaches for constructing an effective 3D Gaussian cloud model consisting of bounding boxes. The proposed solutions were tested, including a comparative analysis of the quality and speed of visualization of a single grass sample, as well as an extensive array containing 4 million instances. The obtained solutions can be used in virtual environment systems, scientific visualization, in the development of a new generation of visualizers based on 3D Gaussians, etc.

**Keywords:** virtual environment, multiobject visualization, point cloud, 3D-Gaussian, ray tracing, GPU

### Введение

Современное виртуальное окружение строится на основе гибридного подхода, при котором традиционные трехмерные полигональные сцены объединяются с процедурными объектами, вычисляемыми в процессе визуализации [1, 2]. В контексте моделирования открытых пространств [3] особенно востребованы процедурные *мультиобъекты*, работающие по принципу «многое в малом» (от лат. *multum in parvo*), которые в реальном времени развертываются во множество экземпляров, в общем случае отличающихся друг от друга. Данная тенденция получила мощную поддержку в архитектуре аппаратно-ускоренной трассировки лучей [4, 5], реализованной в современных графических процессорах (GPUs), в рамках которой появились *трассируемые мультиобъекты*.

Базовая концепция трассируемого мультиобъекта [6] состоит из двух основных компонент (рис. 1). Первая компонента (далее *ядро*) описывает объект-образец и включает в себя его процедурную модель<sup>1</sup>, а также его модель в виде ограничивающих параллелепипедов (axis-aligned bounding boxes, AABBs). Вторая компонента (далее *модификатор*) отвечает за вариативность экземпляра объекта-образца и включает в себя базовый набор его отличительных признаков (положение, ориентация, масштаб). В процессе синтеза кадра визуализации модификаторы экземпляров применяются к ядру мультиобъекта

<sup>1</sup> В данной работе под процедурной моделью объекта-образца понимается совокупность его параметрического описания и алгоритма пересечения с лучом.

на конвейере трассировки лучей GPU (далее **RT-конвейере**), в результате чего формируется изображение множества экземпляров объекта-образца.

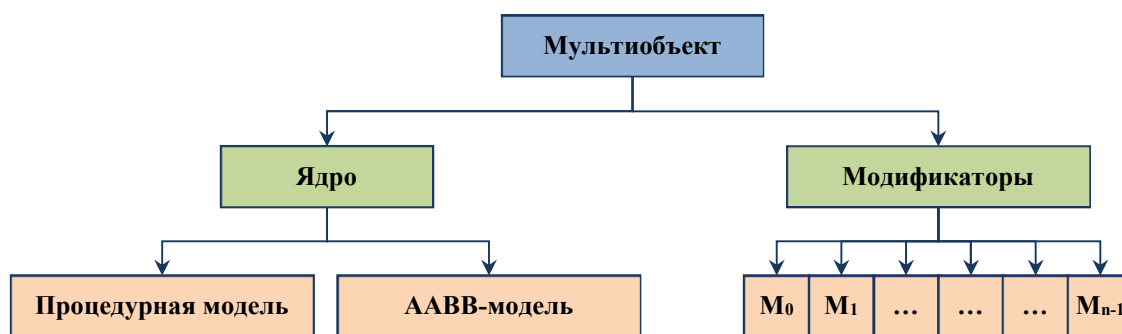


Рисунок 1. Общая структура трассируемого мультиобъекта

Одним из перспективных направлений является исследование и разработка трассируемых мультиобъектов с ядрами, основанными на *облаках точек*. Главным преимуществом таких мультиобъектов является возможность в компактной и универсальной форме описывать целые классы разнообразных объектов окружающей среды, имеющих сложную и уникальную геометрию (например, элементы растительности).

В референсной работе [7] продемонстрирован принцип построения ядра мультиобъекта на примере облака точек, сгенерированного случайным образом. В этой работе процедурная модель облака точек рассматривается как совокупность *примитивных* процедурных моделей его точек (используются сферы и кубы), а ААВВ-модель – как совокупность ААВВs этих примитивных моделей. Такой подход относительно просто реализуется, однако характеризуется тем, что увеличение размера облака точек приводит к аналогичному росту числа ААВВs. Ввиду того что RT-конвейер может одновременно обрабатывать лишь ограниченное число ААВВs (обработка выполняется специальным аппаратным блоком), большой «наплыв» ААВВs создает очередь и приводит к образованию «бутылочного горлышка», тормозящего работу всего RT-конвейера. Кроме того, такой большой рост числа ААВВs приводит к нерациональному расходу видеопамяти, что также является серьезным ограничением.

Описанная проблема исследована в нашей первой работе [8] в контексте задачи визуализации детализированных облаков точек, полученных в результате 3D-сканирования реальных объектов (десятки миллионов точек). В нашем исследовании предложено рассматривать процедурную модель облака точек как совокупность процедурных моделей *групп точек*, соответствующих листьям разреженного воксельного октодеревя. Было установлено, что ААВВ-модель, построенная по таким группам точек, хорошо вписывается в иерархию ограничивающих объемов (BVH) [9], на основе которой работает RT-конвейер, а также содержит существенно меньше ААВВs, чем размер облака точек<sup>2</sup>. По сравнению с референсным подходом [7] разработанное «групповое» ядро обеспечивает прирост производительности до 47 % и сокращение накладных расходов видеопамяти в 1,9-5,4 раз.

Полученные результаты легли в основу *расширенной концепции* трассируемого мультиобъекта (РК), разработанной в нашем втором исследовании [10]. В рамках данной концепции предложено оснастить «групповое» ядро *дополненными модификаторами*, которые в отличие от базовых модификаторов [11] изменяют не только положения, ориентации и масштабы экземпляров объекта-образца, но и их цвета (оттенки). В работе предложен метод задания таких модификаторов, основанный на псевдослучайной выборке из нормального распределения, который обеспечивает эффект уникальности генерируемых экземпляров объекта-образца. В ходе исследования была показана высокая эффективность РК на примере задачи визуализации в реальном времени обширных виртуальных лесных массивов, состоящих из миллионов деревьев (одно дерево содержит порядка миллиона точек).

Настоящая работа посвящена развитию РК в части поддержки работы с *двусторонними объектами-образцами*, как, например, трава или листья. Характерной чертой таких объектов является наличие

<sup>2</sup> На практике один ААВВ охватывает до 4-16 точек.

тонких изогнутых поверхностей, которые проблематично описать с помощью процедурных сфер (для получения визуально-непрерывной поверхности требуется огромное количество сфер малого радиуса). Эффективным путем решения данной проблемы является переход к представлению объекта-образца в виде облака 3D-гауссианов [12]. По сути, 3D-гауссиан представляет собой трехосный эллипсоид, параметры которого подобраны таким образом, чтобы он аппроксимировал определенный участок поверхности объекта. Благодаря тому, что 3D-гауссианы покрывают анизотропные области, с их помощью можно моделировать объекты-образцы любой формы (в том числе, тонкие и изогнутые), затрачивая при этом существенно меньшее количество примитивов по сравнению с облаком сфер. В рамках описанного развития РК в данной работе предлагается метод объединения 3D-гауссианов и аппаратно-ускоренной трассировки лучей, основанный на модификации оригинального «группового» ядра трассируемого мультиобъекта.

### Исходные данные

В нашей работе исходными данными для построения трассируемых мультиобъектов являются облака 3D-гауссианов, заданные с помощью базовой параметрической модели. Данная модель описывает следующие геометрические и визуальные свойства 3D-гауссиана: форму, положение, ориентацию, диффузный цвет и прозрачность.

Геометрические свойства задаются с использованием локальных систем координат 3D-гауссиана и объекта-образца (далее системы GCS и OCS, см. рис. 2а). В частности, форма 3D-гауссиана определяется тройкой коэффициентов  $\{a, b, c\}$  сжатия/растяжения единичной сферы вдоль осей системы GCS, а положение и ориентация задаются относительно системы OCS с помощью координат  $P_c$  центра эллипсоида и единичного кватерниона  $q$ .

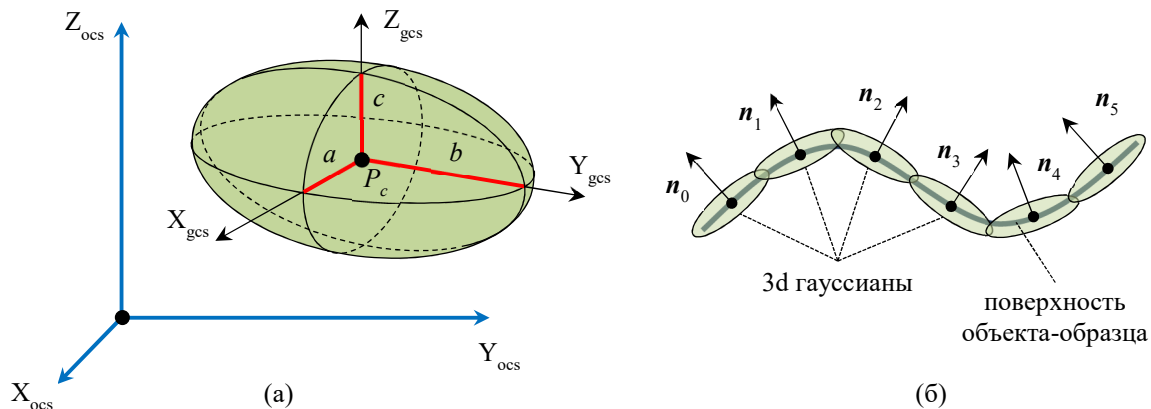


Рисунок 2. Параметры 3D-гауссиана: а) геометрические, б) для расчета освещенности

Визуальные свойства 3D-гауссиана описываются согласно теории переноса излучения [13]. В рамках этой теории цвет 3D-гауссиана задается не в явном виде, а с помощью коэффициентов сферических гармонических функций (гармоник) 0-го, 1-го и более высоких порядков. Такой подход позволяет хранить в компактном виде видозависимые визуальные свойства 3D-гауссиана (блики, отражения, анизотропные материалы), а также обеспечивает непрерывность их изменения в зависимости от угла обзора. В базовую параметрическую модель входит тройка коэффициентов  $\{k_{sh0}, k_{sh1}, k_{sh2}\}$  сферической гармоники 0-го порядка, определяющая диффузный цвет 3D-гауссиана. Данные коэффициенты связаны с rgb-цветом следующим соотношением:

$$(r, g, b) = (k_{sh0}, k_{sh1}, k_{sh2}) Y_0^0 + (0.5, 0.5, 0.5), \quad (1)$$

где  $Y_0^0 = 1/\sqrt{4\pi} \approx 0.282095$  – сферическая гармоника 0-го порядка [14]. Прозрачность 3D-гауссиана задается с помощью привычного коэффициента  $\alpha \in [0, 1]$  [не]прозрачности. Кроме перечисленных выше параметров, в базовую модель 3D-гауссиана также входит нормаль  $n$  к поверхности аппроксимируемого объекта-образца в центре  $P_c$  эллипсоида (см. рис. 2б). Данный параметр необходим для корректного расчета освещенности 3D-гауссиана.

В отличие от расширенных моделей [12, 15], включающих в себя коэффициенты гармоник более высокого порядка, базовая модель 3D-гауссиана занимает существенно меньше памяти, что особенно важно в контексте мультиобъектной визуализации. Также базовая модель поддерживается современными 3D-редакторами [16, 17], что позволяет оценивать корректность получаемых результатов и проводить их сравнительный анализ.

### Постановка задачи

В рамках данного исследования мы будем работать с облаками 3D-гауссианов, полученными из непрозрачных высокополигональных моделей (далее будем называть такие облака *восстановленными*). По сравнению с каноническим итеративным методом извлечения облака 3D-гауссианов из серии снимков [12], процесс восстановления облака 3D-гауссианов по полигональной модели реализуется значительно проще и быстрее, ввиду наличия априорной информации о геометрии объекта. Для подготовки восстановленных облаков мы используем продвинутого конвертер [18], который строит 3D-гауссиановую аппроксимацию в масштабе реального времени ( $< 1$  мс) на основе шейдерной обработки меша на графическом конвейере GPU. Такой подход позволяет быстро создавать детализированные облака 3D-гауссианов (миллионы элементов) для широкого круга объектов-образцов, а также сравнивать качество получаемой аппроксимации с эталонной высокополигональной моделью.

Пусть имеется объект-образец, заданный в виде восстановленного облака 3D-гауссианов, а также задан дополненный модификатор (матрица трансформации и цвет оттенка) некоторого экземпляра этого объекта-образца. Нашей задачей будет отображение данного экземпляра на RT-конвейере с помощью рейкастинга (частный случай обратной трассировки лучей). Чтобы сформулировать эту задачу, вначале кратко рассмотрим особенности реализации рейкастинга непрозрачных процедурных объектов на RT-конвейере.

Как известно, при рейкастинге отображение объекта реализуется на основе поиска ближайшей точки пересечения луча, испущенного из позиции наблюдателя (*первичного луча*), с примитивами, из которых состоит этот объект (рис. 3). В рассматриваемой задаче отображаемый объект состоит из процедурных примитивов (3D-гауссианов), о геометрии которых RT-конвейер ничего не знает, поэтому его работа должна быть дополнена стадией шейдера расчета пересечений (далее *I-шейдер*) [6]. Данная стадия отвечает за расчет параметра  $t_{hit}$  точки пересечения луча с процедурным примитивом и вызывается RT-конвейером автоматически, когда луч пересекает AABV, в котором находится этот процедурный примитив. После обработки I-шейдером всех AABVs, лежащих на пути луча, RT-конвейер автоматически определяет параметр  $t_{chit}$  точки ближайшего пересечения (см. рис. 3) и вызывает для нее стадию шейдера ближайшего пересечения (далее *CH-шейдер*). На данной стадии выполняется расчет цвета луча, который в дальнейшем используется для закраски соответствующего пиксела.

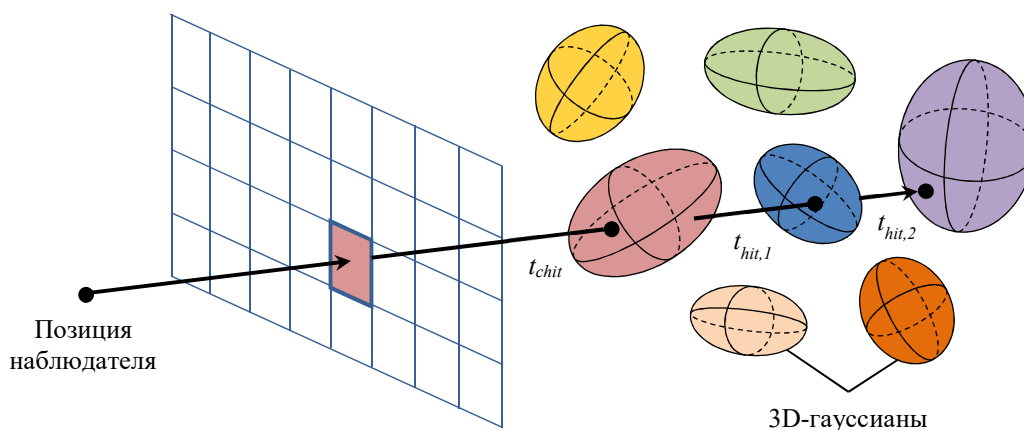


Рисунок 3. Отображение облака 3D-гауссианов с помощью рейкастинга

Исходя из описанного выше, рассматриваемую задачу можно сформулировать в виде следующих трех подзадач: а) построения AABV-модели облака 3D-гауссианов объекта-образца, б) разработки I-шейдера пересечения луча с 3D-гауссианом экземпляра объекта-образца и в) разработки CH-шейдера расчета



цвета луча. Вопросы реализации I-шейдера и СН-шейдера во многом пересекаются с нашим предыдущим исследованием [10] (отличие состоит в использовании в I-шейдере алгоритма пересечения луча с эллипсоидом [19] вместо сферы). Ввиду этого более подробно остановимся на предлагаемом методе решения первой подзадачи.

### Метод построения AABV-модели облака 3D-гауссианов

Построение правильной AABV-модели процедурного объекта является важнейшим фактором, определяющим эффективность мультиобъектной визуализации и аппаратного ускорения трассировки лучей в целом. Если отображаемый объект состоит из небольшого числа процедурных примитивов, то стандартным решением будет заключение каждого такого примитива в свой AABV. Однако, если таких примитивов огромное количество (как в нашем случае) и, что особенно важно, они плотно расположены и пересекаются друг с другом (см. рис. 2б), то реализация стандартного подхода будет приводить к двум серьезным проблемам.

Первой проблемой является избыточное разрастание BVH-иерархии – древовидного представления виртуальной сцены, по которому RT-конвейер «понимает», где и какой экземпляр объекта-образца находится. В рамках базовой концепции трассируемого мультиобъекта [6] размер BVH-иерархии тесно связан с размером AABV-модели объекта-образца и числом его экземпляров. Соответственно, огромное количество AABVs в объекте-образце будет приводить к громоздкой BVH-иерархии и высоким накладным затратам на ее хранение и аппаратную обработку RT-конвейером (см. введение).

Второй проблемой является падение производительности RT-конвейера вследствие повторной обработки процедурных примитивов в местах пересечений их AABVs. Чем больше таких пересечений, тем выше пустая трата вычислительного ресурса RT-конвейера и тем ниже скорость визуализации. Согласно практическим рекомендациям по эффективному использованию аппаратного ускорения трассировки лучей [20], таких ситуаций следует избегать, например, путем объединения пересекающихся AABVs в более крупные.

С учетом вышеизложенного в данной работе предлагается двухэтапный метод построения AABV-модели облака 3D-гауссианов. На *первом этапе* выполняется построение опорной AABV-модели, при которой для каждого 3D-гауссиана облака создается свой AABV (фактически реализуется стандартный подход). На *втором этапе* выполняется оптимизация опорной AABV-модели, при которой реализуется укрупнение пересекающихся AABVs 3D-гауссианов. Рассмотрим эти этапы более подробно.

#### 1. Построение опорной AABV-модели

Задача построения опорной AABV-модели, по сути, сводится к построению AABV трехосного эллипсоида, положение и ориентация которого заданы координатами  $P_c$  центра и единичным кватернионом  $q$ , а полуоси – тройкой значений  $\{a, b, c\}$  (см. раздел «Исходные данные»). Ввиду того что стороны AABV параллельны осям мировой системы координат (далее *системы WCS*), для его построения достаточно найти координаты  $P_{min}$  и  $P_{max}$  его диагональных вершин (рис. 4).

Для простоты примем, что система OCS облака 3D-гауссианов совпадает с мировой системой WCS. Тогда координаты  $P_{min}$  и  $P_{max}$  можно найти через касательные плоскости к эллипсоиду, параллельные плоскостям  $Y_{ocs}Z_{ocs}$ ,  $X_{ocs}Z_{ocs}$  и  $X_{ocs}Y_{ocs}$ :

$$x_{\min, \max} = x_c \mp L_x, \quad y_{\min, \max} = y_c \mp L_y, \quad z_{\min, \max} = z_c \mp L_z, \quad (2)$$

где  $L_x$ ,  $L_y$  и  $L_z$  – это расстояния от центра  $P_c$  эллипсоида до касательных плоскостей  $\{x_{\max}, X_{ocs}\}$ ,  $\{y_{\max}, Y_{ocs}\}$  и  $\{z_{\max}, Z_{ocs}\}$ . Из рисунка 4 легко видеть, что если эллипсоид ориентирован вдоль осей системы OCS, то длины  $L_x$ ,  $L_y$ ,  $L_z$  будут равны полуосям  $a$ ,  $b$ ,  $c$  эллипсоида. Однако в случае произвольно ориентированного эллипсоида вычисление длин  $L_x$ ,  $L_y$ ,  $L_z$  становится нетривиальной задачей. В работе [21] описывается аналитический подход к решению этой задачи в общем виде, основанный на представлении ориентированного трехосного эллипсоида в виде трансформированной единичной сферы, из которого следует

$$L_x = \sqrt{M_{0,0}^2 + M_{0,1}^2 + M_{0,2}^2}, \quad L_y = \sqrt{M_{1,0}^2 + M_{1,1}^2 + M_{1,2}^2}, \quad L_z = \sqrt{M_{2,0}^2 + M_{2,1}^2 + M_{2,2}^2}, \quad (3)$$

где  $M$  – матрица произвольного аффинного преобразования единичной сферы.

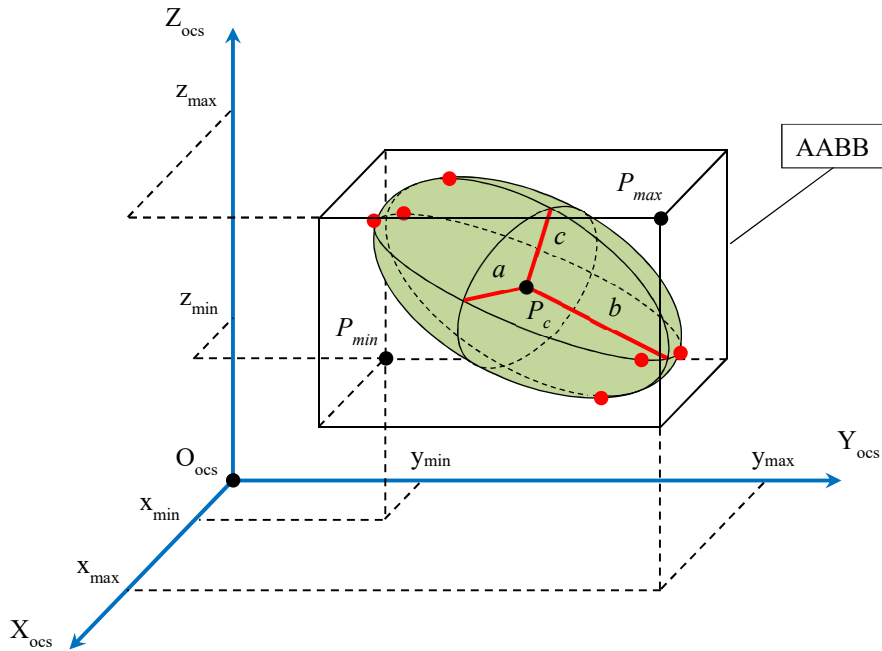


Рисунок 4. Ограничивающий параллелепипед 3D-гауссиана

Разовьем подход [21] применительно к 3D-гауссианам. Для этого вначале определим матрицу  $M$  3D-гауссиана. Согласно теории аффинных преобразований [22] запишем для нее следующее выражение:

$$M = TR_q S, \quad (4)$$

где  $T$  – матрица переноса 3D-гауссиана (пространства) на вектор  $O_{ocs}P_c$ ,  $R_q$  – матрица поворота пространства, соответствующая кватерниону  $q$ , а  $S$  – матрица растяжения/сжатия пространства в  $a$ ,  $b$  и  $c$  раз вдоль осей  $X_{gcs}$ ,  $Y_{gcs}$ ,  $Z_{gcs}$  (см. раздел «Исходные данные»). Выразим матрицу  $R_q$  через компоненты кватерниона  $q$  согласно [22] и раскроем поэлементно произведение матриц (4):

$$M = \begin{pmatrix} a(1-2(y^2+z^2)) & 2b(xy-wz) & 2c(xz+wy) & x_c \\ 2a(xy+wz) & b(1-2(x^2+z^2)) & 2c(yz-wx) & y_c \\ 2a(xz-wy) & 2b(yz+wx) & c(1-2(x^2+y^2)) & z_c \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (5)$$

где  $x, y, z, w$  – компоненты кватерниона  $q$ .

Подставим элементы матрицы  $M$  в выражение (3), приведем подобные члены и упростим полученные выражения  $L_x$ ,  $L_y$  и  $L_z$ , используя свойство единичного кватерниона  $x^2 + y^2 + z^2 + w^2 = 1$ :

$$\begin{aligned} L_x &= \sqrt{a^2 - 4(a^2 - b^2)(x^2 y^2 + z^2 w^2) - 4(a^2 - c^2)(x^2 z^2 + y^2 w^2) - 8(b^2 - c^2)xyzw}, \\ L_y &= \sqrt{b^2 + 4(a^2 - b^2)(x^2 y^2 + z^2 w^2) - 4(b^2 - c^2)(y^2 z^2 + x^2 w^2) + 8(a^2 - c^2)xyzw}, \\ L_z &= \sqrt{c^2 + 4(b^2 - c^2)(y^2 z^2 + x^2 w^2) + 4(a^2 - c^2)(x^2 z^2 + y^2 w^2) - 8(a^2 - b^2)xyzw}. \end{aligned} \quad (6)$$

Нетрудно заметить, что полученное выражение (6) состоит из повторяющихся членов, которые можно предварительно вычислить. Введем следующие предвычисляемые компоненты:  $k_1 = a^2 - b^2$ ,  $k_2 = b^2 - c^2$ ,  $k_3 = a^2 - c^2$ ,  $A = 4k_1(x^2 y^2 + z^2 w^2)$ ,  $B = 4k_2(y^2 z^2 + x^2 w^2)$ ,  $C = 4k_3(x^2 z^2 + y^2 w^2)$ ,  $D = 8xyzw$ . Выполнив соответствующие замены в выражении (6) и подставив его в выражение (2), получим искомое выражение координат вершин AABB через параметры 3D-гауссиана:

$$\begin{aligned} x_{\min, \max} &= x_c \mp \sqrt{a^2 - A - C - k_2 D}, \\ y_{\min, \max} &= y_c \mp \sqrt{b^2 + A - B + k_3 D}, \\ z_{\min, \max} &= z_c \mp \sqrt{c^2 + B + C - k_1 D}. \end{aligned} \quad (7)$$

Отметим, что в случае 3D-гауссианов величины  $a$ ,  $b$  и  $c$  могут принимать крайне малые значения, ввиду чего в подкоренных выражениях могут получаться значения в окрестности 0, что приводит к вырождению AABB. Во избежание таких ситуаций при вычислении выражения (7) мы используем стабилизацию вида  $Q = \max(Q, \delta)$ , где  $Q$  – подкоренное выражение, а  $\delta = 10^{-2}$  – квадрат минимального размера AABB.

## 2. Оптимизация опорной AABB-модели

На этапе оптимизации AABB-модели выполняется слияние AABBs 3D-гауссианов, лежащих близко друг к другу, в более крупные AABBs. Ключевым аспектом данной задачи является разбиение исходного неупорядоченного облака на группы близлежащих 3D-гауссианов. Для этого в данной работе предлагается модифицированная версия метода октантных групп точек [8], созданного нами ранее для решения аналогичной задачи применительно к облакам точек (процедурных сфер).

Суть базового метода состоит в том, что облако точек разбивается на 8 октантов (кубических объемов), по которым распределяются все точки облака (по аналогии с раскладыванием шариков по корзинам). Если число точек в октанте превышает некоторый порог  $K_{max}$ , то этот октант разбивается на 8 подоктантов, между которыми снова распределяются его точки, и так повторяется до тех пор, пока не останутся октанты, содержащие от 1 до  $K_{max}$  точек. На финальной фазе для каждой такой октантной группы точек строится AABB (вычисляются диагональные вершины).

В оригинальной версии метода распределение точек по октантам осуществляется без учета размеров их процедурных примитивов (сфер), т.е. если точка попадает в октант, то считается, что попадает и ее сфера. Это допущение основано на том, что размер процедурных сфер у всех точек облака одинаков и изотропен (не зависит от направления измерения), вследствие чего глубина взаимного пересечения AABBs не будет превышать длину диагонали куба, описанного около процедурной сферы (рис. 5а). В отличие от облака сфер у 3D-гауссианов размеры, в общем случае, неодинаковы и анизотропны, ввиду чего применение оригинального метода группировки может приводить к более глубоким взаимным пересечениям AABBs (рис. 5б), что нежелательно.

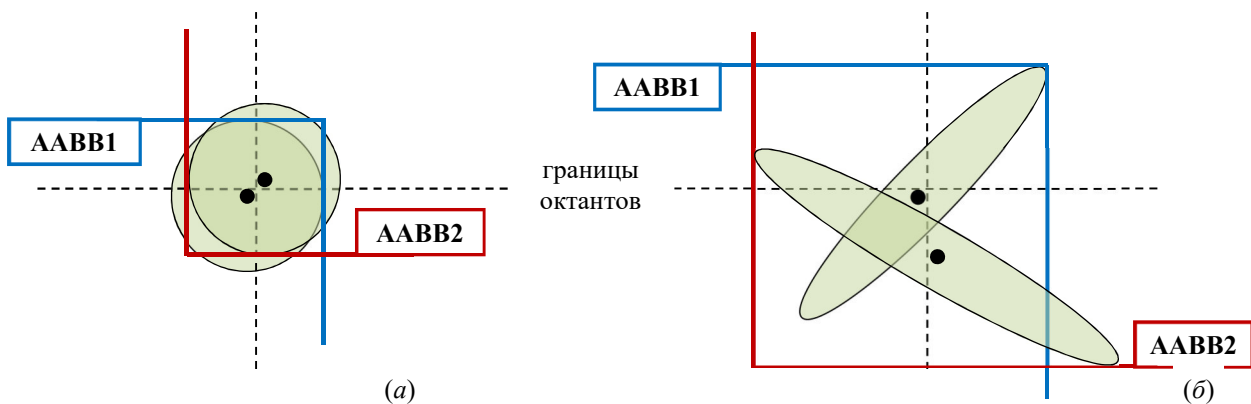


Рисунок 5. Взаимные пересечения AABBs двух групп: а) процедурных сфер; б) 3D-гауссианов

Чтобы уменьшить влияние анизотропности 3D-гауссианов на глубину взаимного пересечения AABBs, в разработанной модификации метода [8] мы исключаем из октантных групп те 3D-гауссианы, которые выходят за границы октанта. Для этого при распределении 3D-гауссианов по октантам мы проверяем не их центры (как в оригинальном методе), а диагональные вершины их индивидуальных AABB, вычисленные по формулам (7) на этапе построения опорной AABB-модели. Те 3D-гауссианы, у которых обе диагональные вершины AABB попали в октант, остаются в группе и участвуют в дальнейшем перераспределении по подоктантам. Если же хотя бы одна диагональная вершина AABB не попадает в октант, то для такого 3D-гауссиана процесс группировки считается завершенным и формируется единичная группа (состоящая из одного элемента). После формирования всех групп 3D-гауссианов (включая единичные) выполняется финальная фаза слияния, при которой индивидуальные AABBs 3D-гауссианов, принадлежащих одной группе, сливаются в общий AABB группы (AABB1 и AABB2 на рис. 6). Отметим, что в случае единичной группы результатом фазы слияния будет индивидуальный AABB (AABB3 на рис. 6). Полученная совокупность групповых и индивидуальных

AABBs образует в результате AABB-модель облака 3D-гауссианов, сбалансированную по количеству / глубине взаимного пересечения AABBs.

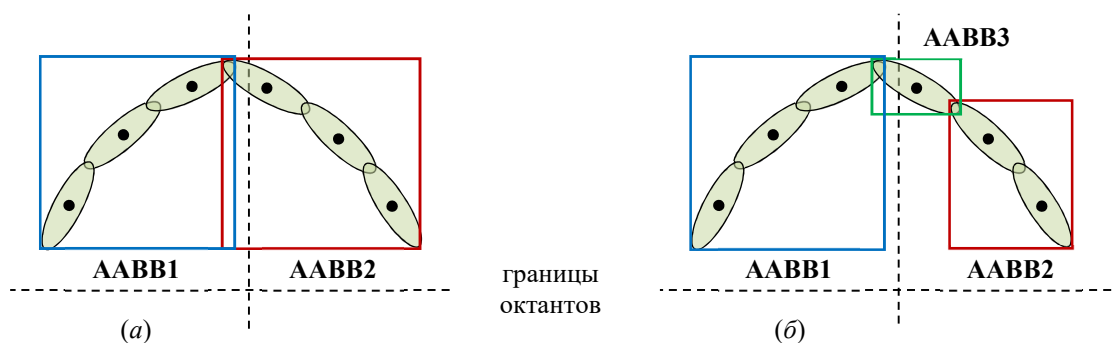


Рисунок 6. Слияние AABBs 3D-гауссианов: а) базовый метод; б) модифицированный

### Результаты экспериментов

Экспериментальная часть данного исследования проводилась на базе созданного нами ранее программного комплекса MultiView мультиобъектной визуализации виртуальной растительности [10]. В состав комплекса входит CPU-модуль, отвечающий за построение ядра мультиобъекта и расчет дополненных модификаторов (см. раздел «Исходные данные»), а также GPU-модуль, реализующий отображение в реальном времени мультиобъекта во множество экземпляров. Программный комплекс написан на языках C++ и GLSL 4.6 с применением библиотеки STL, графических API OpenGL 4.6 и Vulkan 1.3.290 (для программирования аппаратно-ускоренной трассировки лучей).

На основе предложенных методов и подходов в комплексе MultiView была реализована поддержка трассируемых мультиобъектов нового типа – «облако 3D-гауссианов». Данная возможность была экспериментально исследована в ходе апробации комплекса MultiView на персональном компьютере (Intel Core i7-6800K, 16 GB RAM, NVidia RTX 2080, Full HD разрешение). Апробация предложенного решения выполнялась в три этапа.

На **первом этапе** был проведен сравнительный анализ качества визуализации экземпляра мультиобъекта. Для этого была предварительно разработана методика подготовки облака 3D-гауссианов из высокополигональной («high-poly») модели, выступающей в роли эталона. Данная методика включает в себя унификацию локальной системы координат «high-poly» модели с помощью открытого комплекса CloudCompare [23], конвертацию в бинарный GLB-формат с помощью свободного пакета Blender [24] и преобразование в облако 3D-гауссианов с помощью открытого комплекса Mesh2Splat [18]. В ходе сравнительного анализа была выполнена визуализация (рис. 7): а) эталонной «high-poly» модели с помощью комплекса CloudCompare; б) производного облака 3D-гауссианов с помощью интерактивного редактора SuperSplat [17]; в) производного облака 3D-гауссианов с помощью нашего решения; г) производного облака точек (сфер) с помощью решения [10].

На **втором этапе** с помощью комплекса MultiView было исследовано изменение среднего времени синтеза кадра (на примере образца травы из рис. 7) в зависимости от порогового значения  $K_{max}$  на высотах  $h$  наблюдения 0.1, 0.4, 0.8, 1.6 (рис. 8). Высота  $h$  задана в условных единицах (0.1 соответствует высоте образца травы), среднее время измеряется в миллисекундах, а  $K_{max}$  обозначает максимальное число 3D-гауссианов в одном AABB.

На **третьем этапе** с помощью комплекса MultiView была выполнена мультиобъектная визуализация травяного покрова, состоящего из 2000x2000 экземпляров образца травы. Визуализация выполнялась при  $K_{max} = 4$ , базовом цвете оттенка экземпляров (0.5, 0.7, 0.05) и дисперсии r, g, b-компонент базового цвета (0.0, 0.02, 0.01). На рисунке 9 показаны примеры кадров полученной визуализации. В рамках третьего этапа также было измерено среднее времени  $t_{cp}$  синтеза изображений (в мс) при изменении высоты наблюдения от 0.1 до 12.8 (таблица «Среднее время синтеза кадра (массив экземпляров)»).



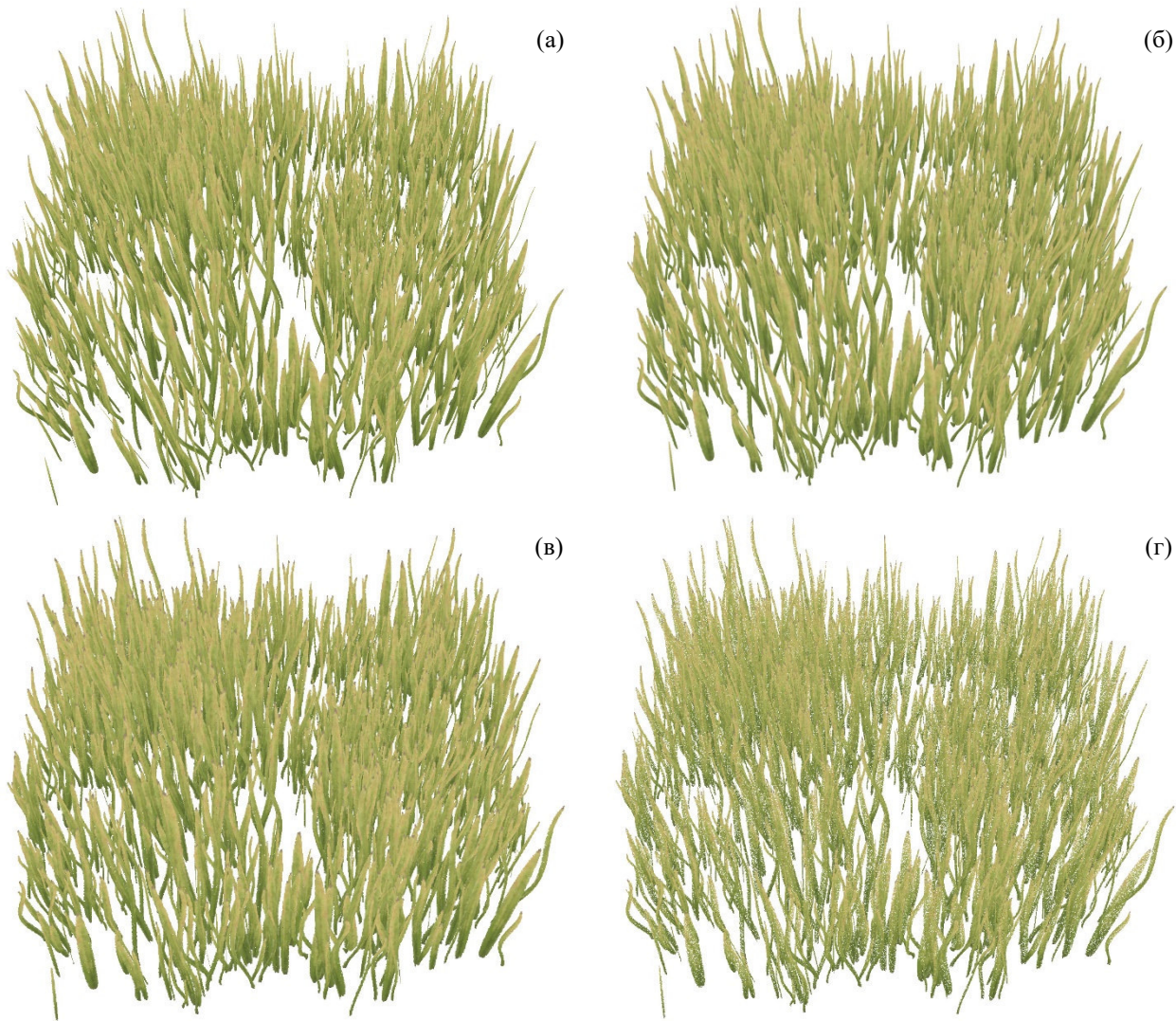


Рисунок 7. Сравнение качества визуализации виртуального образца травы: *а*) эталонная «high-poly» модель, 1085280 полигонов (CloudCompare); *б*) производное облако 3D-гауссианов, 3436498 точек (SuperSplat); *в*) производное облако 3D-гауссианов (наше решение); *г*) производное облако сфер, 3436498 точек (решение [10])

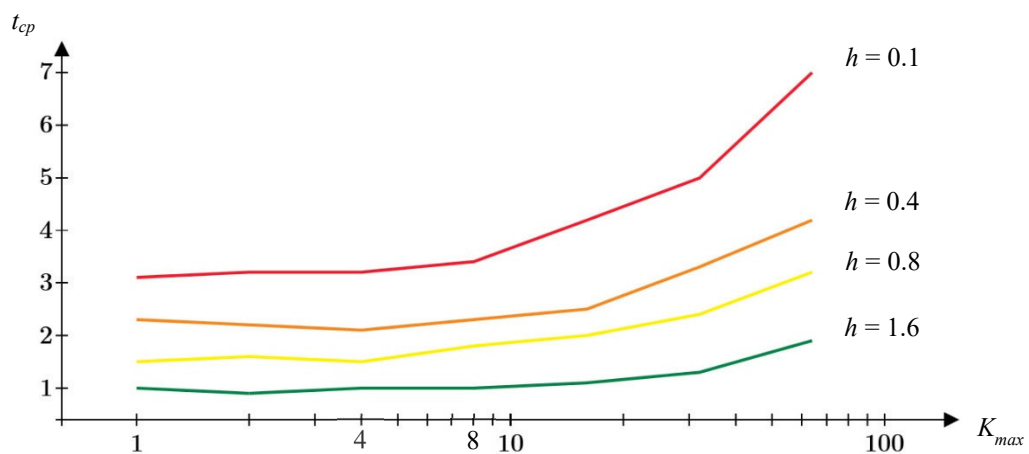


Рисунок 8. Влияние параметра  $K_{max}$  на среднее время  $t_{cp}$  синтеза кадра на различных высотах  $h$  наблюдателя (оценка выполнена для экземпляра образца травы на рис. 7)





Рисунок 9. Примеры кадров мультиобъектной визуализации травяного покрова, состоящего из 4 млн экземпляров, с помощью разработанного решения, основанного на трассируемых облаках 3D-гауссианов

Среднее время синтеза кадра (массив экземпляров), мс

$h$	0.1	0.2	0.4	0.8	1.6	3.2	6.4	12.8
$t_{cp}$	4.6	6.5	8.6	11.0	11.9	12.5	13.1	14.3

#### Обсуждение, выводы и заключение

Из рисунка 7 видно, что редактор SuperSplat и наше решение по качеству визуализации сопоставимы с эталонной моделью, а представление в виде облака сфер, напротив, уступает всем трем решениям ввиду отсутствия информации между сферами. Из этого также следует обратный вывод, что переход на 3D-гауссианы позволяет сократить число элементов в облаке без ущерба качеству визуализации.



Из рисунка 8 видно, что для тестируемого виртуального образца травы баланс между скоростью визуализации и расходом видеопамати (числом AABBs) обеспечивается при значении  $K_{max}$  в диапазоне 4–8. Также в ходе экспериментальной оценки мы попытались на примере этого же виртуального образца сравнить производительность нашего визуализатора (аппаратно-ускоренная трассировка лучей) и визуализатора, входящего в комплекс Mesh2Splat (шейдерная растеризация). Однако в ходе этого испытания было обнаружено, что как только в кадр попадают все 3D-гауссианы, производительность второго визуализатора резко падает до значений 84–93 мс на кадр, независимо от высоты наблюдения. Из этого можно сделать вывод, что для задачи отображения облаков 3D-гауссианов в реальном времени целесообразно использовать аппаратно-ускоренную трассировку лучей, которая более избирательно тратит вычислительные ресурсы GPU и на порядок выигрывает в производительности по сравнению с графическим конвейером.

Из рисунка 9 и данных таблицы следует, что даже при достаточно плотной расстановке экземпляров травы (не заметны места стыковки экземпляров) наше решение обеспечивает визуализацию в масштабе реального времени (не более 40 мс на кадр). Тем не менее полученные значения  $t_{cp}$  превышают на 30–40 % значения, измеренные нами в работе [10] при визуализации массивов облаков точек схожих размеров, что можно объяснить дополнительными взаимными пересечениями AABBs из-за анизотропии 3D-гауссианов.

В заключение отметим, что переход на 3D-гауссианы позволил существенно повысить общий уровень качества визуализации, по сравнению с нашим предыдущим решением [10], основанным на облаках точек, а также сделал возможным визуализацию двусторонних объектов-образцов с тонкой, изогнутой геометрией. В качестве будущей работы планируется распространить полученные результаты на построение виртуальных экосистем, состоящих из образцов растительности различных типов.

#### Финансирование

Публикация выполнена в рамках государственного задания НИЦ «Курчатовский институт» – НИИСИ по теме № FNEF-2024-0002 «Математическое моделирование многомасштабных динамических процессов и системы виртуального окружения».

#### Список литературы

1. UNIGINE: real-time 3D engine. 2005–2025. URL: <https://unigine.com/>.
2. Visualization of Hybrid Virtual Scenes Using Hardware-Accelerated Ray Tracing and Rasterization / Timokhin P.Yu., Mikhaylyuk M.V. // Scientific Visualization. 2024. Vol. 16, no. 3. P. 60–70. <https://doi.org/10.26583/sv.16.3.06>.
3. Immersive landscapes: modelling ecosystem reference conditions in virtual reality / Chandler T., Richards A.E., Jenny B., et al. // Landscape Ecology. 2022. Vol. 37. P. 1293–1309. <https://doi.org/10.1007/s10980-021-01313-8>.
4. Санжаров В.В., Фролов В.А., Галактионов В.А. Исследование технологии Nvidia RTX // Программирование. 2020. № 4. С. 65–72. <https://doi.org/10.31857/S0132347420030061>.
5. NVidia RTX Blackwell GPU Architecture // NVIDIA Corporation. 2025. URL: <https://images.nvidia.com/aem-dam/Solutions/geforce/blackwell/nvidia-rtx-blackwell-gpu-architecture.pdf>.
6. Rusch M., Bickford N., Subtil N. Introduction to Vulkan Ray Tracing // Ray Tracing Gems II. NVIDIA. 2021. P. 213–255. [https://doi.org/10.1007/978-1-4842-7185-8\\_16](https://doi.org/10.1007/978-1-4842-7185-8_16).
7. Lefrançois M.-K. Intersection Shader // NVIDIA Vulkan Ray Tracing Tutorials. 2020–2023. URL: [https://github.com/nvpro-samples/vk\\_raytracing\\_tutorial\\_KHR/tree/master/ray\\_tracing\\_intersection](https://github.com/nvpro-samples/vk_raytracing_tutorial_KHR/tree/master/ray_tracing_intersection).
8. Тимохин П.Ю., Михайлюк М.В. Метод упорядочивания облаков точек для визуализации на конвейере трассировки лучей // Программирование. 2024. № 3. С. 42–53. <https://doi.org/10.31857/S0132347424030054>.
9. Смирнов Л.М., Фролов В.А., Волобой А.Г. Анализ производительности методов обхода двухуровневых BVH-деревьев в трассировке лучей на графических процессорах // Материалы 34-й Международной конференции по компьютерной графике и машинному зрению (Графикон 2024). 2024. С. 147–163. <https://doi.org/10.25206/978-5-8149-3873-2-2024-147-163>.
10. Timokhin P. Y., Mikhaylyuk M. V. Multiobject Visualization of Vast Forests in Virtual Environment Systems // Programming and Computer Software. 2025. Vol. 51, no. 3. P. 198–206. <https://doi.org/10.1134/S0361768825700082>.
11. Lefrançois M.-K. Ray tracing instances // NVIDIA DesignWorks. Vulkan Ray Tracing Tutorials. 2020–2024. URL: [https://github.com/nvpro-samples/vk\\_raytracing\\_tutorial\\_KHR/tree/master/ray\\_tracing\\_instances](https://github.com/nvpro-samples/vk_raytracing_tutorial_KHR/tree/master/ray_tracing_instances).
12. 3D Gaussian Splatting for Real-Time Radiance Field Rendering / Kerbl B., Kopanas G., Leimkuehler T., Drettakis G. // ACM Transactions on Graphics (TOG). 2023. Vol. 42, no. 4. Article no. 139. P. 1–14. <https://doi.org/10.1145/3592433>.

13. Sloan P.J., Kautz J., Snyder J.M. Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments // ACM Transactions on Graphics (TOG). 2002. Vol. 21, iss. 3. P. 527–536. <https://doi.org/10.1145/566654.566612>.
14. Ramamoorthi R., Hanrahan P. An efficient representation for irradiance environment maps // SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 2001. P. 497–500. <https://doi.org/10.1145/383259.383317>.
15. Spec-Gaussian: Anisotropic View-Dependent Appearance for 3D Gaussian Splatting / Yang Z., Gao X., Sun Y., et al. // ArXiv, abs/2402.15870. 2024. <https://doi.org/10.48550/arXiv.2402.15870>.
16. KIRI Engine – 3DGS Render Blender Addon. 2024-2025. URL: <https://www.kiriengine.app/blender-addon/3dgs-render>.
17. SuperSplat – 3D Gaussian Splat Editor. 2023-2025. URL: <https://supersplat.at/editor>.
18. Scolari S. Mesh2Splat: Fast mesh to 3D Gaussian splat conversion. 2025. URL: <https://github.com/electronicarts/mesh2splat>
19. Zhou H. Rendering a Perfect Sphere in OpenGL, then an Ellipsoid. 2023. URL: <https://blog.42yeah.is/rendering/opengl/2023/12/18/perfect-ellipsoid.html>.
20. Sjöholm J. Best Practices for Using NVIDIA RTX Ray Tracing (Updated) // NVIDIA Technical Blog. Jul 25, 2022. URL: <https://developer.nvidia.com/blog/best-practices-for-using-nvidia-rtx-ray-tracing-updated/>.
21. Barnes T. Exact Bounding Boxes for Spheres/Ellipsoids. 2014. URL: [https://tavian.dev/2014/ellipsoid\\_bounding\\_boxes.html](https://tavian.dev/2014/ellipsoid_bounding_boxes.html).
22. Lengyel E. Mathematics for 3D Game Programming and Computer Graphics (Third Edition). Boston, MA: Course Technology PTR, 2012. 624 p.
23. CloudCompare. 3D point cloud and mesh processing software. Open Source Project, 2004-2024. URL: <http://www.cloudcompare.org/>.
24. Blender – The Free and Open Source 3D Creation Suite. 1998-2025. URL: <https://www.blender.org/>.

## Многослойная визуализация дисплея пилота гражданского воздушного судна

Б. Х. Барладян, Н. Б. Дерябин, А. Г. Волобой, Л. З. Шапиро, Е. Ю. Денисов, В. А. Галактионов

Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** Создание системы визуализации дисплея пилота является непростой задачей. С одной стороны, она должна работать под операционной системой реального времени со множеством присущих ей ограничений. С другой стороны, она должна обеспечивать приемлемое время визуализации информации, подготовленной различными разделами бортовой авионики. В данной работе рассматриваются вопросы обеспечения эффективной и безопасной генерации изображений из нескольких разделов операционной системы реального времени (ОСРВ) на один экран. Предполагается использование библиотеки стандарта OpenGL SC 2.0. Нам удалось предложить и реализовать эффективное решение для задачи одновременного вывода на экран информации из нескольких разделов. При этом мы смогли обеспечить приемлемую скорость визуализации 25 кадров в секунду.

**Ключевые слова:** дисплей кабины пилота, бортовая система визуализации, многооконный дисплей, OpenGL SC, ОСРВ, авионика.

## Layered Visualization of a Civil Aircraft Pilot Display

B. Kh. Barladian, N. B. Deryabin, A. G. Voloboy, L. Z. Shapiro, E. Yu. Denisov, V. A. Galaktionov

Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

**Abstract.** Creating a pilot display visualization system is a challenging task. On the one hand, it must operate under a real-time operating system with its many inherent limitations. On the other hand, it must provide an acceptable time for rendering of information prepared by various sections of the onboard avionics. The paper considers issues of ensuring efficient and safe generation of images from several sections of the real-time operating system (RTOS) onto a single screen. It is assumed that the OpenGL SC 2.0 standard library is used. We propose and implement an efficient solution for the task of simultaneously displaying information from several sections on the screen. At the same time, we are able to provide an acceptable visualization speed of 25 frames per second.

**Keywords:** cockpit display, on-board visualization system, multi-window display, OpenGL SC, RTOS, avionics.

### Введение

Современные комплексы бортового оборудования (КБО) проектируются в соответствии с идеологией, известной как «Интегрированная модульная авионика» (ИМА) [1-4]. Эта идеология используется в лайнерах Airbus A320, Boeing 787, Superjet 100 и MC-21. Принципиальным вопросом при разработке бортовой авионики является то, что она относится к системам критическим с точки зрения безопасности (safety critical). Эта специфика не позволяет применять готовые, известные решения для той или иной функциональности. В силу этой специфики КБО должны быть сертифицированы для использования на гражданских авиалайнерах. Сертификация, в частности, требует организации процесса разработки в соответствии с требованиями стандарта DO-178C (в русскоязычной редакции – КТ-178C) [5]. Ключевой особенностью ИМА является возможность исполнения нескольких функциональных приложений (реализующих программную часть той или иной самолетной системы) на одном вычислителе. Необходимым условием при этом является разделение приложений по времени исполнения и доступным ресурсам, т.е. ограничение вытесняющей многозадачности и контроль доступа к памяти для нескольких приложений. Такой режим работы приложений обеспечивается операционной системой реального времени (ОСРВ), что делает ОСРВ неотъемлемой и важнейшей частью любого современного комплекса бортового оборудования. Программный интерфейс ОСРВ, предназначенный для использования на борту воздушного судна, должен соответствовать стандарту ARINC 653 [6].

Для обеспечения безопасности каждое из функциональных приложений должно выполняться в отдельном разделе ОСРВ так, чтобы возможные ошибки в одном разделе имели бы минимальное влияние на работу приложений, работающих в других разделах. Большая часть этих требований обеспечивается ОСРВ, соответствующей стандарту ARINC 653. Однако у графических приложений есть дополнительная специфика, когда на один многофункциональный дисплей одновременно выводится различная информация, генерируемая многочисленными независимыми системами управления полетом.

В общем случае выводимая информация может выводиться как в различные окна на дисплее, так и в общие для различных приложений окна. Для удовлетворения требований по безопасности авиационные приложения используют специальные версии графических библиотек стандартов OpenGL SC. Различные подходы к разработке таких библиотек можно найти в работах [7-9].

В наших исследованиях использовалась OCPB JetOS [4] и собственная реализация графической библиотеки стандарта OpenGL SC 2.0 с использованием аппаратного ускорения на базе пакета MESA [10].

### Постановка задачи

Задача визуализации из нескольких разделов в различные окна одного дисплея, так называемой многооконной визуализации, рассматривалась в работах [11-13]. В этих работах изображения генерировались независимо в каждом разделе, а затем компоновщик, который работал в отдельном разделе, компоновал их в заданных окнах экрана, используя библиотеку кадрового буфера. Однако такой подход не может быть осуществлен при использовании OpenGL с аппаратным ускорением, поскольку в типичной авиационной платформе имеется только один графический процессор и мы не можем использовать отдельные экземпляры OpenGL в каждом разделе. В силу требований безопасности и специфики JetOS графический драйвер, реализующий OpenGL SC, может работать только в одном разделе OCPB.

Следует также отметить, что подход с использованием компоновщика позволяет реализовать только так называемый «многооконный» режим, когда каждое приложение может выводить информацию только в свое окно. На практике, однако, бывает необходимость в некоторых случаях выводить информацию из нескольких приложений в общее окно на дисплее. Поэтому для реализации возможности синтеза графики из нескольких разделов с использованием аппаратного ускорения в OpenGL мы предложили специальный подход, когда все команды OpenGL выполняются в одном специальном разделе операционной системы JetOS. Первые шаги по реализации нашего подхода были сделаны в работе [10] для реализации библиотеки OpenGL SC 1.01 на платформе i.MX6 с GPU Vivante.

### Предлагаемое решение

Работу приложения, использующего библиотеку OpenGL для визуализации, можно разделить на две компоненты:

- 1) подготовка различных данных, необходимых для работы OpenGL: геометрия и ее параметры, различные атрибуты и т.д.;
- 2) непосредственный вызов функций OpenGL с подготовленными параметрами.

Для использования аппаратного ускорения в OpenGL в нашей реализации эти компоненты выполняются в разных разделах операционной системы JetOS. Каждое приложение, подготавливающее данные, выполняется в своем разделе (рисующие разделы). Оно не осуществляет реальные вызовы функций OpenGL, а записывает индексы необходимых для визуализации функций OpenGL с соответствующими параметрами в специальный буфер (массив). Этот массив находится в общей памяти с разделом, который и вызывает функции библиотеки OpenGL (растеризующий раздел). Когда вся информация, необходимая для генерации изображения одного кадра, подготовлена рисующими разделами, то начинает работать растеризующий раздел.

Для работы системы нам в первую очередь необходимо передать данные и команды OpenGL, сформированные в рисующих разделах в растеризующий раздел, где эти команды будут реально выполняться. Для этих целей мы разработали две специальные библиотеки для рисующих и растеризующего разделов: liboglout2 и liboglin2 соответственно. Суффикс 2 в именах этих библиотек означает, что они предназначены для работы с командами OpenGL стандарта OpenGL SC 2.0. В работе [10] рассматриваемый подход первоначально разрабатывался для стандарта OpenGL SC 1.1 и там библиотеки назывались соответственно liboglout1 и liboglin1. С точки зрения предложенной нами технологии разница между стандартами OpenGL SC 1.1 и 2.0 в первую очередь определяется разными наборами функций OpenGL. Кроме того, для стандарта OpenGL SC 2.0 необходимо поддерживать возможность использования нескольких шейдерных программ и шейдерных объектов типа UNIFORM даже в одном разделе OCPB.

Библиотека `liboglout2` содержит набор функций, эмулирующих соответствующие функции стандарта OpenGL SC 2.0, только вместо их выполнения она записывает их идентификаторы в участок общей памяти. В эту же общую память записываются также все параметры и данные, необходимые для выполнения этих функций.

Эти эмулирующие функции относительно просты и имеют ту же нотификацию, что и соответствующие им функции OpenGL. Если такая функция имеет фиксированное число параметров, то в текущую позицию массива записывается индекс вызываемой функции, а за ним подряд значения параметров. Это относится к большинству функций OpenGL: таких, как, например, `glClearColor()`, `glClearDepthf()`, `glClearStencil()`, `glUniform1i()` и т.д. Этот же способ используется тогда, когда передаются массивы известной длины. В более сложных случаях, когда передаются указатели на массивы с неопределенной в момент вызова длиной (например, при вызове `glVertexAttribPointer()`), технология кодирования вызова несколько усложняется. В момент вызова таких функций указатель запоминается, а данные по этому указателю используются позже при вызове функций `glDrawArrays()` или `glDrawRangeElements()`, когда необходимые размерности становятся известными. При этом массивы с запомненными указателями (как правило, координаты вершин, или их текстурные координаты) копируются в зарезервированную область общей памяти. Использоваться они будут в библиотеке `liboglin2` при выполнении реальных функций OpenGL. Библиотека `liboglout2` связывается (линкуется) с рисующим разделом вместо реальной OpenGL SC 2.0. Сам код раздела при этом практически не изменяется. Необходимы только определенные изменения в конфигурационных файлах и некоторый код в `main()`, который пишется по общему шаблону.

Декодирование информации, записанной приложениями с помощью библиотеки `liboglout2`, в вызовы функций OpenGL осуществляется в одном растеризующем разделе с помощью библиотеки `liboglin2`. Оно выполняется последовательно для каждого пакета данных, сгенерированного соответствующими приложениями. Библиотека `liboglin2` состоит из одной функции `ProcessOGL_Input()`, на вход которой подается массив, записанный в рисующем разделе. Она последовательно читает данные из массива, по прочитанному индексу функции OpenGL переходит к участку кода, написанного для данной функции, извлекает из последующих элементов массива параметры функции, включая значения указателей, и вызывает с этими параметрами заданную функцию OpenGL. Библиотека `liboglin2` связывается (линкуется) с растеризующим разделом. Этот раздел также связывается (линкуется) с библиотекой OpenGL SC 2.0. На рисунке 1 представлена схема синтеза графики из нескольких разделов с учетом использования библиотек `liboglout2` и `liboglin2`.

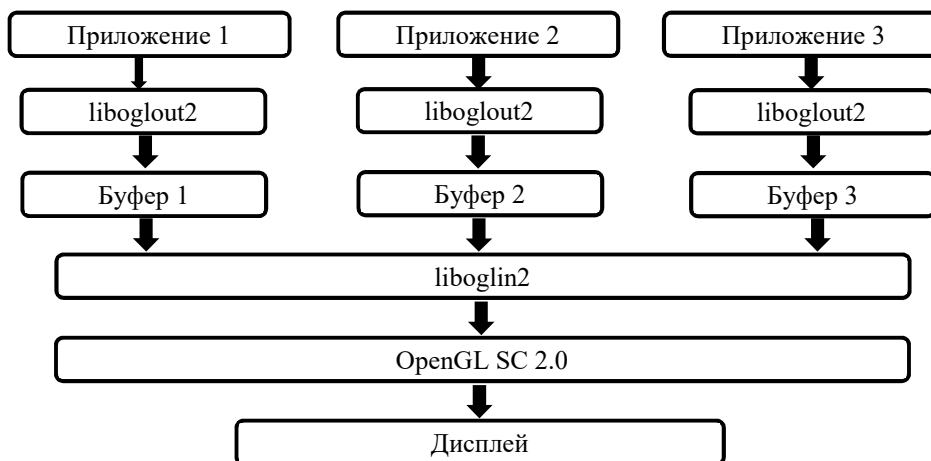


Рисунок 1. Схема работы приложений с использованием библиотек `liboglout2` и `liboglin2`

Проблемы в этой схеме возникают при использовании нескольких шейдерных программ, текстур и шейдерных объектов, создаваемых в разных рисующих разделах. Реальное выполнение команд OpenGL происходит в растеризующем разделе, где и назначаются (либо генерируются) идентификаторы таких объектов. Рисующие разделы не имеют обратной связи с растеризующим разделом и не могут узнать текущий идентификатор объекта. Идентификаторы объектов одного типа

должны быть уникальными и различаться, если они созданы в разных разделах. Кроме того, следует учитывать, что в общем случае код может быть сгенерирован разными версиями пакета SCADe (с помощью которого типично генерируются программы для авионики) или написан вручную. Мы должны учитывать эту проблему при использовании идентификаторов следующих объектов:

- идентификаторы шейдерных программ, которые определяются вызовом функции `glCreateProgram()`;
- идентификаторы шейдерных объектов типа UNIFORM, которые определяются вызовом функции `glGetUniformLocation()`;
- идентификаторы шейдерных объектов типа VERTEX ATTRIBUTE, которые определяются вызовом функции `glGetAttribLocation()`;
- идентификаторы текстур, которые определяются вызовом функции `glGenTextures()`.

Для решения этой проблемы в библиотеке `liboglout2` используется внутреннее собственное определение этих идентификаторов. Когда в `liboglout2` необходимо использовать реальное значение идентификатора при вызове функции OpenGL, то оно определяется с помощью соответствующей таблицы по внутреннему значению идентификатора библиотеки `liboglout2`. Экземпляры библиотек `liboglout2` в каждом разделе используют собственное адресное пространство, поэтому для последовательной нумерации идентификаторов одного типа в разных разделах используются общие переменные – указатели идентификаторов каждого типа, хранящиеся в общей памяти рисующих разделов.

```
static UINT32* prog_ident_ptr;
static UINT32* tex_ident_ptr;
static UINT32* uniform_ident_ptr;
static UINT32* attrib_ident_ptr;
```

Для связи этих переменных с библиотекой `liboglout2`, прилинкованной к данному конкретному рисующему разделу, указатели устанавливаются в библиотеку `liboglout2` с помощью соответствующих функций:

```
VOID SetProgramIdentPtr(GLuint* identPtr)
{
    sProgIdentPtr = identPtr;
}
VOID SetTextureIdentPtr(GLuint* identPtr)
{
    sTexIdentPtr = identPtr;
}
VOID SetUniformIdentPtr(GLuint* identPtr)
{
    sUniformIdentPtr = identPtr;
}
VOID SetAttribIdentPtr(UINT32* identPtr)
{
    sAttribIdentPtr = identPtr;
}
```

Эта установка должна быть сделана до первого обращения раздела к функциям OpenGL. При вызове в рисующем разделе функции OpenGL для получения идентификатора объекта библиотека `liboglout2` возвращает значение по нужному из указанных указателей и увеличивает это значение на 1. Такой подход гарантирует, что значения идентификаторов объектов одного типа не будут повторяться при их получении из разных рисующих приложений. В библиотеке `liboglout2` для определения идентификатора объекта данного типа имеется массив, для которого идентификатор из `liboglout2` рассматривается как индекс, значение массива по которому равно идентификатору, назначенному библиотекой OpenGL.

В каждом разделе теперь потенциально может использоваться собственный программный объект, один или несколько. Эти программные объекты могут создаваться явно с помощью вызова функции



`glCreateProgram()` или неявно внутри библиотеки OGLX пакета SCADe. Предложенный подход в некотором смысле автоматически поддерживает оба варианта создания программных объектов. Перед вызовом функций OpenGL необходимо установить используемый программный объект функцией `glUseProgram(sProgIdent)`. Переменная `sProgIdent` объявлена в библиотеке `liboglout2`. Она содержит идентификатор программного объекта, используемого разделом. Если использовался пакет SCADe, то это единственный способ получить этот идентификатор. В случае явного вызова функции `glCreateProgram()` идентификатор программного объекта будет возвращен ею. Идентификаторы текстур, объектов типа UNIFORM и типа VERTEX ATTRIBUTE используются обычным образом при прямом написании кода раздела, использующего функции OpenGL. В случае использования пакета SCADe в библиотеке OGLX для идентификаторов будут автоматически использоваться внутренние переменные библиотеки `liboglout2`.

### Результаты тестов

На рисунке 2 приведен результат генерации изображений из пяти разделов, одновременно передающих информацию для вывода на экран. Каждый раздел использует один или два собственных программных объекта. В левой верхней части экрана приведено изображение, сгенерированное приложением, подготовленным с помощью пакета SCADe R16 с кодогенератором KCG 6.4.3. В правой верхней части экрана отображается изображение, сгенерированное приложением «Карта перемещения по аэропорту» (AMM Airport Moving Map), написанным по стандарту OpenGL SC 2.0 без использования пакета SCADe. Слева внизу – приложение, использующее программный объект из пакета SCADe R19 с кодогенератором KCG 6.7.5, но не использующее пакет для генерации кода. В центре внизу расположено аналогичное приложение, использующее программный объект из пакета SCADe R16 с другими текстурами. Справа внизу изображение, сгенерированное приложением, которое использует два программных объекта – один из пакета SCADe R16, а второй из пакета SCADe R19. Скорость визуализации изображений из пяти разделов – 25 кадров в секунду.

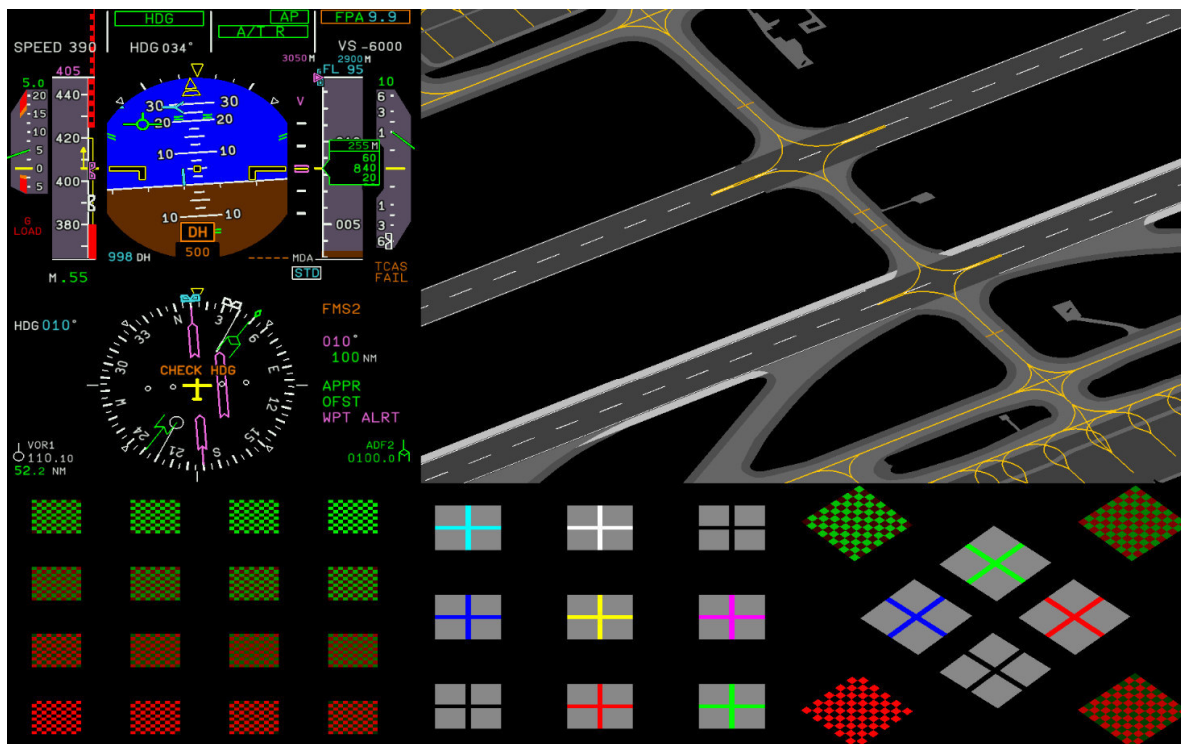


Рисунок 2. Генерация изображений из пяти разделов с использованием разных программных и текстурных объектов

Верхние изображения на рисунке 2 имеют также и смысловое значение для пилота: слева показан основной дисплей полета (Primary flight display), а слева – взлетно-посадочные полосы и рулежные дорожки аэропорта. Однако для тестирования нашего подхода нам было важно, что все пять разделов

генерировали изображения различными средствами, используемыми при разработке приложений для бортового программного обеспечения.

На рисунке 3 приведен результат генерации изображений из трех разделов.

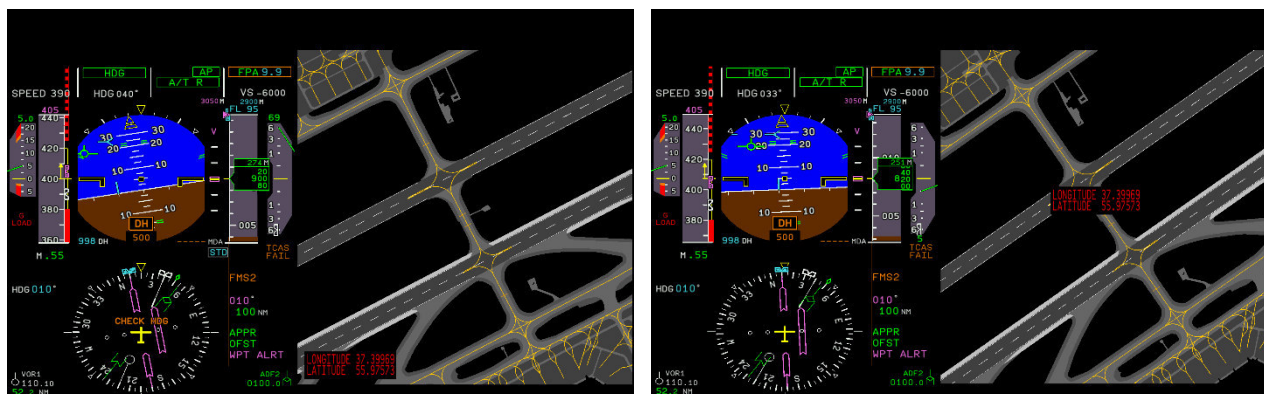


Рисунок 3. Генерация изображений из трех разделов

Левая и правая части рисунка 3 нам уже знакомы – это основной дисплей полета и карта перемещения по аэропорту. Они генерируются в первых двух разделах. Третий раздел добавляет надпись поверх изображения, генерируемого вторым разделом. Положение надписи может варьироваться в зависимости от состояния оборудования самолета или его положения в пространстве, что и показано на двух частях рисунка, соответствующих разным кадрам.

### Заключение

Создание системы визуализации дисплея пилота является непростой задачей. С одной стороны, она должна работать под операционной системой реального времени со множеством присущих ей ограничений. Также она должна быть создана таким образом, чтобы ее можно было сертифицировать (иначе она не может быть использована в бортовом оборудовании). С другой стороны, она должна обеспечивать приемлемое время визуализации информации, подготовленной различными разделами бортовой авионики.

Нам удалось предложить и реализовать эффективное решение для задачи одновременного вывода на экран информации из нескольких разделов. При этом программы приложений, работающих в этих разделах, могут быть написаны с помощью различных средств: как с использованием разных версий пакета SCADe, так и без него. Мы смогли обеспечить при этом приемлемую скорость визуализации 25 кадров в секунду.

### Список литературы

1. Федосов Е.А., Проект создания нового поколения интегрированной модульной авионики с открытой архитектурой // Полет. 2008. № 8. С. 15-22.
2. Применение операционных систем реального времени в интегрированной модульной авионике / Федосов Е.А., Ковернинский И.В., Кан А.В., Солоделов Ю.А. OSDAY 2015. URL: <http://osday.ru/solodelov.html>.
3. Федосов Е.А. Косьянчук В.В., Сельвесюк Н.И. Интегрированная модульная авионика // Радиоэлектронные технологии. 2015. №1. С. 66-71.
4. Солоделов Ю.А., Горелиц Н.К., Сертифицируемая бортовая операционная система реального времени JetOS для российских проектов воздушных судов // Труды ИСП РАН. 2017. Т. 29, № 3. С. 171-178. DOI: 10.15514/ISPRAS-2017-29(3)-10.
5. DO-178C Software Considerations in Airborne Systems and Equipment Certification. URL: [http://www.rtca.org/store\\_product.asp?prodid=803](http://www.rtca.org/store_product.asp?prodid=803).
6. Avionics application software standard interface (ARINC 653). SAE-ITC, 2015.
7. Baek N., Kim K.J. Design and implementation of OpenGL SC 2.0 rendering pipeline // Cluster Comput. 2019. Vol. 22. P. 931–936. URL: <https://doi.org/10.1007/s10586-017-1111-1>.
8. Baek N., Lee H., OpenGL ES 1.1 Implementation Based on OpenGL // Multimedia Tools and Applications. 2012. Vol. 57, iss. 3. P. 669–685.
9. Yoon J., Baek N., Hwanyong L. Graphics Rendering Based on OpenVG and Its Use Cases with Wireless Communications. Wireless Personal Communications. 2017. Vol. 94, iss. 2. P. 175–185. doi: 0.1007/s11277-015-3163-y.

- 
10. Specifics of the Development of an On-Board Visualization System for Civil Aircrafts / B.Kh. Barladian, N.B. Deryabin, A.G. Voloboy, et al. // Programming and Computer Software. 2024. Vol. 50, iss. 3. P. 215-223. DOI: 10.1134/S0361768824700014.
  11. A Safety Critical Compositor for OpenGL SC 1.0.1 and OpenGL SC 2.0. URL: [http://www.coreavi.com/sites/default/files/compositor\\_whitepaper\\_final.pdf](http://www.coreavi.com/sites/default/files/compositor_whitepaper_final.pdf).
  12. EGL\_EXT\_compositor. URL: [http://www.coreavi.com/sites/default/files/coreavi\\_product\\_brief\\_egl\\_ext\\_compositor.pdf](http://www.coreavi.com/sites/default/files/coreavi_product_brief_egl_ext_compositor.pdf).
  13. High speed visualization in the JetOS aviation operating system using hardware acceleration / B.Kh. Barladian, N.B. Deryabin, A.G. Voloboy, V.A. Galaktionov, L.Z. Shapiro // CEUR Workshop Proceedings. 2020. Vol. 2744. P. 107:1-107:9. DOI: 10.51130/graphicon-2020-2-4-3.

## Верификация системы оптического моделирования Lumicept по тестовым сценариям CIE 171:2006

Е. Д. Бирюков<sup>1</sup>, М. С. Копылов<sup>1</sup>, В. Г. Соколов<sup>2</sup>, И. С. Потемин<sup>2</sup>, А. Г. Волобой<sup>1</sup>

<sup>1</sup>Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

<sup>2</sup>Национальный исследовательский университет ИТМО, Санкт-Петербург, Россия

**Аннотация.** Данная статья описывает прохождение проверки системы оптического моделирования Lumicept, вычислительное ядро которой написано специалистами ИПМ и ИТМО, с помощью тестовых сценариев, разработанных Международной комиссией по освещению (CIE) в рамках стандарта CIE 171:2006. В статье кратко рассматриваются другие известные методы верификации систем оптического моделирования, используемых в различных инженерных и научных расчетах. Представлен обзор существующих подходов, включая сравнение с натурными измерениями и использование стандартизированных бенчмарков. Также описаны особенности алгоритмов моделирования в системе Lumicept и их влияние на процедуру тестирования и результаты. Представлены результаты верификации системы Lumicept с использованием тестовых сцен CIE, продемонстрировавшие успешное прохождение тестов и конкурентоспособную точность результатов. В ходе анализа выявлены возможные причины отклонений от эталонных значений в отдельных тестах, что указывает на необходимость дальнейших исследований в области методов верификации и уточнения эталонных данных для спорных тестовых сценариев.

**Ключевые слова:** оптическое моделирование, симуляция освещения, физически корректный рендеринг, трассировка лучей, верификация программного обеспечения, тесты оптического моделирования, Международная комиссия по освещению, CIE 171:2006

## Validation of Lumicept optical modeling system against CIE 171:2006 test cases

E. D. Birukov<sup>1</sup>, M. S. Kopylov<sup>1</sup>, V. G. Solokov<sup>2</sup>, I. S. Potyomin<sup>2</sup>, A. G. Voloboy<sup>1</sup>

<sup>1</sup>Keldysh Institute of Applied Mathematics, Moscow, Russia

<sup>2</sup>ITMO University, St. Petersburg, Russia

**Abstract.** This paper describes the validation of the Lumicept optical modeling software against test cases developed by the International Commission on Illumination (CIE) within the framework of the CIE 171:2006 standard. The paper briefly reviews other known methods for assessment of optical modeling systems used in various engineering and scientific calculations. An overview of<sup>1</sup> existing approaches is provided, including comparison with natural measurements and usage of standardized benchmarks. The features of the simulation algorithms in the Lumicept system and their impact on the testing procedure and its results are also described. The results of validation of the Lumicept system against CIE test cases are presented, demonstrating successful completion of the tests and competitive accuracy of the results. The analysis revealed possible causes of deviations from the reference values in individual tests, which indicates the need for further research in the field of validation methods and refinement of the reference data for controversial test scenarios.

**Keywords:** optical modeling, lighting simulation, physically accurate rendering, raytracing, software validation, optical modeling test cases, International commission on Illumination, CIE 171:2006.

### Введение

Современные методы моделирования распространения света находят широкое применение в разнообразных инженерных и научных дисциплинах. Методы, основанные на принципах физической оптики, позволяют создавать виртуальные модели, точно воспроизводящие взаимодействие света с различными объектами и средами. Одним из наиболее известных и востребованных методов для расчета освещенности сейчас является трассировка лучей. Вторым широко распространенным методом является метод конечных элементов (Radiosity), который, к примеру, был положен в основу широко известного пакета Radiance.

Ключевым преимуществом данных методов является их способность обеспечивать физически корректные результаты симуляций. Это делает их незаменимыми в тех областях, где требуется высокая точность при вычислениях яркости, освещенности, спектрального состава света и других физических величин. Примерами таких задач могут служить разработка современных осветительных приборов, где требуется оптимизация светового потока и минимизация потерь, проверка соответствия производственных и офисных помещений санитарным нормам по освещенности и т.д.

Однако следует отметить, что корректность моделирования в значительной степени определяется рядом факторов, включающих особенности конкретной реализации выбранного метода, сложность геометрии моделируемой сцены, поддержку различных свойств поверхностей и сред, а также правильность задания источников света, включая источники со сложным пространственным распределением параметров. В частности, для точного моделирования естественного освещения необходимо учитывать особенности модели неба, используемой в расчетах, поскольку от этого напрямую зависит распределение яркости и спектрального состава света, поступающего в помещение. Например, небесное освещение можно задать с помощью специальной гониометрической диаграммы, которая содержит информацию об угловом распределении яркости такого освещения, а также в некоторых случаях и об угловом распределении его цвета.

С точки зрения конечного пользователя, будь то инженер, занимающийся разработкой осветительных приборов, специалист по эргономике, оценивающий комфортность рабочей среды, или архитектор, стремящийся создать гармоничное и функциональное пространство, вопрос точности и повторяемости результатов симуляции является первоочередным при выборе того или иного пакета оптического моделирования [1]. Именно от корректности получаемых результатов напрямую зависит применимость выбранной программы (системы) оптического моделирования для решения практических задач. Это, в свою очередь, делает задачу верификации вычислительного ядра системы оптического моделирования освещения, то есть проверки соответствия алгоритмов и моделей, лежащих в основе системы, реальным физическим процессам, весьма важной и актуальной.

### **Методы проверки и тестирования программного обеспечения в области оптического моделирования**

С тех пор как системы оптического моделирования стали применяться в практических расчетах в промышленности, возникла потребность каким-либо образом верифицировать результаты их работы. Это особенно важно в тех случаях, когда такие системы применяются при проектировании устройств, дефекты в которых могут повлечь за собой серьезные негативные последствия, например, автомобильных фар, которые в случае ошибок при проектировании будут слепить водителей встречных транспортных средств. Поэтому было разработано большое количество различных методов проверки, тестирования и сравнения точности работы программ оптического моделирования.

Изначально одним из наиболее распространенных подходов к оценке точности программного обеспечения для моделирования освещения являлось прямое сравнение результатов моделирования с данными натурных измерений. Этот метод предполагает сравнение результатов, полученных с использованием различных программных пакетов, с измерениями, проведенными в реальных помещениях или на физических моделях. Такой подход позволяет оценить точность моделирования в конкретных условиях и выявить факторы, которые могут влиять на результаты. Например, в работе [2] авторы использовали фотографию большого динамического диапазона и данные спектрофотометрических измерений неба в реальных условиях для оценки точности различных моделей неба, используемых при симуляции освещения в библиотеке libRadtran.

Аналогичные исследования проводились и для других программных пакетов, таких как Lark [3], Evalglare [4], INLUX [5]. Это позволило получить более полное представление о сильных и слабых сторонах различных инструментов моделирования освещения и выявить факторы, влияющие на точность результатов.

С развитием технологий и накоплением опыта в области моделирования освещения возникла потребность в стандартизации методов проверки. Это привело к разработке различных бенчмарков, представляющих собой специализированные наборы данных и методик для оценки точности программного обеспечения.

Одним из примеров подобных бенчмарков является BRE-IDMP (Building Research Establishment – International Daylight Measurement Programme), разработанный в 2000 году [6]. Этот бенчмарк представляет собой специализированный набор данных, включающий 3D-модели офисных помещений с различными типами оконных проемов, материалами отделки и расположением мебели, что позволяет имитировать разнообразные сценарии освещения. В отличие от физических моделей,

которые, как показывают исследования, склонны к переоценке освещенности, особенно при пасмурном небе, BRE-IDMP предоставляет данные, полученные в полноразмерных моделях офисных помещений, что повышает надежность результатов. Предоставляя измеренные или аналитически рассчитанные значения освещенности, BRE-IDMP служит стандартизированной основой для сравнения результатов, полученных с использованием различных программных пакетов, и в результате способствует повышению достоверности результатов проектирования освещения.

Другим инструментом для валидации систем оптического моделирования является бенчмарк, разработанный Шрегле и Винольдом в 2004 году [7]. Ключевой особенностью этого бенчмарка является акцент на точной передаче свойств поверхностей за счет использования двунаправленных функций отражательной способности (Bidirectional Reflectance Distribution Function – BRDF). В отличие от упрощенных моделей, использующих однородные коэффициенты отражения, BRDF позволяют учитывать зависимость отражения света от угла падения и угла наблюдения, что особенно важно для материалов с выраженной текстурой или глянцевой поверхностью. Бенчмарк Шрегле и Винольда включает в себя как геометрические модели помещений, так и детальные измерения BRDF для различных материалов, полученные с использованием гониофотометра, а также методы устранения артефактов при измерении BRDF и оценки погрешностей, связанных с использованием BRDF, полученных посредством передискретизации. Валидация, проведенная авторами с использованием методов прямой и обратной трассировки лучей (Photon map и Radiance соответственно), подтвердила, что оба алгоритма решают уравнение рендеринга со сравнимой точностью и что полученные результаты соответствуют измерениям в пределах допустимых погрешностей.

Можно также упомянуть бенчмарк, разработанный при содействии компании Autodesk [8] и изначально предназначенный для тестирования алгоритмов оптического моделирования, также разработанных в данной компании. Этот бенчмарк специализируется на расчетах освещенности помещения естественным светом через окна. Как и предыдущие упомянутые бенчмарки, этот бенчмарк использует в качестве эталона данные, полученные путем реальных измерений некоторых существующих помещений.

Однако на настоящий момент наиболее популярным и признанным бенчмарком является набор аналитических тестов, разработанный под руководством F. Maamarí по заданию Международной комиссии по освещению (CIE) [9], получивший в конечном варианте наименование CIE 171:2006. Особенностью этого бенчмарка является тот факт, что часть его тестов основана на сравнении результатов моделирования не с реальными измерениями, а с эталонными значениями, полученными аналитическим путем. Это позволяет сфокусироваться на проверке соблюдения физических законов освещения в программном обеспечении.

Основной целью данного набора тестов является оценка корректности реализации различных аспектов оптического моделирования в программном обеспечении, таких как корректное задание характеристик источников света, точный расчет прямых и многократных переотражений света от различных поверхностей, и правильный учет влияния геометрии помещения на распределение света. Особое внимание в данном бенчмарке уделяется сложным сценариям, представляющим наибольшую трудность для корректного моделирования, например освещению внутреннего помещения небесной полусферой через небольшое окно. Анализ результатов, полученных для различных программ, показывает, что высокая точность в отношении одного аспекта моделирования не гарантирует точность в других аспектах. Аналитический метод тестирования, то есть использование в качестве эталонных результатов не реальных измерений, а значений, полученных с помощью аналитических вычислений, является особенно важным при валидации моделей дневного освещения, поскольку яркость при дневном освещении зависит от большого количества факторов, в том числе случайных, например количества и расположения облаков, и поэтому добиться повторяемости экспериментальных измерений при дневном освещении очень сложно.

Все эти особенности тестовой модели CIE 171:2006 подчеркивают ее ценность для определения области применимости каждой конкретной программы. Благодаря своей аналитической (или эталонной) природе и акценту на проверке фундаментальных аспектов оптического моделирования, данный бенчмарк является важным инструментом для оценки и сравнения различных программных средств.



Следует отметить, что, несмотря на наличие стандартизированных бенчмарков, сравнение результатов моделирования с данными натурных измерений по-прежнему используется для оценки точности программного обеспечения для моделирования освещения. Однако в рамках данной работы мы сосредоточимся на тестах, разработанных CIE, как наиболее авторитетных и комплексно охватывающих различные аспекты моделирования освещения.

### **Описание тестовых сценариев CIE 171:2006**

Все тестовые сценарии CIE 171:2006 разделены на несколько групп. Первая группа представляет собой набор из 6 простейших экспериментальных тестов (отмеченных индексами 4.1 – 4.6), в которых результаты симуляции в тестируемом программном обеспечении сравниваются с экспериментальными результатами, полученными путем реального воссоздания объектов с заданными свойствами поверхностей, размещения реальных источников освещения и проведения измерений освещенности в заданных точках с помощью физических измерительных приборов. Различия отдельных тестов внутри этой группы сводятся исключительно к разным координатам источников света и несколько отличающимся коэффициентам отражения в свойствах материалов для объектов сцены.

Вторая группа представляет собой так называемые аналитические тесты. Их общее отличие от тестов первой группы сводится к тому, что эталонные значения освещенности были получены не путем реальных физических измерений, а путем аналитических вычислений с использованием базовых формул геометрической оптики. При этом отдельные тесты данной группы предназначены для тестирования самых разных возможностей оптического моделирования, поэтому они имеют большие отличия друг от друга в геометрии объектов, видах и расположениях источников света (искусственные источники света различной формы, а также естественное освещение), свойствах поверхностей (включая не только отражающие, но и преломляющие поверхности), а также в вычисляемых физических величинах, которые включают в себя не только освещенность в заданных точках, но и световой поток, значения светопропускания для материалов поверхностей и т.д. Следует отметить, что по сведениям, полученным от Международной комиссии по освещению, в действительности только в части тестов этой группы (это тесты, отмеченные индексами 5.2–5.8) эталонные значения были получены чисто аналитическим путем. Что касается тестов с индексами 5.9–5.14, то для них эталонные значения были получены с помощью симуляции в существующей системе оптического моделирования, которая успешно прошла другие тесты и поэтому считалась корректной.

Третья группа не содержит конкретных описаний тестовых сцен. В ней содержатся только предложения по сценариям проведения некоторых дополнительных тестов. Эталонные значения для данных сценариев, как и для тестов второй группы, предлагается получать аналитическим путем. При этом главным критерием успешного прохождения теста является симметричность результатов.

CIE разработала общие правила, которые применяются ко всем тестовым сценариям. Они сводятся к максимальному упрощению данных, используемых для тестирования. Геометрия тестовой сцены должна представлять собой простую комнату в форме параллелепипеда без каких-либо внутренних дополнительных объектов. В тестах, рассматривающих дневное освещение, комната должна включать единственное отверстие, через которое дневной свет проникает в нее. В тестах на искусственное освещение количество источников света должно быть минимальным. Свойства поверхностей для стен в комнате должны по возможности обеспечивать только диффузное отражение света, а также желательно, чтобы они были спектрально нейтральными [10].

В настоящей работе мы рассматриваем проведение тестов только из первой группы. Это базовые тесты, и они считаются наиболее надежными. Ранее CIE выдавала официальные сертификаты об успешном прохождении этих тестов компаниям-разработчикам систем оптического моделирования, где упоминались только тесты первой группы.

#### *1. Общее описание тестовых примеров 1 группы*

Тестовый набор состоит из 6 тестов, которые в официальном документе CIE отмечены номерами 4.1–4.6. Как уже было упомянуто, геометрия сцены является одинаковой во всех 6 тестах и представляет собой комнату в форме параллелепипеда с расстояниями между стенами 6.78 м по оси X, 6.72 м по оси Y и высотой от пола до потолка 3.24 м. Также одинаковыми являются координаты точек, в которых производится измерение освещенности. Эти точки расположены на одной плоскости,

параллельной полу и потолку, на расстоянии 0.8м от пола. Точки образуют решетку размерностью 7х7. Различия между тестами сводятся к параметрам материалов стен, пола и потолка комнаты, а также к положению и параметрам источников света.

## 2. Параметры материалов и источников света в тестовых сценах

Для всех 6 тестовых сцен заданы параметры отражательной способности пола, потолка и стен, которые являются постоянными на всей площади соответствующей поверхности. Значения отражательной способности приведены в таблице 1.

Таблица 1. Параметры материалов сцен из группы тестов 4.1–4.6

Поверхность	Средняя отражательная способность	Допустимое отклонение отражательной способности
Тест 4.1		
Потолок	0.70	0.01
Пол	0.06	0.01
Стены	0.40	0.02
Тест 4.2		
Потолок	0.70	0.01
Пол	0.06	0.01
Стены	0.52	0.02
Тест 4.3		
Потолок	0.70	0.01
Пол	0.06	0.01
Стены	0.52	0.02
Тесты 4.4, 4.5 и 4.6 (одинаковые значения)		
Потолок	0.03	0.01
Пол	0.06	0.01
Стены	0.40	0.01

Во всех примерах имеются 4 источника света. Координаты положения их центральных точек по осям X и Y одинаковые (табл. 2).

Таблица 2. Положение источников света в сценах из группы тестов 4.1–4.6

Источник	Координата по оси X	Координата по оси Y
1	1.695м	1.680м
2	1.695м	5.050м
3	5.085м	1.680м
4	5.085м	5.040м

Отличия сводятся к форме источников света и их положению по оси Z:

- *примеры 4.1. и 4.4.* Точечные источники света. Координата всех источников по оси Z составляет 3.14 м. Световой поток каждого источника составляет примерно 2200 лм (точные данные приведены в описании теста);
- *примеры 4.2. и 4.5.* Источники света, имеющие форму круга с диаметром 450 мм. Координата всех источников по оси Z одинаковая – 3.20 м. Световой поток источников находится в пределах от 1800 до 2100 лм (точные данные приведены в описании теста);
- *примеры 4.3 и 4.6.* Источники света, имеющие форму квадрата с размерами 600х600 мм. Координата всех источников по оси Z одинаковая – 3.20 м. Световой поток каждого источника составляет примерно 4100 лм (точные данные приведены в описании теста).

Описания тестов также содержат таблицы с эталонными значениями освещенности в заданных точках и допустимыми пределами отклонения этих значений. Пределы отклонения учитывают все возможные погрешности как при построении эталонной сцены и реальных физических измерениях, так и при симуляции в тестируемой программе.

### **Особенности моделирования в системе Lumiscept и влияние этих особенностей на проведение тестов**

Геометрия сцены в системе Lumiscept [11], вычислительное ядро которой написано специалистами ИПМ и ИТМО, представлена в виде треугольной сетки. Возможность представления геометрии в аналитическом виде (т.е. через задание уравнений, описывающих поверхности различной формы) отсутствует. Такой способ представления геометрии является оптимальным с целью упрощения вычислений при оптическом моделировании. Этот способ является не самым удобным для проектирования геометрических данных «с нуля», однако система Lumiscept предназначена только для моделирования глобального освещения и других оптических и светотехнических расчетов и не предназначена для непосредственного создания геометрических моделей. Геометрия импортируется из файлов популярных систем автоматизированного проектирования (CAD), таких как AutoCAD или CATIA. Если импортируемые данные содержат геометрические объекты, заданные в аналитической форме, то на стадии импорта производится их тесселляция, т.е. разбиение на треугольники с некоторой заданной точностью. Для большинства файловых форматов импорт производится только из заранее сохраненных исходных файлов. Однако для системы CATIA было разработано специальное дополнение (plug-in), представляющее собой ядро системы Lumiscept, встроенное непосредственно в графический интерфейс этой системы. Следует заметить, что для уже загруженной в систему Lumiscept геометрии в виде набора треугольников возможно дополнительное разбиение на треугольники меньшего размера. Координаты промежуточных вершин при этом вычисляются по некоторой интерполяции, при которой принимается, что поверхности, подвергаемые дополнительному разбиению, являются гладкими.

Для получения значений освещенности в некоторых произвольных точках пространства сцены в системе оптического моделирования Lumiscept используются расчет глобального освещения по методу Монте-Карло. Данный метод выдает результат с некоторой точностью, непрерывно увеличивающейся в процессе вычисления. Базовыми критериями для остановки симуляции по методу Монте-Карло являются достижение заданной точности вычислений либо превышение заданного времени симуляции. Кроме того, могут использоваться дополнительные критерии, например получение в процессе симуляции заданного количества лучей света. Следует обратить внимание на то, что вычисления по методу прямой Монте-Карло трассировке лучей не предоставляют возможности получить точное значение освещенности в какой-то конкретной точке пространства. Это вызвано тем, что лучи трассируются в случайных направлениях и невозможно попасть в заданную конкретную точку пространства за конечное время. Вместо этого можно получить среднее значение освещенности на некоторой поверхности, имеющей определенную площадь.

Чтобы получить значение освещенности некоторой поверхности, произвольно расположенной в пространстве, в системе Lumiscept существуют так называемые виртуальные измерительные приборы. Они представляют собой сетки различной формы, состоящие из ячеек. Для ячейки такого прибора вычисляется ее средняя освещенность, равная отношению светового потока, попавшего в данную ячейку, к площади всей этой ячейки. Как правило, в оптических расчетах требуются измерительные приборы, содержащие большое количество ячеек. Однако в нашем случае, когда требуется измерить освещенность в нескольких определенных точках, расположенных на значительном расстоянии друг от друга, приходится использовать отдельный виртуальный измерительный прибор, состоящий из единственной ячейки, для каждой точки измерений. Размер ячейки был подобран опытным путем. Он должен быть значительно меньше расстояния между соседними приборами, но при этом обеспечивать достаточную точность в расчетах по методу Монте-Карло.

### **Результаты тестирования**

Как уже было упомянуто, моделирование на тестовых сценах проводилось по методу Монте-Карло. В качестве основного критерия для остановки вычислений по методу Монте-Карло была выбрана погрешность вычислений, которая в первоначальном варианте составила 1 % для всех тестовых примеров.

После первоначального запуска тестовых сценариев с исходными параметрами оказалось, что в большинстве случаев результаты укладывались в допустимые пределы. Однако в двух тестовых

сценариях значения освещенности в некоторых точках вышли за эти пределы. В данной работе приведены полные таблицы результатов тестирования для двух сценариев – 4.1 (полностью успешный случай) и 4.2 (частичный выход за допустимые пределы). Для остальных тестов будет приведена только общая информация о полученных в результате моделирования значениях средней освещенности сцены и их допустимых пределах, а также о количестве точек, в которых значения вышли за допустимые пределы. В таблицах результатов для тестов 4.1 и 4.2 TE UL / TE LL являются верхним и нижним пределами допустимых значений с учетом полной возможной погрешности, а MB UL / LL – верхний и нижний пределы допустимых значений с учетом только измерительных ошибок.

#### Тест 4.1

Таблица 3. Подробные результаты тестирования для сценария 4.1

Position	Sensor						
	1	2	3	4	5	6	7
TE UL	91	107	115	118	116	107	93
MB UL	85	100	108	110	108	100	87
1	68	79	87	87	87	81	70
MB LL	65	77	83	85	83	77	67
TE LL	59	70	75	77	76	70	61
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	103	124	130	129	129	124	105
MB UL	96	116	122	120	121	116	98
2	78	92	98	96	97	94	78
MB LL	74	89	94	93	93	89	75
TE LL	67	81	85	84	84	81	68
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	112	132	141	141	141	131	113
MB UL	105	123	132	132	132	122	106
3	87	98	105	107	106	100	87
MB LL	81	95	101	102	101	94	81
TE LL	73	86	92	92	92	86	74
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	115	133	143	146	143	133	116
MB UL	108	124	133	137	133	124	108
4	89	101	109	108	108	99	86
MB LL	83	96	103	105	103	96	83
TE LL	75	87	93	96	93	87	76
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	113	132	141	140	141	132	112
MB UL	105	124	131	131	131	123	105
5	87	98	106	107	106	99	86
MB LL	81	95	101	101	101	95	81
TE LL	74	86	92	92	92	86	73
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	103	124	130	127	130	123	104
MB UL	97	116	121	119	121	115	97
6	79	92	98	98	99	93	79
MB LL	74	89	93	92	93	89	75
TE LL	68	81	85	83	85	81	68

Окончание таблицы 3

Position	Sensor						
	1	2	3	4	5	6	7
TE UL	92	108	116	117	115	108	92
MB UL	86	100	108	109	107	100	86
7	69	81	87	89	88	81	69
MB LL	66	77	83	84	83	77	66
TE LL	60	70	76	76	75	70	60

Как можно видеть из таблицы 3, все значения попали в интервал между верхним и нижним пределами допустимых значений, предоставленных CIE. Время выполнения теста составило 1 час 53 минуты 26 секунд, при этом было обработано примерно 9,2 миллиарда лучей.

## Тест 4.2

Таблица 4. Подробные результаты тестирования для сценария 4.2

Position	Sensor						
	1	2	3	4	5	6	7
TE UL	50	68	66	60	66	68	51
MB UL	47	63	62	56	61	63	48
1	36	48	48	45	48	48	37
MB LL	36	49	48	43	47	49	37
TE LL	33	44	43	39	43	44	33
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	65	93	88	77	87	93	67
MB UL	61	87	83	72	81	87	62
2	47	64	63	56	63	65	48
MB LL	47	67	64	55	63	67	48
TE LL	43	61	58	50	57	61	44
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	65	90	87	77	85	90	66
MB UL	61	84	81	72	80	84	62
3	47	63	62	58	63	64	49
MB LL	47	65	62	56	61	65	48
TE LL	42	59	57	50	56	59	43
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	61	79	77	72	77	79	61
MB UL	57	74	72	67	72	73	57
4	44	56	57	55	60	59	47
MB LL	44	57	55	52	55	56	44
TE LL	40	52	50	47	50	51	40
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	66	89	85	75	83	87	64
MB UL	61	83	79	70	78	82	60
5	47	62	63	59	66	68	52
MB LL	47	64	61	54	60	63	46
TE LL	43	58	55	49	54	57	42
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	65	92	85	74	83	89	63

Окончание таблицы 4

Position	Sensor						
	1	2	3	4	5	6	7
MB UL	61	86	80	69	78	83	59
6	47	65	63	59	68	71	52
MB LL	47	66	61	53	60	64	46
TE LL	43	60	56	48	54	58	41
Position	Sensor						
	1	2	3	4	5	6	7
TE UL	50	66	64	57	62	64	48
MB UL	47	62	60	54	58	60	45
7	37	47	49	46	52	52	41
MB LL	36	48	46	41	45	46	35
TE LL	33	43	42	38	41	42	31

В результате выполнения данного теста оказалось, что некоторые значения выходят за пределы допустимых значений с учетом измерительных ошибок, предоставленных CIE. В таблице 4 такие значения отмечены желтым цветом. Возможные причины появления подобных результатов и попытки улучшения вычислительных алгоритмов, которые должны были помочь добиться приведения результатов к допустимым значениям, будут описаны в следующем разделе.

Время выполнения теста составило 1 час 15 минут 20 секунд, при этом было обработано примерно 9,7 миллиарда лучей.

#### Тест 4.3

Таблица 5. Общие результаты тестирования для сценария 4.3

Верхний предел допустимых значений	254.2
Среднее значение освещенности в комнате	234.8
Нижний предел допустимых значений	199.8

Все значения попали в интервал между верхним и нижним пределами допустимых значений (табл. 5), предоставленных CIE. Время выполнения теста составило 55 минут 26 секунд, при этом было обработано примерно 7,5 миллиарда лучей.

#### Тест 4.4

Таблица 6. Общие результаты тестирования для сценария 4.4

Upper limit	37.5
Simulated mean room illuminance	36.0
Нижний предел допустимых значений	29.5

Все значения попали в интервал между верхним и нижним пределами допустимых значений (табл. 6), предоставленных CIE. Время выполнения теста составило 3 часа 21 минуту 39 секунд, при этом было обработано примерно 31,6 миллиарда лучей.

#### Тест 4.5

Таблица 7. Общие результаты тестирования для сценария 4.5

Верхний предел допустимых значений	51.1
Simulated mean room illuminance	43.1
Нижний предел допустимых значений	40.1

Все значения попали в интервал между верхним и нижним пределами допустимых значений (табл. 7), предоставленных CIE. Время выполнения теста составило 5 часов 39 минут 18 секунд, при этом было обработано примерно 27,2 миллиарда лучей.



## Тест 4.6

Таблица 8. Общие результаты тестирования для сценария 4.6

Верхний предел допустимых значений	228.5
Simulated mean room illuminance	212.7
Нижний предел допустимых значений	179.5

Три значения из 49 вышли за допустимые пределы (табл. 8). Время выполнения теста составило 3 часа 30 минут 28 секунд, при этом было обработано примерно 31,5 миллиарда лучей.

**Анализ результатов тестов в случаях выхода за допустимые пределы**

Следует обратить особое внимание на тесты 4.2 и 4.6, результаты которых для некоторых точек оказались за пределами интервала, который считается допустимым. Чтобы улучшить результаты этих тестов, было предпринято несколько попыток изменения различных вычислительных настроек и параметров представления сцены. В данном разделе описываются работы применительно к тесту 4.2, поскольку для теста 4.6 ситуация аналогичная.

Первоначально были предприняты попытки следующих действий для улучшения результатов теста:

- увеличение точности вычислений за счет более длительных вычислений;
- изменение размеров ячеек виртуальных измерительных приборов;
- изменение степени разбиения геометрии сцены.

*1. Увеличение точности вычислений*

Было произведено несколько вычислений со следующими значениями погрешности моделирования: 2 %, 1.5 %, 1 %, 0.75 %, 0.5 %. Время вычисления составило от 19 минут 57 секунд для погрешности 2 %, до 5 часов 1 минуты 59 секунд для погрешности 0.5 %. Поскольку для каждой точки измерения был задан отдельный виртуальный измерительный прибор, на котором достигнутая в определенный момент времени точность вычислений может отличаться от точности, достигнутой в тот же момент времени на других приборах, то было принято решение выбрать в качестве критерия точность (погрешность) вычисления на приборе, находящемся в левом верхнем углу сцены. Эксперимент показал, что увеличение точности в указанном диапазоне слабо влияет на результаты вычислений (табл. 9).

Таблица 9. Проведение теста по сценарию 4.2 с различной точностью вычислений

Погрешность	Номер сенсора	1	2	3	4	5	6	7
2.00 %	1	36	48	48	45	50	48	37
	2	47	65	63	56	63	64	47
	3	47	62	64	56	64	64	50
	4	46	56	57	57	61	59	47
	5	47	62	62	59	66	68	51
	6	47	63	63	58	67	72	52
	7	37	47	47	47	50	52	40
1.50 %	Номер сенсора	1	2	3	4	5	6	7
	1	36	48	47	45	48	48	37
	2	47	64	63	56	64	65	48
	3	47	62	63	57	65	65	48
	4	45	57	58	56	59	59	47
	5	47	62	62	59	66	69	51
	6	47	64	63	58	68	72	52
	7	37	47	48	47	52	52	40

Окончание таблицы 9

Погрешность	Номер сенсора	1	2	3	4	5	6	7
1.00 %	1	36	48	47	45	48	48	37
	2	47	64	63	56	63	65	48
	3	47	63	63	58	64	64	49
	4	45	57	58	56	60	58	47
	5	47	62	62	59	66	69	52
	6	47	64	63	59	68	71	52
	7	37	47	48	46	50	52	40
0.75 %	1	36	48	47	45	48	48	37
	2	47	64	64	57	63	65	48
	3	47	63	63	58	64	64	49
	4	45	56	58	56	60	59	47
	5	47	62	62	59	67	68	51
	6	47	64	63	59	68	71	52
	7	37	48	48	46	51	53	40
0.50 %	1	36	47	48	45	48	48	37
	2	47	64	63	56	63	65	48
	3	47	63	63	58	64	65	49
	4	44	56	58	56	60	59	47
	5	47	62	63	56	66	68	51
	6	47	64	63	59	68	71	52
	7	36	47	48	46	51	52	40
			Значение вне допустимого диапазона					

### 2. Изменение размеров ячеек виртуальных измерительных приборов

Как было описано выше, измерение освещенности в некоторой точке пространстве в системе оптического моделирования Lumisert проводится с помощью виртуальных измерительных приборов, ячейки которых имеют ненулевую площадь. Были предприняты попытки проводить вычисления с различными размерами ячеек измерительных приборов. Использовались следующие размеры: 10x10 мм и 5x5 мм. В результате эксперимента оказалось, что размеры ячеек не влияют на качество результатов (табл. 10).

Таблица 10. Проведение теста по сценарию 4.2 с различными размерами ячеек измерительных приборов (слева – ячейки 10x10, справа – ячейки 5x5 миллиметров)

погрешн.	position\ sensor	1	2	3	4	5	6	7
0.50%	1	36	47	48	45	48	48	37
	2	47	64	63	56	63	65	48
	3	47	63	63	58	64	65	49
	4	44	56	58	56	60	59	47
	5	47	62	63	56	66	68	51
	6	47	64	63	59	68	71	52
	7	36	47	48	46	51	52	40

погрешн.	position\ sensor	1	2	3	4	5	6	7
0.50%	1	37	47	47	45	49	48	37
	2	47	64	62	57	63	65	48
	3	47	63	62	58	63	64	49
	4	44	56	58	56	60	59	47
	5	47	63	63	59	67	68	52
	6	47	64	63	58	68	70	52
	7	37	48	48	46	51	52	40

### 3. Изменение степени разбиения геометрии сцены

Как уже было упомянуто, в системе Lumisert предусмотрена возможность автоматического разбиения геометрии сцены с увеличением количества треугольников. Были проведены тесты с уровнем разбиения 5 % и 1 % от исходного размера треугольников. Эксперимент показал, что увеличение разбиения слабо влияет на результаты вычислений (табл. 11).

Таблица 11. Проведение теста по сценарию 4.2 с различной степенью разбиения (слева – разбиение 1 %, справа – разбиение 5 %)

погрешн.	position\ sensor	1	2	3	4	5	6	7
0.50%	1	36	47	48	44	48	48	37
	2	47	64	63	56	63	65	48
	3	47	63	62	58	64	65	49
	4	44	57	57	56	59	59	47
	5	47	63	63	59	67	68	51
	6	47	64	63	58	68	70	52
	7	37	48	48	46	51	52	40

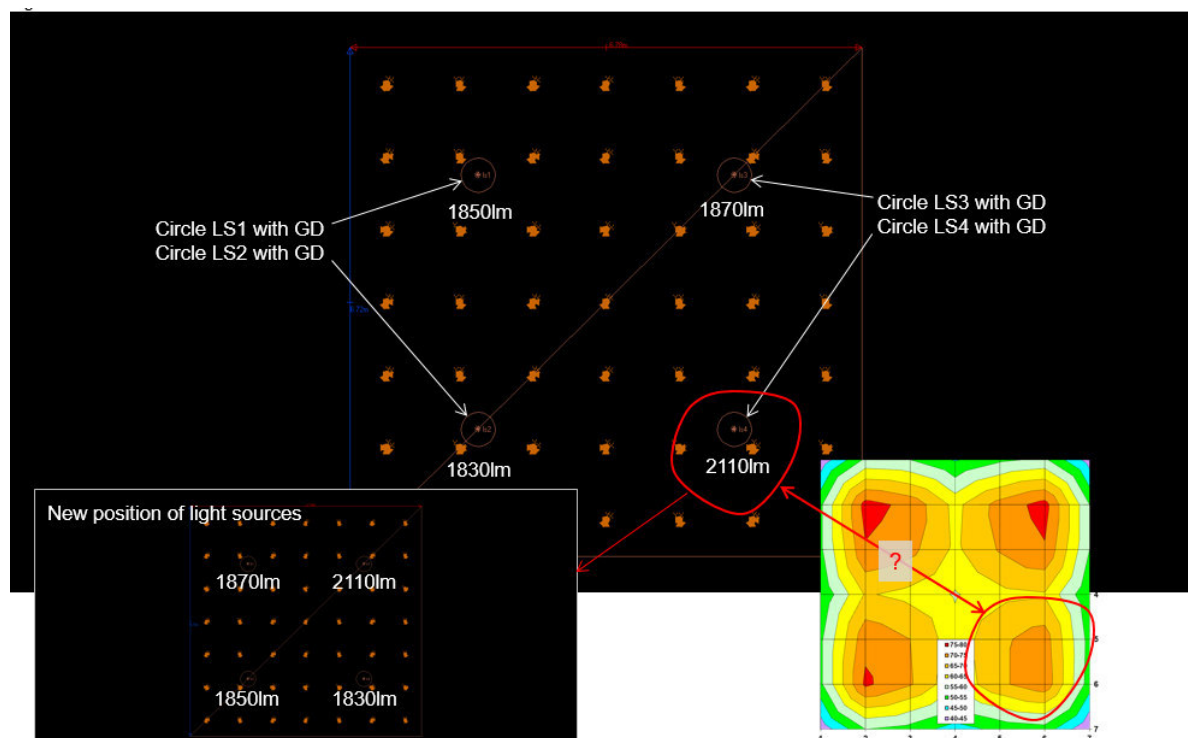
погрешн.	position\ sensor	1	2	3	4	5	6	7
0.50%	1	36	48	48	45	49	48	37
	2	48	64	63	56	63	66	48
	3	47	63	63	58	64	64	49
	4	44	56	58	56	60	59	47
	5	47	63	63	60	67	68	51
	6	47	64	63	58	68	71	52
	7	36	47	48	46	52	52	40

### Возможные варианты изменения параметров сцены

Как показали практические исследования, методы, которые в большинстве случаев помогают улучшить точность вычислений, практически не улучшили ситуацию для данного конкретного теста. Нами был проведен анализ тестов, проведенных по стандарту CIE 171:2006 для других систем оптического моделирования, в результате чего было обнаружено, что именно тест 4.2 из набора CIE 171:2006 является проблемным для очень многих программ оптического моделирования. В частности, подобные результаты возникли в системе AGi32 [12] и Radiance. В результате проведенного анализа мы делаем предположение, что исходные данные CIE для данного конкретного тестового случая являются не в полной мере корректными. Возможно, не были учтены какие-то дополнительные параметры сцены либо даже приведенные данные содержат какие-то ошибки.

Мы также провели анализ действий, которые предпринимались в исследованиях вышеупомянутых систем оптического моделирования для улучшения результатов в данном тестовом примере. В частности, в работе [12] предлагается рассматривать источники света, имеющие форму круга, в виде шестиугольников и дополнительно дробить их на треугольники. Этот подход позволил добиться снижения ошибки с 9.1 до 3.7 %.

Однако нами было замечено, что в описании тестов CIE заметно некоторое противоречие между положением источника света и распределением освещенности. Зона с минимальной освещенностью находится под более мощным источником света. Поэтому мы провели моделирование с измененным положением источников света (см. рисунок).



Симуляция по сценарию на базе теста 4.2 с измененным положением источников света

Здесь источники света поменялись местами друг с другом. После такой переконфигурации источников света, как показано на рисунке, сразу же было обнаружено попадание результатов теста в заданные пределы на большинстве ячеек. Исходя из этого, можно сделать предположение, что Международная комиссия по освещению (CIE) предоставила некорректные исходные данные для данного теста.

### Сравнение результатов тестирования с другими системами оптического моделирования

Таблица 12 показывает сравнение результатов тестирования системы Lumicert с другими известными системами оптического моделирования – Radiance и AGi32. В качестве критерия сравнения используется среднеквадратичная ошибка (RMSD). Поскольку в тестовых примерах отсутствуют эталонные значения, то вместо них использовалось среднее арифметическое верхнего и нижнего пределов допустимых значений с учетом полной возможной погрешности (TE UL / TE LL).

Таблица 12. Результаты тестирования систем оптического моделирования

Тест	Lumicert	Radiance	AGi32
RMSD			
4.1	8.8875	11.8591	9.0805
4.2	6.6238	9.3258	6.6943
4.3	11.0112	10.5480	11.6439
4.4	2.6035	2.5703	2.6721
4.5	3.3489	5.4724	2.9119
4.6	11.4216	11.4994	12.6279

Из данной таблицы видно, что в трех тестовых примерах из шести система Lumicert показала наилучшую точность по сравнению с другими известными системами. В остальных трех случаях результаты системы Lumicert имеют примерно одинаковую точность, как и другие системы.

### Выводы

Система оптического моделирования Lumicert, вычислительное ядро которой создано специалистами ИПМ и ИТМО, успешно прошла верификацию с использованием тестовых сценариев CIE 171:2006. Результаты в целом продемонстрировали корректное моделирование распространения света, при этом их точность не уступала результатам, полученным для других популярных систем. Из всех тестов только в 2 % случаев ошибка превысила рубеж в 10 %, который считается максимально приемлемым в инженерных расчетах. Анализ отклонений от эталонных значений CIE в проблемных тестах выявил высокую вероятность некорректности исходных данных, поскольку именно в этих тестах подобные чрезмерные отклонения также были обнаружены и для других популярных систем оптического моделирования.

Проведенная работа выявила необходимость дальнейшего исследования для определения причин отклонений в отдельных тестовых сценариях. Результаты могут быть использованы для повышения доверия к системе Lumicert и разработки рекомендаций по её применению, а также для совершенствования методов верификации систем оптического моделирования.

### Список литературы

1. Using simulation software calibration tests as a consumer guide—a feasibility study using lighting simulation software / Donn M., et al. 2007.
2. Inanici M., Abboushi B., Safranek S. Evaluation of sky spectra and sky models in daylighting simulations // Lighting Research & Technology. 2023. Vol. 55, no. 6. P. 502-529.
3. Pierson C., Aarts M. P. J., Andersen M. Validation of spectral simulation tools in the context of ipRGC-influenced light responses of building occupants // Journal of Building Performance Simulation. 2023. Vol. 16, no. 2. P. 179-197.
4. Pierson C., Wienold J., Bodart M. Daylight discomfort glare evaluation with Evalglare: influence of parameters and methods on the accuracy of discomfort glare prediction // Buildings. 2018. Vol. 8, no. 8. P. 94.
5. INLUX: A calculation code for daylight illuminance predictions inside buildings and its experimental validation / De Rosa A., et al. // Building and Environment. 2009. Vol. 44, no. 8. P. 1769-1775.
6. Mardaljevic J. The BRE-IDMP dataset: a new benchmark for the validation of illuminance prediction techniques // Lighting Research & Technology. 2001. Vol. 33, no. 2. P. 117-134.

- 
7. Schregle R., Wienold J. Physical validation of global illumination methods: measurement and error analysis // Computer Graphics Forum. 9600 Garsington Road, Oxford, OX4 2DQ, UK.: Blackwell Publishing Ltd., 2004. Vol. 23, no. 4. P. 761-781.
  8. A New Tool and Calculation Methodology for BIM-integrated Rapid Daylight Simulation / Dunn J., et al. (Preliminary Draft for ASHRAE Energy Modeling Conference). 2015
  9. Reliable datasets for lighting programs validation–benchmark results / Maamari F., et al. // Solar Energy. 2005. Vol. 79, no. 2. P. 213-215.
  10. CIE 171:2006 – Test Cases to Assess the Accuracy of Lighting Computer Programs. URL: <https://cie.co.at/publications/test-cases-assess-accuracy-lighting-computer-programs> (accessed 05.07.2025).
  11. Lumicept – Hybrid Light Simulation Software. URL: <http://www.integra.jp/en> (accessed 29.06.2025).
  12. Geisler-Moroder D., Dür A. Validation of radiance against CIE171: 2006 and improved adaptive subdivision of circular light sources //Proceedings 7th International RADIANCE workshop, Fribourg. – 2008.



Секция 2

# ***НАУЧНАЯ ВИЗУАЛИЗАЦИЯ И ВИЗУАЛЬНАЯ АНАЛИТИКА***

---

**SCIENTIFIC VISUALIZATION  
AND VISUAL ANALYTICS**



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия



УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-320-334

## Сравнение пользовательской эффективности традиционных и нейросетевых методов UX/UI-проектирования корпоративных порталов

А. П. Попова<sup>1</sup>, С. С. Бодрунова<sup>2</sup>, П. А. Васев<sup>3</sup>, И. С. Стародубцев<sup>1,3</sup>

<sup>1</sup>Уральский федеральный университет им. Б. Н. Ельцина (УрФУ), Екатеринбург, Россия

<sup>2</sup>Институт «Высшая школа журналистики и массовых коммуникаций» Санкт-Петербургского государственного университета, Санкт-Петербург, Россия

<sup>3</sup>Институт математики и механики им. Н.Н. Красовского УрО РАН (ИММ УрО РАН), Екатеринбург, Россия

**Аннотация.** Современный дизайн веб-порталов, ориентированный на пользователя и его опыт взаимодействия с интерфейсом (UX/UI-дизайн), проходит этап бурного внедрения дизайнерского опыта работы с нейронными сетями и генеративным искусственным интеллектом (генИИ). Внедрение нейросетевого проектирования в коммерческий UX/UI-дизайн позволяет совместить эргономические, эстетические, творческие и маркетинговые задачи в едином процессе проектирования на новом уровне интеграции с учетом опыта, проанализированного и учтенного ИИ. Но при этом внедрение технологии в индустрию происходит несистемно, практически без оценки реальных улучшений опыта пользователя в сравнении с дизайном, созданным традиционными методами. Наше пилотное исследование направлено на сравнительный анализ эффективности пользовательского опыта для экспериментального корпоративного портала кинопроизводственной организации, разработанного как традиционными, так и нейросетевыми методами. На первом этапе мы сопоставляем возможности различных ИИ-агентов в производстве интерфейса по промптам; на втором – сравниваем пользовательский опыт традиционного интерфейса с опытом использования лучшего из полученных ИИ-интерфейсов. Применяя пре- и посттестирование по метрикам успешности задачи, времени выполнения работы и количества ошибок, а также оценку субъективной удовлетворенности пользователя на основе данных от десяти ассессоров, мы показываем, в чем именно опыт взаимодействия с ИИ-интерфейсом может быть успешнее, чем опыт взаимодействия с традиционным дизайном, что открывает двери для грамотного индустриального применения нейросетевой визуализации в коммерческом UX/UI-дизайне.

**Ключевые слова:** UX/UI-дизайн, пользовательский интерфейс, эргономика, нейронные сети, нейросетевая визуализация, искусственный интеллект, генеративный искусственный интеллект, юзабилити-тестирование

## Traditional vs. neural-network-based methods in UX/UI design of corporate web portals: Comparing user efficiency

A. P. Popovskaya<sup>1</sup>, S. S. Bodrunova<sup>2</sup>, P. A. Vasev<sup>3</sup>, I. S. Starodubtsev<sup>1,3</sup>

<sup>1</sup>Ural Federal University, Yekaterinburg, Russia

<sup>2</sup>School of Journalism and Mass Communications, St. Petersburg State University, Russia

<sup>3</sup>Krasovskii Institute of Mathematics and Mechanics, The Urals Branch of the Russian Academy of Sciences, Yekaterinburg, Russia

**Abstract.** Today's user-centered design of web interfaces (UX/UI design) sees perfectioning of user experience as its core target. Within a highly competitive industrial environment, web designers are rapidly implementing layouting techniques that employ neural networks and generative artificial intelligence (genAI). The integration of neural-network technologies into commercial UX/UI design blends the ergonomic, aesthetic, creative, and marketing tasks into an integral process of layouting and composition at a new level of integration that takes into account the experience analyzed and considered by AI. However, the introduction of this technology into the industry has not been systematic, with little assessment of actual improvements in the user experience compared to designs created via traditional instruments. Our pilot study aims at conducting a comparative analysis of user experience efficiency for an experimental corporate portal developed by a film production organization using both traditional and neural network techniques. At the first stage, we compare the capabilities of three AI agents in creating a prompt-based commercial interface. At the second stage, we compare the user experience of a traditional interface with that of the best of the AI-produced layouts we have received. By pre- and post-testing of task success rate, time of task, and user error rate, as well as by evaluating subjective user satisfaction based on data from ten assessors, we show how exactly the experience of interacting with an AI interface could be more successful than that of traditional design. Our results, subject to expansion, provide for more proficient industrial application of AI-based visualization in commercial UX/UI design.

**Keywords:** UX/UI design, user interface, ergonomics, neural networks, neural network visualization, artificial intelligence (AI), generative artificial intelligence (genAI), usability testing

## Введение

В современном мире визуально привлекательный, интуитивно понятный и удобный интерфейс является важным критерием качества цифрового продукта [1]. Именно взаимодействие с интерфейсом формирует первое впечатление пользователя и влияет на его дальнейшую вовлеченность [2]. Создание интерфейса, эргономичного на многих уровнях проектирования [3], является многофакторной проблемой, сочетающей эстетические и навигационные задачи.

При этом для корпоративных вебсайтов особенно важно сохранять и наращивать качество корпоративных порталов при росте их навигационной сложности, поскольку оно оказывает влияние на формирование профессионального имиджа компании, повышение уровня доверия со стороны целевой аудитории, а также способствует поддержанию взаимодействия с пользователями и их повторному обращению к услугам.

Применение нейронных сетей как для визуализации стиля компании, так и для разработки позволяет объединить задачи творческого, эстетического, эргономического и маркетингового характера в единый процесс разработки интерфейса портала на новом уровне с учетом освоенного машинной опыта. В силу этого, а также растущей конкуренции в отрасли UX/UI-дизайна применение нейросетей в дизайне за короткий срок стало «новой нормой» [4], а также начало создавать новые виды неравенства внутри дизайн-сообщества, отделяя ИИ-дизайнеров от тех, кто работает более традиционными методами.

Однако проблема заключается в отсутствии точных данных о том, что применение нейронных сетей в UX/UI-дизайне корпоративного сайта повышает удобство и эффективность интерфейсов по сравнению с традиционным (ручным) методом проектирования.

Таким образом, актуальность нашего исследования обусловлена стремительным развитием технологий искусственного интеллекта и нейронных сетей, которые меняют традиционный подход к проектированию пользовательских интерфейсов, а также лакуной в понимании того, насколько качество интерфейса повышается при применении ИИ и по каким параметрам. Сравнение работы человека с работой нейронных сетей также важно по следующим причинам [5]:

- 1) создание и изучение систем на основе нейронных сетей, а также понимание того, в чем они подобны и чем отличаются от человеческого мышления, позволяет разработать технологии, которые могут лучше соответствовать потребностям и особенностям человека;
- 2) понимание возможностей и ограничений нейронных сетей по сравнению с человеческим интеллектом может привести к более осознанному использованию и развитию искусственного интеллекта.

Наше пилотное исследование направлено на сравнительный анализ эффективности пользовательского опыта для экспериментального корпоративного портала кинопроизводственной организации, разработанного как традиционными, так и нейросетевыми методами (выполнено по заданию заказчика). На первом этапе мы сопоставляем возможности различных ИИ-агентов в производстве интерфейса по промптам; на втором – сравниваем пользовательский опыт традиционного интерфейса с опытом использования лучшего из полученных ИИ-интерфейсов.

Мы проводим двухэтапный эксперимент, в котором сперва разрабатываем ИИ-макет с помощью промптов для трех ИИ-агентов, оцениваем и выбираем лучший из полученных вариантов; затем мы сравниваем ИИ-интерфейс и интерфейс, разработанный действующим дизайнером, в рамках эксперимента по юзабилити-тестированию. На втором этапе для тестирования пользовательского опыта мы используем сочетание количественных и качественных метрик, что позволяет оценить объективный и субъективный опыт пользователя при работе с интерфейсами. Так, мы проводим пре- и посттестирование по таблице Горбова – Шульте (показатель динамики количества ошибок, использующийся сегодня в том числе для оценки когнитивной перегрузки при общении с интерфейсом [6, 7]). В посттесте также задействованы метрики успешности выполнения задачи и времени на ее выполнение. Оценку субъективной удовлетворенности пользователя мы проводим на основе стандартной анкеты, аналогии которой широко используются в юзабилити-тестировании. Пилотное исследование включает данные от десяти ассессоров; как показано в более ранних исследованиях, малые группы способны выявить большой процент проблем в прототипе / интерфейсе [8]. В итоге мы

показываем, в чем именно опыт взаимодействия с ИИ-интерфейсом может быть успешнее, чем опыт взаимодействия с традиционным дизайном. При условии, что эксперимент будет развит до тестирования многих порталов на большем количестве малых групп, он открывает двери для грамотного индустриального применения нейросетевой визуализации в коммерческом UX/UI-дизайне.

Далее мы приводим краткий обзор литературы по влиянию ИИ на современные практики дизайна, подробнее освещаем методику двухступенчатого эксперимента, излагаем результаты и помещаем их в контекст современных исследований в области дизайна, пользовательского опыта и юзабилити (DUXU).

## Обзор литературы и постановка задачи

### 1. Современное развитие веб-дизайна: ориентация на пользовательский опыт

UX/UI-дизайн («дизайн опыта пользователя / дизайн интерфейса» [9, 10] – концепция, подчеркивающая ключевой характер обобщенного пользовательского опыта для разработки цифровых интерфейсов и неразрывную связь логики интерфейса (навигации, связи элементов и т.д.) с культурной, социально-групповой, гендерной и иной вариативностью опыта взаимодействия. Ориентация на опыт пользователя приводит, во-первых, к необходимости адаптации творческой идеи к опыту целевой аудитории, а во-вторых – к необходимости тестирования элементов дизайна и всего интерфейса в целом. UX-дизайн подразумевает весь процесс взаимодействия пользователя с программным продуктом. По мнению некоторых исследователей, тестирование даже является составной частью UX-дизайна как процесса: он состоит из разработки общего пользовательского опыта, включая путь пользователя, информационную архитектуру и юзабилити-тестирование [11]. При этом следует соблюдать основные принципы UX-дизайна, включая удобство использования (эргономичность), доступность и ясность, эффективность, субъективную удовлетворенность пользователя [12].

С появлением ИИ-инструментов для дизайна, включая специально адаптированные под задачи визуализации нейронные сети [13], вопрос соблюдения этих принципов встал особенно остро. Современный набор ИИ-инструментов, доступных веб-дизайнерам и веб-разработчикам, разнообразен по типам и функционалу. Однако пока проводится критически мало исследований, которые проверяли бы соблюдение ИИ-агентами вышеописанных принципов UX-дизайна, а также то, насколько не-человеческий (non-human) агент рынка дизайна лучше или хуже воплощает их по сравнению с дизайнером-человеком.

### 2. Применение нейросетей в UX/UI-дизайне

Сегодня внедрение ИИ в дизайнерскую практику происходит по нескольким функциональным направлениям. Так, в некоторых исследованиях предлагается использовать инструменты ИИ, чтобы помочь дизайнерам более эффективно понимать *контекст* использования своих цифровых решений. Под контекстом при этом чаще всего подразумевается профиль целевой группы и ее поведенческие паттерны. Часто эта информация собирается и анализируется с помощью персоналий или портретов пользователей на основе доступных наборов данных [14], что затем используется для создания прототипов и визуального оформления интерфейсов [15]. Второе направление – автоматизация *перевода* рисованного эскиза в компьютерную форму с одновременным *улучшением* (enhancement) прототипа [16]. Третья возможность – создание *альтернатив* первоначальной версии дизайна [17]. Четвертым направлением может считаться *сотворчество* в виде диалога, в котором дизайнер получает от ИИ-агента дополнительную информацию для понимания проекта, первичные формулировки идей, сопоставление своих идей с уже существующими аналогами и т.д. [18].

Однако наиболее полным участие ИИ в разработке интерфейса становится тогда, когда ИИ-агент создает прототип, интерфейс или иные составляющие цифрового графического объекта самостоятельно на основе полученной формулировки задачи, или *промпта*. Такой тип дизайна получил название *генеративного дизайна* и сегодня является предметом осмысления далеко за пределами медиасферы, охватывая практически все виды дизайна [19]. Такой подход к дизайну основывается на использовании алгоритмов и автоматизации для создания разнообразных и уникальных дизайнерских решений [20]. В результате работы дизайнера с адаптированными

нейросетями и языковыми моделями получается несколько готовых вариантов прототипа, которые отвечают заданным требованиям, но отличаются по внешнему виду.

Метод генерации в дизайне обладает рядом преимуществ [21]:

- достижение конечного результата в более короткие сроки;
- расширение возможностей автора в поиске интересных проектных решений и увеличение тем самым степеней свободы выбора;
- моделирование более сложных объектов.

### 3. Процесс разработки промпта в генеративном дизайне

Однако даже самые мощные и передовые ИИ-инструменты могут быть ограничены неправильной или некорректной информацией, которая им предоставляется. Промпт (от англ. prompt, или «подсказка») – текстовая инструкция, создаваемая для направления работы нейронной сети. Промпты содержат информацию о том, что именно машине нужно сделать с данными и в какой последовательности.

Промпт преобразуется в числовое представление, понятное нейросети. Он включает в себя токенизацию текста, т.е. разбиение его на отдельные слова или символы, и преобразование каждого токена в числовой вектор, который модель может обрабатывать [22]. Затем нейронная сеть получает этот числовой вектор в качестве входных данных и использует его для выполнения задачи. Результатом является ответ, представленный в текстовом формате или в виде иных данных (в зависимости от задачи). Важно отметить, что эффективность результата зависит от составленного промпта, а также качества и разнообразия обучающих данных и архитектуры модели. Проектирование оптимальных промптов уже выросло в отдельную индустрию – промпт-инжиниринг (от англ. prompt engineering).

Промпт может состоять из следующих компонентов:

- инструкция или пошаговое описание конкретной задачи;
- контекст или внешняя информация, которая может направить нейронную сеть на более точные ответы;
- входные данные;
- индикатор вывода.

Рассмотрим несколько существующих стратегий формирования промпта (промптинга):

1) промптинг с нулевой разметкой (от англ. zero-shot prompting). Большие языковые модели (large language models, или LLMs), такие как GPT-3, настроены на выполнение инструкций и обучены работе с большими объемами данных, поэтому они способны выполнять некоторые задачи с «нулевой разметкой», при котором дается только задача или вопрос без примеров правильного выполнения [23];

2) промптинг с несколькими примерами (от англ. few-shot prompting). LLMs демонстрируют впечатляющие возможности промптинга с нулевой разметкой, однако они все еще ограничены в более сложных задачах. Промптинг с несколькими примерами может использоваться для контекстного обучения [24];

3) промптинг с цепочкой мыслей (от англ. chain of thought, или CoT). Эта техника позволяет выполнять сложные рассуждения с помощью промежуточных шагов [24];

4) промптинг с нулевой разметкой и цепочкой мыслей (от англ. zero-shot CoT) – самая новая техника, которая сводится к добавлению инструкции «Давай подумаем шаг за шагом» в исходный промпт [25].

Компания Open AI предоставила подробные инструкции, как сформировать успешный промпт для ИИ [26]:

1) конкретика при написании инструкции и задачи, которую модель должна выполнить. Чем более подробным будет запрос, тем лучше будут результаты. Нет конкретных токенов или ключевых слов, которые привели бы к хорошим результатам; гораздо важнее хороший формат и описательные подсказки;

2) избегание формулировки о том, что не нужно делать. Это способствует большей конкретизации и сосредоточению внимания модели на деталях и ведет к получению более точных результатов моделирования;

3) разбиение задач. При выполнении сложных запросов нейронные сети чаще допускают ошибки, поэтому гораздо эффективнее получить желаемый результат, если попросить нейронную сеть выполнить несколько небольших подзадач по очереди в рамках одной большой задачи.

Этой общей логикой мы будем руководствоваться при создании промптов для генеративного дизайна корпоративного портала. Однако наш подход подразумевает углубленное использование нейронных сетей, а именно – их применение на трех этапах создания дизайн-проекта: 1) при анализе конкурентной среды, создании навигационного каркаса портала и составлении промпта для разработки визуального компонента прототипа; 2) при создании визуального оформления прототипа; 3) при анализе полученного результата. На каждом из этапов используются нейросети / LLMs различных типов. Расширение логики происходит за счет обращения к трудам о фронтенд-разработке порталов и приложений [27]. Это в т.ч. работы о ранних подходах к формализации описания элементов интерфейса, в которых предлагались первые модели рекуррентных сетей для генерации HTML- и CSS-кода [27: 1070, 28], а также исследования, встраивающие ИИ в различные доступные фреймворки для специализированной обработки кода и тестирования полученных результатов [29, 30].

#### *4. Применение нейронных сетей в российской индустрии генеративного медиадизайна*

Насколько можно судить, научная область, оценивающая развитие генеративного медиадизайна в России, пока находится в зачаточном состоянии; можно в целом насчитать не более двух десятков работ, касающихся применения ИИ и/или нейросетей в графическом дизайне, веб-дизайне, рекламе, брендинге. Большинство из них умозрительны и описывают первичный опыт применения ИИ, субъективно выделяя его достоинства и недостатки; часто это студенческие работы, не имеющие доказанных выводов. Некоторые авторы пытаются структурировать применение нейромоделей на различных этапах создания дизайнерского, рекламного, маркетингового цифрового продукта [31, 32, 33]. Единственной обнаруженной нами работой, сравнивающей эффективность ИИ и человека в веб-дизайне, стала работа [27], в которой сравниваются трудозатраты фронтенд-разработчика (программиста, а не дизайнера) для задач автоматической верстки и генерации кода веб-компонентов; указывается на существенное сокращение временных трудозатрат при использовании ИИ-агентов. Также одна магистерская диссертация [34] отражает попытку сравнить качественным методом (опрос пользователей) три типа веб-дизайна рекрутингового портала (классический, генеративный и автономный; мы, однако, предпочтем называть два последних типа единым генеративным). Попыток протестировать эргономичность традиционного и генеративного дизайна количественными методами нами не обнаружено.

Так же мало информации можно найти о том, как ИИ меняет российскую индустрию дизайна и насколько распространена практика генерации ИИ-прототипов; однако из существующих данных видно, что российская индустрия медиадизайна крайне быстро трансформируется под влиянием ИИ. Так, в 2023 году был проведен опрос аналитиками школы дизайна «Contented» и сервиса по поиску работы «Зарплата.ру», в котором приняли участие 1200 дизайнеров в возрасте от 25 до 55 лет, проживающих в российских городах-миллионниках [35]. Всего, по данным опроса, почти половина (45 %) дизайнеров уже использовали в работе нейронные сети; активно пользовался ими каждый шестой респондент (17 %). Чаще всего участники опроса применяли нейронные сети в графическом дизайне (80 %), веб-дизайне (48 %), 3D-графике (30 %) и UX/UI-дизайне (20 %). Мы предполагаем, что эти цифры за 2024–2025 годы только возросли.

Сопоставляя почти полное отсутствие доказательных исследований в области генеративного медиадизайна и бурное внедрение ИИ-агентного прототипирования и визуализации интерфейсов в практический дизайн цифровых медиапродуктов, можно легко убедиться в актуальности и необходимости нашего исследования.

Таким образом, **постановка задачи** исследования такова: 1) произвести путем эксперимента прототип и визуальные компоненты ИИ-интерфейса корпоративного веб-портала; 2) сравнить количественным и качественным методом качество полученного интерфейса с качеством интерфейса, произведенного дизайнером-человеком. Подзадачами исследования стали: 1) тестирование и выбор наилучших ИИ-агентов для производства интерфейса; 2) тестирование юзабилити интерфейсов,

созданных человеком и машиной; 3) определение преимуществ/недостатков ИИ-интерфейса по сравнению с интерфейсом, созданным человеком.

### Дизайн исследования и методы

#### 1. Подготовка эксперимента и ручное проектирование веб-портала

Подготовительный этап эксперимента включал взаимодействие с заказчиком – киностудией «Уралфильм», специализирующейся на съемке документального и рекламного кино, запрос которой состоял в улучшении дизайна корпоративного портала. Во взаимодействии с заказчиком осуществлялись шесть традиционных этапов создания интерфейса портала, завершившиеся подготовкой утвержденной заказчиком версии навигационной карты портала (рис. 1а) и его визуального стиля (рис. 1б). Эти шесть этапов включали:

- 1) аналитика, включающая в себя брифинг с киностудией и анализ целевой аудитории;
- 2) создание пользовательского сценария (user flow);
- 3) выбор концепции дизайна (визуального стиля, цветовой палитры, типографики и UX-подхода);
- 4) определение структуры сайта, состоящей из информационной архитектуры (навигационная карта сайта) и версий высокодетализированного прототипа (high-fidelity wireframes);
- 5) сбор обратной связи с киностудией и утверждение итогового варианта;
- 6) прототипирование страниц с выбранным визуальным стилем.

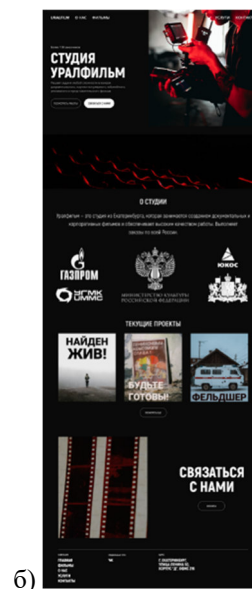
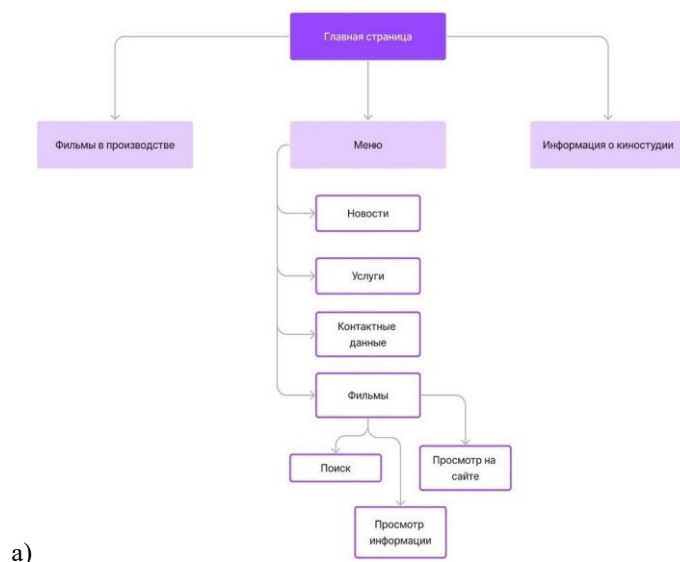


Рис. 1. Утвержденные заказчиком: а) навигационная карта портала; б) wireframe (детализированная репрезентация) главной страницы веб-портала

На основе утвержденной концепции были также разработаны пять полноценных страниц, выполненных в едином стилистическом и композиционном стиле. Страницы обладали кликабельной структурой, соответствующей принципам интерактивного дизайна, что обеспечивало последовательную и логичную навигацию. При проектировании учитывались такие аспекты, как визуальная иерархия, обратная связь с пользователем, согласованность интерфейсных компонентов. Комплексная проработка структуры портала и поведения пользователя в компонентах интерфейса была направлена на формирование цельного пользовательского опыта и повышение вовлеченности посетителей сайта.

#### 2. Первый этап эксперимента: подготовка ИИ-интерфейса по полученному брифу

Как уже сказано выше, логика подготовки ИИ-интерфейса включала трехступенчатое использование ИИ-агентов (нейросетей различной функциональности). В процессе проектирования кликабельного прототипа были использованы следующие типы нейросетевых технологий:



- 1) большие языковые модели (LLM), такие как ChatGPT, Gemini и Claude, применявшиеся для анализа, генерации текстовых описаний и составления промптов;
- 2) генеративно-сопоставительные сети (GAN), а именно – Midjourney, DALL-E, Stable Diffusion, Recraft, Leonardo AI и GPT-4o, использованные для генерации дизайн-концептов интерфейса на основе текстовых описаний (промптов);
- 3) оценочная нейронная сеть Roast My Design для автоматизированной UX/UI-оценки сгенерированных интерфейсов по заданным критериям.

В связи с этим процесс подготовки ИИ-интерфейса изменился и включил следующие этапы:

- 1) ИИ-анализ сайтов киностудий и выявление общего паттерна структурных решений сайта в сфере кино;
  - 2) ИИ-анализ текущей версии сайта «Уралфильм» и генерация подробного описания UX/UI-структуры сайта при помощи трех LLMs;
  - 3) генерация коротких и длинных промптов на основе сформированной LLMs UX/UI-структуры сайта;
  - 4) генерация изображений интерфейсов по полученным коротким и длинным промптам от трех LLMs;
  - 5) оценка дизайна сгенерированных интерфейсов при помощи нейронной сети Roast My Design;
  - 6) выбор наилучшего варианта интерфейса на основании комплексной оценки Roast My Design;
  - 7) проектирование остальных страниц и элементов сайта с сохранением стилистического единства.
3. *Второй этап эксперимента: сравнительное юзабилити-тестирование подготовленных интерфейсов*

После завершения этапа проектирования интерфейсов традиционным (ручным) и генеративным методами был организован этап юзабилити-тестирования, цель которого – получение объективной и субъективной оценки удобства и эффективности предложенных прототипов.

**Выборка ассессоров.** Для этого было привлечено 10 ассессоров (по пять мужчин и женщин в возрасте от 25 до 50 лет), отобранных с учетом гендерного и возрастного разнообразия. Пилотное исследование на малой выборке, несмотря на небольшое количество ассессоров ( $5 \div 10$ ), способно выявить около 90 % потенциальных проблем в пользовательском интерфейсе [8]. С каждым ассессором проводилось тестирование только одного из двух прототипов в день в целях минимизации когнитивной нагрузки, которая могла бы исказить восприятие интерфейса и снизить точность результатов. Такой подход позволяет обеспечить более объективную и сфокусированную обратную связь по каждому из прототипов.

**Этапы юзабилити-тестирования.** В начале встречи каждому ассессору объяснялись задачи исследования и порядок выполнения тестирования, что способствовало их естественному поведению во время взаимодействия с прототипами. Дальнейшие этапы тестирования включали три типа тестов: два объективных и один субъективный. Объективные методы включали количественный пре-/пост-тест и «древесный» метод фиксации успешности прохождения задачи; субъективный метод основывался на заполнении анкеты субъективной удовлетворенности. Сочетание количественных и качественных методов юзабилити-тестирования уже показало высокую продуктивность [36].

Для пре- и посттеста применялся классический тест когнитивной нагрузки, уровня концентрации внимания и скорости переработки информации с применением таблицы Горбова – Шульте [6, 7].

Во время прохождения задания для оценки успешности работы применялся так называемый «древесный» метод (tree-testing method), часто применяемый в индустрии веб-разработки [38]. В отличие от методик на основе «мыслей вслух» (think-aloud techniques), этот метод позволяет фиксировать успешность прохождения задачи в количественных показателях. Во время работы с прототипом ассессорам было предложено пройти тестовые сценарии взаимодействия с прототипами сайта киностудии. Предполагалось прохождение трех задач, специально подобранных с учетом структуры и логики конкретного интерфейса. Задания не сопровождалась подсказками со стороны тестирующего; это позволяло зафиксировать реальный пользовательский путь и потенциальные трудности.

Для интерфейса, спроектированного ручным методом, были заданы следующие сценарии различной степени детализации, направленные на поиск информации фокусной или общей информации: «Вы посмотрели основную информацию о студии и захотели посмотреть последний вышедший у них фильм»; «Вы знаете, что у киностудии вышел фильм, получивший множество наград на кинофестивалях, однако не можете найти его на сайте»; «Вам интересно, чем именно занимается киностудия и какие услуги она предоставляет».

Для интерфейса, спроектированного при помощи нейронных сетей, были заданы альтернативные сценарии сходного характера: «Вы ознакомились с основной информацией о киностудии и захотели посмотреть любой их фильм 2020 года»; «Вам интересно, в каком году была основана киностудия»; «Вы хотите заказать у киностудии съемку проекта – попробуйте найти, как это сделать».

Для оценки действий ассессоров применялись три традиционные метрики «древесного» тестирования:

(1) коэффициент успешности задачи (от англ. «task success rate», TSR) – один из ключевых UX-показателей, который позволяет понять, насколько интуитивно понятен и эффективен пользовательский интерфейс. Эта метрика показывает, как много пользователей смогли выполнить поставленную задачу без посторонней помощи и с достижением желаемого результата. Рассчитывается по формуле:  $TSR = (\text{число ассессоров, успешно завершивших задачу}) / (\text{общее число ассессоров, начавших задачу}) \times 100 \%$ . Высокий показатель (выше 90 %) говорит о хорошей продуманности интерфейса, тогда как низкий (ниже 60 %) сигнализирует «о возможных проблемах в навигации, понимании текста или логике взаимодействия»;

(2) время выполнения задачи (от англ. «time on task», TOT) – среднее количество времени, которое требуется ассессору на завершение задачи. Анализ этой метрики позволяет понять, насколько быстро и эффективно пользователь ориентируется в интерфейсе. Когда выборка составляет 5-15 человек, чаще рассчитывается медиана времени выполнения задач, так как это более устойчивый показатель, не подверженный выбросам. Например, если ассессор сильно задержался:  $TOT = \text{медиана (времени всех участников)}$ ;

(3) уровень ошибок пользователя (от англ. «user error rate», UER) – количество действий, которые не соответствуют целевому сценарию: нажатие не на ту кнопку, попытка перейти по недоступной ссылке, неверный ввод информации и др. Этот параметр позволяет выявить проблемные зоны, в которых ассессоры сбиваются, теряются или делают предположения, не соответствующие заложенной логике интерфейса. Рассчитывается по следующей формуле:  $UER = \text{общее число ошибок} / \text{число ассессоров}$ .

*Анкетирование* включало восемь утверждений – метрик субъективной удовлетворенности по Дж. Бруку [37]: по два в категориях «удобство» (1, 2), «доверие» (3, 4), «лояльность» (5, 6), «эстетика» (7, 8): 1) «сайт прост в использовании»; 2) «на сайте легко ориентироваться»; 3) «я чувствую себя комфортно, находясь на сайте»; 4) «я чувствую себя уверенно, совершая действия на сайте»; 5) «я порекомендую этот сайт другу или коллеге»; 6) «я, вероятно, вернусь на сайт в будущем»; 7) «я считаю сайт привлекательным»; 8) «сайт визуально простой и понятный». Каждое из утверждений ассессор оценивал по шкале Лайкерта от 1 («категорически не согласен») до 5 («полностью согласен»). На основе ответов рассчитывался индекс удовлетворенности пользователя (от англ. «customer satisfaction index», CSI) как субъективная метрика оценки эффективности и качества интерфейса с точки зрения конечного пользователя. Для расчета CSI каждого ассессора использовалась следующая формула:  $CSI = (\sum \text{оценок ассессора}) / (\text{максимальная возможная сумма баллов}) \times 100\%$ .

**Статистический анализ результатов.** По всем шести показателям (TSR, TOT, UER, пре- и пост-тест Горбова – Шульте, CSI) были вычислены средние показатели (Mean) и стандартные отклонения (SD). Также для всех метрик, кроме TSR (ее результаты получились одинаковыми для обоих прототипов) рассчитывались показатели разницы ( $\Delta$ ); для теста Горбова – Шульте рассчитывалась разница между показателями пре- и пост-теста ( $\Delta SG$ ), результаты сопоставлялись. Для метрик TOT, UER, CSI,  $\Delta SG$  было проведено сравнение независимых малых выборок (тест Манна – Уитни) для выявления значимых различий.

## Результаты экспериментов

### 1. Первый этап эксперимента: подготовка ИИ-интерфейса

Как сказано выше, мы использовали трехступенчатую подготовку ИИ-прототипа и его визуального стиля.

**ИИ-анализ конкурентной среды.** Для получения актуального и релевантного перечня кинокомпаний был использован промпт для трех LLMs – ChatGPT, Gemini и Claude: «Составь список из не менее 25 кинокомпаний, специализирующихся только на документальном и рекламном кино». Результаты оказались ограниченными по объему и качеству. Так, ChatGPT смогла сгенерировать список из 25 компаний, но только шесть из них соответствовали заданным критериям; большинство ссылок вели на несуществующие сайты или на ресурсы, не имеющие отношения к киноиндустрии или специализирующиеся на других направлениях кинопроизводства. Gemini предоставила список из семи наименований, из которых только два действительно относились к требуемой категории. Claude, несмотря на кажущуюся полноту списка из 25 компаний, в большинстве случаев предлагала устаревшие или несуществующие организации.

Далее моделям был предложен уточненный промпт для анализа вручную отобранного списка релевантных кинокомпаний: «Проанализируй сайты данных компаний на соответствие принципам UX/UI-дизайна». Однако и здесь наблюдались значительные ограничения. Так, ChatGPT не смогла получить доступ к трем из шести предоставленных сайтов даже при наличии прямых ссылок. Gemini оказалась не в состоянии проанализировать ни один из ресурсов, сославшись на отсутствие доступа или ограничения на стороне сайтов. Claude также не смогла обработать ни один URL-адрес и не предоставила содержательного анализа.

В силу этого на этапе генерации UX/UI-структуры сайта киностудии, основанного на принципах UX/UI-дизайна, был использован единый промпт для трех LLMs: «Проанализируй сайт киностудии «Уралфильм» и на основе его содержания предложи структуру сайта с подробным описанием каждой страницы в соответствии с современными принципами UX/UI-дизайна».

Модели снова показали различный уровень успешности выполнения задачи. ChatGPT показала наиболее полный и логически структурированный результат: был сформирован развернутый перечень всех страниц сайта с описанием их назначения, предполагаемой структуры и функциональных блоков. Ответ включал как рекомендации по визуальному оформлению, так и обоснование логики пользовательских сценариев, что соответствовало требованиям UX/UI-аналитики. Gemini сосредоточилась на выявлении недостатков интерфейса. В ответе были представлены краткие предложения по его обновлению и предложена общая структура редизайна. Однако подробное описание всех страниц, как указывалось в промпте, предоставлено не было, что свидетельствует о частичном выполнении запроса. Claude не смогла получить доступ к содержимому сайта, вероятно, из-за технических ограничений. Несмотря на это, модель сгенерировала детализированную структуру сайта, основываясь исключительно на содержании промпта. Это говорит о том, что модель способна создавать интерфейсную архитектуру в контексте гипотетического сайта, но без учета реального контента и проблем существующего дизайна. Таким образом, критериям успешности в наибольшей степени соответствует результат модели ChatGPT, чье описание было выбрано для последующего составления промптов.

**Генерация промпта для визуального ряда.** Для генерации изображений интерфейса сайта киностудии с помощью GANs необходимо предварительно составить корректный и детализированный текстовый запрос (промпт) с помощью LLMs. От качества и структуры промпта напрямую зависит точность и релевантность полученного изображения; поэтому моделям была поставлена задача сформировать как короткий, так и длинный промпт, способный адекватно описывать предполагаемую UX/UI-структуру главной страницы сайта киностудии «Уралфильм» на английском языке. Выбор английского языка обусловлен тем, что большинство современных генеративных моделей не обладают поддержкой генерации изображений по текстам на других языках, включая русский язык, особенно в специализированных тематических областях, таких как веб-дизайн.

Формулировка исходного запроса состояла из следующего: «На основе предлагаемых изменений домашней страницы «Уралфильм» в соответствии с принципами UX/UI-дизайна составь короткий и

длинный промпт на английском языке для генерации, включающий в себя структуру и расположение элементов». Данная задача была предложена трем LLMs с целью сравнить их подходы к формированию коротких и длинных промптов, а также выявить особенности генерации текста в контексте дизайн-концепций (см. приложение).

При генерации промптов снова обнаружилось расхождение в логике трех LLMs. Так, несмотря на общую логичность и стилистическую стройность формулировок, ChatGPT при создании длинного промпта акцентирует внимание на эмоционально-оценочных характеристиках интерфейса. Например, она использует эпитеты («sleek», «cinematic», «professional»), в то время как конкретное описание структуры страницы и расположения блоков (навигация, баннеры, карточки проектов) осталось непроработанным. Это может повлиять на результат визуальной генерации, так как GANs скорее будут подражать стилю, чем выстраивать структуру. Gemini проявляет склонность к более подробному описанию структуры интерфейса, включая навигационное меню, типы кнопок, расстановку блоков и визуальные элементы. Но даже при этом часть описаний остается достаточно обобщенной и направленной скорее на стилевое восприятие, чем на реалистичное размещение элементов для генерации макета в точной сетке. Claude делает акцент на визуальной стилистике, часто употребляя слова, описывающие атмосферу («cinematographic aesthetics», «elegant», «refined UX»). При этом отсутствует четкая композиция интерфейса относительно компонентов (меню, кнопки, разделы, баннеры), что снижает точность генерации на этапе построения визуального макета.

Несмотря на общее понимание задачи и заданного контекста, все три модели склонны замещать строгую структурную логику эмоционально-оценочными описаниями, что приводит к расплывчатым или излишне обобщенным результатам при передаче задачи GANs. Длинные промпты фокусируются на визуальном и стилистическом описании, но не передают точного расположения блоков, а короткие промпты выполняют роль «обобщенной идеи» или настроения, но почти не применимы к контролируемой генерации структуры.

**Генерация изображений с помощью GANs.** Следующим этапом стало тестирование их эффективности в шести генеративно-сопоставительных сетях (GANs). Каждой из шести нейросетей (Midjourney, DALL-E, Stable Diffusion, Recraft, Leonardo AI и GPT-4o) были переданы как краткие, так и развернутые промпты, сформированные на основе анализа UX/UI-структуры, с целью сравнить качество восприятия и интерпретации текстового описания различными GANs. Все модели произвели прототипирование на основе промптов.

**Оценка качества ИИ-прототипов и их визуального ряда.** Для оценки полученных изображений использовалась специализированная нейросеть Roast My Design, созданная для автоматизированного анализа визуальных интерфейсов и основанная на принципах UX/UI-экспертизы и дизайнерской эвристики. Она обучена на обширной выборке пользовательских интерфейсов и способна проводить полуавтоматический аудит макетов с акцентом на пользовательский опыт. Модель анализирует изображения и выносит оценочные суждения по шести критериям:

- 1) визуальная иерархия (насколько хорошо выстроена композиционная структура, видны ли основные точки фокуса и центры внимания);
- 2) читаемость и типографика (соответствие размеров шрифта, контраста и читаемости);
- 3) сетка и выравнивание (корректность размещения элементов и соблюдение правил визуального ритма);
- 4) интуитивность навигации (понятность структуры и логика переходов);
- 5) современность дизайна (визуальное соответствие текущим тенденциям и лучшим практикам UI/UX);
- 6) целостность композиции (общая гармония и сбалансированность визуальных решений).

Система выдает сводную количественную оценку в баллах и развернутые пояснения, что позволяет сравнивать различные визуальные концепты. Наивысший балл среди всех представленных вариантов получила генерация GPT-4o на основе длинного промпта, составленного с использованием ChatGPT (рис. 2a).

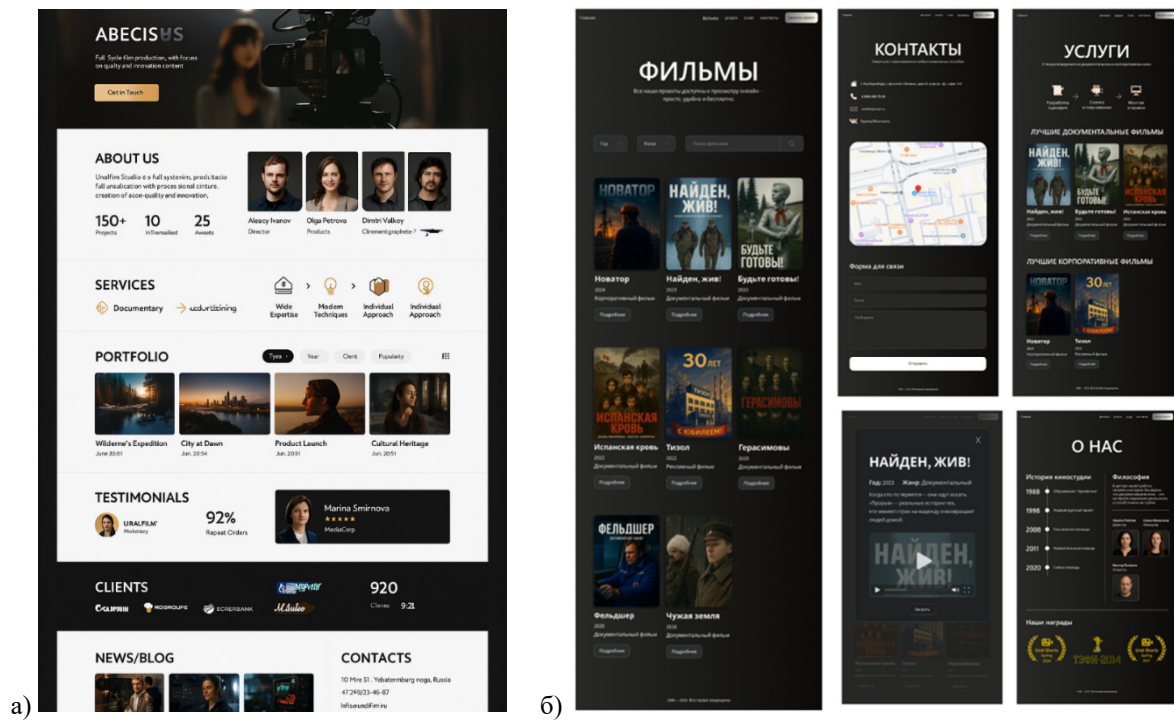


Рис. 2. Генерация: а) интерфейса в GPT-4o на основе длинного промпта от ChatGPT; б) кликабельных страниц корпоративного сайта киностудии с сохранением стилистики

На основе этой генерации была спроектирована главная страница сайта. При этом визуальное решение не было скопировано напрямую, а использовалось в качестве стилистического ориентира. Были адаптированы структура, цветовая палитра и ключевые композиционные принципы, но сохранились индивидуальность макета и его соответствие UX/UI-структуре, сгенерированной LLM (рис. 2б). При помощи функции сохранения визуального стиля, доступной в GPT-4o, удалось обеспечить единый стиль при создании других страниц сайта.

## 2. Второй этап эксперимента: сравнительное юзабилити-тестирование интерфейсов

Результаты тестирования по шести метрикам (TSR, TOT, UER, пре- и посттест Горбова – Шульте, CSI) показаны в таблице 1. Таблица 2 представляет расчет различий ( $\Delta$ ) для всех метрик, кроме TSR.

Таблица 1. Данные юзабилити-тестирования

Интерфейс	Метрика	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Mean	SD
Ручной	TSR	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0,00
	TOT	167	202	21	41	70	54	48	43	53	26	72,50	61,20
	UER	4	4	2	2	2	1	1	0	0	0	1,60	1,51
	Горбов - Шульте (претест)	128	127	55	95	90	84	130	71	70	52	90,20	29,63
	Горбов - Шульте (посттест)	61	119	58	65	71	58	110	63	65	48	71,80	23,38
	CSI	90%	75%	65%	92%	85%	100%	70%	95%	97%	70%	84%	0,13
Сгенерированный ИИ	TSR	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0,00
	TOT	89	165	20	16	18	22	46	17	16	20	42,90	48,61
	UER	2	4	0	0	0	0	2	0	0	0	0,80	1,40
	Горбов - Шульте (претест)	93	114	65	84	86	70	144	70	72	82	88,00	24,33
	Горбов - Шульте (посттест)	89	95	59	66	67	58	92	56	61	53	69,60	16,07
	CSI	92%	60%	87%	92%	90%	100%	45%	80%	97%	82%	83%	0,17

Таблица 2. Расчет показателя различия ( $\Delta$ ) по пяти метрикам

Ассессор	Метрика														
	TOT <sub>h</sub> *	TOT <sub>ai</sub> *	$\Delta TOT$	UER <sub>h</sub>	UER <sub>ai</sub>	$\Delta UER$	GS <sub>PREh</sub>	GS <sub>POSTh</sub>	GS <sub>PREai</sub>	GS <sub>POSTai</sub>	$\Delta GS_h$	$\Delta GS_{ai}$	CSI <sub>h</sub>	CSI <sub>ai</sub>	$\Delta CSI$
A1	167	89	78**	4	2	2	128	61	93	89	67	4	90	92	2
A2	202	165	37	4	4	0	127	119	114	95	8	19	75	60	-15
A3	21	20	1	2	0	2	55	58	65	59	-3	6	65	87	22
A4	41	16	25	2	0	2	95	65	84	66	30	18	92	92	0
A5	70	18	52	2	0	2	90	71	86	67	19	19	85	90	5
A6	54	22	32	1	0	1	84	58	70	58	26	12	100	100	0
A7	48	46	2	1	2	-1	130	110	144	92	20	52	70	45	-25
A8	43	17	26	0	0	0	71	63	70	56	8	14	95	80	-15
A9	53	16	37	0	0	0	70	65	72	61	5	11	97	97	0
A10	26	20	6	0	0	0	52	48	82	53	4	29	70	82	12

Примечание. \* - индексы «h» и «ai» означают ручной (human) и ИИ-интерфейс (AI) соответственно; индексы PRE и POST – пре- и посттест Горбова – Шульте соответственно. \*\* - цвета ячейки обозначают: зеленый – изменения в пользу ИИ-интерфейса, серый – отсутствие значимых изменений, оранжевый – изменения в пользу ручного интерфейса.

Показатели теста Манна – Уитни говорят о значимых различиях между показателями TOT<sub>h</sub> и TOT<sub>ai</sub> (0,029\*\*). К значимому различию приближаются показатели UER<sub>h</sub> и UER<sub>ai</sub> (0,190). Разброс значений не выявил разницу для показателей  $\Delta GS$  и CSI (0,853 и 0,971, соответственно).

Результаты юзабилити-тестирования можно интерпретировать следующим образом. Оба варианта дизайна высокофункциональны: выполнение задачи зафиксировано на уровне 100% во всех случаях. Объективный анализ динамики когнитивной лабильности (тест Горбова – Шульте) не выявил значимого падения показателей при работе с ИИ-интерфейсом; более того, средний показатель ошибки незначительно снизился (с 71,8 до 69,6 пункта), не было зафиксировано ни одного случая снижения лабильности, а в восьми случаях из 10 лабильность мозга повысилась, тогда как для ручного интерфейса это наблюдалось только в половине случаев.

Показатели «древесного» теста (эффективность пользователя во время выполнения задачи) оказались наиболее благоприятными для ИИ-интерфейса. Так, ИИ-прототип резко (в два раза!) и достоверно повысил скорость выполнения задачи, а также в половине случаев позволил снизить число ошибок ассессоров (хотя и незначительно). Семи пользователям из десяти ИИ-прототип позволил пройти поисковый путь без ошибок, тогда как ручной интерфейс дал такую возможность только трем из них.

Однако показатели субъективной удовлетворенности показывают смешанную динамику. Половина пользователей не почувствовали более высокой удовлетворенности от ИИ-интерфейса; двоим он понравился больше, трем – меньше, чем ручной дизайн. Этот результат требует более подробного изучения; требуется «заглянуть внутрь» теста и определить, какие показатели удовлетворенности снижают возможности ИИ-агента.

## Заключение

Общие **выводы** исследования можно сформулировать следующим образом:

1. Пока большие языковые модели не до конца готовы к анализу данных и созданию четких промптов для коммерческого медиадизайна. Это связано с низким уровнем доступности данных для анализа конкурентной визуальной среды, низкой обученностью моделей с точки зрения анализа и разработки информационной архитектуры и структуры порталов, склонностью моделей к описательности и метафоризации промптов.

2. Лучший промпт и лучший ИИ-прототип удалось создать на основе ChatGPT. При этом оценочная дизайнерская нейросеть Roast My Design показала, что может служить многосторонним инструментом независимой оценки генеративного медиадизайна.

3. ИИ-агенты способны предложить вариант прототипирования и визуального оформления портала, который работает как минимум не хуже, а по некоторым параметрам – вдвое лучше ручного дизайна. Так, объективный анализ поведения и когнитивной лабильности ассессоров позволяет



утверждать, что они справляются с различными задачами на поиск целевой и обобщенной информации вдвое быстрее при сопоставимом (и даже несколько более низком) количестве ошибок.

4. Субъективная удовлетворенность пользователя, однако, нуждается в более детальном исследовании. Осталось неясным, за счет каких именно параметров анкеты снизился результат. Можно предположить, что пользователи могут делиться на группы, интуитивно предпочитающие ручной дизайн не-человеческому, и наоборот. Детальное изучение параметров субъективной оценки поможет полнее раскрыть причины относительно низкой удовлетворенности пользователей ИИ-интерфейсом.

В целом наши результаты позволяют утверждать, что ИИ-интерфейсы для индустрии могут работать не хуже, чем созданные человеком, сокращая время пользователей на выполнение задач.

#### Источник финансирования

Работа С. С. Бодруновой поддержана проектом «Накопительное общественное мнение и кумулятивная делиберация: теоретизация, методы анализа и оценка качества современного делиберативного процесса» (2025 – 2027), № 128786104, этап 1.

#### Список литературы

1. Chemerys H., Demirbilek M., Bryantseva H. Fundamentals of UX/UI design in professional preparation of the future Bachelor of Computer Science // AIP Conference Proceedings. Vol. 2453(1). Malang: AIP Publishing, 2022. P. 030025.
2. Шуваев Я. UX/UI дизайн для создания идеального продукта: Полный и исчерпывающий гид. Litres, 2022.
3. Bodrunova S. S., Yakunin A. V. U-index: an eye-tracking-tested checklist on webpage aesthetics for university web spaces in Russia and the USA. In: Marcus, A., Wang, W. (eds), Design, User Experience, and Usability: Theory, Methodology, and Management (DUXU 2017). Held as Part of HCI International 2017, Vancouver, Canada, July 9–14, 2017, Proceedings, Part I. Lecture Notes in Computer Science, vol. 10288. Cham: Springer, 2017. P. 219–233.
4. Малашук Е. В. Влияние искусственного интеллекта на трансформацию сферы веб-дизайна // Актуальные исследования. 2023. № 39(169). С. 17–23.
5. Никулин А., Самыгин Г., Степанцева Т. Практическое применение технологии нейросетей в коммуникационном дизайне на примере высшего учебного заведения // Вестник ГГУ. 2023. № 2. С. 107–113.
6. Yakunin A. V., Bodrunova S. S. Cumulative distortions in usability testing: combined impact of web design, experiment conditions, and type of task and upon user states during internet use. In: Soares, M.M., Rosenzweig, E., Marcus, A. (eds), Design, User Experience, and Usability: UX Research, Design, and Assessment: HCI'2022 Proceedings. Lecture Notes in Computer Science, vol. 13321. Cham: Springer, 2022. P. 526–535.
7. Yakunin A. V., Bodrunova S. S. Cumulative impact of testing factors in usability tests for human-centered web design // Future Internet. 2022. Vol. 14(12), #359.
8. Djulianto M. V., Wardhanie A. P., Candraningrat C. The Role of Perceived Usability, Satisfaction, and Customer Trust in Design and Developing User Loyalty Edutech Website // Business and Finance Journal. 2022. Vol. 7(2). P. 183–194.
9. Nasrullah H. Introduction to UI/UX Design: Key Concepts and Principles. Baku: Baku Engineering University, Information Technologies (SABAH groups), 2023.
10. Combinatorial optimization of graphical user interface designs / A. Oulasvirta, N. Dayama, M. Shiripour, M. John, A. Karrenbauer // Proceedings of the IEEE. 2020. Vol. 108(3). P. 434–464.
11. Светличная К., Ефимова Е. Теория и практика UI/UX дизайна компьютерных приложений: XXI Международная научно-практическая конференция. Ростов-на-Дону, 2021. С. 94–101.
12. Норман Д. Дизайн привычных вещей. М: Манн, Иванов и Фербер, 2022.
13. Shahriar S. GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network // Displays. 2022. Vol. 73, #102237.
14. Buschek D., Anlauff C., Lancher F. Paper2Wire: A case study of user-centred development of machine learning tools for UX designers // Proceedings of Mensch und Computer. 2020. P. 33–41.
15. Todi K., Lu W., Tang X. Sketchplore: Sketch and Explore with a Layout Optimiser // Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS'2016), June 4–8, 2016. ACM, 2016. P. 543–555.
16. Swearngin A., Li Y. Modeling mobile interface tappability using crowdsourcing and deep learning. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, 2019. P. 1–11.
17. Koch J., Paris-Saclay I. Design implications for Designing with a Collaborative AI // Proceedings of the AAAI 2017 Spring Symposium on Designing the User Experience of Machine Learning Systems. AAAI, 2017. P. 4.

18. Generative Agents: Interactive Simulacra of Human Behavior / J. Park, J. O'Brien, C. Cai, M. Morris, P. Liang // Proceedings of the 36<sup>th</sup> Annual ACM Symposium on User Interface Software and Technology. ACM, 2023. P. 1–40.
19. Вэн В. Переопределение генеративного дизайна: теория, история культуры и технологические аспекты: дис. ... канд. искусствоведения. Томск, 2025. URL: [iik.tsu.ru/sites/default/files/ven\\_vey\\_dissertaciya\\_24.06.25\\_g.pdf](http://iik.tsu.ru/sites/default/files/ven_vey_dissertaciya_24.06.25_g.pdf) (дата обращения 25.07.2025).
20. Midjourney Documentation. URL: <https://docs.midjourney.com/> (дата обращения: 26.07.2025).
21. Ruskov M. Grimm in Wonderland: Prompt Engineering with Midjourney to Illustrate Fairytales // Proceedings of the 19<sup>th</sup> Conference on information and research science connecting to digital and library science. 2023. P. 181–190.
22. Brown T., et al. Language models are few-shot learners // Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 1877–1901.
23. Large language models are zero-shot reasoners / T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, Y. Iwasawa // Advances in Neural Information Processing Systems. 2022. Vol. 35. P. 22199–22213.
24. Yuan T. User portrait based on artificial intelligence // Proceedings of the International Conference on Frontier Computing. Singapore: Springer Nature Singapore, 2022. P. 359–366.
25. Christoforakos L., Diefenbach S. Idealization effects in UX evaluation at early concept stages: challenges of low-fidelity prototyping // Proceedings of the International Conference on Applied Human Factors and Ergonomics. Cham: Springer International Publishing, 2018. P. 3–14.
26. Best practices for prompt engineering with the OpenAI API. URL: [help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api](https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api) (дата обращения: 26.07.2025).
27. Благодельский А. С. Применение искусственного интеллекта в фронтенд-разработке: от теории к практике // Вестник науки. 2024. № 3(5(74)). С. 1066–1080.
28. Palas P. The Ultimate Guide to Headless CMS: Everything you need to know to choose the right CMS. Traverse City: Independently published, 2017. 68 p.
29. Application of Artificial Intelligence Technology in Electromechanical Information Security Situation Awareness System / X. Liu, Z. Li, Z. Tang, X. Zhang, H. Wang // Scalable Computing: Practice and Experience, 2024. Vol. 25(1). P. 127–136.
30. Гонатаев Р. Г. Преимущества разработки веб-приложений с применением фреймворков // Тенденция развития науки и образования. 2021. № 70(1). С. 12–15.
31. Щелик С. Ю., Солдатова А. В. Использование нейросетей и возможностей генеративного дизайна в брендинге // Практический маркетинг. 2024. № 7. С. 78–84.
32. Евстафьев В., Тюков М. Искусственный интеллект и нейросети: практика применения в рекламе. Litres, 2024.
33. Малашук Е. В. Методологические аспекты повышения эффективности маркетинга социально значимых организаций посредством применения инструментов искусственного интеллекта в веб-дизайне // Индустриальная экономика. 2023. № 5. С. 174–183.
34. Чернцева Ю. Ю. Использование генеративных инструментов в проектах UX/UI дизайна: магистерская диссертация по направлению подготовки 54.04.01 – дизайн. Томск, 2023. URL: <https://vital.lib.tsu.ru/vital/access/services/Download/vital:19201/SOURCE01> (дата обращения: 26.07.2025).
35. Мамиконян О. 15 % российских дизайнеров переживают, что нейросети смогут их заменить // Forbes.ru. 12 июля 2023 года. URL: <https://www.forbes.ru/forbeslife/492639-15-rossijskih-dizajnerov-perezivaut-cto-nejroseti-smogut-ih-zamenit> (дата обращения: 26.07.2025).
36. Ultes S., Maier W. Blending task success and user satisfaction: Analysis of learned dialogue behaviour with multiple rewards // Proceedings of the 22<sup>nd</sup> Annual Meeting of the Special Interest Group on Discourse and Dialogue. ACL, 2021. P. 403–410.
37. Brooke J. SUS: A quick and dirty usability scale // Usability evaluation in industry. 1996. No. 189(194). P. 4–7.
38. Laubheimer P. Tree Testing. Part 2: Interpreting the Results. URL: [nngroup.com/articles/interpreting-tree-test-results/](https://nngroup.com/articles/interpreting-tree-test-results/) (дата обращения: 26.07.2025).

**Приложение. Версии коротких и длинных промптов для генерации визуального ряда корпоративного портала киностудии «Уралфильм»**

Большая языковая модель (LLM)	Короткий промпт	Длинный промпт
ChatGPT	Modern web interface for «Uralfilm», sleek, cinematic, professional, engaging design.	Design a homepage for a professional video production company specializing in documentaries, commercials. Create a sleek, cinematic interface with high-quality visuals, video banners, and modern typography. Highlight key services, portfolio, and client testimonials with a clean, intuitive layout. Use muted colors, subtle animations, and dynamic sections to reflect creativity, professionalism. Emphasize storytelling and visual impact.
Gemini	Clean modern film studio homepage, prominent header navigation, captivating hero section with film reel or cinematic scene, showcase of recent projects with image previews, services overview, brief about us section, clear «View Portfolio» button, clear «Contact Us» button.	Modern professional film studio homepage for «Uralfilm», minimalist header with «Uralfilm» logo and navigation («About Us», «Portfolio», «Services», «News», «Contact»), large hero section with a high-quality video still or cinematic image related to film production, subtle overlay text «Your Vision, Our Expertise» and a short introduction, prominent rounded buttons «View Portfolio» and «Get in Touch» below the hero text, visually appealing grid of recent film projects with individual image thumbnails and titles, a section outlining services with simple icons (e.g., camera, film strip, editing tools) and brief descriptions (pre-production, production, post-production), a small «About Us» section with a brief text excerpt and a link, a teaser for the latest news or blog post with a title and short preview, clean and balanced layout, readable sans-serif typography, professional and slightly muted color palette, subtle animations on hover, responsive design elements.
Claude	Minimalist UX/UI website homepage for documentary and advertising film studio, clean modern design, black and dark blue palette, professional cinema production interface, soft shadows.	Ultra-modern documentary film studio website homepage in deep black and charcoal color palette, pixel-perfect UX/UI design with asymmetrical grid layout, floating header with minimal navigation, large cinematic showreel background, high-contrast typography, elegant project thumbnails, subtle motion design elements, professional graphic interface with sharp edges, micro-interactions, seamless scrolling experience, cinematographic aesthetic, refined user experience design.

## Влияние графического интерфейса на когнитивный контроль пользователей при решении стрессогенных задач

А. В. Якунин, С. С. Бодрунова

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

**Аннотация.** Использование современных интерактивных систем аналитики, информирования и поддержки принятия решений все чаще осуществляется в условиях многозадачности, при которой пользователь совмещает несколько видов деятельности одновременно. Именно многозадачность выступает ключевым фактором снижения производительности вследствие эмоционального стресса и роста количества ошибок. При этом влияние факторов совмещенной деятельности на интеллект пользователя представляется недостаточно исследованной областью HCI. Предлагаемое исследование посвящено влиянию графического интерфейса на способность пользователя к когнитивному контролю при решении совмещенных задач разной сложности в условиях стресса. В качестве стимулов выступают два сайта с разным качеством юзабилити, на которых тестируются двойные задачи двух типов: с совмещением содержания разных модальностей (вербального и образного) и тождественных по модальности (два типа вербального содержания). В качестве зависимых переменных использованы метрики психофизиологических состояний пользователя – эффективность когнитивного контроля и уровень психоэмоционального напряжения (в показателях ЧСС). Как показали результаты эксперимента, для задач с тождественным по модальности содержанием когнитивный контроль зависит от социального контекста. В случае задач с совмещением содержания разных модальностей более значимо оказывается юзабилити интерфейса.

**Ключевые слова:** эргономика, юзабилити-тестирование, графический интерфейс, модель контекстуальной точности, опыт пользователя, когнитивный контроль, двойная задача, стресс-тестирование.

## The impact of graphical interface on users' cognitive control in solving stressful tasks

A. V. Yakunin, S. S. Bodrunova

School of Journalism and Mass Communications, St. Petersburg State University, St. Petersburg, Russia

**Abstract.** The use of modern interactive systems of analytics, information and decision support is increasingly carried out in the context of multitasking, in which the user performs on several types of cognitive and operational activity simultaneously. It is multitasking that is the key factor in the decrease in productivity due to emotional stress and the increase in the number of errors. At the same time, the impact of factors of combined activity on the user intelligence seems to be an insufficiently studied area of HCI. The proposed study assesses the impact of the quality of graphical user interface (GUI) on the user ability of cognitive control when solving combined tasks of varying complexity under stress. The stimuli are two websites of differing usability quality, on which multitasking is tested with the help of two types of tasks, namely a combination of content of different modalities (textual plus figurative content) for the first task type and a combination of content of identical modality (textual plus textual content). The metrics of the user psychophysiological states used as dependent variables are cognitive control efficiency and the level of psycho-emotional stress (examined via heartbeat rates). As the results of the experiments showed, for tasks with the content identical in modality, cognitive control is shaped by the social environment of task performance. In the case of tasks with modality-diverse content, interface usability becomes a more significant factor.

**Keywords:** ergonomics, user experience, GUI, usability testing, contextual fidelity model, cognitive control, multitasking, combined tasks, stress testing.

### Введение

В последние годы в работах о человек-компьютерном взаимодействии активно обсуждается роль пользовательского интеллекта как фактора взаимодействия человека и интерфейса. Как свидетельствует ряд исследований [1-3], рациональный аспект UX оказывает значительно большее влияние на эффективность интерактивного взаимодействия, чем эстетическое восприятие и эмоциональная оценка функциональности продукта, которые активно изучались ранее. Рациональность информационной архитектуры и очевидность функциональных возможностей, отраженная в графическом интерфейсе, имеют большее значение для производительности, чем эстетическая привлекательность и сенсорные впечатления от визуального аспекта дизайна интерфейса [4-5]. Более того, удобство для интеллекта превращается в один из основных факторов формирования

положительного эстетического впечатления. Как правило, при этом пользователи склонны недооценивать собственные интеллектуальные реакции на качество дизайна продукта, что отражается и на методологии экспертной оценки юзабилити, в которой доминируют субъективные критерии оценки пользовательского опыта на основе воспринимаемого юзабилити и эстетического впечатления. Тот факт, что сама природа эстетического суждения опирается прежде всего на когнитивные процессы («что рационально, то красиво») [3], во многих исследованиях попросту игнорируется.

Между тем понимание психофизиологических механизмов когнитивной деятельности, реализуемой в интерактивном взаимодействии, открывает дорогу для поиска способов успешной адаптации оператора к любым режимам эксплуатации человеко-машинных систем, в том числе экстремальным. Особенно это касается режимов взаимодействия, связанных со стрессом вследствие высокой когнитивной нагрузки, – в случае многозадачности, например. Как показывают некоторые исследования [6-7], влияние психоэмоционального стресса на интеллектуальную производительность пользователя наиболее деструктивно.

Поэтому, на наш взгляд, одним из наиболее перспективных направлений исследований в данной области выступает изучение факторов недирижабельной регуляции когнитивной деятельности в условиях высокой информационной нагрузки. В данном случае воздействие на интеллект пользователя осуществляется только за счет набора факторов, ограниченного ключевыми аспектами процесса интерактивного взаимодействия – свойствами задачи, свойствами продукта (качеством и особенностями дизайна интерфейса), свойствами и состоянием пользователя, а также условиями проведения эксперимента. Эти четыре группы факторов, взятые во взаимодействие, известны как *модель контекстуальной точности* (contextual fidelity model, CFM [8]). Предлагаемое исследование демонстрирует опыт применения данной модели в исследованиях одного из ключевых проявлений активного интеллекта – когнитивного контроля, то есть способности ускоренного рационального фокусного контроля в условиях меняющегося контекста и многозадачности.

Еще одним современным ограничением эргономических исследований является стремление ученых сосредоточить эксперимент на выполнении какой-либо одной задачи, которая изучается; этого требует сам дизайн исследования, предполагающий, что зависимые переменные принадлежат одному изучаемому процессу. Однако в реальности пользователям интерфейсов часто приходится выполнять множественные задачи одновременно и существовать в условиях многозадачности (multitasking). Многозадачность становится критическим фактором в условиях стресса – выполнения срочных, эмоционально вовлекающих, резко усложняющихся задач. Именно такие сочетания условий способны проверить, насколько эффективен когнитивный контроль.

С этой точки зрения представляет безусловный интерес детальное исследование функционирования двойных задач как наиболее распространенных в повседневной практике коммуникации человека и компьютерных систем. При этом распределение когнитивных ресурсов между основной задачей и дополнительной (например, в случае ведения уточняющей переписки одновременно с анализом основной информации) во многом определяется не только сложностью стимулов, но и их перцептивной модальностью: синхронная обработка текста и изображения требует значительно меньше усилий, чем обработка двух текстов или двух изображений одновременно. Поэтому предлагаемое исследование сосредоточено на анализе эффектов, возникающих при исполнении двух типов задач – более легкой, совмещающей обработку текстового и образно-графического контента, и более трудной, совмещающей два типа текстового содержимого.

В силу вышесказанного мы скомбинировали в дизайне эксперимента три условия, которые часто игнорируются в юзабилити-тестировании: рациональный когнитивный контроль, двойную задачу и действие в условиях стресса. Наш дизайн исследования, таким образом, существенно ближе к реальности выполнения критически значимых задач, зависимых от интерфейса.

Ниже мы описываем постановку задачи, приводим более подробный обзор теоретических подходов к анализу усложненных задач.

### Постановка задачи

В предлагаемом исследовании рассматривается взаимодействие и влияние на когнитивный контроль пользователя трех факторов CFM – среды использования продукта, его дизайна (графического интерфейса) и свойств задачи. Применительно к организации исследования это означает измерение воздействия графического интерфейса сайтов с разным качеством дизайна в условиях индивидуального и группового выполнения совмещенных задач двух типов – подразумевающей синхронную работу с текстовыми и образными стимулами («легкой») и работу с двумя типами текстовых стимулов («сложной»). В качестве зависимых переменных используются метрики психофизиологических состояний пользователя: эффективность когнитивного контроля («производительность») и уровень психоэмоционального напряжения (в показателях уровня стресса).

### Теория

К настоящему моменту методы исследования интеллекта как фактора оценки эффективности интерфейсов развиваются по двум направлениям.

Исследования в области когнитивной нагрузки рассматривают зависимость эффективности деятельности пользователя от сложности исполняемых задач – структурной [9] и когнитивной [10]. Основной целью исследований при этом выступает оценка влияния когнитивной сложности задачи на эстетическое восприятие и производительность пользователей. Поэтому анализ факторов, оказывающих на него влияние, является приоритетным для авторов таких исследований [11]. Но, как правило, большинство исследований в данной области исходит только из свойств задачи, игнорируя влияние других факторов, что не отвечает современным представлениям о дифференцированной природе опыта и, в частности, интеллектуальной его компоненты [12].

Второе направление, методы и концепции которого пока не так широко распространены в HCI, – это исследования в области когнитивного контроля [13-14]. Методы данного направления наиболее подходят для изучения взаимодействия «человек – компьютер» в условиях стресса и высокой информационной нагрузки, поэтому их дальнейшее развитие в области оценки интерфейсов и научной визуализации представляется наиболее перспективным.

Главный объект исследований в данном направлении – когнитивный контроль – характеризуется как способность пользователя быстро переключаться с одной задачи на другую и концентрироваться на решении задачи при наличии конкурентных раздражителей, реорганизовывать когнитивную активность в зависимости от меняющегося ситуационного контекста, подавлять процессы и реакции, нерелевантные цели деятельности [15]. Также с данным явлением связана способность пользователя гибко распределять ментальные ресурсы для управления действиями с целью достижения поставленных целей [16]. В ряде исследований когнитивный контроль как фактор производительности пользователя подразумевает систему взаимосвязанных процессов, ответственных за планирование деятельности, когнитивную гибкость, абстрактное мышление, инициацию релевантных и торможение нерелевантных действий, а также селекцию релевантной сенсорной информации [17-19]. В целом на современном этапе актуальные определения когнитивного контроля могут быть сведены к трем ключевым функциональным характеристикам: координации когнитивных процессов, поддержке релевантных цели действий и самоконтролю результатов деятельности.

Фокус современных исследований когнитивного контроля сосредоточен на проблеме ограниченного доступа к когнитивным ресурсам, используемым субъектом в совмещенной (многозадачной) деятельности. В условиях нарастающей сложности задачи индивид вынужден мобилизовывать ресурсы внимания и лабильности интеллекта, что возможно только до определенного предела. Понимание факторов и механизмов такой мобилизации с перспективой ее расширения способно существенно обогатить подходы к проектированию интерактивных информационных систем и веб-приложений, используемых в условиях высокой информационной нагрузки (например, многозадачности в офисной работе и обучении, мультиэкранного и фоновое потребления контента, переключения между интерфейс-зависимыми задачами во время техногенной аварии, управления высокоскоростным транспортом, высокоточным производством и т.д.).



На современном этапе подходы к данной проблеме наиболее полно представлены в концепции Д. Канемана, разработавшего подходы к оценке эффективности работы интеллекта при решении совмещенных задач – основной и дополнительной [20]. Исследования Канемана способны помочь в решении важных проблем человек-машинного взаимодействия, связанных с деятельностью в условиях стресса и многозадачности, в частности – в преодолении эффекта структурной интерференции, согласно которому близость задач с точки зрения перцептивной модальности пропорционально усиливает сложность их одновременного выполнения [20]. Человеку легче выполнить две задачи, одна из которых подразумевает работу с образами, а другая – со словом, чем две задачи, подразумевающие работу со словами. При анализе с помощью компьютерной визуализации сложных процессов в режиме реального времени, нередко осуществляемом в условиях высокой информационной нагрузки, управление эффектом интерференции задач имеет особое значение для снижения ошибочных решений.

В концепции Канемана сформулирован ряд факторов, влияющих на механизм эффективного распределения когнитивных ресурсов [20]:

- структурные характеристики задачи: чем сложнее структура деятельности, тем сильнее мобилизуются ресурсы внимания;
- предварительная оценка пользователем объема ресурсов, необходимых для решения задачи;
- уровень психофизиологической активации субъекта деятельности, определяющий способность к ментальной мобилизации: чем выше утомление, тем ниже способность к активации и производительность.

Характерно, что в концепции Канемана структурные характеристики задачи более важны для эффективности распределения ресурсов, чем сознательное целеполагание пользователя: уровень физиологической активации в большей степени зависит не от мотивации пользователя, но от структурной сложности задачи (наличия синхронных задач, количества критериев оценки деятельности и т.д.).

В дальнейшем другие исследователи дополнили список факторов, включив в него компетентность субъекта (наличие в памяти необходимых знаний) и степень автоматизации его навыков [21].

Таким образом, обзор существующих направлений исследования интеллекта как фактора оценки эффективности интерфейсов позволяет сделать следующие выводы.

Во-первых, оба направления в описании генезиса интеллектуальных реакций и эффектов исходят исключительно из свойств задачи (ее структурной и перцептивной сложности + наличия дополнительных задач) и свойств пользователя (компетентности и уровня активации), игнорируя влияние других факторов: дизайна продукта, выступающего стимулом, и условиями его использования. При этом, констатируя необходимость рационального распределения ресурсов в условиях перегрузки, существующие исследования не раскрывают в полной мере факторы такого распределения, ограничиваясь описанием внутренних стратегий преодоления стресса [22], т. е. опять-таки апеллируя к способностям субъекта деятельности, не пытаясь изменить свойства инструмента деятельности или условия его использования.

Во-вторых, методы и подходы к исследованию когнитивного контроля как интегративной способности пользователя к сохранению интеллектуальной эффективности в неблагоприятных условиях работы представляются перспективными для решения задач человек-компьютерного взаимодействия в условиях высокой информационной нагрузки. При этом следует уделять внимание не только состоянию пользователя и структуре задач, но и влиянию среды взаимодействия, а также качеству микроэргономики интерфейса с продуктом. На сегодняшний день такого внимания в абсолютном большинстве исследований, посвященных произвольной регуляции и когнитивному контролю в стрессогенном окружении, им не уделяется.

В связи с этим особое значение, как представляется, имеет применение описанной выше модели контекстуальной точности юзабилити-тестирования (CFM) для изучения влияния комбинаторных эффектов взаимодействия на когнитивную сферу испытуемого. Анализ воздействия каждого из

факторов CFM на опыт пользователя в различных сочетаниях представляет несомненную ценность для понимания механизмов распределения когнитивных ресурсов при решении совмещенных задач в условиях высокой информационной нагрузки.

## Метод исследования

### 1. Дизайн юзабилити-тестирования

Дизайн эксперимента был основан на версии одного из тестов Д. Канемана, адаптированного для задач медиапотребления. Для моделирования режима совмещенной деятельности мы разработали комбинацию из двух заданий («основное+дополнительное»). Основное задание подразумевало форсированный поиск заголовков новостей по предложенным заранее ключевым словам в течение 20 секунд (общее время теста). В качестве дополнительного использовалось задание на распознавание 15 графических стимулов с интервалом в 1-2 с. В соответствии с теорией структурной интерференции мы разработали два варианта такой составной комбинации:

- *простая задача:* в дополнительном задании периодически синхронно с основным поиском предъявлялся для распознавания графический образ на бумажном носителе;
- *сложная задача:* в дополнительном задании периодически синхронно с основным поиском распознавалась семантическая категория демонстрируемого слова. Именно в этом случае фактор структурной интерференции вступал в силу – у испытуемых ожидалась трудности с синхронной идентификацией стимулов одной и той же (вербальной) модальности.

Форсированный режим поиска и интенсивное вторжение в него дополнительного задания формировали стрессогенную среду эксперимента.

В качестве стимулов для основного задания использовался графический интерфейс двух сайтов с различным качеством юзабилити веб-страницы, определенным по методике расчета разработанного нами ранее индекса юзабилити U-index [23]. Данный индекс позволяет оценить микроэргономику веб-страницы в баллах в диапазоне от 3 (минимальное значение) до 17 (максимум).

#### 1.1. Оценка производительности пользователя

Данный параметр измерялся с помощью двух психодиагностических методов оценки:

- тест на интеллектуальную лабильность «Расстановка чисел» [24: 552–553];
- тест на емкость оперативной памяти.

#### 1.2. Оценка психоэмоционального состояния пользователя

Данный параметр измерялся с помощью фиксации средней частоты сердечных сокращений аппаратным методом (кардиосенсором Xiaomi Smart Band 9).

Все психодиагностические методы оценки применялись до и после эксперимента. Разница между их показателями характеризует влияние факторов тестирования на когнитивный контроль пользователей.

#### 1.3. Оценка эффективности выполнения дополнительного задания

Данный параметр измерялся через фиксацию процента ошибок при распознавании образов (простая задача) и слов (сложная задача) во время проведения тестов. Измерение количества ошибок пользователя выступает аналогом параметра «эффективность выполнения задачи» (task solution rate, TSR) как части «древовидного» юзабилити-теста (tree test) в индустриальном юзабилити-тестировании [25].

### 2. Дизайн эмпирического исследования

Тесты выполнялись в восьми группах (по пять человек в каждой; участие было возможно только в одной группе), четыре из которых выполняли совмещенную задачу на сайте с высоким юзабилити («Коммерсантъ», U-index = 15) и четыре других – на сайте с низким U-индексом (РИА «Новости», U-index = 6). В соответствии с поставленными вопросами мы разработали структуру эксперимента из следующих этапов:

1. Стартовое измерение параметров производительности и психоэмоционального состояния перед основным заданием.

2. Выполнение заданий за отведенное время на сайтах «Коммерсантъ» и РИА «Новости». Измерение процента ошибок.

3. Финальное тестирование параметров производительности и психоэмоционального состояния для определения сдвигов в показателях пользовательского опыта.

### Результаты и их обсуждение

Результаты эксперимента представлены в таблицах 1–3.

Во-первых, обращает на себя внимание значительное влияние группового формата на когнитивный контроль; при этом характер влияния определяется типом задачи. Для сложной задачи социальная среда усиливает негативное влияние сложности на когнитивные процессы вне зависимости от качества стимулов, как можно видеть в таблицах 1–3, в случае сложной задачи для обоих сайтов характерно снижение когнитивного контроля (наиболее значительное – в групповом формате). Это видно по ключевым параметрам производительности (резкий рост интеллектуальной лабильности + снижение оперативной памяти), росту показателя стресса и числа ошибок. Для простой задачи это влияние, напротив, оказывается позитивным, при этом пользователь оказывается чувствителен к свойствам стимула – в тесте на сайте с высоким юзабилити групповой формат усиливает когнитивный контроль, что проявляется в минимальном количестве ошибок и активизации оперативной памяти. Интеллектуальная лабильность при этом не изменяется, что указывает на уровень сложности, недостаточный для мобилизации ресурсов внимания. Но для сайта с низким юзабилити ситуация иная: здесь наблюдается незначительное снижение всех параметров производительности и незначительный рост процента ошибок до средних значений в данном эксперименте. При этом снижение когнитивного контроля все равно сильнее проявляется в групповом формате. Однако среда эксперимента – не единственный фактор влияния на когнитивный контроль в случае простой задачи: как можно видеть из таблицы 3, разница между количеством ошибок на макетах разного качества двукратна. Таким образом, можно говорить как минимум о равном значении свойств продукта и условий его использования для простой совмещенной задачи.

Таблица 1. Результаты для интерфейса с высоким показателем U-INDEX:  
интеллектуальная лабильность, оперативная память, ЧСС

Формат теста	Параметр	Высокий U-index					
		простая задача			сложная задача		
		до теста	после теста	Δ	до теста	после теста	Δ
Групповой	интеллектуальная лабильность (баллы 1-10)	8	8	0	6	10	<b>4</b> значительный рост
	оперативная память	82	85	<b>3</b> слабый рост	83	72	<b>-11</b> значительное снижение
	ЧСС	65	65	0	62	70	<b>8</b> значительный рост
Индивидуальный	интеллектуальная лабильность (баллы 1-10)	9	9	0	8	9	<b>1</b> слабый рост
	оперативная память	90	90	0	85	80	<b>-5</b> умеренное снижение
	ЧСС	65	65	0	65	67	<b>2</b> слабый рост

Таблица 2. Результаты для интерфейса с низким показателем U-INDEX:  
интеллектуальная лабильность, оперативная память, ЧСС

Формат теста	Параметр	Низкий U-index					
		простая задача			сложная задача		
		до теста	после теста	Δ	до теста	после теста	Δ
Групповой	интеллектуальная лабильность (баллы 1-10)	7	6	<b>-1 слабое снижение</b>	4	9	<b>5 значительный рост</b>
	оперативная память	90	90	0	80	70	<b>-10 значительное снижение</b>
	ЧСС	66	66	0	65	69	<b>4 умеренный рост</b>
Индивидуальный	интеллектуальная лабильность (баллы 1-10)	4	5	<b>-1 слабое снижение</b>	7	8	<b>1 слабый рост</b>
	оперативная память	90	88	<b>-2 слабое снижение</b>	90	84	<b>-6 умеренное снижение</b>
	ЧСС	65	65	0	67	70	<b>3 умеренный рост</b>

Таблица 3. Ошибки при выполнении дополнительного задания, %

Формат теста	Высокий U-index		Низкий U-index	
	простая задача	сложная задача	простая задача	сложная задача
Групповой	7	30	15	33
Индивидуальный	10	25	20	25

Во-вторых, результаты эксперимента подтверждают наличие структурной интерференции в случае совмещения задач, близких по перцептивной модальности. В частности, сложная задача, совмещающая вербальные стимулы в основном и дополнительном заданиях, отмечена максимальным снижением когнитивного контроля по всем параметрам производительности (в обоих форматах тестирования) – максимальный процент ошибок распознавания, гиперлабильность ресурса внимания и резкое снижение объема оперативной памяти. Простая задача, совмещающая стимулы разных модальностей (вербальные + образные), напротив, демонстрирует максимальную степень когнитивного контроля для данного эксперимента – об этом свидетельствует минимальный процент ошибок распознавания, отсутствие резких колебаний интеллектуальной лабильности и оперативной памяти. Однако результаты эксперимента демонстрируют зависимость структурной интерференции не только от модальной близости совмещенных задач, но и от внешних условий эксперимента. В связи с этим обратим внимание на экстремальные значения параметров эксперимента именно в групповых тестах.

В-третьих, динамика психоэмоционального стресса, отражаемая частотой сердечного ритма, в целом коррелирует с интенсивностью когнитивных усилий. Так, повышение частоты сердечного ритма соответствуют фазам наиболее активного приложения умственных усилий, отмечая пиковые периоды мобилизации внутренних ресурсов.

Результаты указывают на синхронизацию эмоциональных, физиологических и когнитивных реакций в процессах когнитивного контроля и позволяет отслеживать динамику когнитивной нагрузки по активизации не только через показатели центральной нервной системы (память/лабильность внимания), но и вегетативной, связанной с бессознательной регуляцией физиологических процессов. При этом обращает на себя внимание зависимость уровня стресса от трех факторов:

- *сложности задачи*: колебания сердечного ритма и, соответственно, стрессовая реакция происходят только при решении сложных задач. При этом простые задачи в условиях низкого качества интерфейса (U-index) показывают незначительное снижение лабильности, тогда как сочетанные задачи не должны приводить к такому результату; на интерфейсе высокого качества такого эффекта не наблюдается, а при повышении сложности задачи лабильность растет;
- *качества дизайна*: важными эргономические и эстетические параметры дизайна (U-index) становятся для задач низкой сложности, где низкое качество дизайна повышает процент ошибки вдвое;
- *условий эксперимента*: на сайте с высоким U-index уровень тревоги достигает максимального значения именно в групповом тесте, что позволяет говорить об определяющем влиянии социальной среды на формирование стресса при решении сложных задач.

Таким образом, можно говорить о том, что задачи низкого и высокого уровня сложности различаются по набору ключевых факторов, влияющих на результат ассессоров при прохождении теста. Так, комбинированная задача высокой сложности, выполняемая в присутствии других людей (типичная ситуация производственной аварии, стресса в офисе или при обучении и т.д.), резко повышает уровень стресса, и компенсировать это качеством дизайна практически не удастся – ни в групповом, ни в индивидуальном режиме. Однако при решении задач с более низкой когнитивной нагрузкой качество интерфейса выступает ключевым фактором снижения точности и гибкости реакции.

### Выводы

В экспериментальных данных проявляются следующие паттерны влияния условий эксперимента и качества графического дизайна сайта на решение совмещенных задач.

Во-первых, можно признать, что для сложных задач фактор среды во влиянии на когнитивный контроль оказывается важнее качества стимула, а для простых – напротив, качество веб-дизайна в большей степени определяет когнитивные реакции на задачу в обоих форматах. Сдвиг в значениях параметров между макетами сайтов разного качества оказывается больше, чем между макетами одного качества в разных форматах.

Во-вторых, результаты эксперимента в группах, демонстрирующие экстремальные значения параметров когнитивного контроля, позволяют говорить как минимум о контекстуальной зависимости проявления эффекта структурной интерференции задач. Это является доводом в пользу более сложной природы данного явления, которое не может сводиться только к модальной близости задач и способно менять интенсивность в зависимости от сочетаний факторов тестирования.

### Источник финансирования

Исследование выполнено в рамках проекта «Накопительное общественное мнение и кумулятивная дилиберация: теоретизация, методы анализа и оценка качества современного дилиберативного процесса» Санкт-Петербургского государственного университета, проект №128786104, этап 1, 2025 год.

### Список литературы

1. A generalised semantic cognition account of aesthetic experience / I. Bara, R. J. Binney, R. Ward, R. Ramsey // *Neuropsychologia*. August 2022. Vol. 173, No. 13. 108288.
2. Beyond emoticons: combining affect and cognition in icon design / S. McDougall, I. Reppa, G. Smith, D. Playfoot // *Proceedings of EPCE 2009. LNCS (LNAI)*. Vol. 5639. Heidelberg: Springer, P. 71–80.
3. Freeman M. H. The Aesthetics of Human Cognition // *SSRN*. 2018. 3259227.
4. Otten R., Schrepp M., Thomaschewski J. Visual clarity as mediator between usability and aesthetics // *MuC'20: Proceedings of Mensch und Computer*. 2020. P. 11–15.
5. What Causes the Dependency between Perceived Aesthetics and Perceived Usability? / M. Schrepp, R. Otten, K. Blum, J. Thomaschewski // *International Journal of Interactive Multimedia and Artificial Intelligence*. 2020. Vol. 6. P. 78-85.
6. Yakunin A. V., Bodrunova S. S. Cumulative Distortions in Usability Testing: Combined Impact of Web Design, Experiment Conditions, and Type of Task and Upon User States During Internet Use // *Design, User Experience, and Usability: UX Research, Design, and Assessment: 11th International Conference, DUXU 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Proceedings: Springer Nature*, 2022. P. 526-535.

7. Yakunin A. V., Bodrunova S. S. The Influence of Usability on User Cognitive Activity: A Differential Intelligence Perspective // Social Computing and Social Media: 16th International Conference, SCSM 2024, Held as Part of the 26th HCI International Conference, HCII 2024, Washington, DC, USA, June 29–July 4, 2024: Proceedings. Part I. June 2024. P. 285 – 297.
8. Gilal N. G., Zhang J., Gilal F. G. The four-factor model of product design: scale development and validation // J. Prod. Brand Manag. 2018. Vol. 27, iss. 6. P. 684–700.
9. Sauer J., Sonderegger A. Visual aesthetics and user experience: A multiple-session experiment // International Journal of Human-Computer Studies. 2022. Vol. 165. 102837.
10. Li Q., Liu Z., Wang P. The Impact of Visual Aesthetics of Different Types of APP Icons on User Cognition // Proceedings of the 24<sup>th</sup> International Conference on Human-Computer Interaction, HCII 2022 Virtual Event, June 26 – July 1 2022: Proceedings. Part I. Cham: Springer, 2022. P. 228–237.
11. Bara I., Binney R., Ramsey R. Investigating the role of executive resources across aesthetic and non-aesthetic judgments. 2021. P. 1–4. URL: psyarxiv.com, <https://psyarxiv.com/ydmbr/download?format=pdf>.
12. Лазарева Н. Ю., Савинова А. Д., Чистопольская А. В. Влияние экспериментальных условий на субъективную оценку инсайтности решения // Экспериментальная психология. 2023. № 16(1). С. 23–42.
13. Rayner K., Reingold E. Evidence for direct cognitive control of fixation durations during reading // Current Opinion in Behavioral Sciences. 2015. Vol. 1. P. 107–112.
14. Limitations of Cognitive Control on Emotional Distraction - Congruency in the Color Stroop Task Does Not Modulate the Emotional Stroop Effect / E. Straub, C. Schmidts, W. Kunde, J. Zhang, A. Kiesel, D. Dignath // Cognitive, Affective, & Behavioral Neuroscience. 2022. Vol. 22(1). P. 21–41.
15. Morton J., Ezekiel F., Wilk H. Cognitive Control: Easy to Identify But Hard to Define // Topics in Cognitive Science. 2011. Vol. 3(2). P. 212–216.
16. Cognitive control in autism spectrum disorders / M. Solomon, S. Ozonoff, N. Cummings, C. Carter // International Journal of Developmental Neuroscience. 2008. Vol. 26. P. 239–247.
17. Stuss D., Knight R. Principles of frontal lobe function. Oxford: Oxford University Press, 2002. 630 p.
18. Why people change? The role of cognitive control processes in the onset and cessation of substance abuse disorders / G. Buhringer, H. Wittchen, K. Gottlebe, C. Kufuld, T. Goschke // International Journal of Methods in Psychiatric Research. 2008. Vol. 17. P. 4–15.
19. Gruber O., Goschke T. Executive control emerging from dynamic interactions between brain systems mediating language working memory and attentional processes // Acta Psychologica. 2004. Vol. 115. P. 105–121.
20. Канеман Д. Внимание и усилие. М.: Смысл, 2006. 287 с.
21. Wickens C., MacCarley J. Applied attention theory. Boca Raton, FL: Taylor & Francis, 2008. 248 p.
22. Методы и критерии психофизиологической оценки работников локомотивных бригад для действий в штатных и экстремальных условиях поездной работы: справочно-методическое пособие / С. А. Багрецов, В. А. Бодров, А. П. Войт, М. В. Зотов; под ред. В. М. Львова. М.: ОАО «РЖД», 2007. 47 с.
23. Yakunin A. V., Bodrunova S. S. U-index: An eye-tracking-tested checklist on webpage aesthetics for university web spaces in Russia and the USA // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, 2017. Vol. 10288. P. 219–233.
24. Касьянов С. Психологические тесты. М.: Эксмо, 2006.
25. Laubheimer P. Tree Testing Part 2: Interpreting the Results. URL: [nngroup.com/articles/interpreting-tree-test-results/](https://nngroup.com/articles/interpreting-tree-test-results/) (дата обращения: 26.07.2025).



## О неевклидовой визуализации многомерных функций

А. К. Алексеев<sup>1</sup>, А. Е. Бондарев<sup>2</sup>

<sup>1</sup>РКК Энергия им. С.П. Королева, Королев, Россия

<sup>2</sup>ИПМ им М.В. Келдыша РАН, Москва, Россия

**Аннотация.** Рассмотрены проблемы, связанные с визуализацией данных в многомерном пространстве. В качестве одного из вариантов обсуждено использование Риманова пространства переменной кривизны для моделирования пространства визуализации. Рассмотрено использование уравнений Гильберта-Эйнштейна, Уинслоу, Бельтрами для моделирования геометрии пространства визуализации и пространства восприятия. Применение уравнений Бельтрами позволяет в некоторой степени нивелировать проблемы, связанные с визуализацией многомерной функции. Использование уравнений Гильберта-Эйнштейна осложняется как неясностью трактования априорной информации, так и техническими трудностями. Наиболее перспективным представляется использование уравнений типа Уинслоу, соответствующих построению гармонических координат для уравнений Гильберта-Эйнштейна.

**Ключевые слова:** многомерное пространство, визуализация, Риманово пространство, уравнение Бельтрами, уравнения Уинслоу, уравнения Гильберта-Эйнштейна.

## On a nonequclidean visualization of the multidimensional functions

A. K. Alekseev<sup>1</sup>, A. E. Bondarev<sup>2</sup>

<sup>1</sup> RSC Energia, Korolev, Russia

<sup>2</sup> Keldysh Institute of Applied mathematics of RAS, Moscow, Russia

**Abstract.** The problems of the data visualization in multidimensional spaces are addressed. The application of the Riemann space of variable curvature for the visualization space is discussed. The Hilbert-Einstein, Winslow and Beltrami equations are considered for the visualization and perception spaces. The Beltrami system enables to reduce the problems related with the visualization of multidimensional spaces. The application of the Hilbert-Einstein equations is complicated both the technical problems and problems with treating the a priori information. The application of the Winslow equations, which corresponds to the harmonic coordinates for the Hilbert-Einstein equations is most perspective.

**Keywords:** multidimensional space, visualization, Riemann space, Beltrami system, Winslow equations, Hilbert - Einstein equations.

### Введение

Работа с многомерными данными представляет собой один из наиболее интересных и перспективных разделов научной визуализации. В частности, к ним относится визуальный анализ многомерных функций, таких как плотность распределения вероятности в уравнении Больцмана, многопараметрическое решение уравнений Навье-Стокса, волновая функция в конфигурационном пространстве при решении многочастичных задач квантовой механики и многие другие.

Однако визуализация многомерных функций (даже в случае скалярной функции  $f(x), x \in R^d$ ) при  $d \geq 6$  затруднена как техническими проблемами (необходимые память и быстродействие катастрофически растут при увеличении размерности задачи («проклятие размерности»)), так и геометрическими проблемами, связанными с отличием свойств многомерных пространств от стандартного для нас двух- и трехмерных пространств [1]. В первую очередь, к ним относится эффект концентрации меры, при которой большая часть объема сосредотачивается около поверхности тела. Объем объектов характерного размера  $l \in [0,1]$  (по одной координате) в  $d$ -мерном пространстве изменяется как  $\sim l^d$ . Это влияет как на представление известных функций (особенно в виде двумерных сечений), так и на аппроксимацию неизвестных многомерных функций (при поиске их особенностей, максимумов, минимумов и т.д.).

Многомерные данные обычно задаются на некотором множестве точек пространства, которое соответствует не регулярной многомерной сетке, а некоторому ансамблю точек (с много меньшим числом узлов). Распространенным подходом к визуализации и анализу многомерных данных является понижение размерности и переход к двух- и трехмерным объектам [2]. Понижение размерности можно

выполнить как интуитивным выбором наиболее важных переменных, так и более формальным путем, например, используя анализ основных компонент. Также часто используется набор сечений плоскостями вдоль некоторого пути [3,4], это позволяет находить и визуализировать некоторые особенности данных, например, дыры. Данные мы рассматриваем как набор значений функции на некотором ансамбле точек пространства (гиперкуба, гиперболы). Распространенными являются ситуации, когда многомерная функция может быть задана некоторым ансамблем точек или неявно, с помощью некоторого алгоритма. В обоих этих случаях простейшим кажется подход с построением функции в узлах регулярной сетки. К сожалению, в многомерном случае построение регулярной сетки по сути запрещено проклятием размерности (в 6-мерном случае, используя 100 узлов по каждой координате, мы получим  $100^6 = 10^{12}$  узлов сетки, что создаст серьезные затруднения для стандартной вычислительной техники). Более реалистичным является построение функции в виде значений на случайном ансамбле узлов.

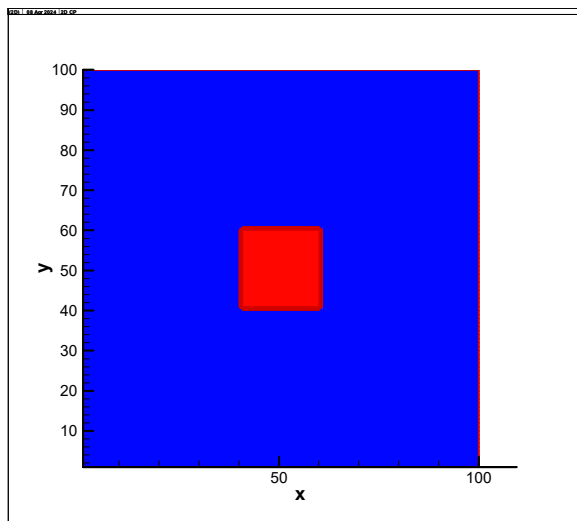


Рисунок 1

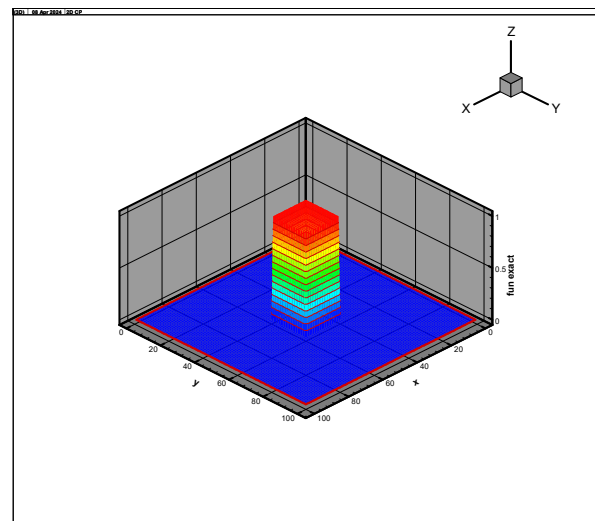


Рисунок 2

Явным недостатком такого подхода является отсутствие учета особенностей, специфичных для многомерных пространств. Например, на рисунке 1 представлено сечение 6-мерного кубика  $Cub \in R^6$  со стороной  $2/10$  по каждой координате плоскостью, которое дает визуальное представление о значимости этого кубика с точки зрения его размеров. К сожалению, это представление абсолютно неадекватно. В реальности отношение объема этого кубика к объему куба, в котором он содержится, равно  $(2/10)^6 = 1/13825 \sim 7 \cdot 10^{-5}$  и достаточно мало, соответствует точке. На рисунке 2 представлена визуализация 6-мерной функции, принимающей единичное значение на этом кубике. Это представление также неадекватно (в шестимерном пространстве рассматриваемая фигура более удачно иллюстрировалась бы некоторой иглой). Таким образом, наивная 2D-визуализация 6-мерной фигуры радикально преувеличивает ее значимость с точки зрения величины объема. При этом погрешность среднего, рассчитанного по этим узлам, не зависит от размерности пространства и определяется числом узлов в ансамбле, что является основным преимуществом метода Монте-Карло при расчете интегралов. Однако при использовании метода Монте-Карло с равномерной выборкой с большой вероятностью деталь, у которой объем невелик (типа вышеописанной функции, рис. 2), обнаружена не будет, так как попадет между узлами. Особенно это относится к разрешению в окрестности центра гиперкуба/гиперсферы.

Рисунки 3 и 4 представляют визуализацию данного сечения с помощью преобразования координат Winslow (), соответствующего наличию априорной информации в форме (10). Эти изображения гораздо лучше отражают важность фигуры с точки зрения занимаемого ей объема в шестимерном пространстве. Здесь мы имеем эффект, обратный эффекту «огромной луны», в котором априорная информация увеличивает некоторый объем в пространстве визуализации.

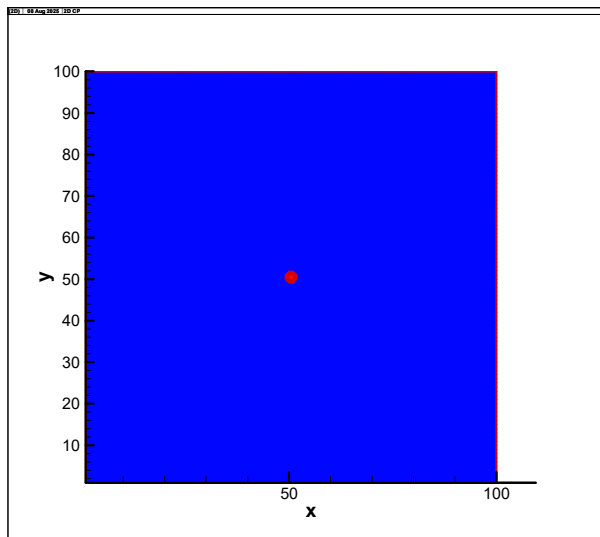


Рисунок 3

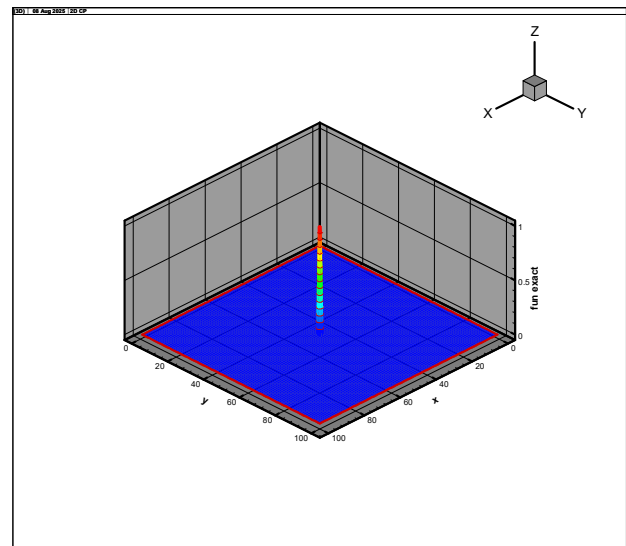


Рисунок 4

### Неевклидова геометрия пространства восприятия

Следует отметить, что пространство человеческого визуального восприятия (*перцептивное пространство* (perception space)) не является линейной проекцией наблюдаемого физического пространства на плоскость сетчатки глаза. Простейшим примером, с которым сталкивался каждый, является «эффект огромной луны», вследствие которого размеры Луны около горизонта выглядят существенно больше, чем около зенита. Такое преобразование возможно с помощью перехода к криволинейной системе координат либо в евклидовом пространстве, либо в неевклидовом пространстве. Существует большой набор работ, посвященных анализу геометрии пространства визуального восприятия. В работе [5] обосновывается, что перцептивное пространство является пространством Лобачевского (имеет постоянную отрицательную кривизну). В работе [6] утверждается, что перцептивное пространство является пространством Римана с положительной кривизной. В работе [7] указывается, что перцептивное пространство является пространством Римана с кривизной переменного знака, плавно изменяющейся по расстоянию от наблюдателя: с отрицательной кривизной на близких расстояниях и с положительной кривизной на дальних расстояниях. В работе [8] представлены экспериментальные данные, подтверждающие предположение о наличии кривизны переменного знака, однако смена знака кривизны по этим данным происходит при достижении высоты уровня глаз (изменяется не по расстоянию, а по высоте). На рисунке 5 представлена иллюстрация [7], демонстрирующая свойства этого пространства.

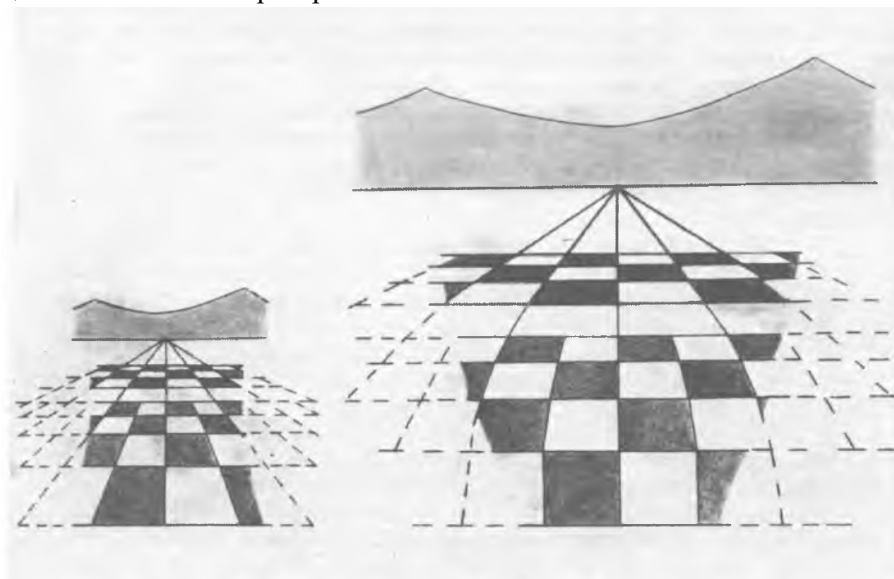


Рисунок 5

Можно считать, что в рамках человеческого зрения на перцептивное пространство (некое трехмерное Риманово пространство переменной кривизны) проектируется некоторая область внешнего трехмерного евклидова пространства.

Таким образом, в рамках человеческого зрения можно выделить три пространства: физическое, пространство визуализации и перцептивное пространство. В рамках научной визуализации (как приема для анализа данных) используется область в некотором промежуточном пространстве, пространстве визуализации, чаще всего двумерном (лист, экран), но в некоторых случаях и трехмерном (стереоочки). При этом возможен переход к новой криволинейной системе координат, задаваемой матрицей Якоби и соответствующим метрическим тензором.

В настоящее время пространство визуализации, как правило (за исключением образцов средневековой живописи [7]), принимается евклидовым. Однако вопрос о целесообразности использования евклидовой метрики для пространства визуализации является нетривиальным. С одной стороны, во избежание дополнительных искажений было бы естественным, чтобы пространство визуализации имело бы ту же метрику, что и перцептивное пространство. С другой стороны, при визуализации многомерных пространств переход от евклидовой метрики к неевклидовой в пространстве визуализации может определяться совсем другими причинами и может оказаться, что метрика визуального пространства не должна совпадать с метрикой перцептивного.

Но в целом переход к неевклидовой метрике для пространства визуализации не кажется противоестественным, так как он в конечном итоге уже выполнен при переходе от трехмерного физического к трехмерному перцептивному.

### Возможности моделирования пространства восприятия с помощью уравнений Гильберта-Эйнштейна

В работе [7] предлагается моделировать перцептивное пространство с помощью уравнений Гильберта-Эйнштейна. Действительно, в достаточно общем случае метрику неевклидова пространства можно описать уравнениями типа Гильберта-Эйнштейна

$$R_{\mu\nu} - 1/2 g_{\mu\nu} R \pm \lambda g_{\mu\nu} = -T^{\mu\nu}, \quad (1)$$

которые связывают тензор Риччи  $R_{\mu k} = R^{\lambda}_{\mu\lambda k}$ ,  $R_{ab} = \partial_m \Gamma^m_{ab} - \partial_a \Gamma^m_{mb} + \Gamma^m_{mn} \Gamma^n_{ab} - \Gamma^m_{an} \Gamma^n_{mb}$

(свертку тензора кривизны 4-го ранга  $R^{\lambda}_{\mu\nu k} = \frac{\partial \Gamma^{\lambda}_{\mu\nu}}{\partial x^k} - \frac{\partial \Gamma^{\lambda}_{\mu k}}{\partial x^{\nu}} + \Gamma^{\eta}_{\mu\nu} \Gamma^{\lambda}_{k\eta} - \Gamma^{\eta}_{\mu k} \Gamma^{\lambda}_{\nu\eta}$ ), кривизну

$R = g^{\mu k} R_{\mu k}$  (Ricci scalar), метрический тензор  $g_{jm}$ , тензор энергии-импульса, символы Кристоффеля  $\Gamma^{\beta}_{\alpha\gamma} = \frac{g^{\beta\delta}}{2} \left( \frac{\partial g_{\alpha\delta}}{\partial x^{\gamma}} + \frac{\partial g_{\delta\gamma}}{\partial x^{\alpha}} - \frac{\partial g_{\alpha\gamma}}{\partial x^{\delta}} \right)$ .

Уравнения Гильберта-Эйнштейна можно использовать для описания Риманова пространства как с положительной, так и с отрицательной кривизной. В общей теории относительности (ОТО) уравнения Гильберта-Эйнштейна получаются из условий стационарности функционала действия Гильберта-Эйнштейна

$$S = \int L_{HE} \sqrt{g} d^4 x = \int (R - \Lambda + L_{phys}) \sqrt{g(x)} d^4 x. \quad (2)$$

На первый взгляд, использование уравнений Гильберта-Эйнштейна для анализа геометрии перцептивного пространства и расчета его метрического тензора выглядит необоснованным. Тем более, что четырехмерное пространство, время и сигнатура псевдоевклидова пространства Минковского, характерные для уравнений Гильберта-Эйнштейна в ОТО (с псевдо-Римановым пространством) в нашем случае не используются. Однако существует обстоятельство, оправдывающее такой подход. Уравнения типа Гильберта-Эйнштейна достаточно универсальны и используются не только в общей теории относительности, но и в информационной геометрии, где метрическим тензором служит информационная матрица Фишера,  $F_{ij} \in R^{n \times n}$  (тензор в пространстве плотностей вероятности случайного распределения  $\rho(u, \theta)$ ) имеющая следующий вид:

$$F_{ij} = \int \left( \frac{-\partial \ln \rho(u, \theta)}{\partial \theta_i} \right) \cdot \left( \frac{-\partial \ln \rho}{\partial \theta_j} \right) \rho du = \int \frac{\partial \rho}{\partial \theta_i} \cdot \frac{\partial \rho}{\partial \theta_j} \frac{1}{\rho} du = 4 \int \frac{\partial \rho^{1/2}}{\partial \theta_i} \frac{\partial \rho^{1/2}}{\partial \theta_j} du. \quad (3)$$

Несимметричное расстояние между двумя плотностями вероятности в одном и том же пространстве параметров можно ввести с помощью меры Кулльбака-Лейблера (Kullback-Leibler), иногда также обозначаемой как кросс-энтропия:  $KL(f, g) = E_f(\ln(f(x)/g(x))) = \int (\ln(f(x)) - (\ln(g(x)))f(x)dx$ .

Мера Кулльбака-Лейблера при малых расстояниях между плотностями распределения описывается информационной матрицей Фишера  $KL(\rho(\theta_i + \Delta\theta_i), \rho(\theta_i)) \approx F_{jk}\Delta\theta_j\Delta\theta_k$ .

Как показано в [9], для этого метрического тензора можно получить уравнение  $R^{\mu\nu}(\theta) - 1/2 g^{\mu\nu} R(\theta) + T^{\mu\nu}(\theta) = 0$ . Следует отметить, что матрица Фишера является обратной ковариационной матрицей и принадлежит Риманову пространству с неположительной кривизной, что не соответствует как ОТО, так и перцептивному пространству.

К сожалению, природа тензора в правой части (аналога тензора энергии-импульса) в этих работах не конкретизируется (в [9] он обозначается как «некоторые статистические ограничения»). По нашему мнению, их можно рассматривать как некоторую априорную информацию о метрике. Если предположить, что геометрия перцептивного пространства зависит от отображенных в нем объектов, то априорная информация о них играет роль, аналогичную тензору энергии-импульса. При этом априорная информация влияет на метрический тензор и на кривизну глобально даже там, где  $T^{\mu\nu} = 0$  (аналогично гравитации).

При обработке информации, осуществляемой при проектировании на перцептивное пространство, использование информации Фишера и информационной геометрии выглядит вполне естественным [10]. В частности, в [10] показано, что подход информационной геометрии (использующий в качестве метрического тензора матрицу Фишера) оказывается применим и к анализу работы сетей нейронов при обработке визуальной информации. В этой работе перцептивное пространство обозначается как «visual space» и рассматривается как статистическое параметрическое пространство, чья геометрия определяется метрическим тензором, определяемым Гауссовым распределением  $(\mu, \sigma)$ . При этом кривизна пространства оказывается отрицательной. Соответствующие выражения для символов Кристоффеля приведены в [11].

Следует отметить, что матрица Фишера определена в параметрическом пространстве достаточно большой размерности, поэтому если она соответствует нейросетям, моделирующим человеческий мозг, то наше трехмерное перцептивное пространство не является следствием архитектуры нашего мозга, а выступает результатом его обучения трехмерными образцами. При обучении многомерными образцами трудности в нашем восприятии могут быть уменьшены или преодолены. В качестве примера можно взять компьютерные игры HyperRouge (2011) и Hyperbolica (2022), которые позволяют обучить наше восприятие на образах из Риманова пространства отрицательной кривизны.

К сожалению, при решении уравнений типа Гильберта-Эйнштейна и последующей координатизации полученного решения встречается огромное количество как принципиальных, так и технических сложностей, что показано в работах по ОТО [12]. Построить координаты по одному метрическому тензору невозможно, да и само построение метрического тензора для уравнения Гильберта-Эйнштейна очень нетривиально. Это связано с тем, что 10 уравнений Гильберта-Эйнштейна в ОТО содержит 14 независимых величин, и для замыкания необходимо еще четыре дополнительных уравнения («координатные условия»). Более того, благодаря соотношениям Бьянки [12] потребуется еще четыре дополнительных уравнения. Следует также отметить, что уравнения этого типа могут иметь сингулярности и содержат нетривиальную физику, зачастую требующую для интерпретации серьезных усилий. В интересном случае двумерного пространства уравнения Гильберта-Эйнштейна вырождены (тензор Гильберта-Эйнштейна исчезает, так как в этом случае  $R_{ij} \equiv 1/2 R g_{ij}$ ).

Из-за перечисленных трудностей надежда на то, что идея Б. Раушенбаха [7] по моделированию перцептивного пространства с помощью уравнений типа Гильберта-Эйнштейна будет реализована,

достаточно слаба. Основной причиной такой формы уравнений является выбор плотности Лагранжиана в действии Гильберта–Эйнштейна. Поэтому поиск более простых подходов к моделированию Риманова пространства переменной кривизны (с более простыми плотностями Лагранжиана и более простыми уравнениями) выглядит перспективным. В частности, на практике вычислительной аэрогазодинамики (при построении расчетной сетки) довольно часто реализуется преобразование координат, удовлетворяющее некоторому функционалу метрического тензора (без его производных). Вместо решения уравнений типа Гильберта–Эйнштейна при этом решаются уравнения Winslow (или аналоги), определяющие как преобразование координат, так и метрические коэффициенты. В стандартном варианте (задачах аэрогазодинамики) эти уравнения задают преобразование координат в евклидовом пространстве. Однако на языке ОТО решение уравнений Winslow соответствует построению гармонических координат в Римановом пространстве. Уравнения типа Winslow позволяют построить криволинейную координатную систему, в некоторых случаях соответствующую Риманову пространству.

### Подход к построению перцептивного пространства с помощью функционалов, зависящих от метрического тензора

В работах по аэрогазодинамике [16-21] для построения неравномерных расчетных сеток часто используется преобразование координат из физического пространства (двух- или трехмерного евклидова) в вычислительное. Здесь мы будем рассматривать двумерный случай, при этом физическая плоскость  $(x, y)$  ( $x_i$ ) преобразуется в вычислительную плоскость  $(\xi, \eta)$  ( $u_i$ ) через решение уравнения Пуассона с источниковыми членами. Рассмотрим использование функционала Winslow [18, 19, 20]

$$I = \int (g_{11}^2 + g_{22}^2) dx dy = \int (\xi_x^2 + \xi_y^2 + \eta_x^2 + \eta_y^2) dx dy = \int \frac{x_\xi^2 + y_\xi^2 + x_\eta^2 + y_\eta^2}{\tilde{J}^2} \tilde{J} d\xi d\eta = \int \frac{\tilde{g}_{11}^2 + \tilde{g}_{22}^2}{\sqrt{\tilde{g}(\xi, \eta)}} d\xi d\eta. \quad (4)$$

Он обеспечивает максимальную гладкость преобразования и при этом запрещает обращение в нуль якобиана  $\tilde{J}$  (и соответствующего метрического тензора) и вырождение соответствующего преобразования. Добавим к стандартной постановке потенциал  $\Phi(\xi, \eta)$

$$I(\xi, \eta, \xi_x, \xi_y, \eta_x, \eta_y) = 1/2 \int (\xi_x^2 + \xi_y^2 + \eta_x^2 + \eta_y^2 + \Phi(\xi, \eta)) dx \cdot dy. \quad (5)$$

Соответствующие уравнения Эйлера–Лагранжа имеют вид уравнений Пуассона с нелинейными источниками (в стандартном варианте веса  $P, Q$  добавляются искусственно).

$$\begin{aligned} \xi_{xx} + \xi_{yy} &= P(\xi, \eta) = \partial \Phi(\xi, \eta) / \partial \xi; \\ \eta_{xx} + \eta_{yy} &= Q(\xi, \eta) = \partial \Phi(\xi, \eta) / \partial \eta. \end{aligned} \quad (6)$$

Далее переменные  $(\xi, \eta)$  принимаются за независимые и на равномерной сетке по  $(\xi, \eta)$  решается система

$$\begin{aligned} \alpha x_{\xi\xi} - 2\beta x_{\xi\eta} + \gamma x_{\eta\eta} &= -\tilde{J}^2 (P x_\xi + Q x_\eta); \\ \alpha y_{\xi\xi} - 2\beta y_{\xi\eta} + \gamma y_{\eta\eta} &= -\tilde{J}^2 (P y_\xi + Q y_\eta). \end{aligned} \quad (7)$$

Система содержит распределенные источники и позволяет рассчитать координаты  $(x, y)$  в узлах равномерной сетки  $(\xi, \eta)$ .

Источниковый член в (6) и (7) часто используется в форме

$$P = Q = b \times \text{sign}(\xi - \xi_0) \times \exp(-d \{(\xi - \xi_0)^2 + (\eta - \eta_0)^2\}^{1/2}), \quad (8)$$

что позволяет сжимать узлы сетки в физическом пространстве  $(x, y)$  около точек  $\xi = \xi_0, \eta = \eta_0$ .

В нашем случае переменные  $(\xi, \eta)$  соответствуют исходным координатам, получаемым простым сечением плоскостью. Переменные  $(x, y)$  соответствуют трансформированным координатам, при этом сетка, в которой задана функция, уже не будет равномерной.

Кроме функционалов Winslow и длины при построении сетки иногда минимизируются другие функционалы [16], напрямую связанные с компонентами метрического тензора. К ним относятся функционалы ортогональности, функционал площади, Liao.

Условиями стационарности этих функционалов служат ЧДУ второго порядка, некоторые из которых достаточно легко решаются на практике и позволяют осуществить преобразование декартовых координат в криволинейные и наоборот. Часть функционалов (площади, ортогональности, Liao) порождают решения, содержащие складки. В качестве лекарства иногда используются суммы этих функционалов [21].

### Построение конформных координат и система Бельтрами

В работах [12, 13] для двумерного случая рассмотрено построение системы координат, реализующих заданную Риманову метрику с помощью решения системы Бельтрами

$$g x_{\xi} = -g_{12} y_{\xi} + g_{11} y_{\eta}, \quad g x_{\eta} = -g_{22} y_{\xi} + g_{12} y_{\eta}, \quad g^2 = g_{11} g_{22} - g_{12}^2 \quad (9)$$

или

$$g_{22} v_{xx} - 2g_{12} v_{xy} + g_{11} v_{yy} + (g_{22,x} - g_{12,y}) v_x + (g_{11,y} - g_{12,x}) v_y + \frac{g_x - g_y}{g} = 0. \quad (10)$$

Решение этого уравнения позволяет вычислить  $v(x, y)$ , далее, имея поле  $u_x, u_y$ , рассчитаем  $u(x, y)$  интегрированием. В численных расчетах использовалась функция априорной важности  $\theta(x, y)$ , при которой деформация расстояния записывается как  $ds^2 = \theta(x, y)(dx^2 + dy^2)$  и дает компоненты тензора  $g_{xx} = \theta$ ,  $g_{yy} = \theta$ ,  $g_{xy} = 0$ . В работах [12,13] используется достаточно

универсальная форма  $ds^2 = \frac{dx^2 + dy^2}{(1 + \delta(x^2 + y^2))^2}$ , пригодная для Римановых пространств как с отрицательной, так и с положительной кривизной.

Использование системы уравнений Бельтрами позволяет осуществить простейший способ построения координат в Римановом пространстве, однако ограничено только двумерным случаем. Уравнения типа Winslow с этой точки зрения более универсальны.

### Функция важности

Если мы ориентируемся на преобразование координат типа (9), (10), основной интерес для нас представляет определение формы и смысла источниковых членов  $(P, Q)$ . Они могут быть связаны чисто с геометрией пространства либо со свойствами функции. Аналогом является сопряженная функция  $\psi(x, y, z, u, v, w)$  [22], которая позволяет визуализировать зоны, где функция  $f(x, y, z, u, v, w)$  оказывает наибольшее влияние на ценный функционал. В случае  $\psi(x, y, z, u, v, w) > 0$  сопряженную функцию можно связать с метрическим тензором  $\psi \sim J = \sqrt{g}$ , что дает априорную информацию о зонах влияния функции. В частности, при эффекте огромной луны  $d\tilde{V} \sim \sqrt{g} dV$  априорная информация искажает объемы. В нашем случае, мы используем конформное отображение  $dx^2 + dy^2 = \rho(\xi, \eta)(d\xi^2 + d\eta^2)$  и функцию (11) в качестве априорной информации. Решая далее уравнения Бельтрами, можно построить преобразование координат, аналогичное представленным на рисунках 3 и 4.

### Обсуждение

Представлено двумерное сечение 6-мерного гиперкуба (см. рис. 1, 2) в евклидовых координатах и в Римановых координатах (см. рис. 3, 4) с априорной метрикой (8) и построением координат решением уравнения Уинслоу (6,7). Оно более информативно в части о тображения объема, занимаемого функцией в сравнении со стандартным сечением гиперкуба плоскостью. Это достигается за счет неевклидовой трансформации пространства визуализации.



В значительной степени визуализация сводится к отображению некоторого физического пространства (чаще всего трехмерного, но в некоторых случаях и многомерного (например, 6-мерное пространство скоростей и координат в уравнении Больцмана) на двумерное (иногда трехмерное) промежуточное пространство—пространство визуализации. Далее пространство визуализации отображается на пространство восприятия (перцептивное пространство), которое является трехмерным пространством Римана с переменной кривизной.

Риманова геометрия пространства восприятия подтверждена как многочисленными современными экспериментами, так и многочисленными примерами средневековой живописи.

Пространство визуализации в современной практике обычно евклидово. В данной работе мы рассматриваем возможности, предоставляемые Римановой метрикой в пространстве визуализации, и проблемы, порождаемые таким подходом.

Как мы уже видели (см. рис. 1, 2), наивная визуализация многомерной функции с помощью сечений плоскостями существенно искажает представление об объеме этой функции. Нелинейное преобразование координат с использованием Римановой метрики позволяет исправить эти искажения (см. рис. 3, 4).

В идеале пространство визуализации должно иметь ту же геометрию, что и перцептивное пространство, это должно уменьшить искажения при проектировании из одного пространства в другое. В случае двумерной визуализации полное устранение искажений вряд ли возможно из-за различия в размерности (художественные попытки этого представлены в [7]). В случае трехмерной визуализации неясно, как технически выполнить построение образа в неевклидовой геометрии. Более того, неясно, насколько геометрия пространства восприятия универсальна при переходе от одного индивидуума к другому.

### Заключение

Для адекватной визуализации многомерных функций можно использовать трансформированную плоскость с Римановой метрикой, где метрический тензор может определяться геометрией и размерностью пространства, априорной информацией о значимости некоторых областей.

Использование уравнений типа Гильберта-Эйнштейна оправдывается аналогией между Римановыми физической и информационной геометриями. Вместо уравнений Гильберта-Эйнштейна для моделирования перцептивного пространства и пространства визуализации проще использовать уравнения типа Winslow или уравнения Бельтрами.

С учетом того, что матрица Фишера определена в параметрическом пространстве достаточно большой размерности, определяемом числом использованных нейронов, наше привычное трехмерное перцептивное пространство не является следствием архитектуры нашего мозга, а является следствием его обучения трехмерными образцами. Вполне вероятен переход к многомерному пространству при обучении многомерными образцами. В качестве аналогии можно рассмотреть компьютерные игры HyperRouge и Hyperbolica, которые позволяют обучить наше восприятие на образах из Риманова пространства отрицательной кривизны.

### Список источников

- [1] Зорич В.А. Многомерная геометрия, функции очень многих переменных и вероятность // ТВП. 2014. Т. 59, вып. 3. Р. 436-451.
- [2] Chen H. et al, Noisy Data Visualization using Functional Data Analysis. arXiv:2406.03396v1 [cs.LG] 5 Jun 2024
- [3] Laa U., Cook D., Buja A., Valencia G. 2020. Hole or grain? A Section Pursuit Index for Finding Hidden Structure in Multiple Dimensions. arXiv:2004.13327v3 [stat.CO] 10 Mar 2022.
- [4] Laa U., Cook D., and Valencia G. 2020. A slice tour for finding hollowness in high-dimensional data. arXiv:1910.10854v1 [stat.CO] 24 Oct 2019.
- [5] Luneburg R.K. Metric methods in binocular visual perception // *Courant Anniversary* / eds K. O. Friedrichs, et al. New York: Interscience, 1948. Pp. 215-240.
- [6] Battro A.M. Riemannian geometries of variable curvature in visual space: visual alleys, horopters, and triangles in big open fields // *Perception*. 1976. Vol. 5. P. 9-23.
- [7] Раушенбах Б.В. Пространственные построения в живописи. М.: Наука, 1980.
- [8] Koenderink J. J., Van Doorn A. J., Lappin J. S. Direct measurement of the curvature of visual space // *Perception*. 2000. Vol. 29, no. 1. P. 69-79.

- 
- [9] Calmet X., Calmet J. Dynamics of the Fisher information metric. 2004. arXiv:cond-mat/0410452v1.
- [10] Mazumdar D., et al. Investigation of the neural origin of non-Euclidean visual space and analysis of visual phenomena using information geometry. arXiv:2505.13917v1, 2025.
- [11] Costa S. I. R., Santos S. A., Strapasson J. E. Fisher information distance: a geometrical reading. arXiv.1210.2354v3[stat.ME], 10 Jan, 2014.
- [12] Arnold D. N. Numerical Problems in General Relativity // *Proceedings of the 3rd European Conference on Numerical Mathematics and Advanced Applications*. P. Neittaanmaki, T. Tiihonen, P. Tarvainen eds. WS, Singapore. P. 3–15.
- [13] Белинский П. П., Годунов С. К., Иванов Ю. Б., Яненко И. К. Применение одного класса квазиконформных отображений для построения разностных сеток в областях с криволинейными границами // ЖВМ. 1975. Т. 5, № 6. С. 1499–1511.
- [14] Уравнения с частными производными / Л. Берс и др. М.: Мир, 1966.
- [15] Knupp P. M., Steinberg S. Fundamentals of Grid Generation. CRC Press, 1993.
- [16] Khattri S. K., Grid Generation and Adaptation by Functionals. arXiv:math/0607388v1 [math.NA] 17 Jul 2006
- [17] Charakhchyan A. A., Ivanenko S. A. A Variational Form of the Winslow Grid Generator // JCP. 1997. Vol. 136. P. 385–398.
- [18] Brackbill J. U., Saltzman J. S., Adaptive zoning for singular problems in two dimensions // JCP. 1982. Vol. 46. P. 342.
- [19] Тишкин В. Ф. Методы построения расчетных сеток. М.: МГУ, 2016.
- [20] Fortunato M., Persson P.-O. High-order unstructured curved mesh generation using the Winslow equations // JCP. 2016. Vol. 303. Pp. 1-14.
- [21] Tinoco J.G., Barrera P., Cortes A. Some Properties of Area Functionals in numerical Grid Generation // *X Meshing Round Table*. Newport Beach, California, USA, 2001.
- [22] Алексеев А.К., Бондарев А.Е. Применение сопряженных уравнений и визуальное представление сопряженных параметров в задачах идентификации и управления течением // Препринт ИПМ им. М.В. Келдыша РАН. 2011. № 50. 24 с.

УДК 004.4'2

DOI: 10.25686/978-5-8158-2474-4-2025-353-366

## Проектирование инструментов визуализации данных на основе интеграции возможностей предметно-ориентированного моделирования и генеративного искусственного интеллекта

А. Д. Джейранян, Я. А. Ларионова, Л. Н. Лядова

Национальный исследовательский университет «Высшая школа экономики», Пермь, Россия

**Аннотация.** Средства визуализации данных представляют собой ключевые инструменты аналитиков, облегчающие выявление зависимостей, тенденций и закономерностей. Эти средства находят применение у широкого круга пользователей (аналитиков, учёных, руководителей предприятий, менеджеров, преподавателей и специалистов других областей, где точное и понятное представление информации критически важно для повышения эффективности анализа данных). Существующие инструменты визуализации ограничивают возможности кастомизации, предлагают лишь базовые настройки параметров, что не всегда удовлетворяет потребностям пользователей, так как некоторые сценарии требуют более глубоких модификаций, включая возможность создания собственных типов диаграмм. Такая глубокая настройка зачастую требует использования языков программирования, что создаёт барьер для пользователей без технической подготовки. Предлагается новая методика разработки пользовательских визуализаций данных, которая объединяет преимущества трёх подходов: языково-ориентированный подход к разработке эффективных моделей визуализации с использованием средств автоматизации разработки языков и генерации кода для реализации визуализаций; расширение существующих библиотек пользовательскими пакетами визуализации данных; автоматическая генерация скриптов для реализации пользовательских визуализаций на основе текстовых запросов к языковым моделям, возможностей генеративного искусственного интеллекта. Представлено описание архитектуры системы визуализации данных, многоаспектной онтологии, являющейся ядром системы. Приведены примеры применения разработанного исследовательского прототипа средств создания пользовательских визуализаций, показывающие практическую значимость предлагаемого подхода. Даны оценки полученных результатов и показаны перспективы развития описанных средств.

**Ключевые слова:** визуализация данных, модели визуализации, классификация методов визуализации, многоаспектная онтология, предметно-ориентированные языки, генерация кода, генеративный искусственный интеллект.

## Design of data visualization tools based on the integrating domain-specific modeling and Generative Artificial Intelligence capabilities

A. D. Dzheiranian, Ya. A. Larionova, L. N. Lyadova

National Research University – Higher School of Economics (HSE University), Perm, Russia

**Abstract.** Data visualization tools are key analytics tools that make it easier to identify dependencies, trends, and patterns. These tools are used by a broad range of specialists (analysts, scientists, business leaders, managers, teachers, and specialists in other fields where accurate and understandable presentation of information is critically important to improve the effectiveness of data analysis). Existing visualization tools limit customization options and offer only basic parameter settings, which do not always meet the needs of users, as some scenarios require more in-depth modifications, including the ability to create custom chart types. Such deep customization often requires the use of programming languages, which creates a barrier for users without technical skills. A new technique for developing custom data visualizations is proposed, which combines the advantages of three approaches: a language-oriented approach to developing effective visualization models using automation tools for developing languages and generating code for implementing visualizations; existing libraries addition with custom data visualization packages; automatic script generation for implementing custom visualizations based on prompts to language models, capabilities of generative artificial intelligence. The article describes the architecture of data visualization system and the multifaceted ontology that is the core of the system. Examples of using the developed research prototype of tools for creating custom visualizations are given, showing the practical significance of the proposed approach. The results obtained are evaluated and the prospects for the development of the tools described are shown.

**Keywords:** data visualization, visualization models, classification of visualization methods, multifaceted ontology, domain-specific languages, code generation, generative artificial intelligence.

## Введение

При принятии решений на основе данных системы визуализации представляют собой ключевой инструмент, позволяющий преобразовать сырые наборы данных в графические представления, с помощью которых можно выявлять скрытые зависимости, тенденции и закономерности. Эти средства находят применение у широкого круга пользователей (руководителей предприятий и менеджеров, учёных и преподавателей и др.), для которых понятное представление информации критически важно для повышения эффективности анализа данных.

Сфера визуализации данных является объектом интенсивного научного интереса, что подтверждается динамикой публикационной активности. Для анализа научных работ был разработан Telegram-бот, обеспечивающий автоматизированное получение данных из крупнейшей базы научных публикаций – Scopus. В результате поиска по ключевому слову «data visualization» было найдено 46 625 научных публикаций за последнее десятилетие, с ростом на 77 %: с 3 232 работ в 2015 году до 5 724 в 2024 году.

В статьях [1, 2, 3] был предложен языково-ориентированный подход к разработке средств визуализации данных, основанный на создании предметно-ориентированных языков для реализации пользовательских моделей визуализации данных. Частотный анализ заголовков найденных статей подтвердил универсальность методов визуализации данных и предметно-ориентированного моделирования (Domain Specific Modeling, DSM), применяемых в различных сферах, включая финансы и бизнес, транспорт, энергетику, здравоохранение и медицину, общественное питание, образование и науку, умные дома и Интернет вещей (IoT) и пр. Однако этот подход пока не находит широкого применения в силу сложности разработки самих языков.

Несмотря на возрастающее значение визуализации, многие пользователи сталкиваются с серьёзными трудностями при создании качественных и информативных графических представлений. Одной из основных проблем является распространение визуализаций низкого качества, которые не только упускают важные аспекты исходных данных, но и могут вводить в заблуждение или искажать информацию [4]. Часто это происходит из-за того, что существующие виды диаграмм ограничены базовыми и чрезмерно обобщёнными геометрическими формами, что приводит к семантическому разрыву между спецификой предметных областей и создаваемыми моделями визуализации.

В связи с этим возникает необходимость внедрения пользовательских спецификаций, позволяющих адаптировать диаграммы к конкретным задачам и особенностям предметной области [5]. Такой подход способствует созданию адаптированных визуализаций, которые учитывают специфические требования пользователей и обеспечивают более эффективное представление данных, тем самым повышая качество управленческих решений и исследовательской работы.

Существующие инструменты визуализации отличаются степенью кастомизации и уровнем реализации спецификаций. В большинстве случаев предлагается лишь базовая настройка параметров, что не всегда удовлетворяет потребностям пользователей. Некоторые сценарии требуют более глубоких модификаций, включая возможность создания собственных типов диаграмм. Такая глубокая настройка зачастую требует использования языков программирования, что создаёт барьер для пользователей без технической подготовки.

Важность совершенствования методов работы с данными также подтверждается официальными документами и национальными программами. Согласно отчёту Gartner «100 предположений стратегического планирования на период до 2028 года», данные и аналитика остаются приоритетными направлениями (<https://www.gartner.com.au/en/doc/over-100-data-and-analytics-predictions-through-2028>).

Внимание многих исследователей сфокусировано на создании диаграмм, предназначенных специально для визуализации данных в контексте их анализа, на проблемах и подходах к их решению [2, 3, 6, 7]. Традиционные инструменты (электронные таблицы и BI-платформы) обеспечивают быстрое создание визуализаций через drag-and-drop-интерфейсы, но возможности кастомизации остаются ограниченными.

Применение языков программирования общего назначения обеспечивает наивысший уровень гибкости и настройки создаваемых моделей, но требует значительной технической подготовки.

Современные исследования [8, 9, 10] и др. демонстрируют потенциал разработки предметно-ориентированных языков (Domain Specific Language, DSL) для визуализации данных, однако существующие решения зачастую ограничиваются небольшим набором стандартных диаграмм (например, круговых диаграмм и гистограмм) или конкретными типами данных (например, геопространственными и т. п.), что не позволяет полноценно настроить диаграмму на разнообразные задачи визуализации.

Наконец, наблюдается активное применение методов искусственного интеллекта (ИИ) для визуализации данных. Современные исследования [11, 12, 13, 14] показывают, что предварительно обученные языковые модели успешно используются для различных этапов анализа данных – от предобработки и генерации рекомендаций до непосредственного построения диаграмм. Однако генерация визуализаций не поддерживает сложные сценарии, ограничиваясь шаблонными решениями [12, 14].

Ещё одна проблема для неопытных пользователей – сложность выбора подходящего типа диаграмм, который позволил бы максимально точно отразить те выводы, идеи, которые хотел бы донести пользователь при визуализации данных из различных источников, даже при использовании стандартных видов диаграмм, предполагающих минимальные настройки параметров. Несмотря на существование различных рекомендаций [4, 5, 15, 16, 17] по выбору типа диаграмм в зависимости от целей анализа, типа данных и характера задач, для неопытных пользователей, не обладающих специальной подготовкой в области визуализации данных, применение этих рекомендаций остаётся затруднительным. В связи с этим возрастает потребность в средствах, которые могли бы поддерживать пользователя в процессе выбора диаграмм [18, 19].

Таким образом, несмотря на наличие широкого спектра подходов и инструментов для визуализации данных, ни одно из существующих решений не сочетает в себе высокую гибкость настройки на конкретные потребности (включая возможность создания собственных моделей визуализации) и доступность для пользователей, не обладающих навыками программирования.

Можно выделить следующие *проблемы*, с которыми сталкиваются разработчики визуализаций:

1. *Использование неэффективных методов визуализации.* Неправильный подбор метода отображения данных может привести к потере их смысла. Определение подходящего типа визуализации остаётся сложной задачей даже для экспертов, поскольку понимание того, какие именно данные отобразить, какую идею передать и как её оформить, является итеративным процессом. Строгие критерии отбора методов требуют глубокого понимания предметной области, организации данных, что для неопытных пользователей ведёт к ошибкам.

2. *Ограниченность существующих методов визуализации.* Существующие методы визуализации данных дают возможность строить, с одной стороны, простые и чрезмерно обобщенные визуализации, которым, с другой стороны, часто не хватает универсальности, средств настройки для применения в различных областях.

3. *Техническая сложность создания эффективных пользовательских визуализаций.* Использование библиотек для разработки нестандартных, интерактивных визуализаций требует навыков программирования, понимания структур данных и ручной работы с кодом, разработчикам необходимо уметь работать с асинхронными событиями и т. д., особенно при работе с распределёнными или потоковыми данными [20].

4. *Риск неверной интерпретации данных.* При создании визуализаций для широкой аудитории необходимо учитывать различные уровни «визуальной грамотности» аудитории (способность интерпретировать изображения) [21] – даже корректно построенная диаграмма может привести к ошибочным выводам [7].

5. *Отсутствие масштабируемости для больших объёмов данных.* При росте размеров наборов данных должны меняться и паттерны отображения; так как существуют трудности в сохранении читаемости, а также в обеспечении производительности при сравнении и кластеризации больших массивов информации.

6. *Отсутствие унифицированных инструментов для оценки качества визуализаций.* Это создаёт трудности оценки, сравнения качества разных визуализаций при их разработке.

7. *Отсутствие устоявшихся подходов к обучению созданию эффективных визуализаций, сформированных принципов обучения и преподавания, формирования «визуальной грамотности» аудитории [22].*

*Объектом исследования, рассматриваемым в данной статье, является процесс создания эффективных визуализаций данных. Предмет исследования – средства создания пользовательских визуализаций данных, обеспечивающих отражение специфики решаемых задач и предметных областей, не требующих от пользователей профессиональных навыков в области программирования.*

*Предлагается новая методика разработки пользовательских визуализаций данных, которая объединяет преимущества трёх подходов: разработка предметно-ориентированных языков (DSL, Domain Specific Language) для создания моделей визуализации, соответствующих потребностям пользователей; расширение существующих библиотек языков программирования пользовательскими пакетами; автоматическая генерация скриптов для реализации пользовательских визуализаций на основе текстовых запросов к языковым моделям (LLM, Large Language Model), возможностей генеративного искусственного интеллекта (GAI, General Artificial Intelligence). Использование многоаспектной онтологии в качестве ядра проектируемой системы позволяет хранить все знания, необходимые для функционирования платформы. В представленном проекте средства разработки пользовательских визуализаций дополняются возможностью подбора подходящих типов визуализации на основе образцов, предлагаемых пользователями.*

### **Постановка задачи**

Основные задачи исследования определяются в ходе анализа требований к средствам визуализации данных, используемым в различных областях.

Для достижения поставленной цели необходимо решить следующие задачи:

- 1) обзор применения визуализации данных, задач и методов визуализации;
- 2) анализ требований к средствам решения задач визуализации данных;
- 3) проектирование средств визуализации данных, управляемых знаниями, интегрированных с возможностями генеративного ИИ, и реализация исследовательского прототипа средств визуализации данных;
- 4) апробация подхода – проведение экспериментов и оценка полученных результатов.

Для проектирования средств визуализации данных необходимо выполнить их *классификацию*, которая должна послужить основой для разработки онтологии языков, используемых для разработки пользовательских визуализаций. Среди известных подходов к классификации можно выделить следующие: по типу данных, по форме представления [23], по цели создания [24], по интерактивности [23] и др. Схема выбора диаграмм (Chart Chooser), предложенная в [1], основана на разделении диаграмм по видам задач анализа (сравнение, распределение, соотношение, структура) и параметрам данных (количеству категорий, переменных и периодов). На основе таксономии методов визуализации данных, включающей пять категорий [25], описанных в [2], разработана расширенная классификация, включающая 22 метода визуализации, объединенных в 7 групп в зависимости от цели создания диаграммы и особенностей данных, что обеспечивает упрощение выбора диаграмм пользователями системы при разработке своих моделей визуализации.

Одним из ключевых трендов при разработке средств визуализации является интеграция визуализации в сторителлинг (повествование) [26]. Этот подход позволяет преобразовать сложные данные в запоминающиеся истории, усиливая эмоциональную вовлечённость аудитории. Основные аспекты дата-сторителлинга:

- *преодоление разрыва между данными и их пониманием.* Это делает информацию доступной для широкой аудитории, например, в медицине для объяснения сложных концепций пациентам [27];
- *повышение эффективности коммуникации.* Визуализации подкрепляют утверждения фактами, улучшают запоминаемость и передачу скрытых в данных смыслов закономерностей [27, 28];
- *иммерсивные технологии.* Применение дополненной и виртуальной реальности создает пространственную связь с данными, усиливая восприятие и вовлечённость [29].

На основе описанных ранее проблем и ключевых трендов разработки можно предъявить *общие требования*, которым должны соответствовать средства визуализации:

1. Обеспечение адаптации графических представлений к специфике предметной области, в том числе возможность комбинирования и расширения существующих типов диаграмм для создания новых пользовательских визуализаций [5].
2. Обеспечение создания и настройки визуализаций без необходимости программирования [20].
3. Обеспечение создания интерактивных представлений данных (с возможностью детализации изображений, их масштабирования, отображения подсказок) [9].
4. Обеспечение редактирования ранее созданных диаграмм, возможность вносить изменения «на лету»: перенастройка параметров, элементов дизайна и структуры диаграммы.
5. Обеспечение широкого набора базовых методов визуализации для представления одних и тех же данных в разных формах, плоскостях и измерениях в зависимости от решаемых задач.
6. Обеспечение рекомендаций по выбору метода визуализации на основе анализа введённых данных и цели создания визуализации [5], на основе образцов, предлагаемых пользователями.
7. Обеспечение достаточного уровня автоматизации при построении диаграмм при сохранении полного пользовательского контроля над настройкой дизайна и параметров визуализации [21].

*Требования к функциональности* средств визуализации данных [2] расширены. Для предоставления *конечному пользователю* возможностей, необходимых для решения специфичных задач аналитики, необходима тесная коллаборация специалистов в разных областях:

- *эксперт в предметной области* обладает специализированными знаниями в своей сфере. Он активно задействуется при разработке онтологии предметной области посредством интервьюирования;
- *DSM-эксперт* способен разрабатывать новые предметно-ориентированные языки, правила трансформации моделей;
- *программный инженер* отвечает за расширение существующих библиотек визуализации пользовательскими пакетами и компонентами;
- *ML-инженер* задействуется на этапе дообучения языковой модели тем пакетам, которые были разработаны ранее программным инженером: он создает специализированный набор данных и обучает на нём модель под разные сценарии для эффективной работы в определённой предметной области.

При использовании прототипа конечный пользователь – разработчик визуализации – может создать стандартную или нестандартную визуализацию. Процесс создания стандартной визуализации может быть выполнен двумя способами: настройкой параметров существующего DSL, разработанного для описания пользовательской модели визуализации, или написанием текстового запроса для генерации скрипта визуализации данных на языке программирования. Эксперт в предметной области отвечает за разработку онтологии, которая служит основой для решения задач пользователями – разработчиками визуализаций. DSM-эксперт занимается созданием метамodelей базовых языков. Генерация новых языков автоматизирована за счёт применения правил отображения онтологии предметной области на метамodelь базового языка.

Создание нестандартной визуализации становится доступным также благодаря написанию текстового запроса, но отличие в том, что модель предварительно обучается ML-инженером на написание кода с помощью кастомных пакетов визуализации, которые разрабатываются программным инженером. Кроме того, нетипичные задачи могут быть решены при помощи разработки нового DSL, соответствующего задачам пользователя. Процесс трансформации моделей в код позволяет реализовать созданные модели. Правила трансформации, необходимые для генерации кода визуализации (в том числе и интерактивной), разрабатываются DSM-экспертом.

### **Теория: моделирование средства визуализации данных**

Проектирование средств визуализации данных основано на разработке многоаспектной онтологии – ядра системы, управляемой знаниями [1, 2, 3, 6]. Она является центральным элементом архитектуры системы.



Многоаспектная онтология – это совокупность онтологий, которые предназначены для независимого моделирования и объединения различных аспектов интеллектуальной системы (рис. 1):

- 1) *онтология источников данных* включает информацию о структуре, типах данных и т. д.;
- 2) *онтология предметных областей* содержит субъективные знания экспертного сообщества: концепты предметной области, их взаимоотношения и ограничения (в том числе правила);
- 3) *онтология DSM-знаний* – ядро логики и визуального конструирования DSL, которое состоит из онтологий, которые содержат описание моделей, визуальных языков, предметных областей, правил трансформации и генерации DSL и т. д.;
- 4) *онтология скриптов* фиксирует результаты генерации: скрипты, используемые промпты (текстовые запросы) и метаданные (версию модели, имя файла источника данных и т. д.);
- 5) *онтология пакетов* описывает расширения существующих библиотек языков программирования общего назначения специализированными пакетами с набором функций и компонентов, необходимых для визуализации нового типа диаграммы или для решения задач в определенной предметной области;
- 6) *репозиторий пользовательских задач* фиксирует задачи визуализации и позволяет находить и переиспользовать существующие решения.

Помимо онтологий, проектируемый граф знаний системы включает также *функции*, реализующие базовые алгоритмы: F1 – функция *автоматизации разработки новых языков* по правилам, заданным пользователем для отображения элементов онтологии предметной области на элементы метамодели выбранного базового языка; F2 – функция *формирования правил трансформации* – генерации кода на целевом текстовом языке по построенной на визуальном языке модели (элементам метамodelей ставятся в соответствие символы языка); F3 – функция *генерации нового DSL* в соответствии с правилами, заданными через F1; F4 – функция *трансформации моделей (генерации кода)* по правилам, заданным через F2; F5 – функция *определения правил интерпретации* для интерактивных моделей; F6 – функция *реализации взаимодействия пользователей с визуальными моделями* в интерактивных визуализациях по правилам, заданным через F5; F7 – функция *связывающая модель предметной области в онтологии с соответствующим ей пользовательским пакетом визуализации*, в котором реализованы пользовательские функции визуализации; F8 – функция *связывания пользовательских задач с источниками данных*, используемыми в решениях; F9 – функция *верификации соответствия скрипта пакету* (сопоставляет сгенерированный скрипт с существующим пакетом визуализации и сохраняет в метаданных скрипта свойство соответствия пакету); F10 – функция *создания в репозитории пользовательских задач объекта «Решение»*, представляющего программный код созданной модели (как результат выполнения функций F4 или F6); F11 – функция *создания нового объекта «Задача» в репозитории пользовательских задач* по текстовому запросу пользователя (автоматически связывает его с порождённым скриптом как с «Решением»).

Проектирование архитектуры системы основано на подходе, ориентированном на знания, где ядром выступает многоаспектная онтология (рис. 1). Платформа разбита на шесть ключевых компонентов (рис. 2).

*Пользовательский интерфейс* – единая точка входа, реализованная через паттерн «Фасад». *Подсистема расширения библиотек визуализации* отвечает за создание новых пакетов над существующими библиотеками (например, Matplotlib, Seaborn, D3 и др.) и обучение языковой модели этим пакетам. *Подсистема генерации скриптов* автоматически превращает естественно-языковые запросы в исполняемые Python-скрипты. *Подсистема предметно-ориентированных языков* – языковой инструментарий для разработки DSL и моделей визуализации с их использованием, а также генерации кода, реализующего визуализации по разработанным моделям [1, 2, 3]. *Модуль интеграции с источниками данных* отвечает за загрузку данных в форматах CSV, XLSX и др. *Подсистема генерации скриптов* – новый компонент – сочетает семантические технологии и большие языковые модели для автоматизации генерации кода визуализации при сохранении истории взаимодействий с пользователем, реализован по принципам многослойной архитектуры, разделяющей функции представления, бизнес-логики и управления данными (рис. 3).

*Уровень представления.* *DashComponent* играет роль корневого контейнера, реализуя пользовательский интерфейс на базе фреймворка *Dash*. Построен по модели реактивного

программирования: обратные вызовы *Dash* связывают действия пользователя с бэкенд-сервисами. Он координирует работу трёх подкомпонентов: 1) *ChatComponent* управляет взаимодействием пользователя с системой в чате в реальном времени через интерфейс *IChatLogic*, отображает историю пользовательских запросов и передаёт ответы модели; 2) *GraphComponent* обрабатывает результаты визуализации, координируя зависимости через три ключевых интерфейса (*IDataParser* – приём и парсинг структурированных данных, *IAIResponse* – отображение сгенерированного ответа модели, *IVisualizationExecutor* – динамическое выполнение сгенерированного кода), реализует интерфейс *IUserRequest* для представления введённых пользователем запросов; 3) *dcc.Store* – это компонент для клиентского хранения состояния прямо в браузере (предоставляет интерфейс *IStateStorage*, позволяющий любым компонентам пользовательского интерфейса читать и записывать временные JSON-данные (например, текущие параметры запроса, настройки графиков или промежуточные результаты). За счёт хранения часто используемого состояния в браузере *dcc.Store* устраняет лишние обращения к серверу и ускоряет работу приложения.

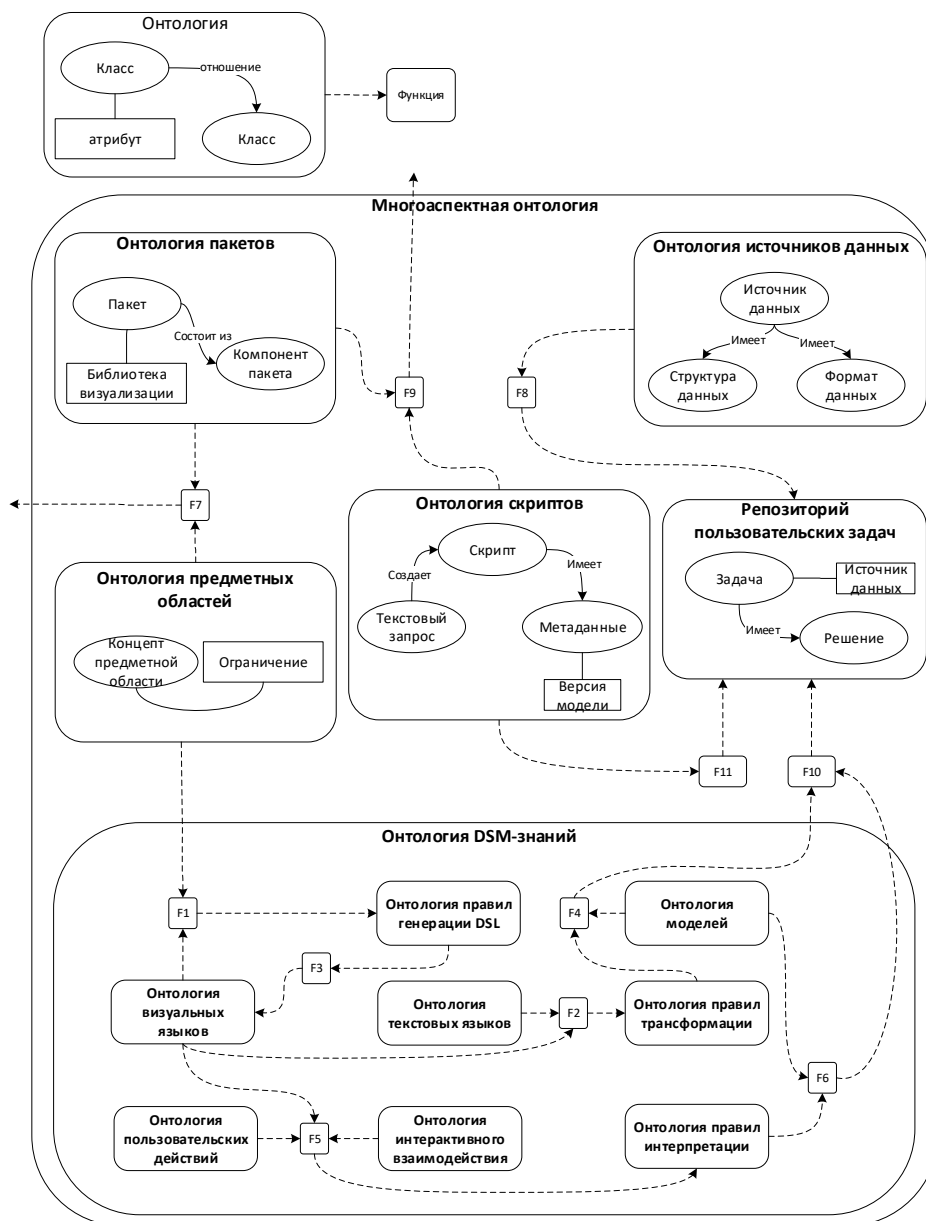


Рис. 1. Структура многоаспектной онтологии

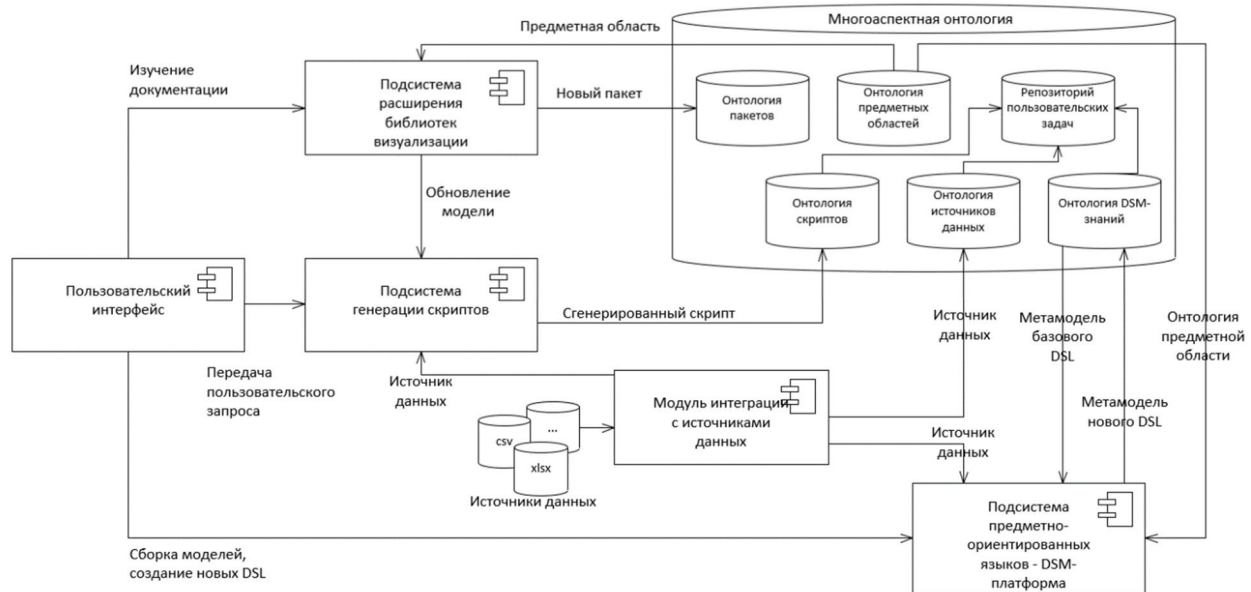


Рис. 2. Обобщённая структура средств визуализации данных, управляемых знаниями

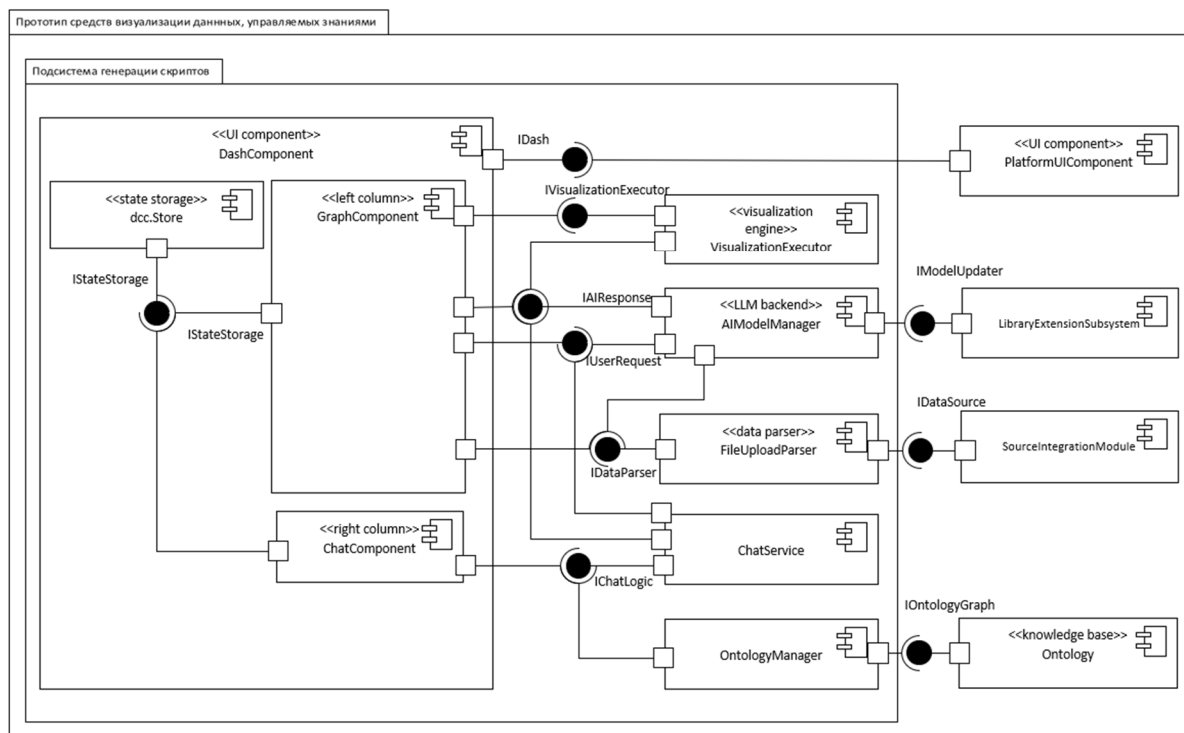


Рис. 3. Архитектура подсистемы генерации скриптов

Уровень бизнес-логики включает пять компонентов, реализующих предметно-ориентированную логику: 1) *FileUploadParser* (*IDataParser*) выполняет парсинг файлов CSV/Excel с помощью *Pandas*, извлекая табличные данные в *DataFrame*; 2) *ChatService* (*IChatLogic*) управляет состоянием сессии чата, обновляет чат при помощи сохраненного контекста через *IStateStorage*; 3) *AIModelManager* (*IAIResponse*) использует пайплайны *LangChain* и API *ChatGroq* (Llama3-70B) для генерации кода визуализации на *Plotly* на основе загруженных данных (*IDataParser*) и пользовательского запроса (*IUserRequest*); 4) *VisualizationExecutor* (*IVisualizationExecutor*) безопасно выполняет сгенерированный Python-код в *sandboxed*-средах; 5) *OntologyManager* реализует операции с онтологией: создание экземпляров сущностей (промптов/скриптов), версионирование моделей и др. Уровень данных (компонент *Ontology*), предоставляющий интерфейс *IOntologyGraph*, выступает в качестве постоянной

базы знаний системы. *Внешние компоненты* системы: *PlatformUIComponent* является основным пользовательским интерфейсом платформы. Он интегрирует *DashComponent* для отображения визуализации и взаимодействия в чате через требуемый интерфейс (*IDash*); *LibraryExtensionSubsystem* отвечает за разработку новых пакетов визуализации и обучение языковой модели этим пакетам. Он реализует интерфейс *IModelUpdater*, через который *AIModelManager* инициирует обновление или селекцию версии LLM-модели при подготовке запроса; *SourceIntegrationModule* передаёт компоненту *FileUploadParser* загруженный в систему файл через интерфейс *IDataSource* для дальнейшей обработки.

Ещё один компонент, который может быть интегрирован в систему, – прототип компонента получения рекомендаций, который на основе сформированной онтологии визуальных языков и образца, представленного пользователем, формирует подсказку по выбору подходящей модели визуализации. Обобщённая структура компонента показана на рисунке 4. Работа компонента основана на использовании описанной онтологии (рис. 1).

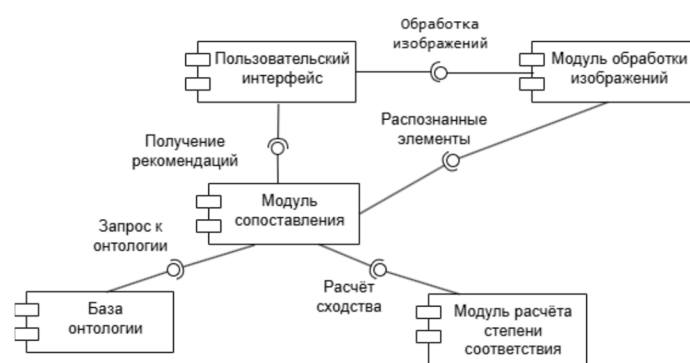


Рис. 4. Структура компонента получения рекомендаций

Для реализации прототипа использованы язык программирования *Python*, библиотека *RDFlib* для работы с онтологиями и модель детекции объектов *YOLOv8*. Все разработки велись в интегрированной среде *PyCharm*. Для извлечения из онтологии информации о типах диаграмм, их характерных элементах и соответствующих весах используются *SPARQL*-запросы.

## Результаты экспериментов

Описанные средства прошли апробацию:

1. Разработаны пользовательские модели визуализации с помощью DSL, созданных на основе базового языка и описания предметной области. С использованием моделей построены визуализации.
2. Проведены эксперименты по генерации скриптов для визуализации данных в соответствии с запросами пользователей различной сложности для данных с разной структурой.
3. Выполнены исследования возможностей подбора подходящих типов визуализации на основе распознавания образцов, предложенных пользователем.

Метамоделей языков формируются как наследники соответствующих классов диаграмм в онтологии. На рисунке 5 показана *мета-модель базового DSL* для создания пользовательских моделей *пузырьковых диаграмм*.

Модель пузырьковой диаграммы, созданная с использованием разработанного языка, приведена на рисунке 6, а. Разработанный язык для создания пузырьковых диаграмм использован в качестве базового для создания нового DSL для визуализации данных о поведении посетителей кафе в зависимости от их предпочтений в кофе и уровня шума в помещении (цель – понять, как шум влияет на выбор кофе и общее время пребывания в кафе). Кастомизированная модель визуализации для этой задачи показана на рисунке 6, б. Диаграммы, разработанные с использованием моделей визуализаций, показанных на рисунке 6, приведены на рисунке 7.

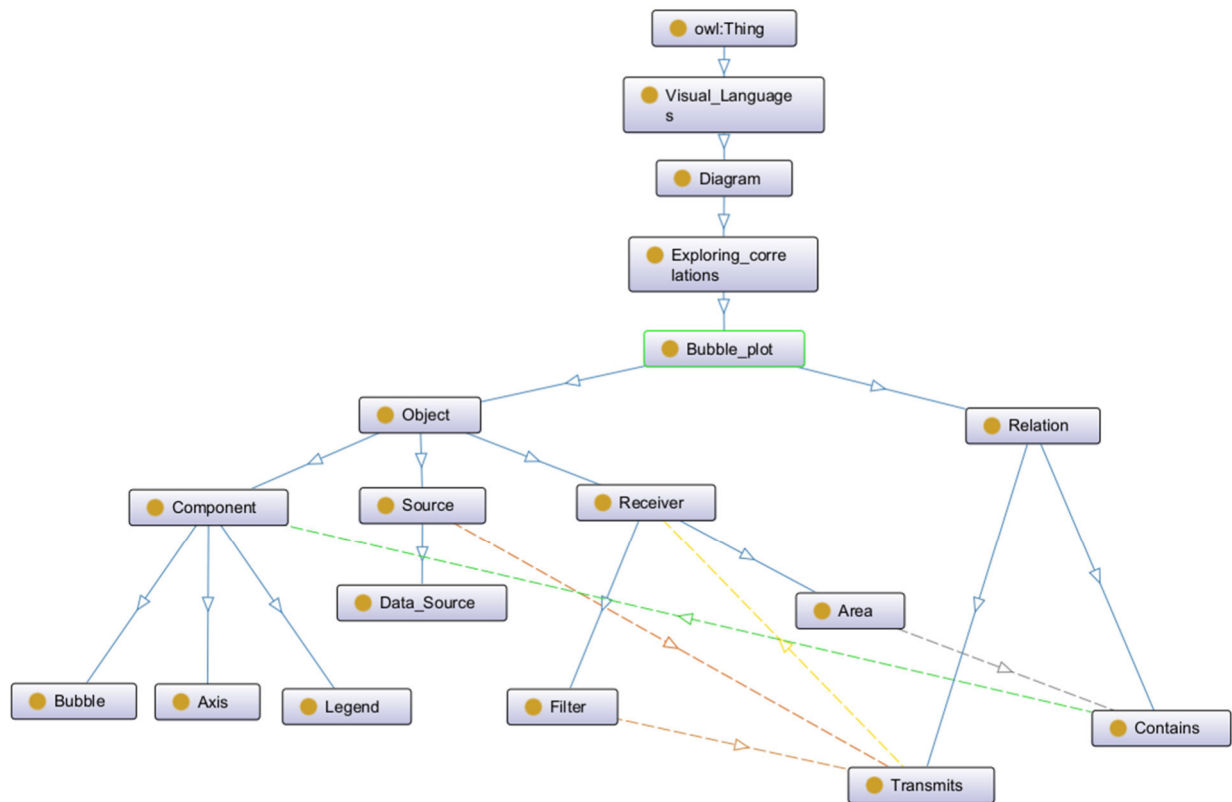


Рис. 5. Представление метамодели языка пузырьковой диаграммы в онтологии (фрагмент онтологии)

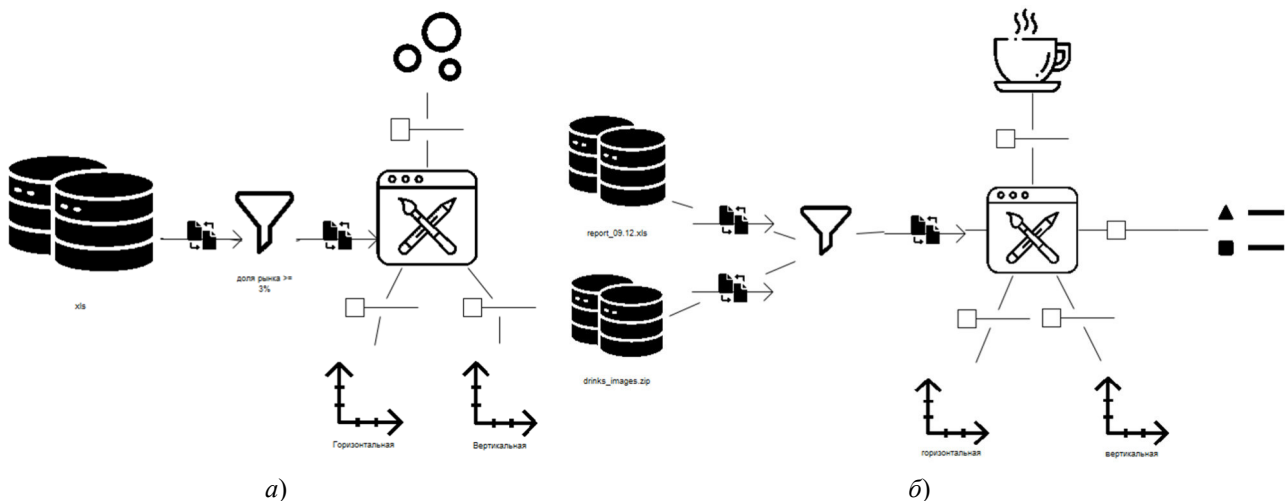


Рис. 6. Использование языка пузырьков диаграмм: а) модель диаграммы о доле рынка, занимаемой отдельными группами товаров, построенная на базовом языке; б) модель диаграммы чашек, созданная на пользовательском DSL

Выполненные при разработке новых DSLs и моделей настройки не требуют от пользователей навыков программирования – для их реализации созданы визуальные конструкторы правил, обеспечивающие как генерацию метамоделей пользовательских DSLs на основе знаний о предметной области, имеющихся в онтологии, так и генерацию кода для реализации визуализаций (в том числе и интерактивных).

С использованием разработанных компонентов проведена серия экспериментов с обработкой данных, имеющих различную структуру, и запросов различной сложности. В качестве примера рассмотрим сложную задачу с выполнением запроса без указания типа диаграммы: «Покажи мне визуализацию, из которой можно сделать много выводов». На рисунке 8 показан пользовательский

интерфейс ИИ-ассистента для генерации визуализаций. Левая панель объединяет загрузчик файлов с поддержкой форматов CSV/Excel; интерактивную таблицу для предварительного просмотра загруженных данных; текстовое поле для ввода запросов на естественном языке и кнопку для запуска обработки ИИ; динамическое отображение графиков Plotly и сгенерированных фрагментов кода. Правая панель отображает историю чата для текущей сессии с разным оформлением для пользовательских сообщений и ответов модели. Приложение на основе введенных данных формирует контекстно-ориентированный промпт. После формирования запрос отправляется на обработку через API Groq с использованием модели LLaMA. Далее код, сгенерированный ИИ, преобразуется в исполняемые рабочие процессы визуализации на Python. Компонент истории чата обеспечивает сохранение диалога.

Учитывая открытый формат задачи, модель сгенерировала точечную диаграмму (рис. 9). Эта визуализация демонстрирует взаимосвязь между предложенными данными: продолжительностью сна, уровнем физической активности и уровнем стресса для каждого участника. Вместе со сгенерированным программным кодом в ответе предложены вопросы для облегчения интерпретации, что помогает анализу данных пользователем.

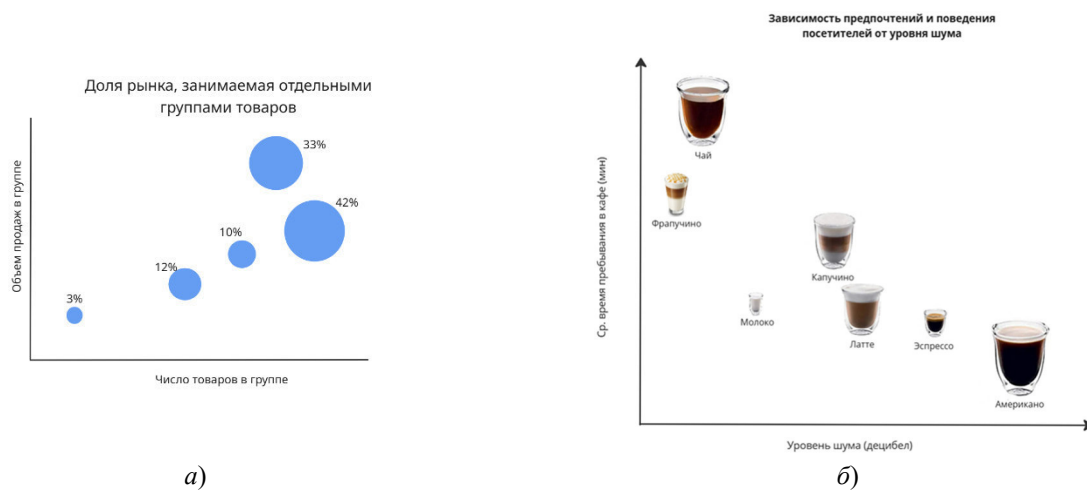


Рис. 7. Использование моделей пузырьковых диаграмм: а) реализация модели, созданной на базовом языке; б) реализация модели, созданной на пользовательском DSL диаграммы чашек

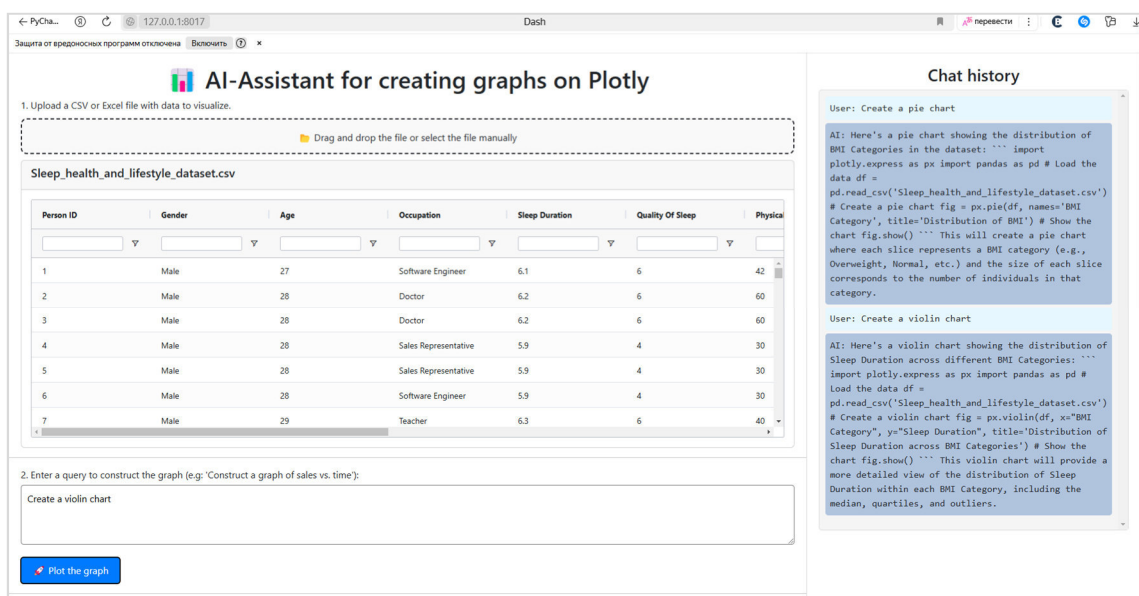


Рис. 8. Пользовательский интерфейс ИИ-ассистента



Рис. 9. Визуализация, сгенерированная по запросу ИИ-ассистентом

Анализ онтологии после проведённых экспериментов подтверждает, что она динамически обновлялась в ходе всех взаимодействий: цепочка «Запрос → Сгенерированный скрипт → Метаданные» сохраняется в виде экземпляров классов, что свидетельствует о надёжной интеграции артефактов в онтологию прототипа.

Была выполнена также *апробация подхода к созданию рекомендательной системы по выбору модели для эффективной визуализации данных*. Процесс определения типа диаграммы по целевому образцу, предоставленному пользователем, включает несколько этапов, при выполнении которых происходит обнаружение ключевых элементов диаграмм-образцов (решение основано на модели YOLOv8), сопоставление результатов распознавания элементов целевой диаграммы с типами диаграмм, описанными в онтологии, и расчёт степени соответствия. Результаты отображаются в виде текста, где перечислены все подходящие типы диаграмм с указанием степени их соответствия образцу и распознанных элементов, повлиявших на оценку.

### Обсуждение результатов

Результаты проведённых ранее экспериментов по разработке пользовательских визуализаций на основе языково-ориентированного подхода [1, 2, 3] показали его практическую значимость при наличии средств автоматизации разработки DSLs и средств трансформации моделей типа «Модель–Текст» (генерации кода по визуальным моделям). Проведённые эксперименты в рамках предложенной методики [6] по интеграции ранее разработанных средств и возможностей ИИ показали хорошие результаты. Благодаря использованию онтологии обеспечивается сохранение и повторное использование метамodelей языков, моделей визуализации и сгенерированного кода, что позволяет существенно сократить трудозатраты при создании новых визуализаций и гарантирует воспроизводимость полученных результатов. В отличие от традиционных инструментов, где глубокая настройка диаграмм требует ручного написания кода, разработанный подход сокращает время создания визуализаций и снижает требования к навыкам пользователей. Разработанный в рамках исследования ИИ-ассистент автоматизирует процесс формирования и исполнения Python-скриптов с использованием библиотеки Plotly. В ходе тестирования ИИ-ассистент успешно справился с различными задачами. При использовании ИИ-ассистента упрощается процесс выбора графического представления, подходящего для конкретных данных или целей анализа. Интеграция ИИ-ассистента с онтологией обеспечивает масштабируемое накопление знаний и расширяет возможности коллективной аналитики. При этом следует помнить, что применение ИИ требует внимательного отношения к вопросам качества и достоверности результатов, так как возможны искажения, неточности в сгенерированных визуализациях.



## Заключение

Предложена и прошла апробацию новая методика визуализации данных, которая, объединяя преимущества нескольких подходов, обеспечивает создание кастомизированных визуализаций без необходимости глубоких знаний в программировании. *Научной новизной* обладают следующие результаты: *интегрированная методика создания средств визуализации данных*, построенная на сочетании различных подходов; *модульная архитектура* системы, управляемой знаниями, обеспечивающая расширяемость средств визуализации, их настройки на новые сценарии использования; *новая классификация методов визуализации данных* – основа для разработки визуальных языков для создания визуализаций; *онтология скриптов*, аккумулирующая пользовательские запросы, генерируемые скрипты и метаданные, что позволяет применять шаблоны кода в решении аналогичных задач и расширяет возможности совместной аналитики.

*Практическая значимость* подтверждается результатами проведения серии экспериментов.

*Направления дальнейших исследований* включают разработку пользовательских пакетов для визуализации данных с функциями тонкой настройки диаграмм на конкретные предметные области, обучение языковой модели для автоматической генерации кода с использованием этих пакетов. Актуальны также исследования возможностей автоматизации разработки онтологий.

## Список литературы

1. Designing Data Visualization System Based on Language-Oriented Approach / A. D. Dzheiranian, I. D. Ermakov, K. A. Proskuryakov, L. N. Lyadova // Proceedings of ISP RAS. 2024. Vol. 36(2). P. 127–140. DOI: 10.15514/ISPRAS-2024-36(2)-10.
2. An Approach to Developing Data Visualization Tools Based on Domain Specific Modeling / A. D. Dzheiranian, I. D. Ermakov, K. A. Proskuryakov, L. N. Lyadova // Scientific Visualization. 2024. Vol. 16(4). DOI: 10.26583/sv.16.4.08.
3. Разработка инструментов визуализации данных на основе предметно-ориентированного моделирования / А. Д. Джейрянян, И. Д. Ермаков, К. А. Проскуряков, Л. Н. Лядова // Материалы 34-й Международной конференции по компьютерной графике и машинному зрению (GraphiCon), Омск, 17–19 сентября 2024 г. С. 300–314. DOI: 10.25206/978-5-8149-3873-2-2024-300-314.
4. Sawicki J., Burdukiewicz M. VisQualdex : a Comprehensive Guide to Good Data Visualization // Scientific Visualization. 2023. Vol. 15(1). P. 127–149. DOI: 10.26583/sv.15.1.11.
5. Making Data Visualization More Efficient and Effective : a Survey / X. Qin, Y. Luo, N. Tang, G. Li // The VLDB Journal. 2020. Vol. 29. P. 93–117. DOI: 10.1007/s00778-019-00588-3.
6. Джейрянян А. Д., Лядова Л. Н. Преодоление ограничений современных инструментов визуализации данных: новая методология // Big Data and Advanced Analytics : сборник научных статей XI Международной научно-практической конференции, Минск (Республика Беларусь), 23–24 апреля 2025 г. С. 386–396.
7. Исаев Р. А., Подвесовский А. Г. Когнитивная ясность графовых моделей: подход к пониманию идеи и способ выявления влияющих факторов с использованием визуального анализа // Научная визуализация. 2022. Т. 14(4). С. 38–51. DOI: 10.26583/sv.14.4.04.
8. Smeltzer K., Erwig M. A Domain-Specific Language for Exploratory Data Visualization // GPCE 2018: Proceedings of the 17<sup>th</sup> ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences. 2018. P. 1–13. DOI: 10.1145/3278122.3278138.
9. Morgan R., Grossmann G., Schrefl M., Stumptner M., Payne T. VizDSL: A Visual DSL for Interactive Information Visualization // Advanced Information Systems Engineering. CAiSE 2018. Lecture Notes in Computer Science(). Vol. 10816. Springer, Cham. P. 440–455. DOI: 10.1007/978-3-319-91563-0\_27.
10. ModelViz: A Model-Driven Engineering Approach for Visual Analytics System Design / A. Khakpour, A. Vazquez-Ingelmo, R. Colomo-Palacios, F. J. Garcia-Penalvo, A. Martini // IEEE Access. 2024. Vol. 12. P. 42667–42682. DOI: 10.1109/ACCESS.2024.3379268.
11. Calliope: Automatic Visual Data Story Generation From a Spreadsheet / D. Shi, X. Xu, F. Sun, Y. Shi, N. Cao // IEEE Transactions on Visualization and Computer Graphics. 2021. Vol. 27(2). P. 453–463. DOI: 10.1109/TVCG.2020.3030403.
12. VIS+AI: Integrating Visualization with Artificial Intelligence for Efficient Data Analysis / X. Wang, Z. Wu, W. Huang, Y. Wei, Z. Huang, M. Xu, W. Chen // Frontiers of Computer Science. 2023. Vol. 17. 176709. DOI: 10.1007/s11704-023-2691-y.
13. Ai4vis: Survey on Artificial Intelligence Approaches for Data Visualization / A. Wu, Y. Wang, X. Shu, D. Moritz, W. Cui, H. Zhang, D. Zhang, H. Qu // IEEE Transactions on Visualization and Computer Graphics. 2021. Vol. 28(12). P. 5049–5070. DOI: 10.1109/TVCG.2021.3099002.

14. Maddigan P., Susnjak T. Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, Codex and GPT-3 Large Language Models // IEEE Access. 2023. Vol. 11. P. 45181–45193. DOI: 10.1109/ACCESS.2023.3274199.
15. Midway S. R. Principles of Effective Data Visualization // Patterns. 2020. Vol. 1(9). 100141. DOI: 10.1016/j.patter.2020.100141.
16. Midway S. R., Brum J. R., Robertson M. Show and Tell: Approaches for Effective Figures // Limnology and Oceanography Letters (L&O Letters). 2023. Vol. 8(2). P. 213–219. DOI: 10.1002/lol2.10288.
17. Schwabish J. A. Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks. New York: Columbia University Press, 2021.
18. Vizml: A Machine Learning Approach to Visualization Recommendation / K. Hu, M. A. Bakker, S. Li, T. Kraska, C. Hidalgo // CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019). 2019. Glasgow, Scotland UK. ACM. 18 p. DOI: 10.1145/3290605.33003581.
19. Kubernátová P., Friedjungová M., van Duijn M. Constructing a Data Visualization Recommender System // Computer and Information Science. Vol. 862. Springer, 2019. Pp. 1–25. DOI: 10.1007/978-3-030-26636-3\_1.
20. Singh A., Goyal C., Parashar J. M. Democratizing Data Visualization and Insights Extraction with Pandas, Generative AI, and CSV Data // International Journal of Scientific Research in Engineering and Management. 2024. Vol. 8(5). P. 1–5. DOI: 10.55041/IJSREM33437.
21. Dhawka P., Dasgupta S. The Social Construction of Visualizations: Practitioner Challenges and Experiences of Visualizing Race and Gender Demographic Data // Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25). 2025. Yokohama. Available from: <https://arxiv.org/html/2502.09048v2>.
22. Challenges and Opportunities in Data Visualization Education: A Call to Action / B. Bach, M. Keck, F. Rajabiyazdi, T. Losev, I. Meirelles, J. Dykes, S. Carpendale // IEEE Transactions on visualization and computer graphics. 2023. Vol. 30(1). P. 649–660. DOI: 10.1109/TVCG.2023.3327378.
23. Classification and Analysis of Techniques and Tools for Data Visualization Teaching / J. J. Cuadrado-Gallego, Y. Demchenko, M. A. Losada, O. Ormandjieva // Proceedings of the IEEE Global Engineering Education Conference (EDUCON). 2021. P. 1593–1599. DOI: 10.1109/EDUCON46332.2021.9453917.
24. Analysis of Current Visualization Techniques and Main Challenges for the Future / M. P. Cota, M. D. Rodríguez, M. R. González-Castro, R. M. M. Gonçalves // Journal of Information Systems Engineering & Management. 2017. Vol. 2(3). P. 19. DOI: 10.20897/jisem.201719.
25. Kirk A. Data Visualization: a Successful Design Process. Birmingham : Packt Publishing LTD, 2012. 206 p.
26. Storytelling and Visualization: An Extended Survey / C. Tong, R. Roberts, R. Borgo, S. Walton, R. S. Laramée, K. Wegba, X. Ma // Information. 2018. Vol. 9(3). 65. DOI: 10.3390/info9030065.
27. The Art of Storytelling in Patient Blood Management: a Narrative Review on Bridging the Gap Between Complex Data and Understandable Information / R. Dubey, M. Auron, C. L. Christenson, Q. Eichbaum, M. Friedman, D. Tolich, C. Bocquet // Annals of Blood. 2024. Vol. 9. DOI: 10.21037/aob-24-24.
28. Peng Z. Telling a Story with Graphs: How to Apply Narrative Visualization Strategies and Visualization Techniques in Journalistic Storytelling // The Palgrave Handbook of Global Digital Journalism. Palgrave Macmillan, Cham. 2025. P. 283–306. DOI: 10.1007/978-3-031-59379-6\_18.
29. Immersive Data-Driven Storytelling: Scoping an Emerging Field Through the Lenses of Research, Journalism, and Games / J. Méndez, W. Luo, R. Rzaev, W. Büschel, R. Dachsel // IEEE Transactions on Visualization and Computer Graphics. 2025. DOI: 10.1109/TVCG.2025.3531138.

## Сравнительная оценка солверов пакета OpenFOAM при моделировании цепи косых скачков сжатия

А. Е. Бондарев, А. Е. Кувшинников  
ИПМ им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В работе представлен сравнительный анализ точности различных солверов OpenFOAM при моделировании формирования цепи косых скачков уплотнения. Исследование фокусируется на оценке способности солверов верно воспроизводить сложную структуру потока, характеризующуюся многократными ударными волнами. Особое внимание уделяется нахождению наиболее точного солвера с учетом изменения параметров: скорости входного потока и угла клина, что соответствует методике обобщенного вычислительного эксперимента. Результаты показывают, что солвер rhoCentralFoam демонстрирует лучшую точность, что делает OpenFOAM эффективным инструментом для изучения и подобных явлений.

**Ключевые слова:** сравнительная оценка точности, вычислительная газовая динамика, цепочка скачков уплотнения, OpenFOAM.

## Comparative assessment of OpenFOAM solvers for simulating oblique shock train

A. E. Bondarev, A. E. Kuvshinnikov  
Keldysh Institute of Applied Mathematics Russian Academy of Sciences, Moscow, Russia

**Abstract.** The paper presents a comparative analysis of the accuracy of various OpenFOAM solvers in modeling the formation of an oblique shock train. The study focuses on assessing the ability of solvers to accurately reproduce a complex flow structure characterized by multiple shock waves. Particular attention is paid to finding the most accurate solver taking into account the change in parameters: the inlet flow velocity and the wedge angle, which corresponds to the methodology of a generalized computational experiment. The results show that the rhoCentralFoam solver demonstrates the best accuracy, which makes OpenFOAM an effective tool for studying similar phenomena.

**Keywords:** comparative assessment of accuracy, computational fluid dynamics, oblique shock train, OpenFOAM.

### Введение

Исследование аэродинамических явлений, связанных с распространением ударных волн, представляет собой фундаментальную задачу в области газовой динамики и вычислительной гидродинамики. Особый интерес вызывают сценарии, где взаимодействие ударных волн с геометрическими препятствиями приводит к формированию сложных волновых структур, таких как отраженные скачки уплотнения и цепочки ударных волн. Такие явления имеют критическое значение для понимания и проектирования широкого спектра инженерных систем. Численное моделирование этих процессов становится незаменимым инструментом, позволяющим детально изучать динамику потока, распределение параметров и характеристики ударных волн, которые сложно или невозможно исследовать экспериментально [1–3].

В последние десятилетия вычислительный эксперимент стал неотъемлемым инструментом в арсенале инженеров и исследователей, занимающихся аэродинамическими задачами. Возможность моделировать сложные течения на компьютере позволяет проводить детальный анализ, который часто недоступен или чрезмерно дорог при использовании одних лишь лабораторных или натурных экспериментов. Вычислительный эксперимент позволяет исследовать широкий диапазон параметров, варьировать геометрию, изучать нестационарные процессы и получать детальную информацию о параметрах потока в любой точке расчетной области. Однако моделирование высокоскоростных течений с ударными волнами представляет собой особенно сложную задачу для численных методов. Ударные волны характеризуются резкими градиентами, что требует применения высокоточных численных схем, способных корректно разрешать эти разрывы без существенных артефактов, таких как осцилляции или искусственное сглаживание. Неправильный выбор численного подхода может привести к неверному определению положения ударной волны, ее интенсивности, а также к искажению других важных характеристик потока. Это подчеркивает важность тщательного подбора и тестирования солверов, используемых для таких задач [4, 5].

Численное моделирование таких процессов требует применения надежных и точных вычислительных методов, способных адекватно описывать разрывные решения, характерные для ударных волн. Программный пакет OpenFOAM [6, 7], являясь мощным инструментом с открытым исходным кодом, предоставляет широкий спектр солверов, каждый из которых обладает своими особенностями в плане точности, устойчивости и вычислительной эффективности при решении различных типов задач. Выбор подходящего солвера является критически важным для получения достоверных результатов, особенно в задачах с ударными волнами, где требуются высокая разрешающая способность и корректное воспроизведение фронтов ударных волн. Понимание преимуществ и недостатков каждого солвера в контексте данной задачи позволит рекомендовать наиболее подходящие инструменты для дальнейших исследований в областях аэродинамики, где возникают ударные волны.

Данная работа является продолжением цикла работ авторов. Ранее в качестве референтных задач рассматривались задача об образовании скачка уплотнения [8], задача об образовании двумерной волны разрежения [9] и задача обтекания конуса, имеющего сферическое затупление [10]. Благодаря применению обобщённого вычислительного эксперимента [11, 12], расчетчики получают возможность уверенно ориентироваться в широком спектре разработанных численных методов. Это позволяет им подбирать для своих расчетов наиболее точные и эффективные решения. Сама методика подразумевает исследование задачи путём дискретизации её определяющих параметров в пределах определённого диапазона, последующего параметрического анализа и визуализации многомерных результатов.

### Постановка задачи

Высокоскоростной поток газа с числом Маха  $M$  течет слева направо (рис. 1).

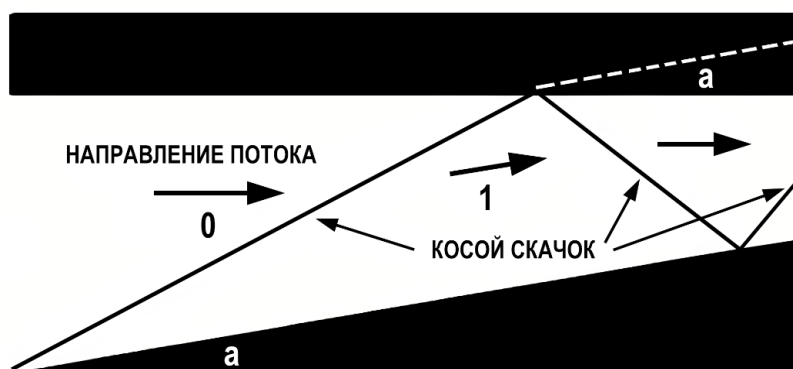


Рисунок 1. Схема течения

Обозначим область свободного потока зоной «0». Поток встречает клин «а» с углом наклона  $\beta$  и создает косой скачок уплотнения, зона за этим скачком уплотнения обозначена как зона «1». Течение в зоне «1» параллельно клину «а», а условия задаются соотношениями для косого скачка уплотнения, приведенными, к примеру, в [13]. Затем косой скачок ударяется о твердую стенку и отражается от нее, создавая новый скачок уплотнения. Течение за отраженным скачком уплотнения обозначено как зона «2». Поскольку поток в зоне «1» параллелен клину «а», он падает на сплошную стенку под углом «а», как показано белой пунктирной линией. Поток в зоне «2» параллелен сплошной стенке, а условия в зоне «2» определяются соотношениями для косого скачка уплотнения, при этом условия перед ним соответствуют условиям в зоне «1». Отраженный скачок уплотнения сам отражается от клина, создавая цепочку скачков уплотнения в канале, образованном клином и сплошной стенкой. При прохождении каждого скачка уплотнения и отражения число Маха потока уменьшается. В конце концов, число Маха в какой-то зоне становится слишком низким для формирования косого скачка уплотнения, и формируется конечный, прямой скачок уплотнения.

Определяющими параметрами задачи в терминах обобщенного вычислительного эксперимента здесь служат число Маха  $M$  и угол клина  $\beta$ . Число Маха менялось от 1.8 до 2.0 с шагом 0.1, угол клина  $\beta$  от  $5^\circ$  до  $10^\circ$  с шагом  $2.5^\circ$ .

### Организация расчётов

В сравнении участвовали четыре солвера: два стандартных солвера — rhoCentralFoam и sonicFoam, а также два авторских — pisoCentralFoam [14] и QGDFoam [15]. Последние два солвера разработаны коллективами Института системного программирования РАН и Института прикладной математики им. М.В. Келдыша РАН.

Эти солверы сильно отличаются: rhoCentralFoam использует схему Курганова – Тадмора, центрально-противопотоковую схему годовновского типа [16], sonicFoam использует алгоритм PIMPLE, включающий в себя метод расщепления [17]. pisoCentralFoam — гибридный метод, использующий как схему Курганова – Тадмора, так и метод PIMPLE [18], QGDFoam основан на квазигазодинамической системе уравнений [19, 20].

Расчётная область разбивается на ячейки. Пакет OpenFOAM для решения требует задания граничных и начальных условий. На входной границе «inlet» задаются параметры невозмущенного набегающего потока (давление  $P = 101325$  Па, температура  $T = 300$  К,  $x$ -компонента скорости  $U_x$  изменяется от 625.05 м/с до 694.5 м/с,  $y$ -компонента скорости  $U_y$  равна 0 м/с). На выходной границе «outlet» задаются граничные условия равенства нулю производных газодинамических функций по нормали к границе. На границе клина «wedge» и на верхней границе «wall» для давления и температуры задается условие нулевого градиента, для скорости задается условие «slip», соответствующее условию непротекания для уравнений Эйлера. Для передней «front» и задней «back» границ используется специальное условие «empty». Это условие задается в случаях, когда вычисления в заданном направлении не проводятся, так как мы решаем двумерную задачу. Схема расчётной области для клина с углом  $\beta = 10^\circ$  представлена на рисунке 2. Стоит отметить, что на указанном изображении для наглядности сетка крупнее, чем в реальных расчётах.

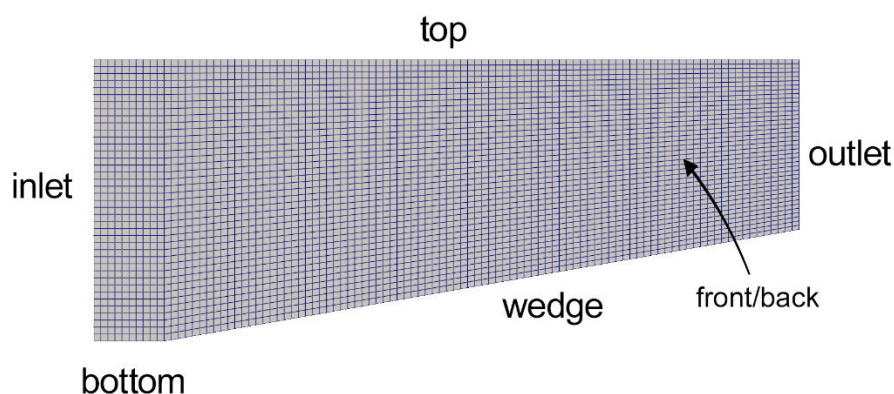


Рисунок 2. Схема расчётной области

Число ячеек сетки зависит от угла наклона, так как от него зависит и число отражений косого скачка. 50000 для  $\beta = 5^\circ$ , 32500 для  $\beta = 7.5^\circ$  и 25000 для  $\beta = 10^\circ$ . Начальные условия соответствуют граничным на грани inlet, то есть в качестве начальных условий используются параметры набегающего потока. В солвере QGDFoam в качестве начальных условий также указывался коэффициент сглаживания  $\alpha = 0.1$  на всей расчетной области. Также задавались значения молярной массы  $M = 28.96$  и удельной теплоемкости при постоянном давлении  $C_p = 1004$ .

OpenFOAM выделяется среди прочих программных пакетов тем, что управление симуляциями осуществляется через текстовые файлы. Такой метод даёт значительную гибкость: он позволяет без труда автоматизировать запуск расчётов, корректировать параметры моделирования и проводить анализ полученных данных.

Единый подход к проведению расчётов крайне важен для сравнения солверов. Он гарантирует, что все они будут тестироваться в идентичных условиях, что делает оценку их производительности и точности объективной. Когда методики расчётов, сетки, граничные условия и физические модели стандартизированы, полученные результаты становятся сопоставимыми и заслуживают доверия. Это даёт исследователям возможность изолировать влияние посторонних факторов и сфокусироваться на особенностях конкретного решателя. Более того, единообразие в проведении тестов помогает лучше

понять преимущества и недостатки каждого солвера, что, в свою очередь, облегчает выбор наиболее подходящего инструмента для решения конкретной инженерной задачи.

В работе с OpenFOAM мы применили те же настройки конфигурационных файлов fvSchemes и fvSolution, что и в своих предыдущих работах.

### Результаты экспериментов

Картинки течения представлены на рисунках 3 и 4 в виде распределения давления и плотности в расчетной области. Представленное распределение давления получено с помощью солвера *rhoCentralFoam*. Разрушения решения не наблюдалось ни для одного из солверов, что свидетельствует о высоких стабилизирующих свойствах всех солверов, участвующих в исследовании.



Рисунок 3. Поле давления установившегося течения для солвера *rhoCentralFoam*,  $\beta = 10^\circ$



Рисунок 4. Поле плотности установившегося течения для солвера *rhoCentralFoam*,  $\beta = 10^\circ$

Построим оценки отклонения от точного решения для всей расчетной области в аналоге нормы  $L_2$ . Для этого определим относительную погрешность  $Err$  для аналога нормы  $L_2$  следующим образом:

$$Err = \sqrt{\sum_m |y_m - y_m^{exact}|^2 S_m} / \sqrt{\sum_m |y_m^{exact}|^2 S_m}.$$

Здесь  $y_m$  — значение исследуемой величины (давления или плотности),  $V_m$  — объём ячейки. Значения  $y_m^{exact}$  получены из аналитического решения задачи [1, 3]. В анализе сравнительной точности участвовали солверы *sonicFoam*, *QGDFoam*, *rhoCentralFoam* и *pisoCentralFoam*. Далее в таблицах для солверов используются сокращенные обозначения: rCF (*rhoCentralFoam*), pCF (*pisoCentralFoam*), sF (*sonicFoam*), QGDF (*QGDFoam*). Значения отклонения от точного решения по всей расчетной области рассчитаны для давления  $p$  и плотности  $\rho$  приведены в таблицах 1–3. Жирным шрифтом выделены наименьшие значения в каждой строке.

Таблица 1. Ошибки для  $M=1.8$ 

Величина	Угол клина	rCF	pCF	sF	QGDF
Давление	5°	<b>0.017524</b>	0.018371	0.030166	0.019833
	7.5°	<b>0.020764</b>	0.021894	0.032143	0.021187
	10°	<b>0.023818</b>	0.025093	0.032409	0.027019
Плотность	5°	<b>0.012136</b>	0.012755	0.020317	0.013353
	7.5°	<b>0.014737</b>	0.015372	0.022233	0.014753
	10°	<b>0.016814</b>	0.017542	0.021391	0.018078

Таблица 2. Ошибки для  $M=1.9$ 

Величина	Угол клина	rCF	pCF	sF	QGDF
Давление	5°	<b>0.020155</b>	0.021083	0.032724	0.022784
	7.5°	<b>0.023237</b>	0.024518	0.035507	0.025689
	10°	<b>0.030908</b>	0.031736	0.041045	0.032071
Плотность	5°	<b>0.014082</b>	0.014973	0.021985	0.015372
	7.5°	<b>0.016177</b>	0.017247	0.023442	0.017129
	10°	<b>0.021463</b>	0.022354	0.027843	0.021769

Таблица 3. Ошибки для  $M=2.0$ 

Величина	Угол клина	rCF	pCF	sF	QGDF
Давление	5°	<b>0.020062</b>	0.021357	0.034088	0.022215
	7.5°	<b>0.023178</b>	0.024817	0.036381	0.025272
	10°	<b>0.027503</b>	0.028304	0.035913	0.027800
Плотность	5°	<b>0.013722</b>	0.014783	0.022694	0.014851
	7.5°	<b>0.016411</b>	0.017425	0.024615	0.017274
	10°	0.019611	0.020549	0.024687	<b>0.019223</b>

Для анализа таблиц визуализируем данные в виде поверхностей ошибок. Результат представлен на рисунках 5 и 6.

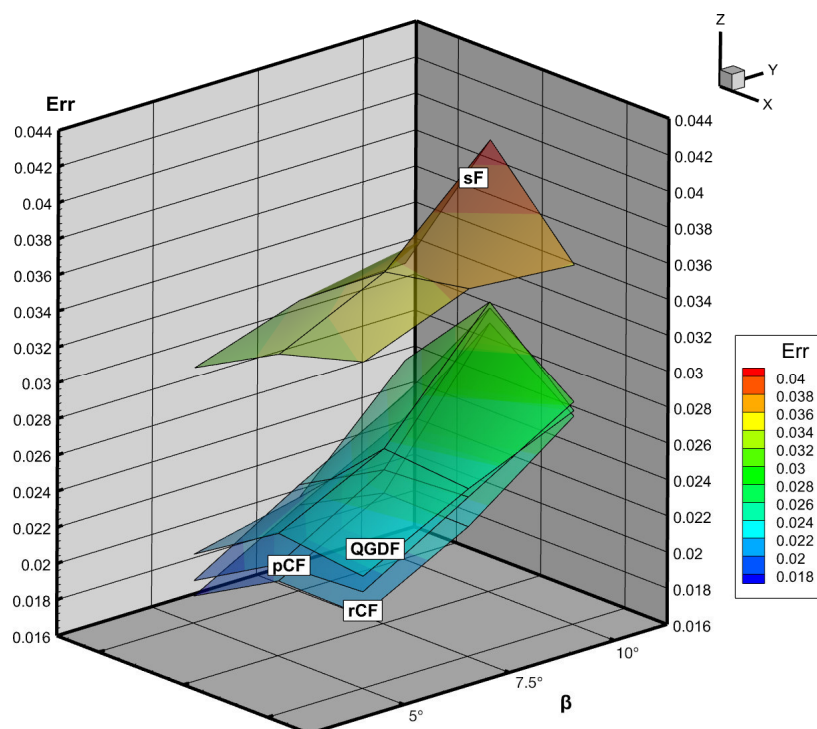


Рисунок 5. Поверхности ошибок рассматриваемых солверов для давления



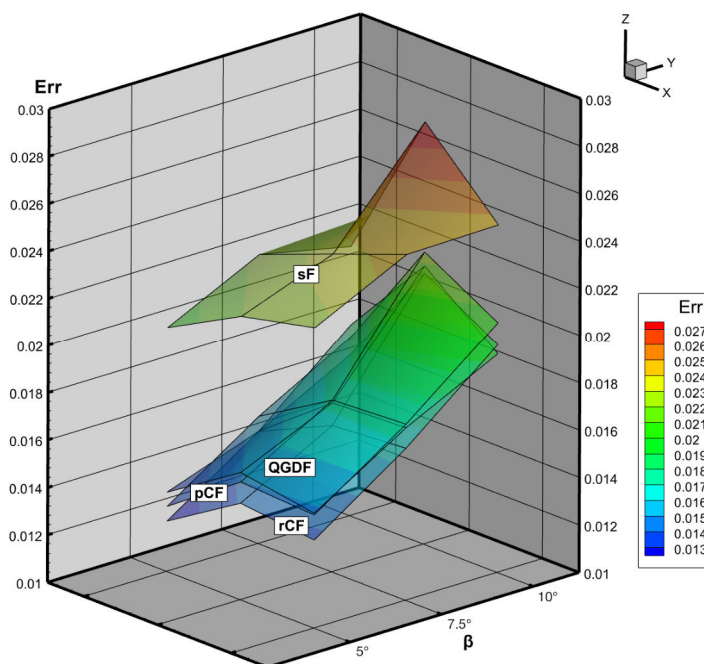


Рисунок 6. Поверхности ошибок рассматриваемых солверов для плотности

Анализ представленных таблиц позволяет сделать ряд важных выводов относительно их точности и зависимости от входных параметров. В первую очередь, следует отметить, что солвер rhoCentralFoam последовательно демонстрирует наименьшие значения ошибок как для расчета давления, так и для расчета плотности во всех без исключения рассмотренных сценариях, что однозначно указывает на его превосходство в точности по сравнению с другими исследованными солверами. Это превосходство проявляется в абсолютных значениях ошибок, но не в относительной устойчивости к изменению условий задачи. Рассматривая зависимость точности солверов от угла клина, можно четко проследить общую тенденцию: с увеличением угла клина ошибки численного решения для всех солверов имеют тенденцию к росту. Это закономерное явление, обусловленное физикой течения: увеличение угла клина приводит к повышению градиентов параметров потока, формированию более интенсивных ударных волн и, как следствие, к усложнению задачи для численных методов. Солвер sonicFoam, однако, демонстрирует наименьший абсолютный прирост ошибки при увеличении угла клина. Например, при  $M=1.8$ , переход от  $5^\circ$  к  $10^\circ$  углу клина для давления приводит к увеличению ошибки rhoCentralFoam на 0.006294 (с 0.017524 до 0.023818), тогда как для sonicFoam этот прирост составляет 0.002243 (с 0.030166 до 0.032409), а для QGDFoam — 0.007186 (с 0.019833 до 0.027019). Это подчеркивает его большую устойчивость к геометрическим изменениям. Аналогичная картина наблюдается и для расчета плотности, где sonicFoam также демонстрирует наименьший рост ошибки.

Влияние числа Маха на точность солверов тоже является существенным. Как видно из таблиц, повышение числа Маха ведет к увеличению ошибок для всех солверов. Это обусловлено возрастанием роли сжимаемости потока, что требует более точных численных схем для адекватного описания явлений, таких как ударные волны, особенно при высоких скоростях. Солвер rhoCentralFoam, несмотря на рост абсолютных значений ошибок с увеличением числа Маха, как и в случае с углом клина, сохраняет свое преимущество в точности.

Сравнивая остальные солверы, можно отметить, что pisoCentralFoam показывает результаты, очень близкие к rhoCentralFoam, но, как правило, с несколько более высокими ошибками. Он также демонстрирует схожую тенденцию к увеличению ошибок с ростом числа Маха и угла клина. QGDFoam занимает промежуточное положение, его результаты часто сопоставимы с pisoCentralFoam, но в некоторых случаях он может уступать ему в точности. Солвер sonicFoam систематически демонстрирует самые высокие ошибки среди всех исследованных солверов, особенно при расчете давления. Важно также отметить, что ошибки при расчете плотности в целом оказываются ниже, чем при расчете давления, для всех солверов и при всех условиях. Это может быть связано с тем, что

изменение плотности меньше изменения давления (для  $M = 2$  и  $\beta = 10^\circ$  давление увеличивается в 2.8 раза, а плотность в 2.08 раза).

Таким образом, на основании детального анализа представленных данных можно сделать вывод, что солвер *gCF* является наиболее надежным и точным выбором для решения задачи обтекания клина в рассматриваемом диапазоне параметров, благодаря своей стабильно низкой ошибке и меньшей чувствительности к увеличению числа Маха и угла клина.

### Заключение

Данная работа подтвердила возможность успешного применения OpenFOAM для моделирования цепочки косых скачков уплотнения. Сравнительная оценка солверов показала, что *rhoCentralFoam* является наиболее предпочтительным выбором для достижения высокой точности, однако и другие исследованные солверы также решают поставленную задачу. Результаты исследования вносят вклад в понимание применимости CFD-методов в аэродинамических расчетах. Полученные данные имеют важное практическое значение для инженеров и исследователей, работающих в области аэродинамики. Это исследование поможет им принимать обоснованные решения при выборе численных методов для своих задач и открывает путь к дальнейшему развитию и применению OpenFOAM в передовых областях науки и техники.

### Благодарности

Вычисления проведены с помощью гибридного суперкомпьютера K100, установленного в Суперкомпьютерном Центре коллективного пользования ИПМ им. М.В. Келдыша РАН.

### Список литературы

1. Anderson J. D. Modern Compressible Flow: With Historical Perspective. McGraw-Hill Education, 2021. 778 p.
2. Anderson J. D. Fundamentals of Aerodynamics. McGraw-Hill Education, 2016. 1152 p.
3. Чёрный Г. Г. Газовая динамика. М.: Наука, 1988. 424 с.
4. Toro E. F. Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction. Heidelberg: Springer Berlin, 2009. 724 p.
5. Ferziger J. H., Perić M., Street R. L. Computational Methods for Fluid Dynamics. Springer Cham, 2009. 724 p.
6. OpenFOAM Foundation. URL: <http://www.openfoam.org> (дата обращения: 02.07.2024).
7. Jasak H. OpenFOAM: Open source CFD in research and industry // Int. J. Nav. Archit. Ocean Eng. 2009. Vol. 1. P. 89–94.
8. Alekseev A. K., Bondarev A. E., Kuvshinnikov A. E. Comparative analysis of the accuracy of OpenFOAM solvers for the oblique shock wave problem // Mathematica Montisnigri, 2019, Vol. XLV. P. 95–105.
9. Bondarev A. E., Kuvshinnikov A. E. Analysis and Visualization of the Computational Experiments Results on the Comparative Assessment of OpenFOAM Solvers Accuracy for a Rarefaction Wave Problem // Scientific Visualization. 2021. Vol. 13, no. 3. P. 34–46.
10. Bondarev A.E., Kuvshinnikov A.E. Integrating Scientific Visualization in the Assessment of OpenFOAM Solvers for the Flow Around a Spherically Blunted Cone // Scientific Visualization. 2024. Vol. 16, no. 4. P. 25–36.
11. Bondarev A.E. On the Construction of the Generalized Numerical Experiment in Fluid Dynamics // Mathematica Montisnigri. 2018. Vol. XLII. P. 52–64.
12. Bondarev A.E., Galaktionov V.A. Generalized Computational Experiment and Visual Analysis of Multidimensional Data // Scientific Visualization. 2019. Vol. 11, no. 4. P. 102–114.
13. Дейч М. Е. Техническая газодинамика. Москва–Ленинград: Госэнергоиздат, 1961. 670 с.
14. United collection of hybrid Central solvers — one-phase, two-phase and multicomponent versions. URL: <https://github.com/unicfdlab/hybridCentralSolvers> (дата обращения 10.07.2025).
15. OpenFOAM framework for simulation of fluid flows using regularized (QGD/QHD) equations. URL: <https://github.com/unicfdlab/QGDSolver> (дата обращения 10.07.2025).
16. Kurganov A., Tadmor E. New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations // J. Comput. Phys. 2000. Vol. 160, no. 1. P. 241–282.
17. Issa R. Solution of the implicit discretized fluid flow equations by operator splitting // J. Comput. Phys. 1986. Vol. 62, no. 1. P. 40–65.
18. Kraposhin M. V., Banholzer M., Pfitzner M., Marchevsky I. K. A hybrid pressure-based solver for nonideal single-phase fluid flows at all speeds // Int. J. Numer. Meth. Fluids. 2018. Vol. 88, no. 2. P. 79–99.
19. Истомина М. А. О реализации одномерного квазигазодинамического алгоритма в открытом программном комплексе OpenFOAM // Препринты ИПМ им. М.В. Келдыша. 2018. № 001.
20. Четверушкин Б. Н. Кинетические схемы и квазигазодинамическая система уравнений М.: МАКС Пресс, 2004. 328 с.

УДК 004.9, 620.1, 621.7

DOI: 10.25686/978-5-8158-2474-4-2025-374-384

## Оценка корректного восстановления технологии штамповки при помощи инструментов цифрового проектирования и моделирования

М. А. Петров, Е. В. Крутина, М. А. Р. Баходирий

Московский политехнический университет, кафедра «Обработка материалов давлением и аддитивные технологии», Москва, Россия

**Аннотация.** Метод разработки и улучшения технологии может быть реализован на основании имеющегося готового изделия, реализуя методику обратного (реверсивного) инжиниринга. В статье рассмотрено поэтапное восстановление технологии неизвестной поковки, напоминающей деталь типа «Фиттинг» или элемент трубопроводной арматуры, изготовленной на гидравлическом прессе с сервоуправлением. Выполнено определение марки материала, создана геометрическая копия поковки и подготовлен деформирующий инструмент для численного моделирования процесса в различных постановках задачи, определены силовые показатели процесса. Предложена таблица объективной шкалы оценки и контроля корректности разрабатываемой технологии.

**Ключевые слова:** обратное проектирование, CAD-CAE-CAI, МКЭ, корректность технологии, D16, 3D-сканер

## Assessment of metal forming technology reconstruction using digital tools for design and simulation

M. A. Petrov, E. V. Krutina, M. A. R. Bakhodiriy

Moscow Polytechnic University, Department "Material Forming & Additive Technologies", Moscow, Russia

**Abstract.** The method of development and technology improvement can be carried out based on the existing (hemi) finished part, implementing the technique of reverse engineering. The paper considers the step-by-step technology reconstruction of an unknown forging, similar to a "Fitting" or a pipeline element manufactured on a hydraulic press with servo control. The material grade was determined, a geometric twin of the forging was created and a forming tool was prepared for numerical process simulation for several various problem formulations. The force characteristics of the process were determined. A table of an objective scale for assessing and monitoring the correctness of the technology have been developed as well.

**Keywords:** reverse engineering, CAD-CAE-CAI, FEM, correctness of technology, AA 2024, 3D-scanner

### Введение

Изготовление новых деталей для перспективных изделий подразумевает выполнение ряда конструкторских и технологических работ. Интенсивное развитие изделий имеет недостаток, заключающийся в утрате технологической информации, необходимой для изготовления конкретной детали. Необходимость восстановления утраченной технологии приводит к появлению различных вариантов производственных технологий. В конечном итоге вместо разработки технологии инженерам и исследователям приходится заниматься инженерной археологией, целью которой является выбор и обоснование правильного варианта реализации утраченной технологии изготовления детали изделия, подготовкой конструкторской документации. Наиболее целесообразным инструментом в этом случае является не экспериментальный метод «проб и ошибок», а виртуальное проектирование, включающее проектирование геометрии, выбор материала и технологическое моделирование. В современном технологическом укладе преимущества разработки цифрового прототипа детали являются неоспоримыми, а используемые для этого инструменты хорошо зарекомендовали себя. Под инструментами в данном случае подразумеваются программные средства и программно-аппаратные комплексы, работающие с гарантированной точностью.

Работы по разработке конструкторской документации на основе имеющейся детали получили название обратное проектирование или реверсивный инжиниринг [1–4]. Ошибочным является мнение о том, что знания исследователя и инженера в этом случае являются поверхностными. Наоборот, объём знаний удваивается, так как прежде чем ввести информацию в компьютер, её выбор необходимо обосновать, в противном случае введённая некорректная информация приведёт к получению

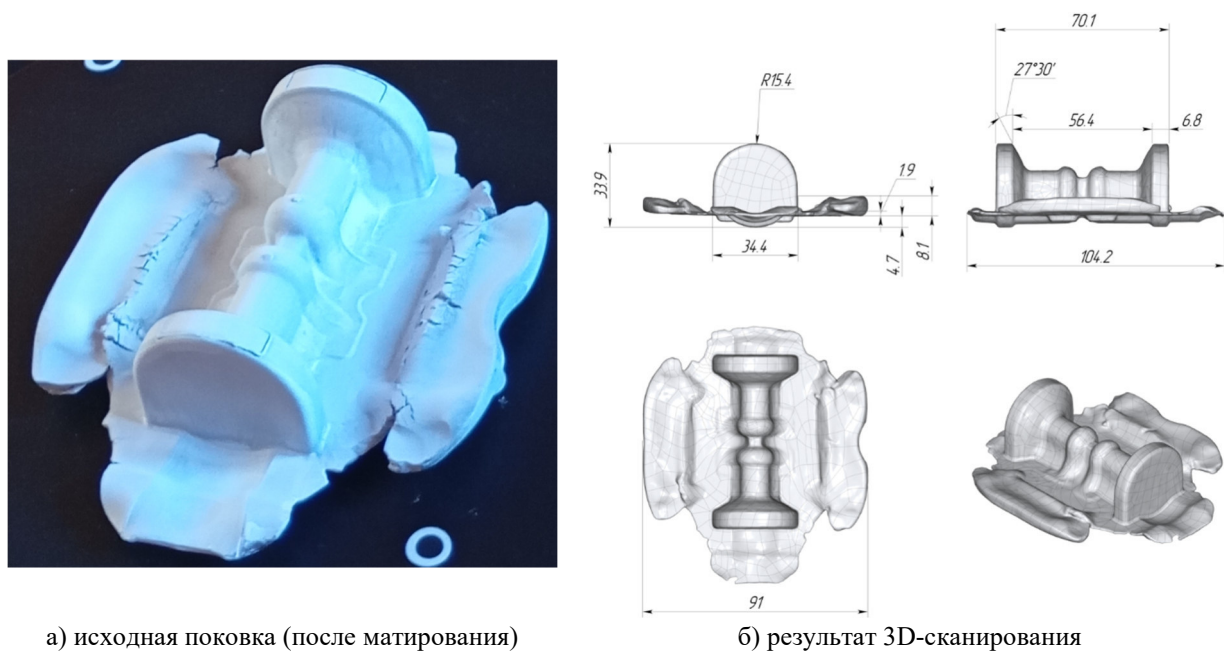
некорректных результатов и затянет разработку конструкторской документации. Подход обратного проектирования является многоуровневым, включающим не только копирование геометрии детали [5–7], но и улучшение технологии, позволяющей её изготовить [8–11].

### Постановка задачи

Объектом исследования является поковка, выполненная из деформируемого алюминиевого сплава Д16, имеющая полезную часть, используемую для получения требуемой детали, и излишек металла (облой), удаляющийся операцией обрезка или механическим способом при помощи фрезерования. Химический состав был определён при помощи портативного спектрометра Melytec ED4 (JPSpec, КНР) и базы данных материалов, результаты измерения представлены в таблице 1. Также известно, что поковка получена на гидравлическом прессе с сервоуправлением. Отсутствует информация о том, выполнялась деформация в горячем или холодном состоянии. Исходя из этого численное моделирование технологического процесса выполнялось для условий холодной и горячей деформации. Общий вид и чертёж поковки представлены на рисунке 1а. Опорные размеры были получены по контактной технологии прямого измерения с использованием цифрового штангенциркуля (точность 10 мкм), а общая топология поверхности определялась бесконтактной технологией обратного проектирования с использованием системы на основе структурированного инфракрасного подсвета RangeVision Neopoint v1 (заявленная точность 50 мкм). Исходной заготовкой под штамповку является цилиндрическая заготовка из прессованного прутка с наиболее подходящим диаметром, согласно размерному ряду по ГОСТ 21488-97, высота которой определялась по объёму полученной 3D-модели. Контрольным показателем являлась масса поковки, определявшаяся с точностью до грамма на электронных весах.

Таблица 1. Химический состав сплава Д16

Элемент	%	Элемент	%
Al	89,48	Ni	0,2004
Cu	3,07	Cr	0,0654
Si	1,503	Mn	0,0227
Fe	0,2344		



а) исходная поковка (после матирования)

б) результат 3D-сканирования

Рис. 1. Общий вид поковки (а), 3D-модель и чертёж поковки с указанием габаритных размеров (б)

Целью исследования является получение корректной, обоснованной технологии штамповки изучаемой детали. Обозначенная цель достигается выполнением следующих задач: 1) геометрическое

проектирование инструмента по поковке; 2) выбор математического описания материала; 3) численное моделирование технологии штамповки для выбранного температурного диапазонаковки и штамповки; 4) сравнение полученных результатов численного моделирования с результатами 3D-сканирования и построение карты отклонений геометрических размеров; 5) введение условной оценочной шкалы по наиболее важным критериям оценки.

### Геометрическое моделирование

Рассматриваемая поковка имеет сложный геометрический профиль, воссоздать который является непростой задачей. Наиболее правильным способом получения информации об объекте будет являться бесконтактное 3D-сканирование и контроль точек, определяющих наиболее значимые размеры, например, габаритные.

Результатом 3D-сканирования являются отдельные сканы, представляющие собой облака точек, на основе которых с помощью разных алгоритмов строится полигональная сетка (ПС). Для корректного получения ПС предварительно выполняется поиск нормалей в каждой точке к предполагаемой поверхности. Неправильное направление нормали приводит к существенному искажению ПС. На рисунке 2 показаны варианты построения ПС в программе MeshLab. Рабочий вариант ПС для дальнейшего проектирования деформирующего инструмента был подготовлен в программе RV 3D Studio, в которой для создания ПС применяется алгоритм Пуассона. Контроль правильности построения ПС выполнялся в программе Geomagic Wrap. Итоговая модель насчитывала 269897 треугольных элементов (фасеток) и 134957 точек (вершин). На следующем шаге ПС была представлена в точных поверхностях, а результат экспортирован в формат твердотельной модели \*.step. Последующая работа по проектированию инструмента выполнялась в программе T-Flex CAD инструментами твердотельного и поверхностного моделирования, в которой были спроектированы пуансон, матрица, боковые плунжеры и рассчитаны заготовки.

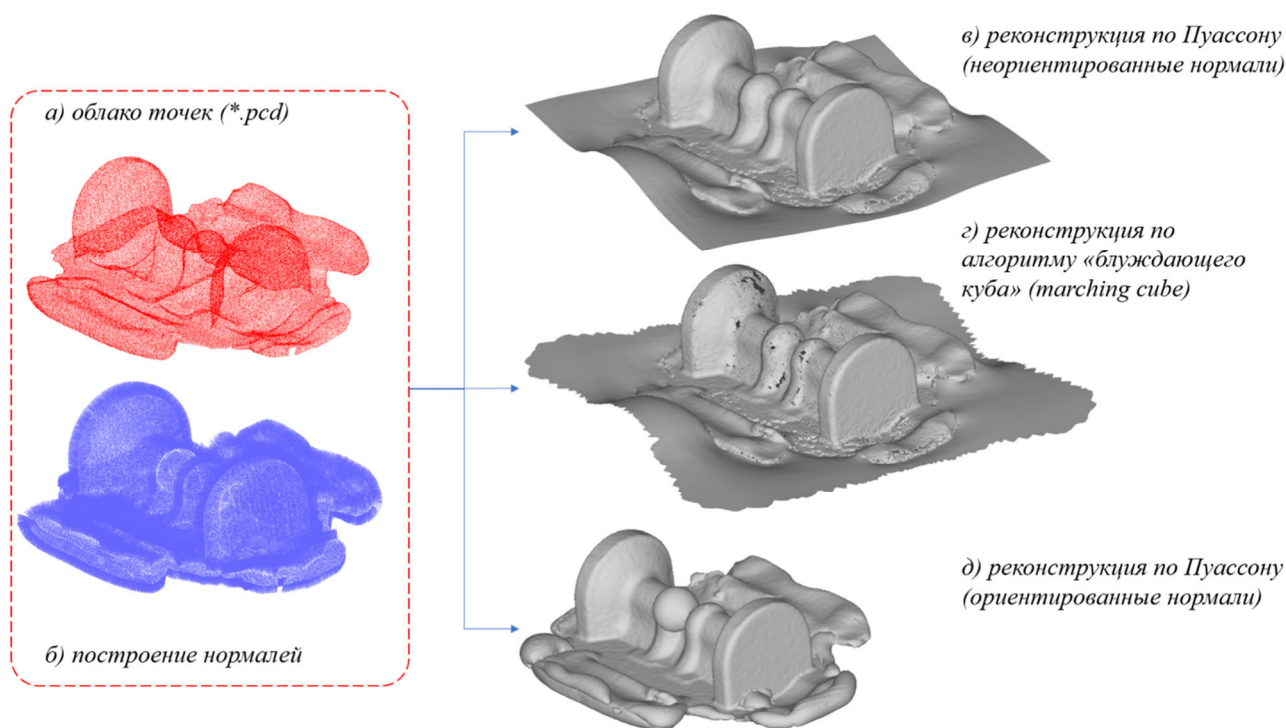


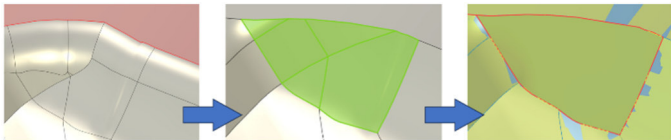
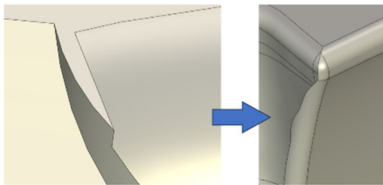

Рис. 2. Особенности построения ПС

При построении опорных поверхностей, относительно которых выполнялось построение деформирующего инструмента, было проведено предварительное выравнивание поковки относительно базовой плоскости (БП) в глобальной системе координат, а затем построены плоскости. Выбор плоскости построения инструмента выполнялся сравнением углов отклонений плоскостей ( $\alpha$ ),



построенных в САД-программе на верхней и нижней гранях облоя (индексы «верх» и «низ» соответственно) в области облойного мостика, с учётом направления перемещения деформирующего инструмента в САЕ-программе. Построение плоскостей на гранях выполнялось по трём произвольным точкам, образованными узлами полигональной сетки ( $\alpha_3$ -УПС), а также по одной точке и условию параллельности базовой горизонтальной плоскости ( $\alpha_1$ -УПС\_БП). В таблице 2 представлены результаты определения углов между плоскостями, а также массы поковки (М). Расстояние между плоскостями с индексами «1-УПС\_БП\_верх-низ» составляло 1,495 [мм]. Эта величина была принята за толщину облойного мостика. Наиболее подходящими плоскостями являются плоскости с минимальными значениями углов. Для построения инструментов была выбрана средняя плоскость, располагающаяся между плоскостями с индексами «1-УПС\_БП\_верх/низ». Дальнейшие корректировки геометрии выполнялись при помощи инструментов твердотельного моделирования. Следует отметить, что были выявлены участки с наложением фасеток. Для исправления геометрии такие участки удалялись с последующим затягиванием области по криволинейному закону. Перед технологическим моделированием выполняется скругление рёбер. Некоторые рёбра не скруглялись, так как это приводило к появлению более мелких рёбер, затрудняющих подготовку КЭ-сетки в САЕ-программе.

Таблица 2. Определение углов наклона плоскостей, массы и критических элементов геометрии

Показатель	Ед. измерения	Значение	
$\alpha_3$ -УПС_верх-низ	[град]	1,023	
$\alpha_3$ -УПС_верх_1-УПС_БП_верх		0,403	
$\alpha_3$ -УПС_низ_1-УПС_БП_низ		0,664	
$\alpha_3$ -УПС_верх_БП		0,403	
$\alpha_3$ -УПС_низ_БП		0,664	
М <sub>весы</sub>	[кг]	0,1350	а) пример локальной замены нескольких фасеток на одну 
М <sub>САД</sub>		0,1368	

Общая ошибка геометрического моделирования ( $\Delta_{\text{гм}}$ ) может быть представлена в виде суммы ошибок реконструкции модели на разных этапах:

- 3D-сканирования ( $\Delta_{\text{ск}}$ ), включает оптическую ошибку устройства сканирования ( $\Delta_{\text{ус}}$ ), связанную с разрешением матрицы 3D-сканера, ошибку, связанную с неравномерным матированием поверхности ( $\Delta_{\text{м}}$ ) и, соответственно, отсутствием информации о плохо видимой 3D-сканером области, ошибку финального совмещения ( $\Delta_{\text{фс}}$ ), связанную с начальной калибровкой 3D-сканера ( $\Delta_{\text{нк}}$ ) и промежуточной точностью совмещения сканов в ручном ( $\Delta_{\text{срр}}$ ) и автоматическом ( $\Delta_{\text{сар}}$ ) режимах);
- построения ПС ( $\Delta_{\text{пс}}$ ) (ошибка объединения фасеток ( $\Delta_{\text{оф}}$ ), закрытие отверстий ( $\Delta_{\text{зо}}$ )).

Ошибки матирования, начальной калибровки, а также ручного совмещения сканов можно отнести к категории человеческого фактора ( $\Delta_{\text{чф}}$ ). Общая ошибка может быть представлена уравнением (1).

$$\Delta_{\text{гм}} = \Delta_{\text{ск}} + \Delta_{\text{пс}}, \quad (1)$$

где  $\Delta_{\text{ск}} = \Delta_{\text{ус}} + \Delta_{\text{м}} + \Delta_{\text{фс}}$ ;  $\Delta_{\text{фс}} = \Delta_{\text{нк}} + \Delta_{\text{срр}} + \Delta_{\text{сар}}$ ;  $\Delta_{\text{пс}} = \Delta_{\text{оф}} + \Delta_{\text{зо}}$ ;  $\Delta_{\text{чф}} = \Delta_{\text{м}} + \Delta_{\text{нк}} + \Delta_{\text{срр}}$ .

Гарантированная ошибка устройства сканирования определяется производителем системы и составляет по спецификации 50 [мкм]. Значение ошибки, вызванной человеческим фактором, требует анализа. Так, для неравномерного матирования поверхности оценку можно провести по отсутствующей геометрической информации, отнесённой к общей площади поверхности. На рисунке 3 показана область поковки с явным искажением геометрии (область 3). Общая площадь поверхности составляет 19934,97 [мм<sup>3</sup>], из которых 29,25 [мм<sup>3</sup>] отводится на интересующую область, что составляет ~0,15%. Кроме того, на ПС отсутствуют трещины, что указывает на неполноту собранной информации. Такая степень детализации не требуется при проектировании поковки, поэтому можно считать, что матирование было выполнено на достаточном уровне. Величина ошибки ручного и автоматического совмещения составила 0,261 [мм], а ошибка начальной калибровки – не превышает 50 [мкм]. Ошибка при финальном совмещении составила 0,14 [мм]. Данная величина представляет собой

среднеквадратичное отклонение точек одного скана до ближайших к ним точек другого совмещаемого скана. Значение ошибки построения ПС при высокой и средней детализации, как правило, весьма мала. При большом значении ошибки этого типа геометрия сильно искажается, что отчётливо проявляется на результате. Чтобы определить её, необходимо построить карту отклонения геометрических размеров (КОГР) для всего рассматриваемого объёма. Так как сложные поверхности описываются с помощью сплайнов, получение точных измерений при помощи контактной техники (штангенциркуль, линейка и т.д., ошибка которых составляет  $\frac{1}{2}$  цены деления прибора) теряет смысл. Однако качество перевода полигональной модели в твердотельную может иметь смысл. Так, результат сравнения модели поковки с обломом, полученный в программе GOM Inspect, после 3D-сканирования и конвертации в твердотельную модель с целью дальнейшей разработки инструмента представлен на рисунке 4. Видно, что отклонение почти всей поверхности лежит в интервале значений  $\pm 20$  [мкм]. Имеются локальные области, где отклонения достигают значений  $\pm 50$  мкм, сопоставимых с заявленной точностью системы 3D-сканирования. Таким образом, величина ошибки геометрического преобразования моделей не превышает 50 [мкм].

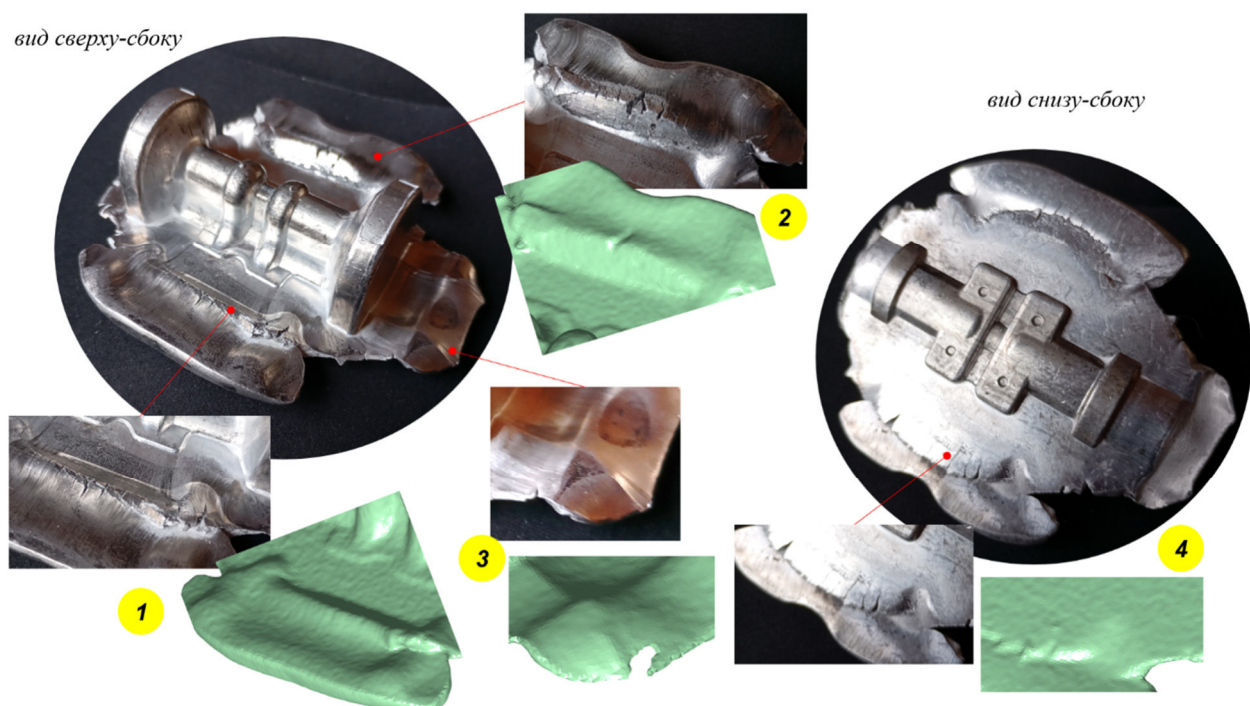


Рис. 3. Сравнение локальных несоответствий в модели геометрии поковки

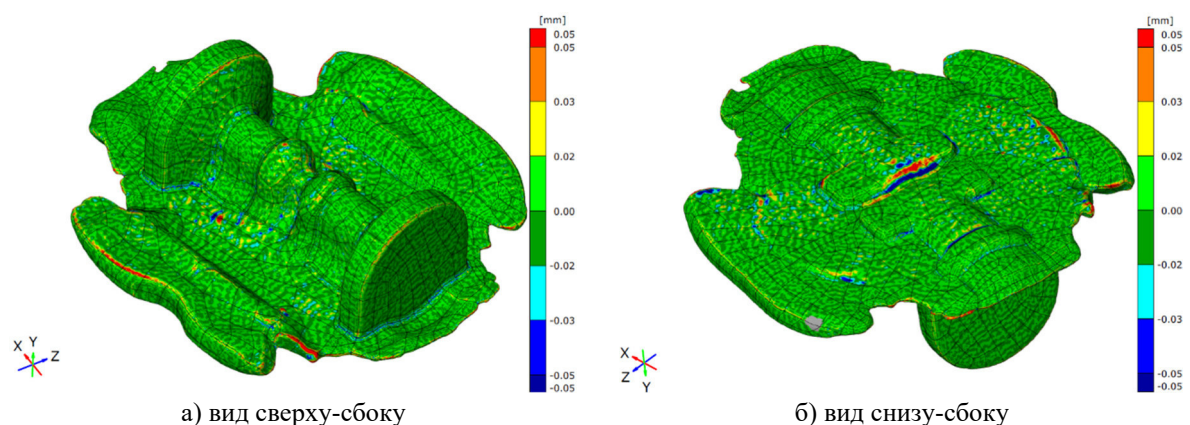


Рис. 4. Построение КОГР для сравнения поволоков после твердотельного преобразования



### Технологическое моделирование

Численное моделирование технологического процесса для случаев холодной объёмной штамповки (ХОШ) и горячей объёмной штамповки (ГОШ) выполнялось в программе QForm, работающей по методу конечных элементов (МКЭ). Расчёт включал в себя несколько стадий или операций, которые обозначены как вариант 1 и 2. В варианте 1 на первой операции выполняется штамповка основным ползуном с закреплённым на нём пуансоном, боковые ползуны (плунжеры) – неподвижны. На второй операции основной ползун останавливается и задействуются боковые плунжеры. Материал заготовки – Д16 (по ГОСТ) или АА2024 (по AISI), материал инструмента – 4Х5МФС (по ГОСТ) или H11 (по AISI). Исходные данные для моделирования представлены в таблице 3. Моделирование задачи осуществлялось в трёхмерной постановке, без назначения плоскостей симметрии. Трение на контактной поверхности материала инструмента и заготовки определялось по закону Леванова, согласно уравнению (2).

$$\tau = mk \left( 1 - e^{-b \frac{\sigma_n}{\sigma_s}} \right), \quad (2)$$

где  $\tau$  – касательное напряжение на контактной поверхности, [МПа];  $m$  – фактор трения, [-];  $k$  – макс. касательное напряжение при сдвиге, [МПа];  $b$  – экспериментальный коэффициент Леванова, [-];  $\sigma_n$  – нормальное контактное давление, [МПа];  $\sigma_s$  – напряжение текучести или сопротивление деформации, [МПа].

Таблица 3. Исходные данные для моделирования

Параметр	Ед. измерения	Значение
Температура:	[°C]	
- заготовки		20 (ХОШ) и 500 (ГОШ)
- инструмента		20 (ХОШ) и 250 (ГОШ)
Макс. сила деформирования:	[МН]	
- пуансона		120
- бокового плунжера		50
Фактор трения:	[-]	0,15 (ХОШ) и 0,3 (ГОШ)
Макс. перемещение ползуна:	[мм]	31
Макс. перемещение боковых плунжеров:	[мм]	2

#### 1. Выбор оборудования

Особенностью работы на прессах с сервоуправлением является возможность задавать закон движения выходного звена, что позволяет решать большой круг технологических задач на одном оборудовании. В классических прессах циклограммы имеют вид синусоиды. В прессах с сервоуправлением возможны установки выстоя в нижних крайних положениях и задействование дополнительных вспомогательных устройств. На рисунке 5 показаны циклограммы перемещения изучаемых процессов для случаев ХОШ с задействованием боковых ползунов.

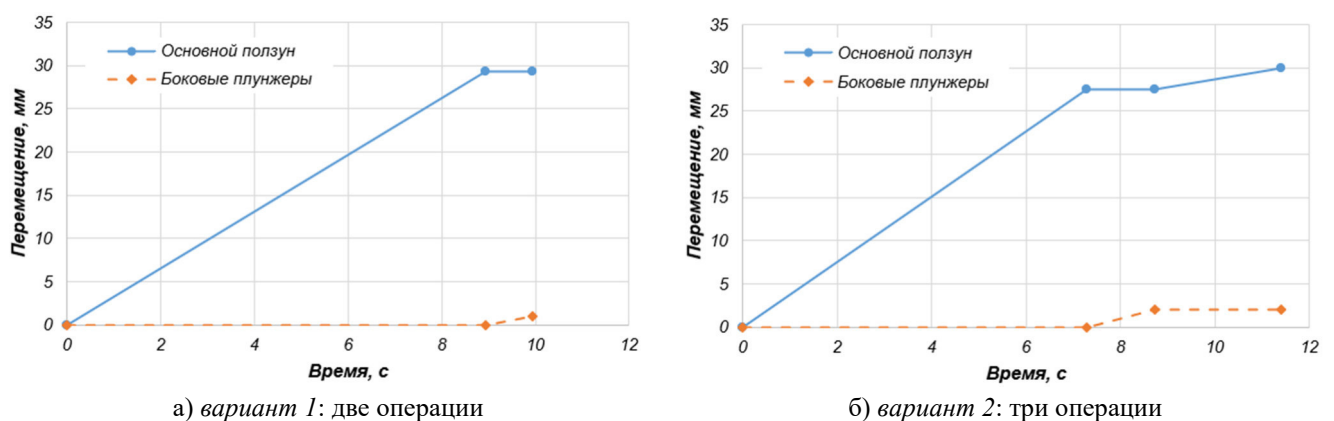


Рис. 5. Циклограммы перемещения подвижных (выходных) звеньев

## 2. Выбор заготовки

Для точного вычисления размеров заготовки была выполнена параметризация размеров (рис. 6). Наиболее подходящими диаметрами прессованных алюминиевых заготовок, согласно ГОСТ 21488-97, были выбраны  $\varnothing 35$  и  $\varnothing 50$  мм, для которых рассчитывались высоты. Моделирование выполнялось для заготовки  $\varnothing 35$ , ориентированной вдоль наибольшей стороны поковки.

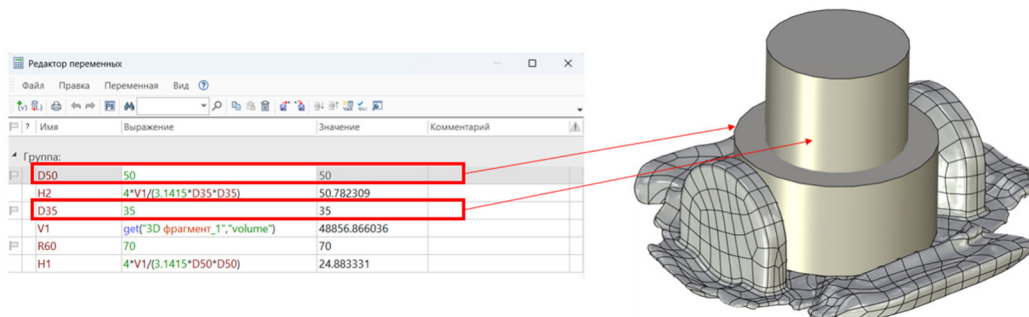


Рис. 6. Определение размеров исходных заготовок

## Обсуждение результатов

На рисунке 7а показаны графики сила-перемещение ползунов, а на рисунке 7б – применение дополнительных боковых ползунов на промежуточной второй операции. Видно, что на третьей операции (область «Б») имеется место повышения силы деформирования, а максимальное значение силы составляет 5,5 МН. Сила же деформирования на боковых ползунах не превышает 0,2 МН. Штамповка в горячем состоянии намного выгоднее с силовой точки зрения. Однако энергия деформирования уравнивается энергией, затрачиваемой на нагрев заготовки. Значение силы деформирования резко увеличивается в тот момент, когда зазор между инструментами уменьшается, скорость истечения в облойном мостике увеличивается, материал испытывает затруднённое течение по свободной поверхности, преодолевая растущие силы трения. Одновременно с этим направление течения на границе облой-гравюра штампа меняется и материал медленно заполняет большие выступы поковки. Номинальная сила оборудования (табл. 3) по сравнению с расчётной силой весьма высока и скорее всего не будет достигнута. Это позволяет выбрать оборудование с меньшими силовыми характеристиками.

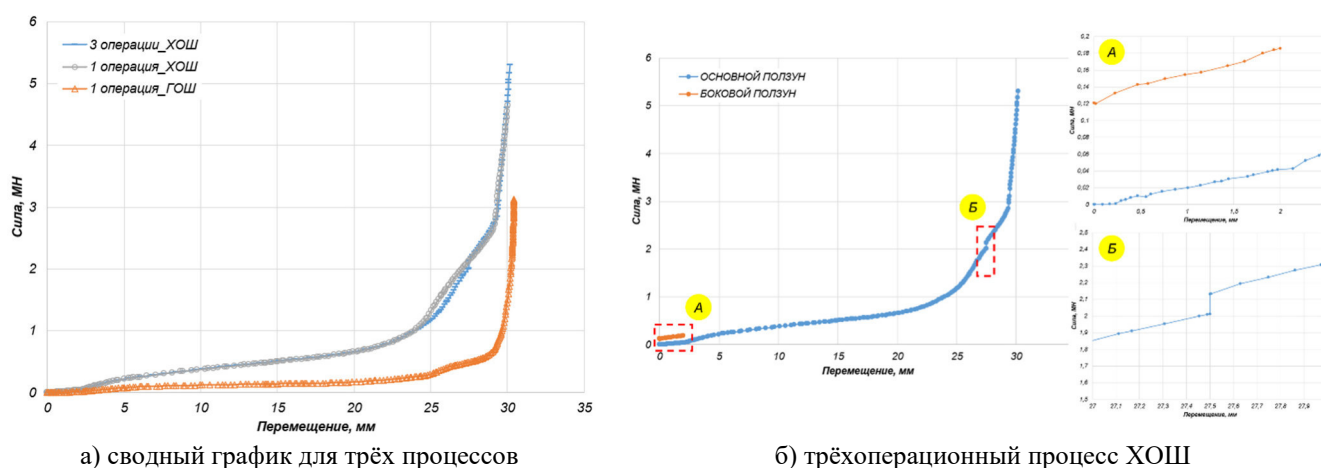


Рис. 7. Графики сила-перемещение выходных звеньев пресса

На рисунке 8 показаны результаты моделирования штамповки до величины хода 30 мм. Наибольший процент заполнения материалам наблюдается для ХОШ с одной операцией. В области больших выступов имеются незаполнения.

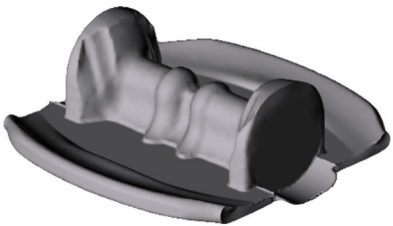
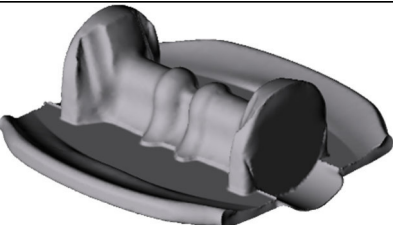
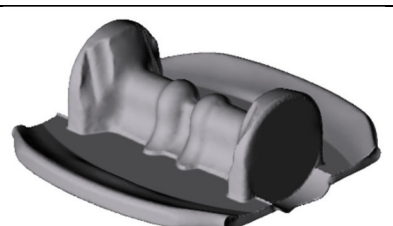
Процесс	Результат	Продолжительность процесса и объём поковки
ХОШ (одна операция)		а) $t = 9,677$ [с], $V = 48677,34$ [мм <sup>3</sup> ]
ХОШ (три операции)		б) $t = 11,26$ [с], $V = 48670,96$ [мм <sup>3</sup> ]
ГОШ (одна операция)		в) $t = 0,6696$ [с], $V = 48641,36$ [мм <sup>3</sup> ]

Рис. 8. Общий вид поверхностей поковок

Для полного заполнения гравюры штампа необходимо увеличить перемещение основного ползуна до 31...31,5 [мм], а также подобрать момент включения боковых ползунов, так как результаты ХОШ для одной и трёх операций лишь немногим отличаются друг от друга. Температурный интервалковки и штамповки для Д16 составляет 250 °С...550 °С. На протяжении всего процесса моделирования температура находится в указанном температурном интервале, лишь на последней стадии температура на наиболее удалённой области облоя снижается до 230 °С, что является несущественным.

Полученная после технологического моделирования полигональная модель была сопоставлена с твердотельной моделью, полученной на основании 3D-сканирования (рис. 9).

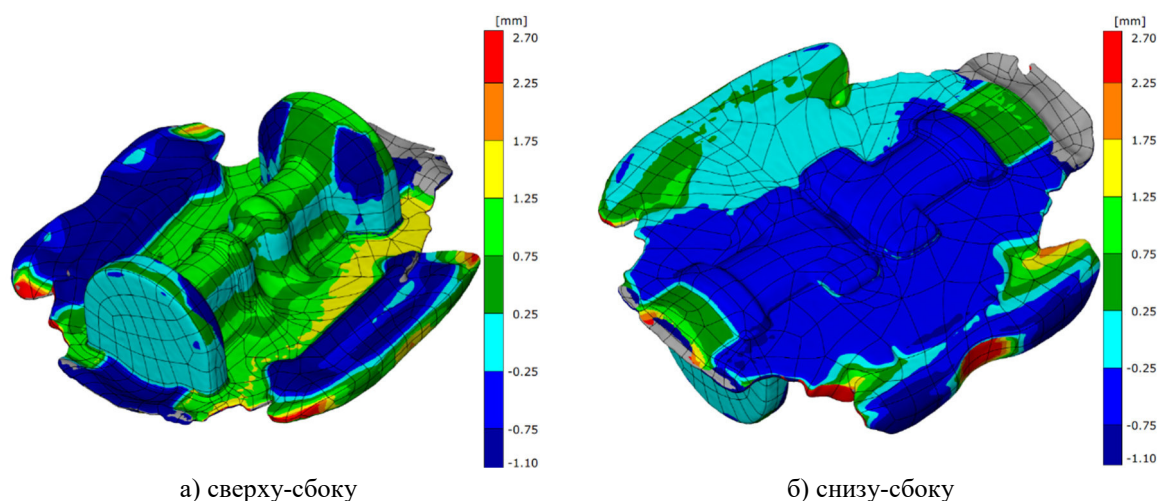


Рис. 9. Сравнение поковок и построение КОГР

Отклонение по точности при предварительном совмещении моделей составило 0,8707 [мм]. Принимая во внимание ГОСТ 7505-89 как обобщённое руководство по проектированию поковок, можно определить, что для степени сложности С4, классу точности Т4 (индекс поковки 10) или Т5 (индекс поковки 12), величина предельных отклонений линейных размеров составит +2,7 (верхнее)

и -1,1 (нижнее). Для этого интервала тело поковки имеет завышенные значения только в местах неполного заполнения инструмента, в области больших выступов. Более 85 % размеров области облоя лежат вне допустимого интервала значений. Так как облой удаляется на последующей операции, то с технологической точки зрения поковка является годной при условии, что заполнены большие выступы.

Возможна также разработка и альтернативных вариантов реализации технологии. Например, задача может быть представлена в упрощённом виде как симметричная относительно фронтальной плоскости. На рисунке 10 показаны полученные результаты для случаев безоблойной (слева) и облойной штамповки (справа). В области большого выступа также видны небольшие незаполненные материалом области.

Сравнение объёмов исходных заготовок и коэффициент использования материала (КИМ) приведён в таблице 4. КИМ рассчитывался как избыточный на основании наименьшего из предложенных объёмов заготовки, при этом наименьшее значение принималось за базовый идеальный расход 100 %.

Таблица 4. Определение расхода материала

Вариант	Значение объёма заготовки, мм <sup>3</sup>	КИМ
Технология на основе реверс-инжиниринга	48856,86	+49,79%
Разработанная технология (безоблойная штамповка)	32616,732	100%
Разработанная технология (облойная штамповка)	38730,81	+18,75%

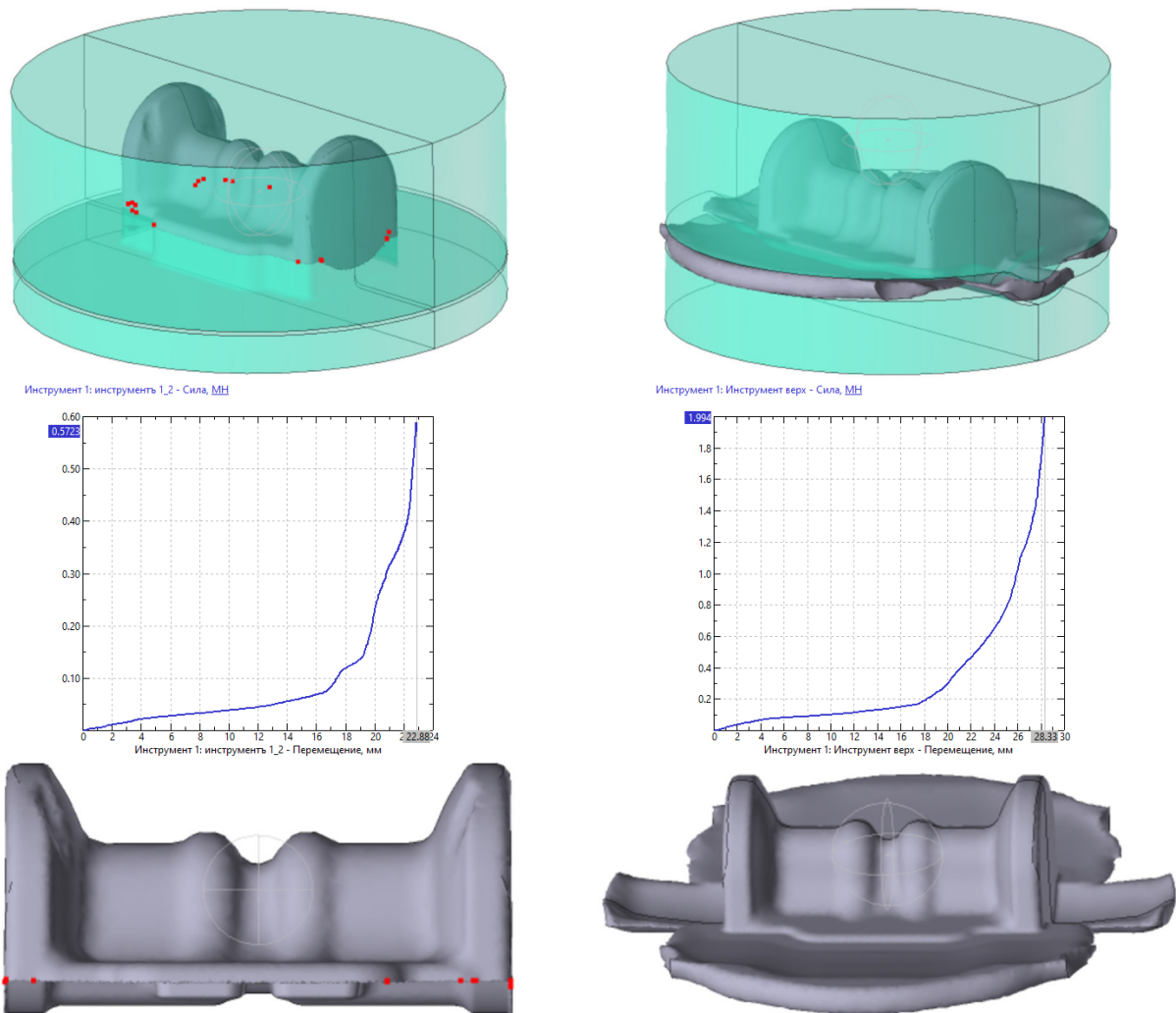


Рис. 10. Упрощённая постановка задачи: деформирующий инструмент (*вверху*); графики сила-перемещение ползуна (*посередине*) и полученные поковки (*внизу*)

На основании выполненных работ обратного проектирования технологии объёмной штамповки можно предложить следующий метод оценки её корректного выполнения. На рисунке 11 показаны критерии с оценочной шкалой от 1 до 5. Выделены три оценочные области, объективно отражающие качество выполненных работ. По результатам выполненных работ достигнута высокая точность реконструкции технологии.

№	Критерий	1	2	3	4	5
1	Точность воспроизведения геометрии (CAD)					
2	Точность массовой и объёмной характеристики поковки					
3	Точность допусков поковки (CAE)					
4	Циклограмма прессы обеспечивает полное заполнение гравюры инструмента					
5	Реологическое описание свойств соответствует оригиналу					

5...15 - низкая точность реконструкции технологии

16...20 - средняя точность реконструкции технологии

21...25 - высокая точность реконструкция технологии

Рис. 11. Предлагаемая оценочная шкала

## Выводы

В статье представлена методика оценки корректности восстановления технологии штамповки сложной детали из алюминиевого сплава. Выполнено геометрическое проектирование инструмента по поковке на основании полигональной модели, полученной при помощи 3D-сканирования, определён материал на основе анализа его химического состава и выбрано реологическое описание течения материала, проведено численное моделирование технологии штамповки для условий холодного и горячего деформирования, определены параметры оборудования и циклограммы движения гидравлического прессы с боковыми ползунами с сервоприводами, выполнена оценка точности реконструкции геометрии для проектирования деформирующего инструмента и построены КОГР для результатов численного моделирования с проверкой вхождения в интервал размерных допусков и, наконец, введена условная оценочная шкала по наиболее важным критериям работы.

Таким образом, обратное проектирование технологии носит комплексный характер, не ограничивается геометрической реконструкцией объекта. Для его выполнения требуются практические навыки работы на компьютере, а также объективные практические знания в профильной области.

## Список литературы

1. Várady, T., Martin, R. R., & Cox, J. (1997). Reverse engineering of geometric models – an introduction. *Computer-Aided Design*, 29(4), 255–268. DOI: 10.1016/s0010-4485(96)00054-1
2. Raja, V., & Fernandes, K. J. (Eds.). (2008). *Reverse Engineering*. Springer Series in Advanced Manufacturing. DOI: 10.1007/978-1-84628-856-2
3. Wang, W. (2010). *Reverse Engineering: Technology of Reinvention* (1st ed.). CRC Press. DOI: 10.1201/EBK1439806302
4. Schnakovsky, C., Ganea, B., Raveica, C., & Herghelegiu, E. Reverse engineering for automotive industry. *Annals of the Oradea University. Fascicle of Management and Technological Engineering*, Volume VII (XVII), 2008. DOI: 10.15660/AUOFMTE.2008.1203
5. Benkő, P., Martin, R. R., & Várady, T. (2001). Algorithms for reverse engineering boundary representation models. *Computer-Aided Design*, 33(11), 839–851. DOI: 10.1016/s0010-4485(01)00100-2
6. Kjellander, J. A. P., & Rahayem, M. (2009). An Integrated Platform for 3D Measurement with Geometric Reverse Engineering. *Computer-Aided Design and Applications*, 6(6), 877–887. DOI: 10.3722/cadaps.2009.877-887
7. Henry P. Moreton and Carlo H. Séquin. 1992. Functional optimization for fair surface design. In: *Proceedings of the 19th annual conference on Computer graphics and interactive techniques (SIGGRAPH '92)*. Association for Computing Machinery, New York, NY, USA, 167–176. DOI: 10.1145/133994.134035
8. Chang, K.-H., & Wang, Y. (2013). Evaluation of Sheet Forming Simulation for an Integrated Reverse Engineering System. *Computer-Aided Design and Applications*, 10(5), 767–777. DOI: 10.3722/cadaps.2013.767-777



- 
9. Schoene, C, Stelzer, R, Schmidt, U, & Suesse, D. Reverse Engineering for Spotting of Sheet Metal Forming Parts. *Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Volume 1: 36th Design Automation Conference, Parts A and B. Montreal, Quebec, Canada. August 15–18, 2010. pp. 407–414. ASME. DOI: 10.1115/DETC2010-28208
  10. Wang, J., Goel, A., Yang, F., et al. Blank optimization for sheet metal forming using multi-step finite element simulations. *Int J Adv Manuf Technol* 40, 709–720 (2009). DOI: 10.1007/s00170-008-1383-y
  11. Петров М.А., Уржумов П.Н. Компьютерное моделирование процесса горячей объёмной штамповки поковки «крышка» и топологическая оптимизация штампа // *Машиностроение: сетевой электронный научный журнал*. 2022. Т. 9, № 2. С. 3–9. DOI: 10.24892/RIJE/20220201

## Low-dimensional embedding for exploring the phylogenetic characteristics of Tick-borne Encephalitis Virus

Majid Forghani<sup>1,2</sup>, Kazem Forghani<sup>3</sup> and Mikhail Bolkov<sup>4,\*</sup>

<sup>1</sup> N.N. Krasovskii Institute of Mathematics and Mechanics of the UB RAS, St, Yekaterinburg, Russia

<sup>2</sup> Institute of Natural Sciences and Mathematics, Ural Federal University, Yekaterinburg, Russia

<sup>3</sup> Iran University of Science and Technology (IUST), Tehran, Iran

<sup>4</sup> Institute for the Study of Aging, Russian Gerontological Scientific and Clinical Center, N.I. Pirogov Russian National Research Medical University of the Ministry of Health of the Russian Federation, Moscow, Russia

**Abstract.** The development of a model for predicting virus evolution requires an accurate determination of virus characteristics. TBEV Analyzer is an analytical platform for characterizing the tick-borne encephalitis virus, which provides the evolutionary history of the virus in a hierarchical graph form called the clusteron structure. In this work, we explore three-fold characteristics of phylogenetic analysis, including subtype-lineage-clusteron, represented by clusteron structure through embedding the viruses into a 2D numerical space. The embedding is carried out by computing genetic distances and projecting viruses from genetic sequence space into the two-dimensional numerical vector space using dimensionality reduction techniques. Further, a phylogenetic tree is mapped into 2D space to describe the relationship between viruses. The 2D coordinates of the tree's inner nodes are optimized by gradient descent. To characterize each clusteron, we estimate and visualize its distribution in the final low-dimensional space. To the best of our knowledge, this is the first time that such a visualization has been made for tick-borne encephalitis virus that includes viruses, clusteron distribution, and a phylogenetic tree in one scene. Our preliminary results indicate that the approach has the potential to serve as an exploratory and complementary tool in studying and modeling the virus evolution in TBEV surveillance.

**Keywords:** visualization, embedding, TBEV Analyzer, phylogenetic tree, t-SNE.

### Introduction

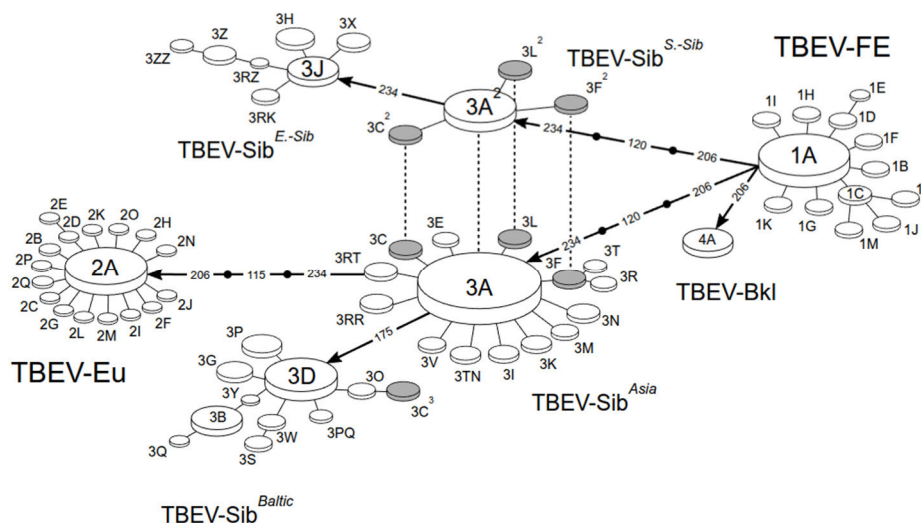
Viruses are an integral part of human life, some of which can lead to serious and severe diseases. During the evolution, they change their characteristics, part of which is due to evading the immune response of the host. The tick-borne encephalitis virus (TBEV) is a pathogen that can cause serious damage to the central nervous system, leading to disability or death in humans. The geographical distribution of TBEV forms a belt that includes the Russian Federation. Generally, TBEV includes four subtypes, namely, TBEV-FE (Far Eastern), TBEV-Bkl (Baikalian), TBEV-Sib (Siberian), and TBEV-Eu (European) [1]. The general distribution of TBEV is patchy. The distribution is locally restricted to a geographical area known as the focus, an environment in which a specific kind of TBEV circulates. This leads to the formation of micro- and macro-foci, which strongly depend on various factors, including the TBEV characteristics [2].

In 2019, an analytical platform titled TBEV Analyzer was introduced [3] that provides the phylogenetic characteristics of TBEV at three hierarchical levels, namely, subtype, lineage, and clusteron. The platform can directly fetch a query from GenBank [4] and analyze it, supply an interactive map for visualization of TBEV distribution, and offer an application programming interface (API) [5]. The recent version, TBEV Analyzer 3.0, has been announced, followed by theoretical and practical enhancements, including updating the phylogenetic model called clusteron structure (CS) (see Fig. 1), which is the core of the platform, by adding eleven novel clusterons [6]. The results of previous studies indicate that the platform can accurately determine the phylogenetic characteristics of a query virus.

The TBEV Analyzer relies on the clusteron approach (CA), which provides the overall picture of TBEV evolution in a hierarchical representation in graph form. The term “clusteron” stems from “cluster” and “clone.” Indeed, it is the smallest unit of viral population in the CA framework. A clusteron refers to the group of viruses featured by the three following conditions [7]:

- First, the viruses have an identical amino acid signature of the glycoprotein E fragment.
- Second, they are phylogenetically related.
- Third, they are characterized by a specific type of spatial distribution.





**Figure 1.** The current clusteron structure of TBEV. The homoplastic clusterons are connected with dashed lines. The connected clusterons have the same amino acid signature, but they belong to different subtypes/lineages. The CS has four dominant subtypes, among which the Siberian features four lineages

The computational pipeline of the platform mainly consists of preprocessing and verification of the query; construction of the phylogenetic tree inference for determining the subtype/lineage of the query; and finally, the search for matches based on clusteron-specific amino acid signatures. The pipeline requires specific configurations and carefully selected algorithms to obtain reliable results. This is the basic motivation behind the development of a unified analytical platform for TBEV.

The traditional way to represent the evolutionary history of a set of taxa is by constructing the phylogenetic tree. The tree accounts for the similarities and differences between genetic sequences to describe the sophisticated evolutionary process in a graphical, human-readable fashion. There are various applications of phylogenetic trees in viroinformatics. As an example, the information of phylogenetic trees can directly be incorporated in modeling the influenza antigenicity [8]. Previously, we suggested several approaches to visualize the phylogenetic tree of the influenza virus in a 3D numerical vector space. In [9], we suggested clustering taxa according to the physicochemical properties of amino acids of protein sequences. By incorporating the additional dimension into the 2D representation of the tree, the information of taxa clusters enriches the visualization, where each cluster can be independently studied according to its phylogeny.

Inspired by Rubik's cube solving algorithms [10], we proposed a visualization of phylogenetic trees by plotting the leaf-to-root path [11]. Since only the sequences of tree leaves are known, the algorithm of reconstructing the ancestral sequences was employed to obtain the sequence of every inner node. Further, the genetic sequences were encoded by one-hot encoding and embedded by multidimensional scaling (MDS) [12] or t-Distributed Stochastic Neighbor Embedding (t-SNE) [13] into 2D or 3D space for visualization. This approach was enhanced by applying reduced amino acid alphabets (RAAA) to embed other aspects of evolution along the 2D tree representation and embedding by Word2Vec [14] instead of evolutionary distance embedding [15, 16].

Considering the aforementioned works, a question may arise: "How can the relationships between viruses in the clusteron structure be represented in a low-dimensional vector space?" Perhaps the answer to this question may help us better understand and explore the evolutionary process. In addition, this may facilitate understanding those boundaries that form a clusteron in terms of the physicochemical properties of genetic sequence, which may reveal the hidden mechanism of the evolutionary process.

The aim of this paper is to generate a visualization of TBEV viruses by employing a two-step embedding from the space of genetic sequence into the 2D numerical space. Our contribution to this paper is twofold:

- Performing the clusteron approach analysis of all TBEV viruses registered in GenBank, determining their characteristics using the TBEV Analyzer platform, and creating a unified dataset of TBEV.
- Proposing a new visualization approach for TBEV by embedding the viruses from genetic space into 2D numerical vector space. This visualization includes three components: viruses, the phylogenetic

tree of clusteron representatives, and finally supplying the kernel density estimation (KDE) plot for each clusteron in the 2D visualization scene to explore the underlying structure of clusterons.

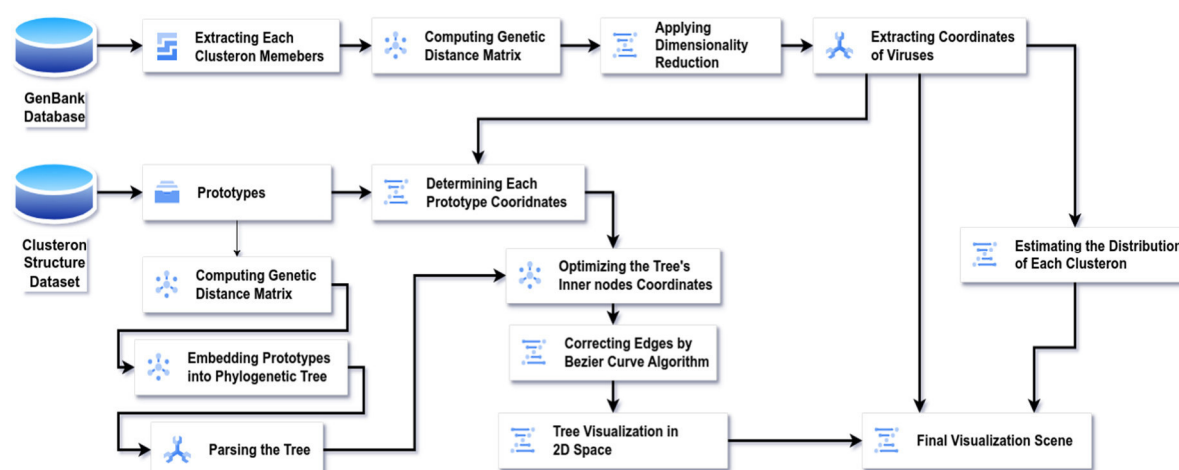
To the best of our knowledge, this is the first attempt to visualize relationships between TBE viruses by combining three components: viruses, a phylogenetic tree, and clusteron distributions.

The remaining parts of this paper are organized as follows: Section 2 explains the materials and methodology in more detail. Section 3 presents the computational experiment setup and the results. Finally, we give conclusions in Section 4.

## Methodology

The overall schema of the proposed approach is illustrated in Fig. 2. The approach consists of the following steps:

- Analyzing GenBank records for TBEV and determining three-fold characteristics of viruses.
- Computing the genetic distance matrix of viruses collected in the previous step using both coding (nucleotide) and protein sequences.
- Embedding the viruses by applying dimensionality reduction techniques to the genetic distance matrix and obtaining the coordinate of each virus in the 2D space.
- Constructing the phylogenetic tree for representatives of clusterons.
- Embedding the tree by optimizing the coordinates of the tree's inner nodes in 2D space through gradient descent.
- Computing the distribution of clusteron members.
- Visualizing the obtained results, including all viruses, the distribution of clusterons, and the phylogenetic tree in 2D space.



**Figure 2.** The overall schema of our pipeline. The procedure provides two-step embedding. The 2D coordinates of viruses are obtained by computing the pairwise distance matrix and applying a dimensionality reduction technique.

Further, a phylogenetic tree is constructed on prototype sequences and embedded into the 2D space. Finally, we estimate the distribution of each clusteron and plot it in the visualization scene

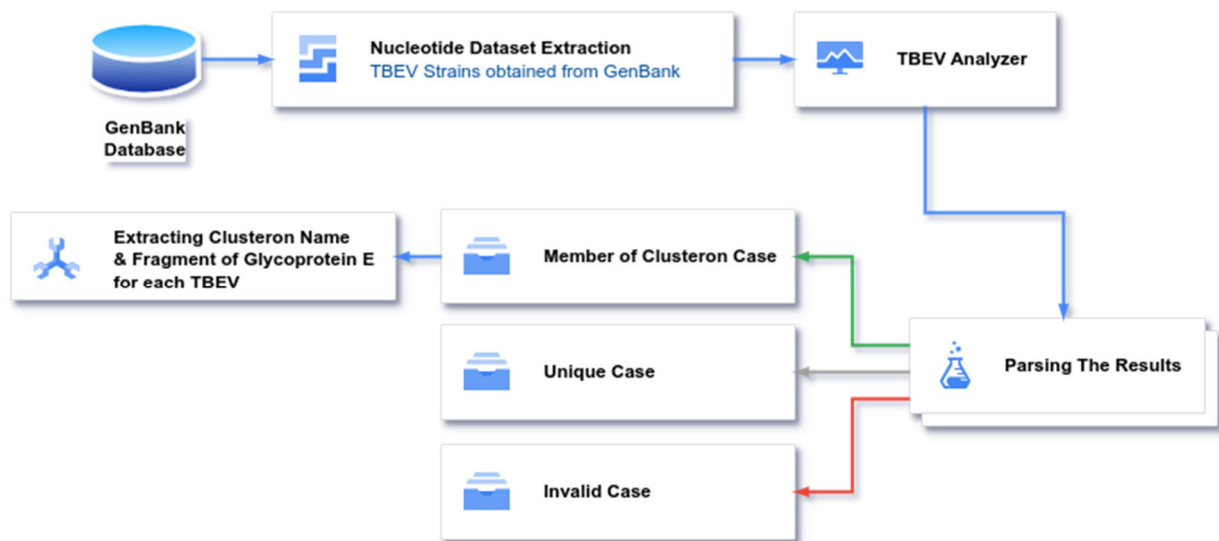
## 1. Data Collection

Typically, the TBEV Analyzer requires both coding (nucleotide) and amino acid sequences to infer the phylogenetic characteristics of a query virus. On the one hand, the coding sequence is essential to determine the virus characteristics at the global scale, i.e., subtype and lineage. On the other hand, the amino acid sequence is required to search for a match regarding the clusteron-specific amino acid signature within the lineage or subtype, in the result of which the platform detects the type of clusteron for the query. Algorithmically, the TBEV Analyzer works with two types of data: coding and amino acid sequences. In practice, the platform requires a fragment of the genetic sequence of glycoprotein E as input data. The fragment includes 454 nucleotides (positions 309–762) in gene E. Since the amino acid sequence can easily be achieved by translating the nucleotide sequence codons, we restrict our work to the coding sequences.

In the TBEV Analyzer framework, a clusteron has three hierarchical characteristics: subtype, lineage, and clusteron. A clusteron in the framework has three elements: a name (or label), a prototype, and a specific amino acid signature. The prototype coding sequence is a key to identifying the phylogenetic lineage or subtype of a query, whereas the specific amino acid signature is the primary component to determine the type of clusteron. A prototype is the representative of a clusteron. It plays a role similar to the centroid of a cluster, where this centroid is assigned from the biological aspect.

To compile the database for our experiments, we extract viruses using the query “TBEV” or “tick-borne encephalitis virus” in FASTA format from the nucleotide core (nucore) of the GenBank. The records are then fed into the TBEV Analyzer, in the results of which three data sets are obtained (see Fig. 3). A virus record can have one of the following cases:

- The record does not have the target fragment sequence of gene E completely. Such a query is assigned to the “invalid set.”
- The record has the complete sequence of the target fragment, and it belongs to a known clusteron within the CA framework. Such a record is assigned to the “clusteron members set,” denoted by  $S_c$ .
- The final case happens when the record has the complete sequence of the target fragment, but it does not belong to any known clusteron. The collection of such records is called the “unique set.”



**Figure 3.** Database preparation for the proposed pipeline. We search GenBank by query “TBEV” or “tick-borne encephalitis virus.” All obtained records are sent to the TBEV Analyzer. After parsing the results of analysis, the members of each clusteron are determined. Each entry of the final collected dataset includes three fields: the GenBank ID, the specific fragment of the coding sequence of glycoprotein E, and the label (i.e., the clusteron name)

After parsing TBEV Analyzer's results, we focus on the clusteron members set, where each member has its own label and the sequence of the target fragment. Thanks to the TBEV Analyzer, all members have sequences of the same length, acquired from the results of the platform analysis. Therefore, there is no need for preprocessing steps such as alignment.

In addition to the set of clusteron members, we require information about the CS prototypes. As stated earlier, each prototype also contains the coding sequence of the target fragment. Hence, each member of the prototype set has a label (or the name of the clusteron) and a corresponding coding sequence. This set is referred to as the “prototype set” and denoted by  $S_p$ . Since each prototype represents a virus sequence taken from GenBank, it follows that  $S_p \subset S_c$ .

## 2. Computing genetic distance

There are several genetic distance metrics, each of which has its own advantages and disadvantages. For the sake of simplicity, we employ p-distance, or uncorrected distance. The distance is computed using the following formula for both coding and protein sequences:

$$d(S1, S2) = \frac{\text{diff}(S1, S2)}{L},$$

where  $S1$  and  $S2$  are two sequences with the same length  $L$ ,  $d(S1, S2)$  is their distance, and  $diff(S1, S2)$  is the number of sites with different nucleotides/amino acids. We compute the pairwise distance matrix for the set of all viruses in  $S_c$  in two formats: coding and protein sequences. These two matrices are further combined through a weighted average to leverage the information of both coding and protein sequences. The obtained pairwise distance matrix for the set of clusteron members, i.e.,  $S_c$ , is denoted by  $D_c$ .

### 3. Embedding GenBank viruses

A key aspect of our approach is the representation of three types of information (viruses, a tree, and clusterons) in the low-dimensional space to facilitate the exploration of viral relationships in the 2D space. To achieve this, the t-SNE technique is applied to the distance matrix  $D_c$  to obtain the coordinates of each virus within  $S_c$  in the low-dimensional space. t-SNE is a non-linear dimensionality reduction technique that employs the distances between objects in high-dimensional space, representing them in conditional probabilities, and minimizes Kullback-Leibler divergence between distributions of similarity in high- and low-dimensional spaces. The embedding can mainly be controlled by perplexity, which is related to the number of nearest neighbors used in the algorithm, and early exaggeration, which affects the space between clusters.

We apply t-SNE and obtain the coordinate of each virus in the low-dimensional space. In such a manner, the coordinates of each prototype in the embedding space can be extracted, and the set of these coordinates is denoted by  $X_{pro}$ .

### 4. Reconstructing the phylogenetic tree

Algorithms for reconstructing a phylogenetic tree from a set of aligned sequences can be categorized into three main groups: distance-based methods (e.g., neighbor-joining), character-based methods (such as maximum parsimony and maximum likelihood), and Bayesian inference methods. Among these, the neighbor-joining (NJ) algorithm is widely used in molecular evolution studies that can provide a rough phylogenetic tree from a pairwise distance matrix of sequences. While we adopt this algorithm for its simplicity, our approach works with any tree reconstruction method that provides branch lengths.

A binary phylogenetic tree is constructed by applying the NJ algorithm to the pairwise distance matrix of prototypes (denoted by  $D_p$ ). The prototypes are located in the leaves of the tree, while the inner nodes express the inferred similarity between them. As another component of the final visualization, we embed this phylogenetic tree into the low-dimensional space gained in Subsection 2.3.

### 5. Embedding phylogenetic tree in low-dimensional space

Here, we embed the tree structure in the low-dimensional space. To achieve this, we first need to establish a correspondence between the objects in the tree structure and those in the low-dimensional space. This bridge can be depicted by prototypes. Subsection 2.3 gives the coordinates of prototypes, i.e.,  $X_{pro}$ , in the low-dimensional space, while prototypes are located in the tree leaves. Accordingly, in order to embed the tree, we need to optimize the position of inner nodes in the low-dimensional space in such a way that the edge lengths of the tree are preserved as much as possible.

Zhang et al. [17] proposed a visualization algorithm for genetic sequences called PhyloMap. In their method, they optimized the inner node positions using the gradient descent and the following objective function, which is very similar to “Sammon's mapping”:

$$E = \frac{1}{\sum_{i < j} s \cdot d_{i,j}^*} \sum_{i < j} \frac{(s \cdot d_{i,j}^* - d_{i,j})^2}{s \cdot d_{i,j}^*}, \quad (1)$$

where  $d_{i,j}^*$  and  $d_{i,j}$  are the distances between objects  $i$  and  $j$  in the tree and the low-dimensional space, respectively. The parameter  $s$  in Eq. 1 acts as a scaling factor and compensates for the differences between spaces when the distances in the phylogenetic tree and low-dimensional space come from different natures. Their algorithm uses the gradient descent on the inner nodes in order to minimize error expressed by Eq. 1. To optimize the inner node coordinates, we use the central difference formula for approximating the partial derivative of the objective function numerically. Given that our phylogenetic tree obtained from the previous step is binary and rooted, each inner node (excluding the root) is connected to three nodes (two children and one ancestor). Consequently, we consider only these three distances in the objective function for each inner node. In the case of the root node, we use only two edges of its children.

A phylogenetic tree may not be perfectly embedded in the low-dimensional space. For instance, the straight-line distance between two nodes in the low-dimensional space can be shorter than their corresponding edge length in the tree. Zhang et al. tackled this issue by introducing the application of the Bezier curve [18], where the straight edge is replaced by a smooth curve. In this case, the  $d_{i,j}$  in Eq. 1 is replaced by  $d_{i,j}^b$ , which is the length of the Bezier curve between nodes  $i$  and  $j$ . Here, we employ the quadratic Bezier curve, which requires at least three points to draw a curve. The curve points are computed as follows:

$$B(t) = (1-t)^2 P_0 + 2(1-t)t P_1 + t^2 P_2,$$

where  $P_0$ ,  $P_1$ , and  $P_2$  are the control points, and  $t \in [0,1]$ . The interval  $[0,1]$  is discretized to  $m$  values, where  $m$  determines the resolution of the curve.

## 6. Estimating the distribution of clusteron members

As an additional layer of information in our visualization, we compute and plot the distribution of clusteron members for each clusteron individually. One critical component of the clusteron definition is its specific geographical distribution. Since we do not have access to the exact geographical location of the sample for some viruses in  $S_c$ , we use the 2D coordinates from Subsection 2.3 to obtain an estimation of distribution for each clusteron in the embedding space. The distribution is estimated using the kernel density estimation (KDE) method [19]. Given a set of  $n$  points  $X = \{x_1, x_2, \dots, x_n\}$ , the KDE value for a point  $x$  can be computed by:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

where  $h > 0$  is the bandwidth parameter that controls the smoothness of the estimated distribution, and  $K(\cdot)$  is the kernel function. In our case, a Gaussian kernel is employed that produces a smooth and continuous density estimate without artifacts participating. The coordinates of the clusteron members in the low-dimensional space are employed by the KDE method to compute the distribution, which is then visualized as a part of the final 2D scene.

## 7. Visualization scene

In this study, the dimension of the visualization scene is set to two. The final scene comprises three main components:

- A 2D visualization of all viruses within the set  $S_c$ , where each member is associated with a specific clusteron.
- A phylogenetic tree, in which each leaf corresponds to a clusteron prototype. Since  $S_p \subset S_c$ , the coordinates of the tree leaves are fixed, while the positions of inner nodes are achieved by solving optimization tasks through gradient descent.
- The KDE-based spatial distribution plot for each clusteron that represents the density of its members in the embedding space.

The algorithmic overview of the approach is presented in Algorithm 1.

---

### Algorithm 1. Visualization of TBEV

---

**Input:** The set of sequences for TBEVs from GenBank:  $S$ , the set of prototype sequences:  $S_p$ , error:  $e$ , the number of iterations:  $iter\_num$ .

**Output:** A 2D map of viruses and their characteristics.

- 1 Filtering the set  $S$  by the TBEV Analyzer and obtaining the set  $S_c$ .
- 2 Computing the pairwise distance matrix  $D_c$  for objects in  $S_c$ .
- 3 Embedding the objects of  $S_c$  into 2D space by applying t-SNE to  $D_c$ .
- 4 Determining the coordinates of prototypes, i.e.,  $X_{pro}$ , from step 3.
- 5 Computing pairwise distance matrix  $D_p$  for  $S_p$ .
- 6 Constructing the phylogenetic tree  $T$  from  $D_p$ .
- 7 Fixing the coordinates of leaves in  $T$  using  $X_{pro}$ .
- 8 Initializing the 2D coordinates of inner nodes in  $T$ .

---

```

9  while  $E_i < \delta$  or  $i < iter\_num$  do
10   for inner node  $x$  do
11     if  $i \% 5 == 0$  then
12       Update the coordinate of  $x$  by gradient descent.
13     else
14       Update the coordinate of  $x$  by gradient descent when at least one of the edges connected to the  $x$ 
        has the condition  $d_{i,j} > s \cdot d_{i,j}^*$ .
15    $E_i = error$  by Eq. 1

16 for each edge with  $d_{i,j} < s \cdot d_{i,j}^*$  do
17   Correct the edge by computing the Bezier curve between nodes  $i$  and  $j$  such that  $d_{i,j} \approx s \cdot d_{i,j}^*$ .

18 for each clusteron  $C_k$  do
19   Determine its members from  $S_c$ .
20   Obtain the coordinates of elements in  $S_c$  from the results of step 3. The set of coordinates is denoted by
      $X_{c_k}$ .
21   Compute the distribution for  $X_{c_k}$ .
22   Plot  $X_{c_k}$  and its distribution in the 2D scene.

23 Plotting the edges of the tree in the 2D scene.
:
:

```

---

### Experiments & Results

We downloaded all TBE viruses from the nucleotide core of the GenBank database. After analyzing them with the TBEV Analyzer (available at [tbev.viroinformatics.com](http://tbev.viroinformatics.com)), a total of 895 viruses passed the analysis, which formed  $S_c$ . Other viruses were excluded from the experiment due to lack of a complete sequence (of the target fragment) or of a known clusteron. Thus, each member of  $S_c$  has the coding sequence of the target fragment of the gene E and its label, i.e., the clusteron to which it belongs. As mentioned earlier, we applied the p-distance to compute the distance matrices of coding and protein sequences. A weighted average was applied to obtain the pairwise distance  $D_c$ , where the weight for coding and protein distance matrices was 0.4 and 0.6, respectively. The distance matrix was then fed into the t-SNE algorithm, available via the scikit-learn package [20], to map viruses into the 2D numerical space. After embedding, each member of  $S_c$  is characterized by three properties: a label, a coding sequence, and a 2D coordinate.

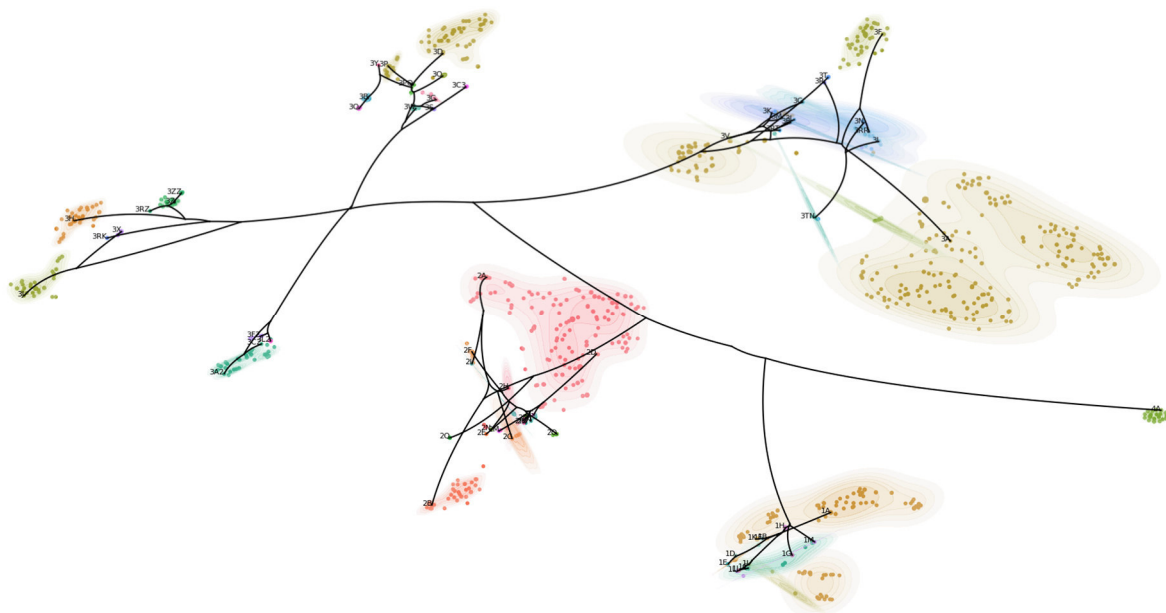
We used the Kimura 2-parameter distance [21] to compute  $D_p$  and construct the phylogenetic tree by applying the NJ algorithm. The tree was built from the sequences of prototypes  $S_p$ , using MEGA software version 11.0.13 [22]. MEGA facilitates the tree parsing by allowing the tree to be saved in a tabular format, providing access to each edge along with its length and endpoints. Since distances  $d_{i,j}^*$  and  $d_{i,j}$  in Eq. 1 have different natures, we employed distance scaling parameter  $s$  to embed the tree in low-dimensional space. Each prototype (located at a leaf of the tree) is also a member of  $S_c$ , and its 2D coordinate was available. To preserve edge length in the 2D space, we optimized the coordinates of the inner nodes by numerically approximating the partial derivative and applying gradient descent to minimize the objective function defined in Eq. 1. A quadratic Bezier curve was further used to correct the edges on the 2D map.

Totally, there are 68 clusterons within the CA framework. The study of each clusteron as a significant unit of CA is essential. We customized the visualization of each clusteron by a unique color. We plotted all objects of  $S_c$  in the 2D visualization scene, while each object is colored regarding its label. The  $S_c$  may contain duplicate entries with identical coding sequences of the target fragment, leading to the same 2D coordinates. To highlight the number of viruses with the identical sequence, the radius of the plotted marker in the 2D scatter plot can be adjusted by the number of duplicates. We labeled each prototype point in the scene with the name of its clusteron. The set of edges, refined using Bezier curve correction, was also plotted into the scene. The last component is the distribution of the clusteron members. The distribution was computed by the function



*kdeplot* from the Python package Seaborn [23]. The visualization was enhanced by plotting the contours of the 2D density for each clusteron, coloring it according to the clusteron label, and decreasing the alpha value to see overlaps between clusterons.

The final results are presented in Fig. 4. Here, we clearly observe four distinct subtypes: TBEV-FE (Far-Eastern) in the bottom, TBEV-Sib (Siberian) with its lineages, TBEV-Bkl (Baikalian) at the right side of the figure, and TBEV-Eu (European) in the center of the figure. An interesting observation arises with the clusteron founders, where their distributions are clearly larger than derivative clusterons. Note that a TBEV can be characterized at three phylogenetic hierarchical levels: subtype, lineage, and clusteron. The lineage is currently defined only for the Siberian subtype. Both the subtype and lineage of a query virus can be simply inferred from the coding sequence of the target fragment of gene E. In contrast, the clusteron is determined using a clusteron-specific amino acid signature. If the query's signature matches that of a known clusteron within the query's subtype and lineage, the virus belongs to that clusteron. As observed in Fig. 4, some clusterons have a wide distribution while others are more compact. Such dispersion in distribution is the result of the embedding using nucleotide-level information. The t-SNE algorithm controls the visualization through two parameters: perplexity and early exaggeration. Since the perplexity is related to the number of nearest neighbors, it is expected that leveraging more viruses for derivative clusterons can improve their distribution and visualization quality.



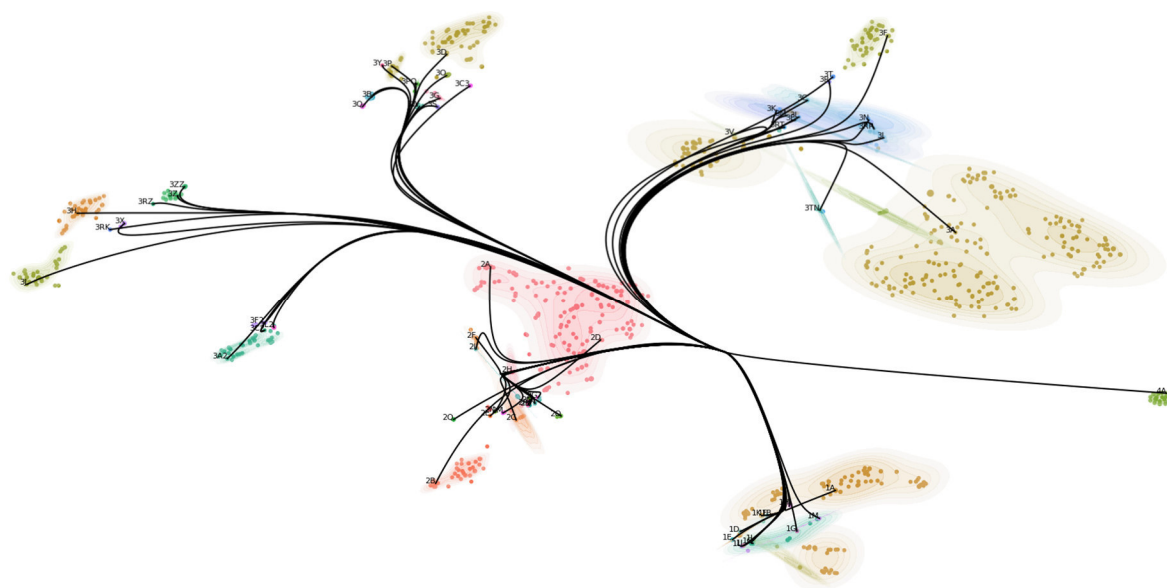
**Figure 4.** The visualization of TBE viruses, accepted by the framework of the clusteron approach in the 2D numerical space. This scene demonstrates three key elements: the viruses, their clusteron distribution, and the phylogenetic tree of prototypes. Each point in the plot is colored according to its clusteron label, and the same color scheme is used for the corresponding clusteron distribution. The location of each prototype is labeled by its clusteron name

In our previous work [16], we suggested the visualization of the leaf-to-node path as a unified curve. This can be advantageous when working with a large number of viruses, in studying pivotal events of evolution, checking the variability among viruses, observing the main trend, and monitoring the direction of evolution. Taking this idea, we generated a smoothed version for visualizing the evolution of TBEV (see Fig. 5).

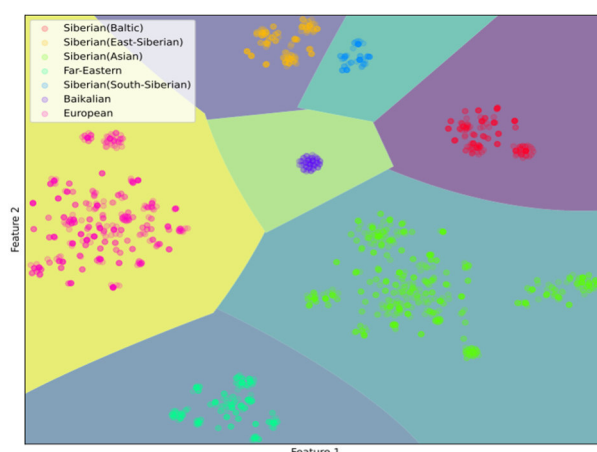
A decisive factor of our approach is the distance metric, which participated in the computation of 2D coordinates. The definition of distance can be customized by application of reduced amino acid alphabets [24] or by introducing the amino acid physicochemical properties [25]. As a supplementary experiment, we used both the coding and amino acid sequences of each virus and embedded them by Word2Vec into a numerical vector space. Further, the objects were mapped from this high-dimensional space into 2D space by t-SNE. Fig. 6 shows the visualization of viruses leveraging the information of coding sequences. Fig. 7 illustrates the visualization of the same viruses by adding information on amino acid sequences along with the coding sequences. The subtype and lineage can be clearly recognized in Fig. 6. In Fig. 7, we customized the marker



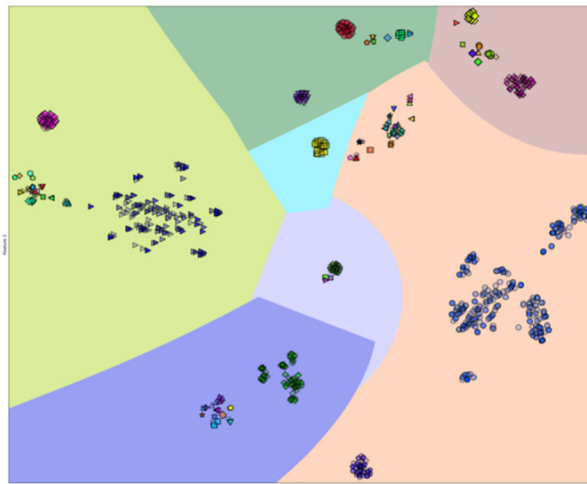
and color of viruses to provide a unique visual shape for members of each clusteron. It can be seen that clusterons with a higher number of members achieve better representation. On the contrary, the clusterons with a lower number of members are close together and more compact in the region, which makes their recognition difficult. This may also be due to the t-SNE settings. Unfortunately, replacing t-SNE by MDS does not generate a desirable result. We believe that controlling the influence of coding and amino acid sequence information involved in embedding calculations using a weighted average technique has a direct impact on the quality of visualization. Note that although the embedding of high-dimensional vectors into a 2D space using t-SNE is not isometric, it effectively captures the structure of virus clusterons at a coarse level and provides visually meaningful separation for most clusterons at a fine level. It should be better to mention that we also applied other dimensionality reduction techniques such as MDS, Isomap, and Locally Linear Embedding, but the best results were obtained by t-SNE. However, the choice of distance metric and dimensionality reduction technique strongly affects the final visualization, which requires further investigation.



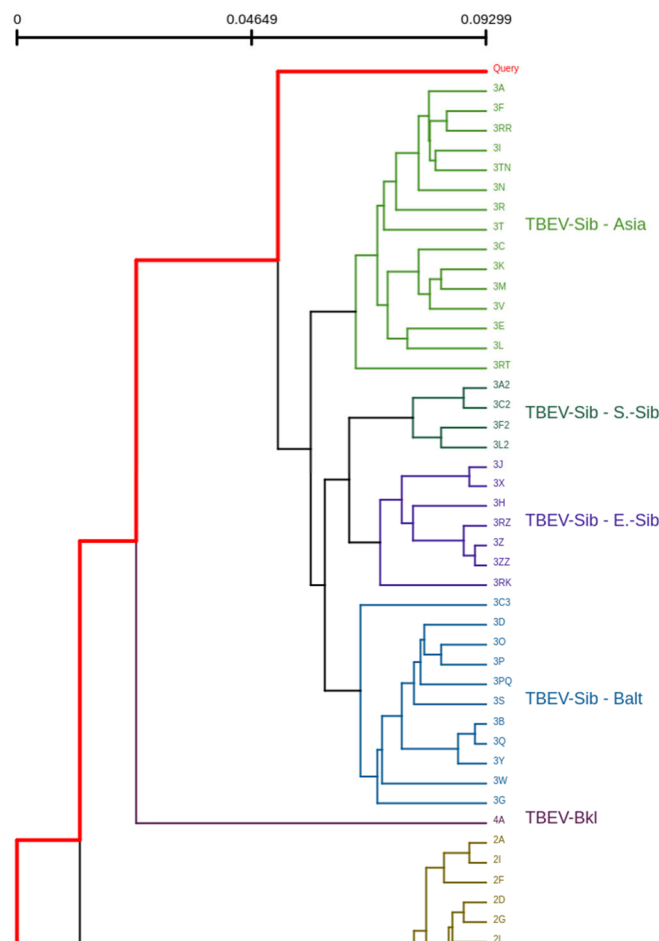
**Figure 5.** The visualization of the leaf-to-root path for TBEVs. Each path begins at a leaf node, passes through inner nodes, and terminates at the root. The leaf, root, and their inner nodes along the path are employed as control points to generate a smooth curve by the Bezier curve



**Figure 6.** Visualization of the TBEV subtype and lineage. The visualization was obtained by the coding sequence embedding of viruses into a high-dimensional numerical vector space, followed by applying t-SNE to generate 2D coordinates. Virus colors are customized based on its subtype/lineage. Note that the boundary approximation was obtained using a support vector machine classifier



**Figure 7.** Visualization of TBEV's threefold phylogenetic characteristics. The visualization was obtained by embedding both coding and amino acid sequences of viruses, computing the distance matrix, and then applying t-SNE. Virus colors and markers are customized based on their clusteron. Regions are highlighted according to subtype/lineage. Note that the boundary approximation was obtained using a support vector machine classifier



**Figure 8.** An example of a TBEV virus recognized as unique by the TBEV Analyzer. In this partial visualization, the virus appears to have a close relationship with the Siberian subtype within the phylogenetic tree. Recognizing such unique sequences in our visualization scene requires defining boundaries for the formation of clusterons

Considering the proposed visualization, a challenge may arise when the model encounters viruses that are recognized as unique by the TBEV Analyzer. The design and development of algorithms in the TBEV Analyzer ensure that a unique virus maintains a distant relationship with known clusterons within the phylogenetic framework. This guarantees that a virus falling outside the scope of known lineages/subtypes is

definitively tagged as unique (see Fig. 8). While this can be clearly expressed in the context of genetic distance by the CA framework, there is no simple boundary yet in the proposed visualization that defines when a virus is unrelated or unique. This challenge is critical for designing a robust model to characterize TBEV and to define thresholds or boundaries for the identification and formation of a clusteron, which also provides room for further investigation.

## Conclusions

In this pilot study, we attempt to explore and characterize the TBEV clusterons by embedding them in a 2D numerical space. To the best of our knowledge, this is the first time that such a customized visualization has been generated for TBEV. The main advantage of the proposed approach is embedding three types of information into the 2D visualization scene: the spatial representation of viruses on the map, the distribution of clusterons, and the relationship between clusteron representatives in a phylogenetic tree. A desirable fact in Fig. 4 is the distinct representation of subtypes/lineages in the visualization scene with dominant distribution of clusteron founders, i.e., clusterons 1A, 2A, 3A, 3A<sup>2</sup>, 3J, 3D, and 4A. It should be noted that this approach does not claim to be a replacement for classical representation of phylogenetic trees.

From a technical perspective, a drawback of our approach is handling a large number of viruses or embedding a large phylogenetic tree in the 2D space, causing a complex visualization scene. This issue becomes particularly serious when the distance for obtaining the 2D coordinates and that for constructing the phylogenetic tree are different. Switching from 2D to 3D space may help refine and simplify the visualization. In comparison with our previous work, the coordinate of an inner node in the phylogenetic tree is achieved through an optimization task instead of applying the ancestral reconstruction algorithms and encoding the inner node sequence directly. Although we use the t-SNE technique to map from high-dimensional to 2D spaces, it can easily be replaced by any dimensionality reduction technique. Also, the phylogenetic tree algorithm can be replaced by any algorithm that generates a tree with branch lengths.

Our results indicate that such embedding can be advantageous for characterizing TBEV at the coarse level by providing its subtype/lineage only using the information obtained from the coding sequence. The three-fold phylogenetic characteristics (i.e., subtype-lineage-clusteron) of a virus cannot be determined without incorporating the amino acid signature. Adding the information of amino acid sequences improves the quality of visualization. This can be observed for clusteron founders 3A and 3A<sup>2</sup>. While these founders have the same clusteron-specific amino acid signature, they belong to different phylogenetic lineages. That is why their representations are far from each other in Fig. 4.

Further investigation is required to develop a reliable algorithm to accurately identify the phylogenetic characteristics of TBEV, especially in the case of unique viruses (see a unique case in Fig. 8). This helps us determine the physicochemical and phylogenetic boundaries that define a clusteron in the CS. Moreover, the search for an appropriate distance metric is of interest. During the data collection process, we have identified several unique viruses with the same signature that can be analyzed for forming new potential clusterons in the CA framework, which will be reported to the TBEV Analyzer team. Taken all together, our results reveal that the approach can serve as an exploratory and complementary tool in studying and modeling evolution in TBEV surveillance. We expect that extending the genetic database of TBEV will lead to an expansion of the clusteron member set and improve our understanding of TBEV evolution. We plan to take advantage of the ZADU package [26] to assess embeddings using various metrics. Further research can focus on investigating the correlation between geographical, phylogenetic, and embedding distances in order to extract key insights about hidden mechanisms underlying viral evolution.

## Acknowledgments

The work was performed as part of research conducted in the Ural Mathematical Center with the financial support of the Ministry of Science and Higher Education of the Russian Federation (Agreement number 075-02-2025-1549).

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

- [1] S. Kovalev, T. Mukhacheva, Reconsidering the classification of tick-borne encephalitis virus within the Siberian subtype gives new insights into its evolutionary history, *Infection, Genetics and Evolution* 55 (2017) 159–165. doi:10.1016/j.meegid.2017.09.014.
- [2] A. Wallenhammar, R. Lindqvist, N. Asghar, S. Gunaltay, H. Fredlund, A. Davidsson, S. Andersson, A. K. Overby, M. Johansson, Revealing new tick-borne encephalitis virus foci by screening antibodies in sheep milk, *Parasites & Vectors* 13 (2020) 1–12. doi:10.1186/s13071-020-04030-4.
- [3] M. Forghani, S. Kovalev, P. Vasev, M. Bolkov, TBEV Analyzer: A platform for evolutionary analysis of tick-borne encephalitis virus, in: 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), IEEE, 2019, pp. 0397–0402. doi:10.1109/SIBIRCON48586.2019.8958021.
- [4] D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt, E. W. Sayers, Genbank, *Nucleic acids research* 46 (2018) D41–D47. doi:10.1093/nar/gkx1094.
- [5] M. Forghani, S. Kovalev, M. Bolkov, M. Khachay, P. Vasev, TBEV Analyzer platform for evolutionary analysis and monitoring tick-borne encephalitis virus: 2020 update, *Biostatistics & Epidemiology* 6 (2022) 57–73. doi:10.1080/24709360.2021.1985392
- [6] M. Forghani, S. Kovalev, M. Khachay, E. Ramsay, M. Bolkov, P. Vasev, Identifying new clusterons: Application of TBEV Analyzer 3.0, *Microorganisms* 11 (2023) 324. doi:10.3390/microorganisms11020324.
- [7] S. Y. Kovalev, T. A. Mukhacheva, Clusteron structure of tick-borne encephalitis virus populations, *Infection, genetics and Evolution* 14 (2013) 22–28. doi:10.1016/j.meegid.2012.10.011.
- [8] W. T. Harvey, D. J. Benton, V. Gregory, J. P. Hall, R. S. Daniels, T. Bedford, D. T. Haydon, A. J. Hay, J. W. McCauley, R. Reeve, Identification of low-and high-impact hemagglutinin amino acid substitutions that drive antigenic drift of influenza A (H1N1) viruses, *PLoS pathogens* 12 (2016) e1005526. doi:10.1371/journal.ppat.1005526.
- [9] M. Forghani, P. Vasev, V. Averbukh, Three-dimensional visualization for phylogenetic tree, *Scientific Visualization* 9 (2017) 59–66.
- [10] C. A. Steinparz, A. P. Hinterreiter, H. Stitz, M. Streit, Visualization of Rubik's cube solution algorithms., in: *EuroVA@ EuroVis*, 2019, pp. 19–23.
- [11] M. Forghani, P. Vasev, E. Ramsay, A. Bersenev, Visualization of the evolutionary path: an influenza case study, in: *CEUR Workshop Proc.–CEUR-WS*, volume 3027, 2021, pp. 358–368.
- [12] M. A. A. Cox, T. F. Cox, *Multidimensional Scaling*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 315–347. doi:10.1007/978-3-540-33037-0\_14.
- [13] L. v. d. Maaten, G. Hinton, Visualizing data using t-SNE, *Journal of machine learning research* 9 (2008) 2579–2605.
- [14] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013). doi:10.48550/arXiv.1301.3781.
- [15] M. Forghani, A. Firstkov, P. Vasev, E. Ramsay, Visualization of the evolutionary trajectory: Application of reduced amino acid alphabets and word2vec embedding, in: *Graphicon-Conference on Computer Graphics and Vision*, volume 32, 2022, pp. 275–287.
- [16] M. Forghani, P. Vasev, M. Bolkov, E. Ramsay, A. Bersenev, PhyloTraVis: A new approach to visualization of the phylogenetic tree, *Programming and Computer Software* 48 (2022) 215–226. doi:10.1134/S0361768822030045.
- [17] J. Zhang, A. M. Mamlouk, T. Martinetz, S. Chang, J. Wang, R. Hilgenfeld, Phylomap: an algorithm for visualizing relationships of large sequence data sets and its application to the influenza A virus genome, *BMC bioinformatics* 12 (2011) 248. doi:10.1186/1471-2105-12-248.
- [18] L. Shao, H. Zhou, Curve fitting with Bezier cubics, *Graphical models and image processing* 58 (1996) 223–232. doi:10.1006/gmpip.1996.0019.
- [19] Y.-C. Chen, A tutorial on kernel density estimation and recent advances, *Biostatistics & Epidemiology* 1 (2017) 161–187. doi:10.1080/24709360.2017.1396742.
- [20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [21] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, *Journal of molecular evolution* 16 (1980) 111–120. doi:10.1007/BF01731581.
- [22] K. Tamura, G. Stecher, S. Kumar, MEGA11: molecular evolutionary genetics analysis version 11, *Molecular biology and evolution* 38 (2021) 3022–3027. doi:10.1093/molbev/msab120.
- [23] M. L. Waskom, Seaborn: statistical data visualization, *Journal of Open Source Software* 6 (2021) 3021. doi:10.21105/joss.03021.
- [24] M. Forghani, A. Firstkov, M. Alyannezhadi, D. Danilenko, A. Komissarov, Reduced amino acid alphabet-based encoding and its impact on modeling influenza antigenic evolution, *Russian Journal of Infection and Immunity* 12 (2022) 837–849. doi:10.15789/2220-7619-RAA-1968.
- [25] S. Kawashima, P. Pokarowski, M. Pokarowska, A. Kolinski, T. Katayama, M. Kanehisa, AAindex: amino acid index database, progress report 2008, *Nucleic acids research* 36 (2007) D202–D205. doi:10.1093/nar/gkm998.
- [26] H. Jeon, A. Cho, J. Jang, S. Lee, J. Hyun, H.-K. Ko, J. Jo, J. Seo, ZADU: A python library for evaluating the reliability of dimensionality reduction embeddings, in: *2023 IEEE Visualization and Visual Analytics (VIS)*, IEEE, 2023, pp. 196–200. doi:10.1109/VIS54172.2023.00048.

## Visualization of phylogenetic self-similarity of genomic sequence

Kazem Forghani<sup>1</sup>, Majid Forghani<sup>2,3</sup> and Mikhail Bolkov<sup>4,\*</sup>

<sup>1</sup> Iran University of Science and Technology (IUST), Tehran, Iran

<sup>2</sup> Institute of Natural Sciences and Mathematics, Ural Federal University, Yekaterinburg, Russia

<sup>3</sup> N.N. Krasovskii Institute of Mathematics and Mechanics of the UB RAS, Yekaterinburg, Russia

<sup>4</sup> Institute for the Study of Aging, Russian Gerontological Scientific and Clinical Center, N.I. Pirogov Russian National Research Medical University of the Ministry of Health of the Russian Federation, Moscow, Russia

**Abstract.** The phylogenetic analysis is one of the fundamental analyses in bioinformatics, aiming at determining the evolutionary relationship between a set of taxa. While recent advances in sequencing technology have led to decreased costs and increased availability of whole-genome sequences, there are various studies focusing on employing a partial sequence as a genome representative in phylogenetic studies. This situation can often be encountered in the analysis of viral evolution, where the analysis is more biased towards antigen evolution and variation can be tracked through one or more coding subsequences of the genome. Such a representative may drastically decrease the cost of computation. The lack of a tool for finding a genome representative motivated us to investigate the application of distance metrics of tree space in order to highlight candidates for the role of phylogenetic representative. In this preliminary study, we propose a pipeline to compute the phylogenetic self-similarity of the genome by using a tree distance and visualizing the results in a 2D map. Though our approach is at an early stage of development, the obtained results indicate that the approach has potential as an exploratory tool in phylogenetic studies.

**Keywords:** visualization, phylogenetic tree, similarity, tree distance, genome.

### Introduction

Phylogenetic analysis plays an important role in many studies within evolutionary biology [1]. It is one of the fundamental analyses in bioinformatics, aiming at determining the evolutionary relationship between a set of taxa. Reconstructing a phylogenetic tree from molecular data can lead to a better understanding of evolution among different groups of organisms [2]. Specifically, in the case of virus studies, it provides a valuable view of the evolutionary history of viruses and the transmission behavior of diseases, which is urgent for designing a strategy to combat them [3]. Recent advances in sequencing lead to a rapid growth of large-scale genomic data, which highlights that estimating phylogeny requires novel, fast, and accurate methods and pipelines [4].

Embedding genetic sequences into tree space is a good way of interpretation since the raw sequences are not human-readable in terms of phylogenetic characteristics. Although visualizing phylogenetic relationships as a rooted acyclic graph is simple, interpretable, and facilitates hypothesis formulation and testing, such representation cannot cover all evolutionary events [5]. Recombination, horizontal transfer, and hybridization are examples of events that are poorly described in the content of rooted acyclic graphs. Therefore, a generalization of tree structure called a phylogenetic network is employed to express the reticulate events [6, 7].

An efficient way to quickly gain overall insight about phylogeny is selecting phylogenetic representatives to approximate a phylogenetic tree of the full genome. Previous studies have also demonstrated that some regions of the genome have the potential to be selected as its phylogenetic representative. This can be visited especially in viruses, where the majority of the evolution is directed at events occurring in specific antigens. For example, a fragment of glycoprotein E (gE) from the Tick-Borne Encephalitis Virus (TBEV) complete genome is used for genotyping and classifying its isolates [8]. The fragment consists of 151 amino acid residues (aa positions 104–254), encoded by 454 nucleotides (positions 309–762) in gene E [9].

Similarity measurement deeply underpins many areas of bioinformatics, including classical topics, such as sequence alignment [10], as well as modern topics like protein function prediction and even protein folding [11]. Various types of similarity measures have been defined for genomic sequences. While some are directly related to phylogeny, others are taken from fields such as natural language processing (NLP) and fractal analysis. In this work, we focus on self-similarity, which can reflect the evolutionary behavior of a continuous fragment in relation to the entire genome. Here, we provide a brief review of fractal-based methods and those based on NLP techniques.

To the best of our knowledge, the first fractal-like visualization of genomic sequences is the Chaos Game Representation (CGR), proposed by Jeffrey [12]. The CGR method draws a unit square, with each corner representing a nucleotide (A, C, G, and T). The algorithm begins by setting the center of the square as the initial point. At each iteration, the algorithm selects the nucleotide at the position corresponding to the iteration index. It then assigns a new point that lies midway between the point associated with the previous iteration (i.e., the previous nucleotide) and the corner corresponding to the current nucleotide. A practical proof of the potential application of CGR for assessing self-similarity was provided by plotting the CGRs of 21 human chromosomes [13]. Hao et al. [14] have observed fractal-like patterns across long DNA sequences and employed them to develop a deterministic method for visualization of DNA in a 2D portrait format. As stated by the authors, the method can be used to highlight the evolutionary relatedness of species. The proposed visualization illustrates the regularities in DNA, which can be beneficial in measuring self-similarity. Recently, Durán-Meza et al. [13] studied the multifractal properties of DNA and proposed a visualization of long DNA sequences by plotting the multifractal spectra representing the level of multifractality in the genomic sequence. Although fractal-based approaches have typically been developed to measure similarity between sequences, for example, in classification tasks [15], we believe they also have potential for measuring self-similarity within genomic sequences. Despite these promising results, fractal-based approaches remain difficult to interpret.

Generally, many mathematical tools for measuring similarity require a numerical representation of data. A genomic sequence can be represented in a numerical vector space using encoding or embedding approaches. These methods typically break the sequence into k-mers while preserving their order and replace each k-mer with its numerical representation. A simple example of DNA encoding is one-hot encoding, where each nucleotide is represented by a binary vector. More advanced approaches are adapted from NLP, such as the Word2Vec framework [16]. Another popular method is positional embedding [17], a key component of transformer models. This type of embedding not only captures the relationships between words within the sequence context but also incorporates their positional information, thereby enhancing model performance.

Accordingly, various embeddings for genomic sequences have been proposed, such as ProtVec [18] and Gene2Vec [19]. These embeddings are typically trained on large datasets to capture general patterns within the data or are generated for specific tasks using specialized datasets, for example, immunological sequences. Since some species have a unique type of genome organization and specific genetic patterns, employing Word2Vec-based embeddings to identify a genome representative may not be adequate. Despite their promising results on similar tasks, these models also require training on genomic sequences of each species individually to effectively measure self-similarity within its genome. In other words, the model must be specifically retrained and adapted for each species. In addition, numerical vector embeddings of genomic sequences often increase the sequence length, leading to higher computational cost.

Our aim is to leverage phylogenetic self-similarity across a set of genome sequences to highlight potential regions of candidates for the role of genome representative. The absence of a dedicated tool motivates the development of our approach. To maintain simplicity and interpretability, the proposed method operates in the space of phylogenetic trees.

Our contribution in this paper is as follows:

- Proposing a pipeline for computing the phylogenetic self-similarity along the genome and visualizing the results in a 2D map.
- Experimentally investigating the extent to which phylogenetic representatives can approximate the phylogenetic behavior of the complete genome.

To the best of our knowledge, no such pipeline has been proposed so far. The pipeline consists of four main steps:

- Estimating a reference phylogenetic tree.
- Estimating partial phylogenetic trees based on fragmented genome sequences.
- Comparing the partial phylogenetic trees with the reference tree using a tree distance metric.
- Visualizing the resulting similarity measures along the genome.

The pipeline accepts an aligned *FASTA* file as its input, processes it, and generates the final output as a 2D visualization scene. We also investigate the phylogenetic characteristics of a coding (nucleotide) sequence of

glycoprotein E as a genome representative for tick-borne encephalitis virus (TBEV). This fragment serves as a typical sequence for determining the TBEV phylogeny and is the key element in the clusteron approach [8, 20] and its implementation in the TBEV Analyzer platform [21].

The remainder of this paper is organized as follows: Section 2 presents the methodology of the pipeline, where each step is described in detail. In Section 3, we outline the computational experiment setup and present the results. The pipeline is evaluated through numerical experiments and visualization of the TBEV genome. Finally, conclusions and directions for future work are provided in Section 4.

## Methodology

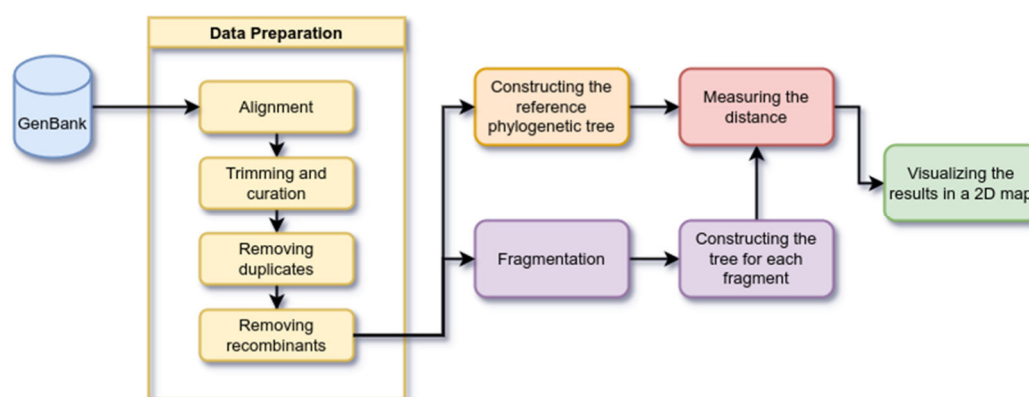
Figure 1 illustrates the overall schema of the proposed pipeline. It consists of five main steps:

- Data preparation.
- Estimation of the reference phylogenetic tree from the complete sequences.
- Fragmentation of the complete sequences and construction of a phylogenetic tree for each fragment.
- Measurement of the distance between each fragment tree and the reference tree.
- Visualization of the results in a 2D map, where the x-axis represents the position of the fragment within the complete sequence, and the y-axis indicates the fragment length.

Each step of the pipeline is explained in detail in the following subsections.

### 1. Data preparation

The goal of this step is to prepare a FASTA-format file of genomic sequences that meet three criteria: they must be aligned, free of recombinant sequences, and free of duplicates. We begin with alignment. Sequence alignment is a crucial step in most evolutionary analyses. The purpose of alignment is to arrange genomic sequences by identifying regions of similarity that indicate the type of relationship, such as evolutionary history. The output of an alignment algorithm consists of sequences of equal length. For more details on sequence alignment, the interested reader is referred to [22]. It is worth noting that, although alignment remains a foundational technique in bioinformatics, there is a growing trend in the scientific community toward developing alignment-free methods for various bioinformatics tasks [23].



**Figure 1.** Schema of the proposed pipeline. The pipeline mainly consists of five steps: data preparation; construction of the reference phylogenetic tree; sequence fragmentation; construction of a phylogenetic tree for each fragment; measurement of the distance between the reference tree and each fragment tree; and visualization of the final results in a 2D map. Note that FastTree is used for phylogenetic tree construction

The output of an alignment process usually requires post-processing steps, such as trimming, to tackle gaps. Hence, alignment results often need manual correction and expert validation. Among the available alignment tools, we chose MAFFT (Multiple Alignment using Fast Fourier Transform) [24] due to its speed, accuracy, and efficiency in handling long sequences, large datasets, and gap-rich regions. Note that the quality of alignment significantly affects the accuracy and reliability of inferred phylogenetic trees [25].

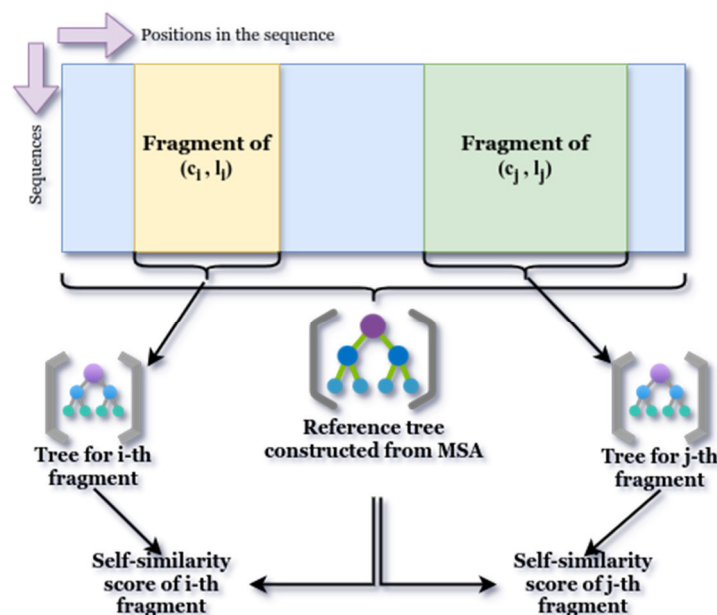
After trimming and handling gaps in the aligned sequences, duplicates are identified and removed. As mentioned earlier, certain evolutionary events, such as recombination, cannot be accurately represented in a rooted acyclic graph. Therefore, it is necessary to exclude the sequences of recombinants. Recombination



events within the sequences are identified using the Recombination Detection Program (RDP) [26] along with other methods within RDP5 (GENECONV, BootScan, MaxChi, Chimaera, SiScan, and 3Seq), applying the default parameters and a significance threshold of p-value of 0.05. Sequences flagged as potential recombinants by at least two methods are excluded from further analysis. The final curated multiple sequence alignment (MSA) is then converted to *PHYLIB* format using the *Biopython* package [27].

## 2. Reference phylogenetic tree estimation

The second step involves estimating a reference phylogenetic tree from the aligned sequences obtained in the previous step. This inferred tree serves as a reference for evaluating the similarity between a selected fragment and the whole sequence. A fragment, in this context, refers to a continuous subsequence of the genome. Algorithms for reconstructing a phylogenetic tree from a set of aligned sequences can be grouped into three main categories: distance-based methods (e.g., neighbor-joining), character-based methods (such as maximum parsimony and maximum likelihood), and Bayesian inference methods [2]. Maximum likelihood (ML)-based methods are generally evaluated as the most accurate, but they are computationally intensive [4]. Since searching for the optimal tree under the ML framework is an NP-hard problem, many algorithms employ heuristics to manage the trade-off between speed and accuracy.



**Figure 2.** The process of fragmentation and measuring the phylogenetic tree. A fragment is a shifting window along the multiple sequence alignment. The fragment is varied by its center and length. The self-similarity score is obtained by measuring the distance between the fragment and reference trees

There are various widely used tools for phylogenetic tree inference based on maximum likelihood, e.g., RAxML [28], IQ-TREE2 [29], and FastTree 2 [30]. RAxML (Randomized Axelerated Maximum Likelihood) is one of the leading programs based on ML estimation and is capable of handling large-scale datasets through various strategies to accelerate computation. In terms of speed and performance, FastTree is recognized as one of the fastest tools for approximate phylogenetic tree inference using the ML approach, offering reasonable accuracy in topology estimation despite slightly lower accuracy in branch length estimation [4, 25, 3]. In our pipeline, we use FastTree 2 with its default parameters for phylogenetic tree construction due to its computational efficiency. For simplicity, we refer to FastTree 2 as FastTree throughout the remainder of this paper. We also include RAxML in our numerical experiments to demonstrate that FastTree provides a good approximation of RAxML's results. It is worth mentioning that all trees generated by these programs are subsequently rooted at the midpoint using the *ETE3 Toolkit* [31].

## 3. Fragmentation

A fragment is a multiple sequence alignment that represents a subset of the full-genome MSA. It consists of aligned subsequences defined by two parameters: center and length. Conceptually, a fragment can be viewed

as a sliding window along the complete sequences, where the window's position and size are determined by its center and radius (see Fig. 2). To measure self-similarity, we generate multiple fragments from the full-genome MSA by varying their center and length, construct a phylogenetic tree for each fragment, and compute the distance between the fragment tree and the reference tree. A parallel computation strategy is employed to accelerate the pipeline's speed during the distance measurement step. Notably, all settings used for phylogenetic tree construction in this step are identical to those used for estimating the reference tree.

### Comparison of phylogenetic trees

Since a similarity measure can be viewed as the converse of a distance measure, we frequently switch between them in the rest of this paper. To evaluate the phylogenetic similarity between each fragment and the complete sequences, we use the topology of their corresponding phylogenetic trees. The genetic sequences of a fragment are embedded into tree space, and then the distances are calculated over this metric space. There are various metric distances defined for tree space according to tree clusters, splits, or sets of taxa (labels). The Robinson-Foulds (RF) distance [32] is one of the most common and simple metrics still used for phylogenetic tree comparison. Although it can be computed fast in linear time, it suffers from some shortcomings, e.g., imprecision compared to other methods, overestimation, and underestimation in some cases [33]. The distance is constructed based on two functions,  $\alpha$  and  $\alpha^{-1}$ , which are called contraction and decontraction operators, respectively. The  $\alpha$  operator contracts an edge by merging its endpoints and creating a new vertex labeled by the union of the endpoint's labels. On the contrary, the  $\alpha^{-1}$  performs the inverse operation by generating a new edge and converting a vertex into two vertices representing the endpoints of the new edge, while the label set of the old vertex is split in any fashion between the two new vertices. Given two trees,  $T_1$  and  $T_2$ , defined on the same set of taxa, the  $\alpha$  operator contracts all edges from  $T_1$  that are not present in  $T_2$  and results in  $T_1 \wedge T_2$ . The  $\alpha^{-1}$  expands the  $T_1 \wedge T_2$  structure by adding those edges that are in  $T_2$  but are not included in  $T_1$  to reconstruct  $T_2$ . The RF distance between two trees,  $T_1$  and  $T_2$ , is defined as the minimum number of operations ( $\alpha$  and  $\alpha^{-1}$ ) to convert  $T_1$  into  $T_2$  [32]. The distance ranges from zero to  $2(n - 3)$  for the case of unrooted trees with  $n$  taxa. Therefore, it can be normalized as follows:

$$d_{NRF}(T_1, T_2) = \frac{d_{RF}}{2(n - 3)}.$$

There are various versions of RF distance by its generalization, normalization, or the method of calculating the similarity score between pairs of splits [34, 35, 36]. Smith [35] investigated the generalized RF metrics for comparing phylogenetic trees, in the results of which he recommends the application of clustering information (CI) distance due to intuitiveness and meaningfulness. The CI distance is a kind of entropy distance. Taking the same definition and notation provided by Smith, we explain the basics of this distance metric.

A split of a tree with taxa set  $X$  into a bipartition  $A$  and  $B$  leads to the formation of two disjoint clusters of taxa associated with  $A$  and  $B$ . The entropy of split  $S$  is calculated as follows:

$$E(S) = -P_{Cl}(A) \log P_{Cl}(A) - P_{Cl}(B) \log P_{Cl}(B),$$

where  $P_{Cl}(A)$  is the probability that a randomly selected taxon (label) belongs to cluster  $A$  and is defined as  $P_{Cl}(A) = |A| \div |X|$ . The same holds for  $P_{Cl}(B)$  accordingly. The CI distance corresponds to the mutual clustering information (MCI) concept. Given two splits,  $S_1$  and  $S_2$ , the MCI measures the amount of information that is shared between splits. In other words, it reflects how knowledge about disjoint clusters  $A_1$  and  $B_1$  obtained from split  $S_1$  decreases the uncertainty about disjoint clusters  $A_2$  and  $B_2$  in  $S_2$ . MCI is calculated as follows [34]:

$$\begin{aligned} I_{Cl}(S_1; S_2) = & P_{Cl}(A_1, A_2) \log \frac{P_{Cl}(A_1, A_2)}{P_{Cl}(A_1)P_{Cl}(A_2)} + \\ & + P_{Cl}(A_1, B_2) \log \frac{P_{Cl}(A_1, B_2)}{P_{Cl}(A_1)P_{Cl}(B_2)} + \\ & + P_{Cl}(B_1, A_2) \log \frac{P_{Cl}(B_1, A_2)}{P_{Cl}(B_1)P_{Cl}(A_2)} + \\ & + P_{Cl}(B_1, B_2) \log \frac{P_{Cl}(B_1, B_2)}{P_{Cl}(B_1)P_{Cl}(B_2)}, \end{aligned}$$

where  $P_{Cl}(A_1, A_2) = |A_1 \cap A_2| \div |X|$  represents the probability of a point occurrence in both clusters  $A_1$  and  $A_2$  from  $S_1$  and  $S_2$ , respectively. The metric reflects the degree of agreement between two trees on exhibiting the relationship of taxa by grouping them. The CI distance employed in this paper is based on MCI and normalized, so it ranges between zero and one. The interested reader is referred to Smith's paper [34] for a detailed description of the distance.

## 2D map visualization

The fragmentation yields a 2D map, with the X-axis and Y-axis representing the center and length of fragments, respectively. The map is constructed based on a grid, where each point corresponds to a fragment. The coordinates of each point are determined by striding over user-specified ranges for the center and length parameters, and the value at each point represents the distance between the fragment's phylogenetic tree and the reference tree. Thus, each fragment is associated with a 3D coordinate  $(x, y, z)$ , where  $x$  and  $y$  denote the fragment's center and length, and  $z$  indicates the tree distance. To enhance the visualization quality, we apply interpolation and generate a contour plot, in which the color of each point reflects the similarity between the fragment and the full MSA. Colors are assigned using a blue-to-red gradient palette, where blue indicates high similarity and red indicates low similarity.

## Experiments & Results

Here, we discuss the details of two experiments. In the first one, we reconstruct the 2D self-similarity map using the expensive method RAxML, the result of which is called the *reference map*. Although we believe this map has a very good approximation of self-similarity, we try to find a desirable approximation that has a better trade-off between the accuracy and cost. This new approximation is obtained by applying FastTree, and the resulting map is compared with the one obtained from RAxML. That is why we refer to the RAxML's map as the reference map.

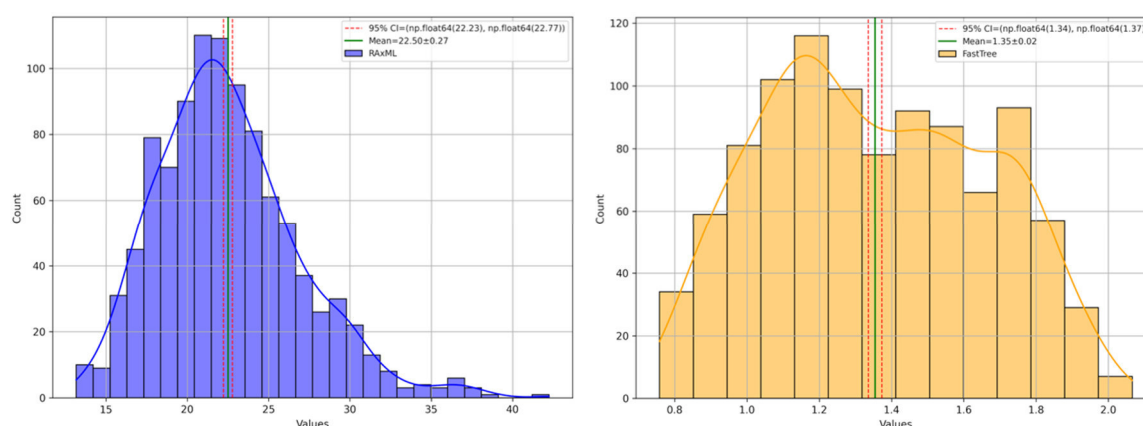
We use the tick-borne encephalitis virus (TBEV) as a case study. This virus has an approximately 11-kilobase genome. We downloaded 493 records from the GenBank nuccore database [37] by setting the minimum length of sequence to nine kilobases. The obtained FASTA file from GenBank was fed into the MAFFT (v7.490) program to conduct alignment. The result of alignment was manually trimmed and curated in MEGA 11 (version 11.0.13) [38]. The curated FASTA file was checked for any duplicate sequence by the SeqKit toolkit [39]. We removed the recombinants using the RDP5 program [26], as stated in Subsection 2.1. The final FASTA file has 300 viruses that are further used for constructing the self-similarity map.

Before starting the experiments, we need to determine the grid, based on which the 2D map will be constructed. We set the minimum and maximum of the fragment length ( $L_{min}$  and  $L_{max}$ ) to 300 and 600, respectively. The stride value for both length and center was set to three since a codon is defined by three nucleotides. The fragment center varies within the range  $[\lfloor \frac{L_{min}}{2} \rfloor, L - \lfloor \frac{L_{min}}{2} \rfloor]$ , where  $L$  is the length of the complete genome sequence. Considering these parameters, a total set of 334209 fragments was generated during each experiment.

To conduct the first experiment, it is required to construct the reference tree. This tree is obtained by applying RAxML to the whole genome sequences of TBEVs. We measure the similarity of a fragment tree according to the reference tree. Note that all parameters for generating the (reference and fragment) trees are fixed during each experiment. Once the reference tree was achieved, the fragments were fed into RAxML in a parallel fashion to construct their phylogenetic tree under the GTRGAMMA model. We employed the rpy2 package [40], which allows accessing R packages in Python, to use the TreeDist and ape packages for loading trees and measuring the distance between a fragment tree and the reference one. The distance was computed by the *ClusteringInfoDistance* function in a normalized fashion [34, 36]. Note that all trees were rooted at the midpoint using the *ETE3 Toolkit* [31] before distance measurement. The fragmentation forms a grid, the axes of which represent the length and center of fragments. Each row of the grid includes the fragments with the same length while the center is striding. To provide a contour visualization, we applied the unstructured triangular grid followed by interpolating to the grid using Matplotlib [41].

The only difference between the first and second experiments lies in the phylogenetic tree reconstruction program. The second experiment utilizes FastTree 2 instead of RAxML to boost the speed of the pipeline. To

evaluate the speed of FastTree in comparison to RAxML, we carried out a supplementary experiment in which 1000 out of 334209 fragments were randomly selected and given to both programs to construct the phylogenetic tree while we tracked the execution time for them. Figure 3 shows the distributions of tree generation for fragments. The result indicates that FastTree is at least 20 times faster than RAxML. Here, a question arises whether this acceleration gives a good approximation of the RAxML results. Figure 4 demonstrates the results of 2D map visualization for both programs.



**Figure 3.** The execution time distribution for constructing the phylogenetic tree of randomly selected 1000 fragments. The left and right histograms are the results for RAxML and FastTree, respectively. FastTree outperforms RAxML in terms of speed by about 20 times. The X-axis represents time in seconds

In recent years, a platform for analyzing the characteristics of TBEV called TBEV Analyzer has been developed [21]. The platform utilizes a specific fragment of the coding (nucleotide) sequence of the gene E and returns the hierarchical phylogenetic characteristics of the query virus. This fragment includes 454 nucleotides (positions 309–762) encoding 151 amino acid residues (positions 104–254 aa) of the viral genome (gene E). As mentioned in [42], there are five reasons due to which this fragment is chosen to characterize TBEV:

- It includes both conservative and variable regions.
- It contains unique amino acid substitutions at positions 175, 206, and 234 that are critical for identifying the subtypes and phylogenetic lineages.
- The majority of the GenBank TBEV records cover this fragment.
- The fragment length is a compromise between sufficient information content and the possibility of its amplification, which allows studying the maximum number of virus samples obtained per season from natural foci of TBEV, almost in real-time mode.
- The results of phylogenetic analysis based on this fragment are quite comparable in informativeness with those of the complete genome sequence.

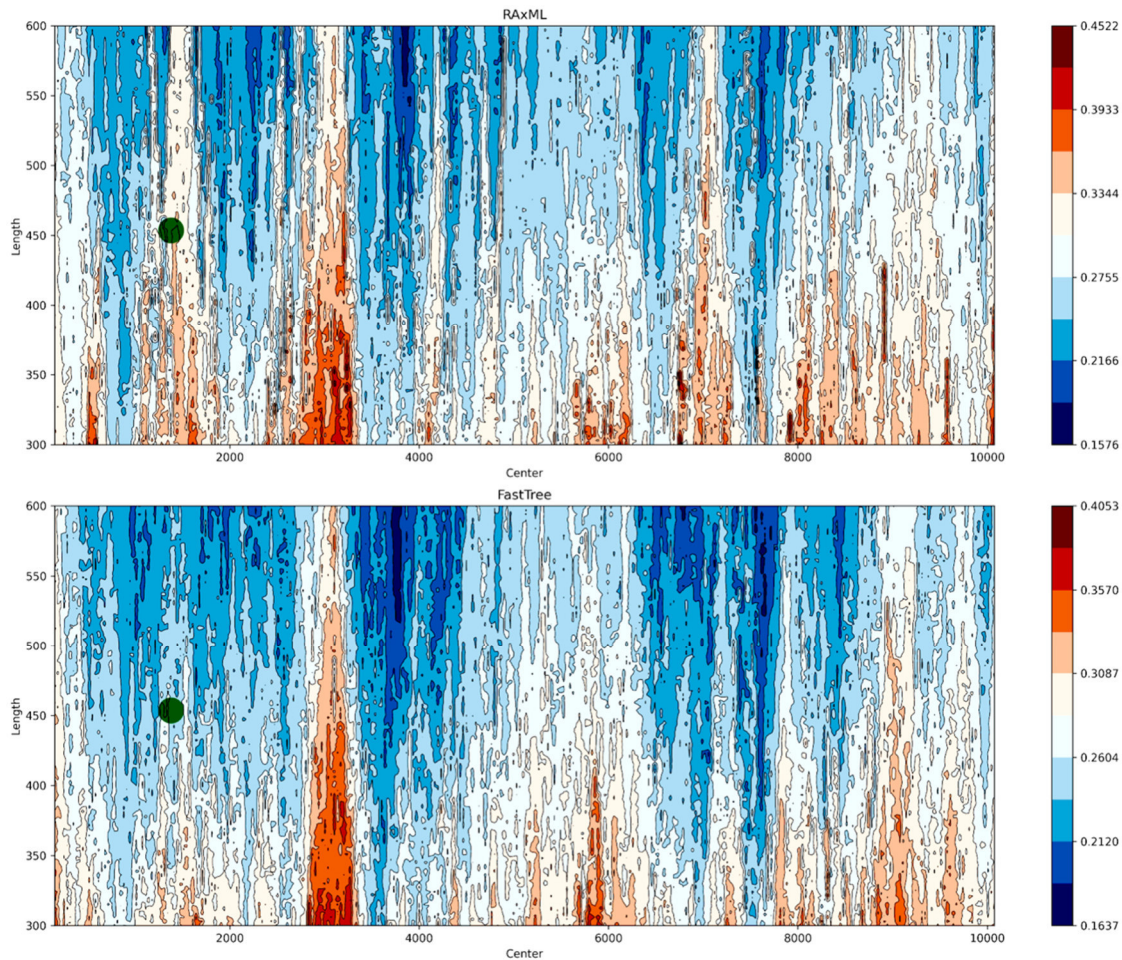
The last reason indicates that it has a role to be a representative for the TBEV genome. The green point in Fig. 4 indicates the position of this fragment along the grid. The overall distance of this fragment is about 0.3 and 0.25 for RAxML and FastTree 2, respectively. It seems that the fragment has a moderate degree of similarity. Obviously, there are several regions with dark blue indicating the high degree of similarity with the full genome. It should be better to point out that the 2D map is used to infer a set of candidates; however, choosing a representative requires additional information and verification that is beyond the pipeline's functionality.

Altogether, we recommend using our pipeline with FastTree to save time and resources. FastTree provides a good approximation of RAxML's results. Generally, the results of FastTree are smoother and more optimistic than those of RAxML. The pipeline mainly has two drawbacks. The first drawback is the resolution of the grid. Increasing the length of the fragment provides more information about the genome, leading to decreasing the fragment tree distance. We try to determine a proper resolution for the fragmentation grid. A high resolution in both center and length causes the number of fragments to increase in a quadratic manner, which adds more cost to computation. The appropriate choice is currently determined experimentally and depends on the length of the genome. Another drawback is that the choice of grid parameters requires prior knowledge about the



minimum and maximum length of the fragment. Indeed, the fragment should have enough length to reflect the phylogenetic variation.

To the best of our knowledge, this is the first time that a visualization of phylogenetic self-similarity of the genome has been proposed. The primary application of our pipeline is to identify a set of candidates of partial sequences aiming to search for a genome representative. The choice of candidates depends on the task in which the representative is employed and requires prior knowledge about, e.g., biological characteristics of candidates. Application of the genome representative may dramatically decrease the cost of computation in phylogenetic studies.



**Figure 4.** The 2D map visualization of phylogenetic self-similarity for the TBEV genome. The top plot is the results obtained from the RAxML program, while the bottom one is for FastTree. It is worth mentioning that FastTree gives a good approximation of the RAxML map while it is almost 22 times faster than RAxML within our experiments. The blue and red colors exhibit the low and high distances to the reference phylogenetic tree, respectively. The green point shows the result for the coding sequence of the gene E fragment used in the TBEV Analyzer platform

## Conclusions

In this paper, we conducted a pilot study on computation and visualization of phylogenetic self-similarity for the genome. The main goal of the proposed pipeline is identifying the potential candidates that are more similar to the genome regarding the phylogenetic characteristics. The set of candidates can further be processed by an expert to select a representative of the genome. Such a representative may drastically decrease the cost of computation in phylogenetic studies. A desirable representative should be informative, biologically significant, and available in a public database. A good example of such a representative is the fragment of TBEV's gene E that is employed as input for the TBEV Analyzer platform. Besides its advantages, we demonstrated that this fragment has a moderate degree of similarity to the genome.

We plan to improve the pipeline speed by employing the VeryFastTree program [43] instead of FastTree. However, this requires comparing the map obtained from VeryFastTree with those of FastTree and RAxML

to determine the quality of approximation. Another improvement can be achieved by the application of heuristics for searching candidates in the grid space. This can avoid computing the similarity for the whole grid and lead to implementing an optimized algorithm. The recent advances in the field of constructing phylogenetic trees suggest the application of deep learning [44]. We expect that deep learning has the potential to facilitate the task of computing the phylogenetic self-similarity and speed up the pipeline while maintaining the quality of approximation in the visualization scene.

### Acknowledgments

M. Forghani was supported by the Ministry of Science and Higher Education of the Russian Federation, project FEUZ-2023-0022.

### Declaration on Generative AI

The authors have not employed any Generative AI tools.

### References

- [1] P. Kapli, Z. Yang, M. J. Telford, Phylogenetic tree building in the genomic age, *Nature Reviews Genetics* 21 (2020) 428–444. doi:10.1038/s41576-020-0233-0.
- [2] R. Godini, H. Fallahi, A brief overview of the concepts, methods and computational tools used in phylogenetic tree construction and gene prediction. *Meta Gene*, 2019; 21: 100586, 2019. doi:10.1016/j.mgene.2019.100586.
- [3] C. Young, S. Meng, N. Moshiri, An Evaluation of Phylogenetic Workflows in Viral Molecular Epidemiology, *Viruses* 14 (2022) 774. doi:10.3390/v14040774.
- [4] P. Zaharias, T. Warnow, Recent progress on methods for estimating and updating large phylogenies, *Philosophical Transactions of the Royal Society B* 377 (2022) 20210244. doi:10.1098/rstb.2021.0244.
- [5] D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies, *Molecular biology and evolution* 23 (2006) 254–267. doi:10.1093/molbev/msj030.
- [6] K. T. Huber, V. Moulton, T. Wu, Transforming phylogenetic networks: Moving beyond tree space, *Journal of theoretical biology* 404 (2016) 30–39. doi:10.48550/arXiv.1601.01788.
- [7] M. Hellmuth, D. Schaller, P. F. Stadler, Clustering systems of phylogenetic networks, *Theory in Biosciences* 142 (2023) 301–358. doi:10.1007/s12064-023-00398-w.
- [8] S. Kovalev, T. Mukhacheva, Reconsidering the classification of tick-borne encephalitis virus within the Siberian subtype gives new insights into its evolutionary history, *Infection, Genetics and Evolution* 55 (2017) 159–165. doi:10.1016/j.meegid.2017.09.014.
- [9] M. Forghani, S. Kovalev, M. Bolkov, M. Khachay, P. Vasev, TBEV analyzer platform for evolutionary analysis and monitoring tick-borne encephalitis virus: 2020 update, *Biostatistics & Epidemiology* 6 (2022) 57–73. doi:10.1080/24709360.2021.1985392.
- [10] M. S. Rosenberg, *Sequence alignment: methods, models, concepts, and strategies*, Univ of California Press, 2009.
- [11] K. Yan, J. Wen, J.-X. Liu, Y. Xu, B. Liu, Protein fold recognition by combining support vector machines and pairwise sequence similarity scores, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 18 (2020) 2008–2016. doi:10.1109/TCBB.2020.2966450.
- [12] H. J. Jeffrey, Chaos game visualization of sequences, *Computers & Graphics* 16 (1992) 25–33. doi:10.1016/0097-8493(92)90067-6.
- [13] G. Durán-Meza, J. López-García, J. L. del Río-Correa, The self-similarity properties and multifractal analysis of DNA sequences., *Applied Mathematics & Nonlinear Sciences* 4 (2019). doi:10.2478/AMNS.2019.1.00023.
- [14] B.-I. Hao, H.-C. Lee, S.-y. Zhang, Fractals related to long DNA sequences and complete genomes, *Chaos, Solitons & Fractals* 11 (2000) 825–836. doi:10.1016/S0960-0779(98)00182-9.
- [15] Y. Li, B. Jiang, H. Chen, X. Yao, Symbolic sequence classification in the fractal space, *IEEE Transactions on emerging topics in computational intelligence* 5 (2018) 168–177. doi:10.1109/TETCI.2018.2876528.
- [16] T. Mikolov, W.-t. Yih, G. Zweig, Linguistic regularities in continuous space word representations, in: *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies, 2013*, pp. 746–751.
- [17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* 30 (2017).
- [18] E. Asgari, M. R. Mofrad, Continuous distributed representation of biological sequences for deep proteomics and genomics, *PloS one* 10 (2015) e0141287. doi:10.1371/journal.pone.0141287.
- [19] J. Du, P. Jia, Y. Dai, C. Tao, Z. Zhao, D. Zhi, Gene2vec: distributed representation of genes based on co-expression, *BMC genomics* 20 (2019) 82. doi:10.1186/s12864-018-5370-x.
- [20] S. Y. Kovalev, T. A. Mukhacheva, Clusteron structure of tick-borne encephalitis virus populations, *Infection, Genetics and Evolution* 14 (2013) 22–28. doi:10.1016/j.meegid.2012.10.011.
- [21] M. Forghani, S. Kovalev, M. Khachay, E. Ramsay, M. Bolkov, P. Vasev, Identifying new clusterons: application of TBEV analyzer 3.0, *Microorganisms* 11 (2023) 324. doi:10.3390/microorganisms11020324.

- [22] Y. Zhang, Q. Zhang, J. Zhou, Q. Zou, A survey on the algorithm and development of multiple sequence alignment, *Briefings in bioinformatics* 23 (2022) bbac069. doi:10.1093/bib/bbac069.
- [23] A. Zielezinski, H. Z. Girgis, G. Bernard, C.-A. Leimeister, K. Tang, T. Dencker, A. K. Lau, S. Röhling, J. J. Choi, M. S. Waterman, et al., Benchmarking of alignment-free sequence comparison methods, *Genome biology* 20 (2019) 144. doi:10.1186/s13059-019-1755-7.
- [24] K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Molecular biology and evolution* 30 (2013) 772–780. doi:10.1093/molbev/mst010.
- [25] J. A. Lees, M. Kendall, J. Parkhill, C. Colijn, S. D. Bentley, S. R. Harris, Evaluation of phylogenetic reconstruction methods using bacterial whole genomes: a simulation based study, *Wellcome open research* 3 (2018). doi:10.12688/wellcomeopenres.14265.2.
- [26] D. P. Martin, A. Varsani, P. Roumagnac, G. Botha, S. Maslamoney, T. Schwab, Z. Kelz, V. Kumar, B. Murrell, RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets, *Virus evolution* 7 (2021) veaa087. doi:10.1093/ve/veaa087.
- [27] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kau, B. Wilczynski, et al., Biopython: freely available python tools for computational molecular biology and bioinformatics, *Bioinformatics* 25 (2009) 1422. doi:10.1093/bioinformatics/btp163.
- [28] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics* 30 (2014) 1312–1313. doi:10.1093/bioinformatics/btu033.
- [29] B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, *Molecular biology and evolution* 37 (2020) 1530–1534. doi:10.1093/molbev/msaa015.
- [30] M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—approximately maximum-likelihood trees for large alignments, *PloS one* 5 (2010) e9490. doi:10.1371/journal.pone.0009490.
- [31] J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: reconstruction, analysis, and visualization of phylogenomic data, *Molecular biology and evolution* 33 (2016) 1635–1638. doi:10.1093/molbev/msw046.
- [32] D. F. Robinson, L. R. Foulds, Comparison of phylogenetic trees, *Mathematical biosciences* 53 (1981) 131–147. doi:10.1016/0025-5564(81)90043-2.
- [33] H. Folkertsma, A. Mittal, Comparing phylogenetic trees: an overview of state-of-the-art methods, 16th SC@RUG 2018-2019 (2019) 14.
- [34] M. R. Smith, Information theoretic Generalized Robinson-Foulds metrics for comparing phylogenetic trees, *Bioinformatics* 36 (2020) 5007–5013. doi:10.1093/bioinformatics/btaa614.
- [35] M. R. Smith, Robust analysis of phylogenetic tree space, *Systematic Biology* 71 (2022) 1255–1270. doi:10.1093/sysbio/syab100.
- [36] M. R. Smith, TreeDist: Distances between Phylogenetic Trees. R package version 2.9.2, 2020. doi:10.5281/zenodo.3528124.
- [37] D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt, E. W. Sayers, Genbank, *Nucleic acids research* 46 (2018) D41–D47. doi:10.1093/nar/gkx1094.
- [38] K. Tamura, G. Stecher, S. Kumar, MEGA 11: molecular evolutionary genetics analysis version 11, *Molecular biology and evolution* 38 (2021) 3022–3027. doi:10.1093/molbev/msab120.
- [39] W. Shen, B. Sipos, L. Zhao, SeqKit2: A swiss army knife for sequence and alignment processing, *Imeta* 3 (2024) e191. doi:10.1002/imt2.191.
- [40] L. Gautier, rpy2, 2023. URL: [https://github.com/rpy2/rpy2/releases/tag/RELEASE\\_3\\_5\\_11](https://github.com/rpy2/rpy2/releases/tag/RELEASE_3_5_11).
- [41] J. D. Hunter, Matplotlib: A 2D graphics environment, *Computing in Science & Engineering* 9 (2007) 90–95. doi:10.1109/MCSE.2007.55.
- [42] M. Forghani, S. Kovalev, P. Vasev, M. Bolkov, TBEV analyzer: A platform for evolutionary analysis of tick-borne encephalitis virus, in: 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), IEEE, 2019, pp. 0397–0402. doi:10.1109/SIBIRCON48586.2019.8958021.
- [43] C. Piñeiro, J. M. Abuín, J. C. Pichel, Very Fast Tree: speeding up the estimation of phylogenies for large alignments through parallelization and vectorization strategies, *Bioinformatics* 36 (2020) 4658–4659. doi:10.1093/bioinformatics/btaa582.
- [44] A. Suvorov, J. Hochuli, D. R. Schrider, Accurate inference of tree topologies from multiple sequence alignments using deep learning, *Systematic biology* 69 (2020) 221–233. doi:10.1093/sysbio/syz060.



## Современные технологии визуализации для создания и управления строительной документацией

Л. А. Максименко

Сибирский государственный университет геосистем и технологий, Новосибирск, Россия

**Аннотация.** В работе рассмотрены роль и значимость технологий визуализации данных в современном информационном обществе в контексте разработки и ведения строительной документации, развития методик оценки эффективности визуализационных инструментов в рамках строительных процессов и образования. Также освещены регуляторные аспекты применения методов визуализации, включая вопросы классификации, безопасности, конфиденциальности и этики использования данных. В исследовании темы «Современные технологии визуализации в создании и управлении строительной документацией» сделаны акценты на классификацию технологий визуализации по назначению; рассмотрены вопросы интеграции технологий в процессы создания, хранения, обмена и использования документации, включая современные методы визуализации, такие как BIM/ТИМ-моделирование, виртуальная и дополненная реальность и другие, что позволяет выделить их роль в строительной отрасли. Рассмотрена концепция создания цифрового двойника строительного объекта, которая дополнена наблюдениями за техническим состоянием строительных объектов в период их эксплуатации, с акцентом на роль геодезического обеспечения и инструментального мониторинга. Показано, что визуализации для сбора, хранения и обработки данных становятся все более доступными и масштабируемыми с развитием облачных вычислений и цифровых потоков 3D-моделирования.

**Ключевые слова:** строительная документация, визуализация данных, классификация, методы визуализации, BIM-модель здания, цифровой двойник, инженерный BIM, информационные ресурсы, геоинформационные системы, документооборот, генеративный дизайн.

## Modern visualization technologies in the creation and management of construction documentation

L. A. Maksimenko

Siberian State University of Geosystems and Technologies, Novosibirsk, Russia

**Abstract.** The article examines the role and importance of data visualization technologies in the modern information society in the context of the development and maintenance of construction documentation. Regulatory aspects of the application of visualization methods are also highlighted, including issues of classification, security, confidentiality and ethics of data use. The study of the topic "Modern visualization technologies in the creation and management of construction documentation" focuses on the classification of visualization technologies by purpose; The issues of technology integration into the processes of creating, storing, exchanging and using documentation, including modern visualization methods such as BIM modeling, virtual and augmented reality, and others, are considered, which makes it possible to highlight their role in the construction industry. The concept of creating a digital twin of a construction site is being considered, which is complemented by observations of the technical condition of construction sites during their operation, with an emphasis on the role of geodetic support and instrumental monitoring. It is shown that visualizations for data collection, storage and processing are becoming more accessible and scalable with the development of cloud computing and digital 3D modeling streams.

**Keywords:** construction documentation, data visualization, classification, visualization methods, BIM building model, digital twin, engineering BIM, information resources, geoinformation systems, document management, generative design.

### Введение

Актуальность темы «Современные технологии визуализации в создании и управлении строительной документацией» обусловлена быстрым развитием информационных технологий и их внедрением в строительную отрасль, что в условиях постоянного усиления требований к точности, эффективности и прозрачности процессов проектирования, строительства и эксплуатации объектов, становится необходимым для повышения качества работы и сокращения временных затрат. В условиях глобальной цифровизации и автоматизации работы в строительстве применение современных технологий визуализации становится ключевым фактором повышения конкурентоспособности и инновационности компаний, а также обеспечивает более эффективное взаимодействие между проектировщиками, строителями и заказчиками. Новизна и актуальность темы обусловлены значительным прогрессом в области информационных технологий и их внедрением в строительную

отрасль. В отличие от ручных методов, которые основываются на двумерных чертежах, бумажной документации и устаревших подходах, современные технологии позволяют создавать точные трёхмерные модели, использовать виртуальную и дополненную реальность, автоматизированные системы генерации документации и облачные платформы для совместной работы. В условиях перехода строительной отрасли на цифровые методы ведения проектов актуальность работы определена потребностью комплексной подготовки специалистов не только для работы с современным программным обеспечением, но и связана с разработкой и апробацией в образовательном процессе методик и инструментов, основанных на технологиях визуализации, для формирования у обучающихся навыков работы с цифровой строительной документацией. Дальнейшая разработка темы позволит не только повысить технологическую эффективность строительных процессов, но и значительно улучшить качество, прозрачность и управление проектами, что актуально в условиях современного строительства и цифровой трансформации отрасли.

### **Постановка задачи**

Цель исследования состояла в определении роли и эффективности современных технологий визуализации в процессе создания и управления строительной документацией различного назначения, а также оценке их влияния на эффективность, точность и качество строительных процессов. В задачи исследования входили вопросы анализа существующих классификаций технологий визуализации, объединения их в единую схему для разработки рекомендаций по выбору методов визуализации на различных этапах оформления строительной документации. Рассмотрены вопросы интеграции технологий в процессы создания, хранения, обмена и использования строительной документации, включая современные методы визуализации, такие как BIM/ТИМ-моделирование, виртуальная и дополненная реальность и другие, что позволяет выделить их роль в строительной отрасли.

Актуальность работы в условиях перехода строительной отрасли на цифровые методы ведения проектов связана с существующими запросами на подготовку специалистов не только для работы с программным обеспечением, но и для приобретения навыков оформления и документирования результатов работы в соответствии с требованиями отечественных стандартов. В связи с введением распоряжения Правительства РФ от 21 декабря 2021 г. № 3759-р «Об утверждении стратегического направления в области цифровой трансформации науки и высшего образования» [1] рассматриваются способы применения метода визуализации информации с использованием программных комплексов генеративного дизайна с элементами искусственного интеллекта, диаграмм и дашбордов в рамках обновления и расширения учебного процесса.

### **Теория**

Строительная документация, как исторически сложилось, представляет собой набор графических и текстовых документов, которые описывают все этапы строительства объекта. В последнее время появилась концепция строительства без бумажных технологий, основанная на компьютерных визуализациях, но, как показывает опыт, практическая реализация проекта невозможна без чертежей, расчетов, смет, планов по организации работ и других документов. Переход к безбумажным технологиям предъявляет высокие требования не только к материальной базе, но и требует переосмысления в мировоззрении и восприятии электронной документации. Строительная документация обеспечивает контроль над всеми этапами строительства, помогает согласовывать действия всех специалистов и служб и гарантирует, что на каждом этапе строительства соблюдаются стандарты качества. По проектной документации все участники проекта работают с одной и той же информацией. К основным видам строительной документации относят проектную документацию, генеральный план, архитектурные решения, конструктивную часть, разделы по инженерным системам, сметную документацию, проект организации строительства (ПОС), проект производства работ (ППР). В настоящее время большое внимание отводится разработке эксплуатационной документации, в качестве обязательного документа требуется разработка электронного паспорта.

Применявшаяся ранее традиционная подготовка всех видов проектной документации, разрабатываемая вручную или с минимальной помощью цифровых инструментов постепенно

вытесняется новыми инструментами и технологиями. Основными проблемами «ручного» подхода были большие временные затраты, отсутствие единого информационного пространства, трудности с актуализацией данных и другие. Следует отметить, что эскизирование, технический рисунок и некоторые другие формы «ручной» графики до сих пор актуальны и востребованы.

С развитием компьютерной графики и техники воспроизводства строительной документации все большее развитие получают методы, позволяющие преобразовывать данные, информацию или абстрактные концепции в визуальные формы, такие как графики, диаграммы, карты, трёхмерные модели и другие изображения. Совокупность таких методов можно определить как технологии визуализации, основная цель которых – упростить восприятие сложных данных, выявить закономерности, тенденции и связи, а также сделать информацию более доступной и понятной для пользователей. Общая классификация визуализации на современном этапе представлена на рисунке 1.

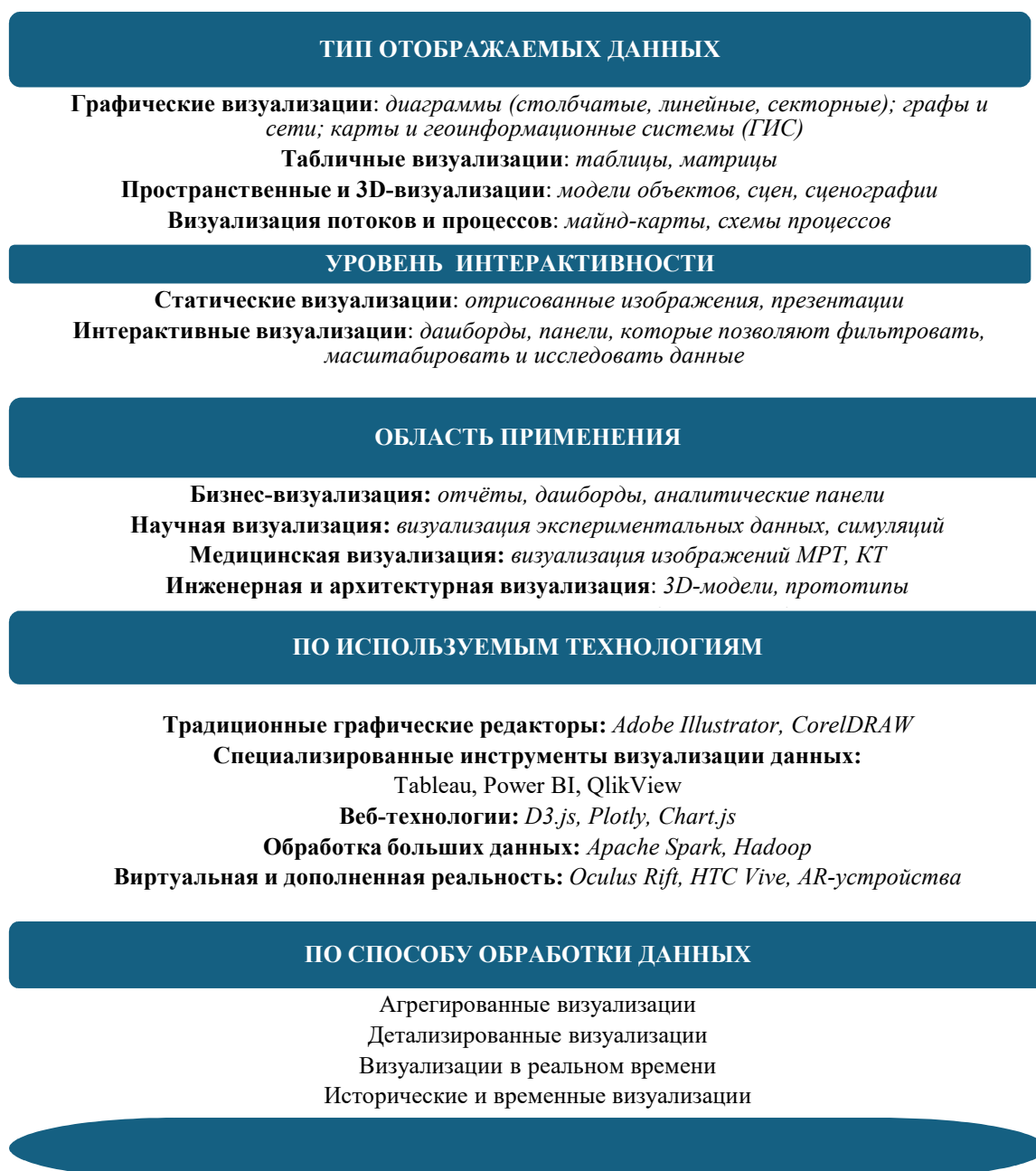


Рисунок 1. Классификация методов визуализации

Приведенная классификация методов визуализации помогает понять разнообразие методов и инструментов, выбрать наилучший вариант для решения задач в конкретной области.

### Результаты экспериментов

Отечественные инструменты цифровой визуализации для разработки и мониторинга строительной документации создаются в программных комплексах. BIM-модель здания содержит всю необходимую документацию о проекте в 3D-формате, что позволяет объединить все инженерные системы и конструктивные решения, управлять замечаниями и комментариями к проекту прямо в модели, согласовывать материалы и принимать работы прямо на стройплощадке. Вся строительная документация (проектная, рабочая, исполнительная, сметная, расход материалов и другие документы) оказывается собрана в едином информационном пространстве.

Выбор программного обеспечения для создания чертежей зависит от задач и опыта пользователя. Существуют как платные, так и бесплатные программы, а также онлайн-сервисы. Как уже отмечалось выше, при переходе строительной отрасли на цифровые методы ведения работы появляется необходимость разработки и апробации в образовательном процессе методик и инструментов, основанных на технологиях визуализации, для формирования у обучающихся навыков работы с цифровой строительной документацией.

Кадастровая деятельность представляет собой информационное пространство, наполненное в том числе архитектурно-строительными данными об объектах недвижимости. Продвигаемые идеи создания 3D-кадастра до сих пор не находят конкретных решений в законодательстве, поскольку проблемы регистрации прав и учета объектов недвижимости в большинстве случаев успешно решаются по 2D-строительной документации.

При подготовке специалистов по направлению «Землеустройство и кадастры» применяются пространственные и 3D-визуализации, реализованные в отечественном программном обеспечении. Наиболее востребованным из них является NanoCAD – российская проприетарная САПР-платформа для Windows, разработанная компанией «Нанософт разработка». Она позволяет создавать рабочую документацию и чертежи, поддерживает формат DWG благодаря интерфейсу, схожему с AutoCAD, и имеет открытый API для разработки специализированных решений. NanoCAD предлагает широкий спектр полезных и уникальных функций, делающих его привлекательным выбором для пользователей. Помимо основного функционала, платформа nanoCAD включает в себя несколько независимо лицензируемых встроенных модулей: СПДС, Механика, Растр, Организация, Топоплан, Строительство, 3D-инструменты и другие. Программный комплекс nanoCAD Инженерный BIM состоящий из шести профессиональных инструментов, предназначенный для инженеров-проектировщиков, активно внедряется и в учебный процесс. Структура комплекса nanoCAD Инженерный BIM представлена на рисунке 2, в состав комплекса входит платформа nanoCAD.

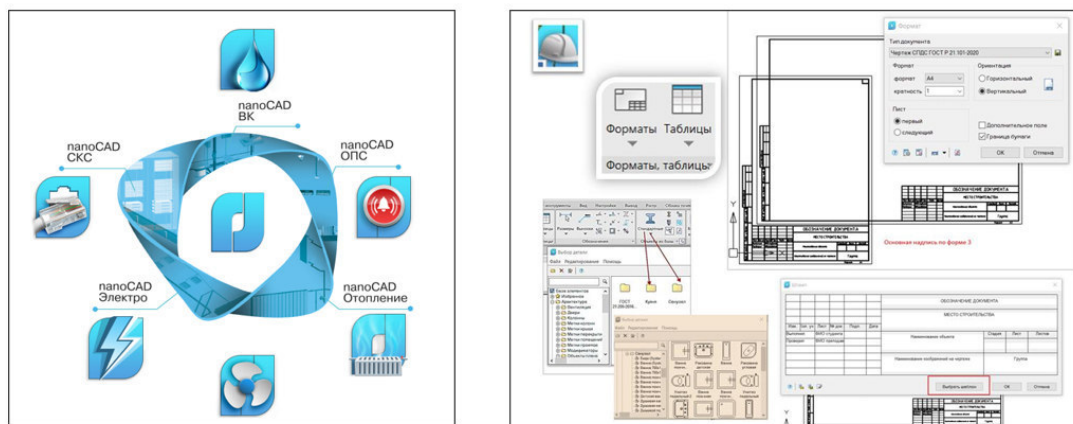


Рисунок 2. Структура комплекса nanoCAD Инженерный BIM и СПДС

Графическая визуализация поэтажных планов является частью цифровой модели здания, служит связующим элементом между концепцией проекта, его реализацией и последующим использованием здания в эксплуатационный период. Выполнение поэтажных планов в графических редакторах имеет свои особенности. В NanoCAD, для выполнения строительных чертежей разработан лицензируемый встроенный модуль СПДС. Следует отметить, что модуль СПДС представляет пользователям эффективные инструменты, позволяющие выполнять эту рутинную работу за очень небольшой промежуток времени. Работа в редакторе NanoCAD предполагает соблюдение определенной последовательности действий при выполнении строительной документации. Графический редактор позволяет создавать трехмерные модели, включая модели строительных объектов. В настоящее время чертежи часто выполняются с использованием различных рабочих пространств редактора. Наиболее распространенный и простой способ – черчение непосредственно в пространстве модели, где размеры элементов соответствуют их реальным размерам. В этом случае необходимо масштабировать элементы оформления, такие как линии, размерные стили и рамки с основной надписью.

Российская BIM-система Renga также востребована в образовательной среде для интегрированного проектирования. ПО Renga обладает необходимой функциональностью, интуитивно-понятным интерфейсом и доступной стоимостью. Документация, создаваемая в программе, соответствует используемой в России нормативно-технической документации. Информационная модель объекта строительства используется на всем его жизненном цикле. Благодаря Renga Collaboration Server, входящему в состав системы, все участники проектной группы могут совместно работать над единым проектом в режиме реального времени. Визуальная цифровая информационная модель, созданная в Renga, может быть использована на протяжении всего жизненного цикла объекта, включая этапы проектирования, прохождения экспертизы, строительства и эксплуатации.

В настоящее время в связи с появлением на рынке информационных услуг большого количества программного обеспечения становится актуальной задача обмена данными между различными программами. Наиболее распространенным становится формат данных IFC. IFC (Industry Foundation Classes) — открытый формат файлов, не зависящий от производителя и обеспечивающий совместимость отдельных приложений в строительной отрасли. Это схема или спецификация, которая предоставляет общую информацию для сбора всей BIM-информации на протяжении всего жизненного цикла здания. Основные характеристики формата IFC – совместимость с практически всеми известными САПР-программами; структуризация данных и стандартизация требований. Объекты принадлежат к различным классам и имеют четкую иерархию; формат IFC позволяет создавать индивидуальные копии BIM-моделей, отражающие её состояние на определённый момент времени. Эти копии (контрольные точки) могут быть привязаны к конкретным пользователям, что упрощает отслеживание изменений и ответственности, анализ данных. Формат IFC при выгрузке формирует файлы журнала, в которых скрыто много полезной информации; осуществляет контроль потоков информации. Формат позволяет контролировать информацию, время и сроки строительства, материальные средства. IFC – широко распространенный формат данных, совместимый с большинством популярных САПР-систем. Это позволяет обмениваться данными при работе в разных программах.

Интеграция данных в формат IFC позволяет производить обмен информацией между программами разных разработчиков, поддерживающими технологию информационного моделирования зданий. Процессы генерации 2D-чертежей из 3D-моделей широко используются в практике проектирования для подготовки документации к производству или реализации проектов. Современные САД-программы содержат встроенные инструменты, которые позволяют строить 3D-модели и автоматически генерировать плоские 2D-чертежи, связанные с 3D-моделью. Автоматизированные процессы снижают риски ошибок при создании изображений, а также 2D-чертежи, связанные с 3D-моделью, облегчают редактирование и изменение документации.

Бумажные оригиналы строительной документации с помощью цифровой фотографии и сканирования могут быть легко и быстро переведены в электронные форматы. Задача преобразования растровых планов зданий в векторные форматы на сегодняшний день остается актуальной и востребованной при подготовке графической части технических планов зданий, подготовке

технических паспортов для объектов недвижимости различного назначения. Этот процесс является также важным этапом в архитектурном проектировании, строительстве и управлении объектами недвижимости. Преобразование растровых планов зданий в векторные форматы является востребованным для повышения эффективности работы с инвентаризационной документацией, автоматизации процессов и обеспечения высокого качества проектных и эксплуатационных документов и изображений. Растровые планы обычно получают путём сканирования бумажных чертежей или фотографий, что ограничивает их дальнейшее использование и редактирование.

Процесс конвертации изображений, основанных на пикселях, в векторные графические форматы, использующие математические описания линий, кривых и фигур, требует, как правило, проведения дополнительных работ, таких как, обрезка, очистка и коррекция исходного растрового изображения (например, удаление шума, повышение контрастности) для лучшей точности распознавания. Очевидные преимущества векторных изображений – масштабируемость, редактируемость, совместимость и интеграция, упрощение обмена информацией. Векторные файлы занимают меньше места и лучше подходят для долгосрочного хранения и документации. При преобразовании растровых изображений в векторные форматы помимо автоматических настроек распознавания изображений, таких как запуск функции автоматического векторизатора, которая анализирует контуры, цвета и границы на изображении, возможна ручная корректировка для повышения точности и устранения ошибок.

Эксплуатационный период жизненного цикла объекта недвижимости, как правило, сопровождается изменениями. Изменения объекта недвижимости без документального отражения приводят к недостоверности и устареванию технической документации. Цифровая модель, дополненная актуальной информацией о текущем состоянии, выходит на новый качественный уровень визуализации путем создания цифрового двойника объекта. Актуальность исследования возможностей применения цифрового двойника (англ. digital twin) привлекает всё больше внимания, так как даёт возможность создавать интегрированную цифровую копию физического здания, которая собирает данные в реальном времени и моделирует его поведение для мониторинга, анализа и управления. Современные методы наблюдений позволяют дать объективную и точную оценку состояния здания на различных стадиях его эксплуатационного периода. Разработка и внедрение мониторинговых систем инженерных конструкций повышает эффективность использования информации. Общий вид системы мониторинга здания [5] представлен на рисунке 3.

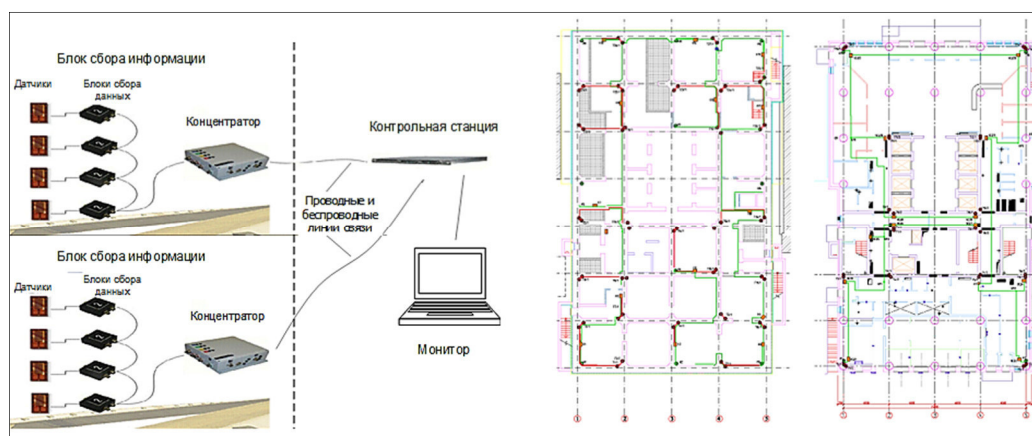


Рисунок 3. Система поэтажной расстановки датчиков

Геодезические наблюдения имеют критически важное значение для здания и состоят в периодическом измерении высотных положений нивелировочных марок относительно реперов и в сопоставлении результатов измерений, проводимых в различные периоды времени. Обоснованием для геодезического мониторинга являются нормативные документы [9]. Одним из эффективных и современных методов создания цифрового двойника объекта на определенный период времени является лазерное сканирование. При этом облако точек передает информацию о точной геометрии

здания (фасады, стены, перекрытия, колонны, лестницы, инженерные сети), происходит обнаружение деформаций, трещин, участков износа и неплотностей. В результате обработки данных реализуется возможность as-built BIM-модели и извлечения чертежей (планы, разрезы, фасады) и 3D-модели для анализа. Состояние здания на момент обследования сохраняется в цифровом архиве. На кафедре ГиИН СГУГиТ в настоящее время происходит активное внедрение технологий лазерного сканирования в учебный процесс. Разработана методика фасадной съемки на базе современных цифровых тахеометров Trimble, Nikon, Leica и других.

Одним из источников формирования технической документации для обновления цифрового двойника объекта выступают результаты проведения технической инвентаризации. К обновляемой информации в период эксплуатации здания относятся сведения о перепланировках и переоборудовании помещений, стоимость и перечень услуг по управлению имуществом в многоквартирных домах, информация о жилищном фонде и другие сведения. Предполагается перевод эксплуатационной документации в электронный вид на портал ГИС ЖКХ. Но в настоящее время наблюдается ограниченное применение современных российских геоинформационных технологий для обработки пространственных данных и, как следствие, существуют различия в подходах управления пространственными данными, а также несогласованность при их учете, использовании и управлении. В связи с этим предлагается добавить, помимо технической документации, и картографическую подложку для отображения данных на ней, чтобы портал ГИС ЖКХ стал полноценным геоинформационным сервисом.

В результате рассмотрения основных аспектов визуализации технической документации на объекты недвижимости и плюсов применения геоинформационных систем в цифровой документооборот можно сделать вывод о том, что такая интеграция с учетом программ генеративного искусственного интеллекта является перспективным направлением. Внедрение ГИС, карт и интерактивных панелей в управление технической документацией недвижимости служит для наглядного представления данных об объектах недвижимости, их состояния и связей, упрощает восприятие и анализ информации, может стать важным шагом в совершенствовании процессов управления недвижимостью, поможет повысить эффективность работы с данными, улучшит принятие решений и упростит доступ к необходимой информации для заинтересованных сторон.

### **Обсуждение результатов**

Современные технологии визуализации в формировании строительной документации предоставляют возможность создания единой информационной среды, где вся документация интегрируется и синхронизируется в реальном времени. Такие решения, как информационное моделирование зданий (ТИМ/BIM), облачные хранилища, автоматизированные системы расчетов и электронные платформы для согласования документов, помогают значительно сократить время на подготовку, облегчить выполнение документации по строительству, снизить вероятность ошибок. Технологии визуализации играют ключевую роль в создании и управлении строительной документацией, повышая точность, эффективность и качество проектных решений.

Наиболее актуальными и перспективными технологиями, используемыми в строительстве, являются:

- блокчейн и цифровые двойники;
- 3D – 6D-моделирование (BIM – Building Information Modeling);
- виртуальная и дополненная реальность (VR и AR);
- фотограмметрия и лазерное сканирование (LiDAR);
- облачные платформы и совместная работа;
- искусственный интеллект и генеративный дизайн в визуализации строительных проектов;
- интерактивные презентации и визуализации.

Эти технологии позволяют не только ускорить процесс формирования строительной документации, но и повысить ее качество, снизить риски ошибок и обеспечить более эффективное управление проектами на всех этапах строительства. Несмотря на существующие вызовы, такие как необходимость в специализированном программном обеспечении и обучение персонала, преимущества применения 3D-моделей значительно перевешивают недостатки. Стремительное



развитие технологий искусственного интеллекта откроет новые возможности для их применения в различных областях, в том числе и 3D-кадастре.

### Заключение

Настоящее исследование подтвердило актуальность применения современных технологий визуализации в процессе создания и управления строительной документацией различного назначения. Анализ существующих методик и инструментов позволил сформировать схему выбора способов визуализации на разных этапах оформления документации. Рассмотрение вопросов интеграции современных технологий в процессы создания, хранения, обмена и использования данных продемонстрировало потенциал BIM/ТИМ-моделирования, виртуальной и дополненной реальности и других методов для повышения эффективности, точности и качества строительных работ. Расширение применения BIM/ТИМ-моделирования, VR/AR и связанных визуализационных инструментов в образовательной деятельности и реальном секторе повышает уровень подготовки специалистов к цифровой трансформации, соответствует требованиям стандартов и распоряжению правительственных документов [13]. Также можно отметить, что необходимо развивать практики подготовки кадров, адаптировать учебные программы под современные методики визуализации и продолжать исследовательскую работу по взаимодействию BIM, GIS и визуализационных платформ, включая оценку пользовательского опыта и производительности систем. Целесообразно сосредоточиться на углубленном анализе совместимости стандартов данных и развитии методик оценки эффективности визуализационных инструментов в рамках строительных процессов и образования.

### Список литературы

1. Распоряжение Правительства РФ от 21 декабря 2021 г. № 3759-р «Об утверждении стратегического направления в области цифровой трансформации науки и высшего образования». Доступ из справ.-правовой системы «Консультант Плюс».
2. Об информации, информационных технологиях и о защите информации: Федеральный закон от 27.07.2006 № 149-ФЗ (ред. от 14.07.2022). Доступ из справ.-правовой системы «Консультант Плюс».
3. Стандарты графических символов, используемые для технической документации на продукцию. <https://www.muayene.com/ru/standartlar/genellemeler-terminoloji-standardizasyon-dokumantasyon-standartlari/grafik-semoller-standartlar/teknik-urun-dokumantasyonlari-icin-kullanilan-grafik-semoller-standartlari>
4. Национальный стандарт Российской Федерации ГОСТ Р 56907-2016. Бережливое производство. Доступ из справ.-правовой системы «Консультант Плюс».
5. Единая система конструкторской документации. Общие требования к ГОСТ 2.301-2.321. Доступ из справ.-правовой системы «Консультант Плюс».
6. Единая система конструкторской документации. Общие требования к текстовым документам. ГОСТ 2.105-2019. Доступ из справ.-правовой системы «Консультант Плюс».
7. Интеграция геоинформационных систем в управление технической документацией недвижимости / Л. А. Максименко, Е. С. Агеев, А. С. Лаптева, И. М. Макарихина // Наука и социум : материалы научно-практических конференций АНО ДПО "СИППИИС", Новосибирск, 15 мая 2025 года. Новосибирск: Сибирский институт практической психологии, педагогики и социальной работы, 2025. С. 67-73. EDN IUKUJS.
8. Каталог визуализации данных. Режим доступа: <https://datavizcatalogue.com/RU/>.
9. Система проектной документации для строительства. Основные требования к проектной и рабочей документации. ГОСТ Р 21.101-2020. Дата актуализации: 01.01.2021. Доступ из справ.-правовой системы «Консультант Плюс».
10. Федотова В. С. Профессиональная подготовка будущих специалистов в области кадастрового учета и регистрации прав к графической визуализации объектов недвижимости // Педагогика. Вопросы теории и практики.
11. Максименко Л. А. Сбор и обработка кадастровой информации в сфере управления недвижимым имуществом. DOI: 10.33764/2411-1759-2024-29-1-118-126 // Вестник СГУГиТ (Сибирского государственного университета геосистем и технологий). 2024. Т. 29, № 1. С. 118–126.
12. Приказ Минцифры России от 18.01.2023 № 21 «Об утверждении Методических рекомендаций по переходу на использование российского программного обеспечения, в том числе на значимых объектах критической информационной инфраструктуры Российской Федерации, и о реализации мер, направленных на ускоренный переход органов государственной власти и организаций на использование российского программного обеспечения в Российской Федерации».

## Моделирование социальных сетей: визуализация и расчет метрик

М. А. Решетников, Е. Б. Замятин

Пермский государственный национальный исследовательский университет (ПГНИУ), Пермь, Россия

**Аннотация.** В статье рассматриваются методы визуализации и алгоритмы вычисления метрик в ходе моделирования социальных сетей. Реализованы алгоритмы поиска наиболее влиятельных узлов социальной сети. Проведено моделирование процесса распространения информации на примере модели SIR: с помощью вычисленных метрик центральности выявлены ключевые узлы наиболее влиятельных пользователей и проведено исследование, которое определяет влияние изменения количества этих узлов на скорость распространения сообщений в сети. Результаты моделирования иллюстрируются диаграммами и графами, полученными с помощью системы моделирования Triad.Net.

**Ключевые слова:** социальные медиа, визуализация графов, метрики центральности, моделирование распространения информации.

## Simulation of social media: visualization and metrics calculation

M. A. Reshetnikov, E. B. Zamyatina

Perm State National Research University, Perm, Russia

**Abstract.** The paper discusses visualization methods and algorithms for calculating metrics during the simulation of social networks. Algorithms for searching for the most influential nodes of a social network are implemented. The process of information dissemination is simulated using the SIR model as an example: using the calculated centrality metrics, key nodes of the most influential users are identified, and a study is conducted that determines the impact of changing the number of these nodes on the speed of message dissemination in the network. The simulation results are illustrated by diagrams and graphs obtained using the Triad.Net simulation system.

**Keywords:** social media, graph visualization, centrality metrics, information diffusion modeling.

### Введение

Социальные сети прочно вошли в повседневную жизнь. Современный человек потребляет огромное количество информации из социальных сетей, таких как Twitter, Вконтакте, Facebook, Одноклассники и т.д. В интернете они являются средством быстрого распространения информации, позволяют людям объединяться в группы по интересам и могут влиять на общественное мнение.

На основе данных из социальных сетей можно понять, кого из политиков поддерживает народ, одобряют ли люди действия и позицию властей. Проанализировав структуру социальных сетей, можно выявить, как необходимо распространять информацию, чтобы она имела наибольший охват аудитории. Таким образом, социальные сети можно успешно использовать и для проведения маркетинговых операций. Благодаря анализу контента, который потребляет пользователь, можно сказать, какая реклама будет ему интересна, а какая нет [1–3].

Социальные сети также являются средством массовой информации, где каждый пользователь может распространять новости от своего имени или имени сообщества, из-за чего возникает необходимость дополнительной проверки новостей на их достоверность [4].

По сути дела, исследование социальных сетей и поведение пользователей в сети является актуальным, поскольку необходимо решить задачу скорейшего распространения нужной информации (для маркетологов, например). Важной является и обратная задача: выявление и сдерживание распространения нежелательной информации [4–6].

При решении этих задач используют методы Social Mining [7]. Основными задачами Social Network Mining являются: (а) анализ информационных потоков: выявление наиболее «весомых» узлов в социальной сети, а именно людей, через которых проходит большой поток информации и чье мнение оказывает существенное влияние на другие узлы (применяется в таких сферах, как интернет-маркетинг, реклама и безопасность пользователей); (б) персонификация предложений: фильтрация контента, потребляемого пользователем, в соответствии с интересами этого пользователя. Например, показ рекламы автомобилей человеку, который собирается покупать машину в ближайшее время, повышает удовлетворенность пользователя от использования ресурса, а реклама становится более эффективной; (в) поиск аномалий, компьютерных ботов и мошенников: необходимо вовремя выявлять пользователей, чьи действия негативно влияют на социальную сеть.

При изучении социальных сетей зачастую не учитывают временные характеристики объектов исследования в социальных сетях. В настоящей статье рассматривается динамическое моделирование социальных сетей, учитываются моменты выполнения тех или иных событий в социальной сети, их последовательность. Моделирование выполняется с использованием программной системы Triad.Net [8]. Особенности реализации системы имитационного моделирования Triad.Net рассматриваются ниже.

Очень часто при изучении социальных сетей используют виртуальные сети, синтетические данные, модели случайных графов [9], модели процессов распространения информации. Модели распространения также основываются на теории графов: пользователи – узлы (вершины) графа, а связи между ними – ребра графа. Существующие модели распространения информации разделяются на следующие категории: независимые (а) каскадные модели, (б) линейные пороговые модели, (в) модели эпидемии. При этом последние различаются по количеству возможных состояний: модели эпидемии, описывающие процесс распространения на основе трех состояний узлов (SIR), эпидемиологические модели, включающие четыре состояния, а также модели распространения эпидемии, подразумевающие наличие множества состояний.

Наряду с виртуальными социальными сетями исследуют характеристики пользователей и каналы связей между ними в реальных социальных сетях [10, 11].

В моделях социальных медиа часто требуется выявить наиболее влиятельных участников и проследить пути распространения информации в социальной сети. Для этого используют алгоритмы, определяющие метрики центральности, в частности метрику PageRank [12], и биоинспирированные алгоритмы для поиска лидеров мнений. В настоящем исследовании были выбраны биоинспирированные алгоритмы Firefly и Grey Wolf Optimization (GWO) [13, 14].

Одновременно важна качественная визуализация структуры сети и результатов моделирования, поскольку наглядное представление графа упрощает анализ сложных связей между узлами.

Особенностью работы является создание инструмента, позволяющего не только отображать сеть, но и активно вмешиваться в её структуру для мгновенной оценки последствий на моделях распространения (SIR). Для этого используется онтологический подход к хранению данных, обеспечивающий семантическую согласованность и связь с предметной областью и подход для выявления наиболее влиятельных участников через биоинспирированные алгоритмы.

В статье представлены расширения модуля TriadNSim (модуль, предназначенный для исследования социальных медиа в системе моделирования Triad.Net), интегрирующего методы поиска лидеров мнений, моделирования распространения информации и визуализации на графе социальной сети.

### Постановка задачи

Цель исследования – разработка прикладного инструмента для вычисления метрик центральности и визуализации результатов моделирования социальных сетей. В рамках этой цели ставились следующие задачи:

- исследовать методы определения влиятельных узлов в сети, включая классические метрики центральности и биоинспирированные алгоритмы оптимизации;
- реализовать алгоритмы PageRank, Firefly, GWO для оценки влияния участников;
- выполнить визуализацию социальной сети в виде графа, где параметры узлов (цвет, размер, шрифт) отображают вычисленные метрики;
- разработать пользовательский интерфейс для управления визуализацией, выбором алгоритма и анализа результатов;
- обеспечить доступ к онтологии с данными о пользователях и сообществах с помощью возможностей Triad.Net;
- провести экспериментальное моделирование информационных процессов (SIR-модель распространения новости) в социальной сети из 20 пользователей и проанализировать влияние выявленных лидеров на время распространения информации.

### Теория

Задачи поиска ключевых узлов и оптимизации информационных потоков важны для понимания динамики социальных процессов, таких как вирусное распространение новостей или воздействие на общественное мнение. Для их решения используются методы анализа графов, позволяющие

количественно оценивать значимость узлов и устойчивость сети. Классические метрики центральности служат фундаментальным инструментом для ранжирования узлов по их структурной важности. Эти показатели позволяют обнаружить узлы с большим потенциалом распространения информации и понять роль каждого участника в сети.

Наряду с метриками центральности применяются биоинспирированные алгоритмы оптимизации, способные выявлять оптимальные стратегии распространения и выбор лидирующих узлов. Алгоритм светлячков (Firefly) и алгоритм серых волков (GWO) имитируют поведение природных систем поиска оптимальных решений и могут быть использованы для подбора набора узлов, обеспечивающих максимальный охват сети. Сочетание метрик центральности и метаэвристических алгоритмов обеспечивает мощный арсенал методов для анализа социальных сетей и моделирования информационных процессов. В последующих подпунктах будут рассмотрены теоретические основы некоторых упомянутых метрик и биоинспирированных алгоритмов.

- *Алгоритм PageRank*

PageRank — алгоритм ранжирования узлов графа, основанный на вероятностной модели перехода между вершинами. Узел считается тем более значимым, чем больше на него ссылаются другие значимые узлы. В контексте социальной сети это позволяет выявить участников, на которых «ориентируются» другие, что может означать авторитет или популярность узла.

- *Алгоритм светлячков (Firefly Algorithm)*

Метаэвристика – основана на поведении светлячков, у которых яркость соответствует качеству решения. В социальной сети каждый узел интерпретируется как светлячок, чья яркость зависит от метрик центральности. Узлы перемещаются (влияют друг на друга) в сторону более ярких, тем самым алгоритм итеративно усиливает влияние лидеров. На рисунке 1 представлена блок-схема алгоритма светлячков.

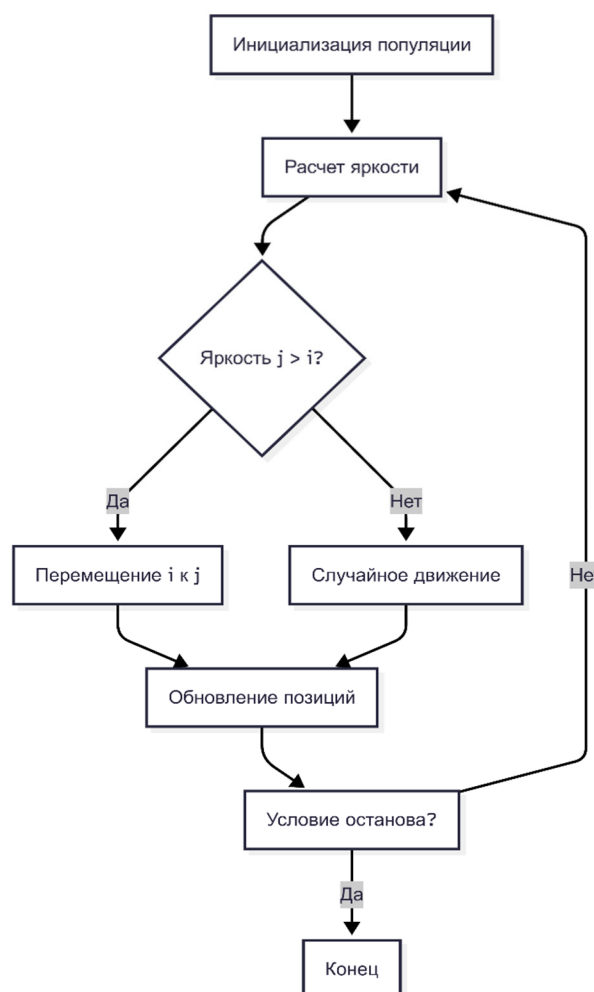


Рисунок 1. Блок-схема алгоритма светлячков

- *Алгоритм серых волков (GWO)*

Метаэвристический алгоритм моделирует иерархическую охоту стаи волков, блок-схема которого представлена на рисунке 2. Узлы-волки стремятся к «добыче» (оптимальному решению), следуя за тремя главными (альфа, бета, дельта). GWO показал лучшую устойчивость в динамических сетях по сравнению с Firefly.

- *SIR-модель распространения информации*

Модель SIR (Susceptible–Infected–Recovered) адаптирована под задачу информационного распространения. Агенты (узлы) могут быть в одном из трёх состояний:

- 1) S — уязвимы к получению информации,
- 2) I — обладают информацией и могут передавать её другим,
- 3) R — вышли из процесса распространения.

Передача информации осуществляется с вероятностью  $\beta$ , переход в состояние R — с вероятностью  $\gamma$ . Эта модель позволяет эмулировать динамику распространения новостей, слухов или вирусного контента в социальных сетях.

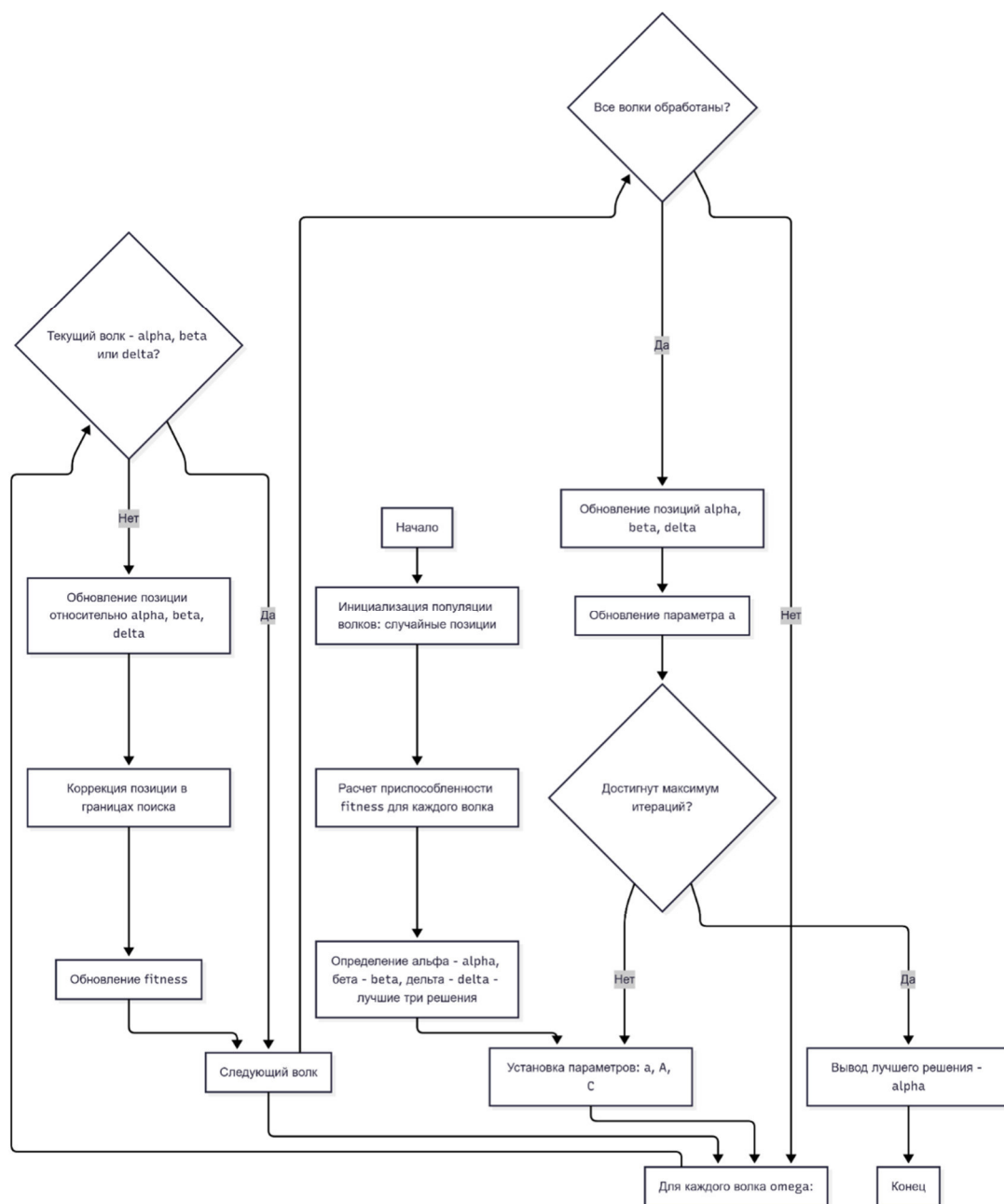


Рисунок 2. Блок-схема алгоритма серых волков

### • Система моделирования Triad.Net

Triad.Net — это программная среда имитационного моделирования, предназначенная для построения, анализа и визуализации сложных динамических систем. Она широко применяется в задачах, связанных с анализом поведения программных агентов, алгоритмов маршрутизации в компьютерных сетях, информационными процессами в социальных сетях и т.д.

Для загрузки данных в онтологию используется программный агент, реализующий сбор информации из социальной сети «ВКонтакте» в реальном времени. Агент отслеживает действия пользователей и связи между сущностями, сохраняя их в OWL-онтологию через RabbitMQ-очереди. Онтология отражает актуальное состояние сети, позволяет Triad.Net визуализировать не только топологию, но и семантические паттерны, пример ее структуры представлен на рисунке 3. Онтология включает в себя:

- динамические атрибуты пользователей (пол, возраст, интересы);
- историю действий (лента новостей, просмотры постов);
- граф связей с весами (частота взаимодействий).

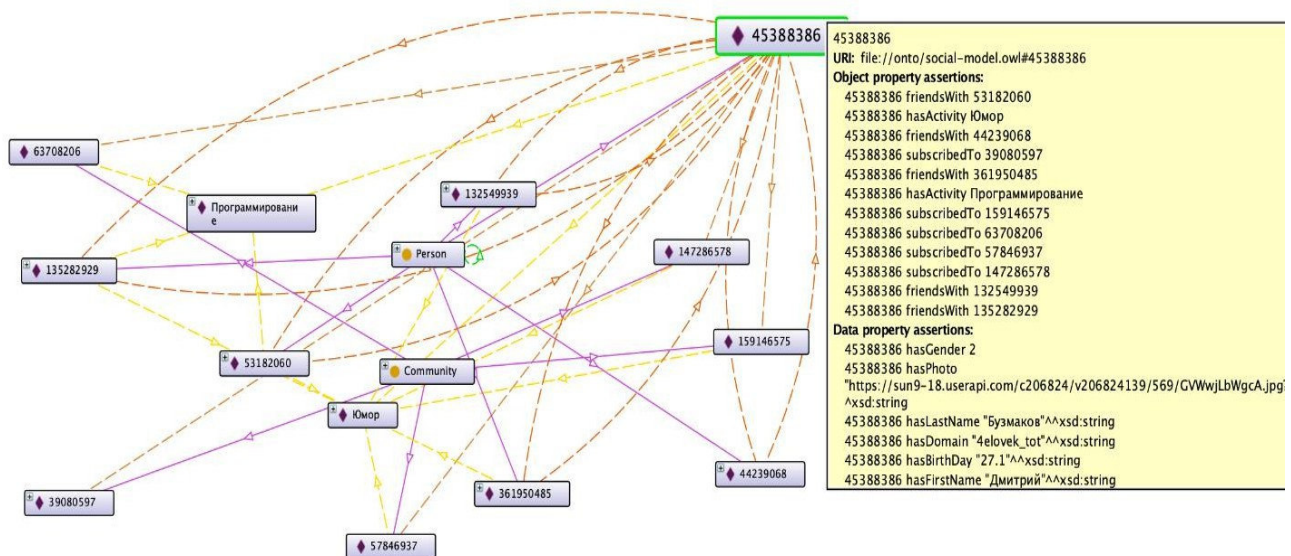


Рисунок 3. Онтология пользователя социальной сети

В Triad.Net принято трехуровневое представление имитационной модели:  $M = (STR, ROUT, MES)$ , где STR – слой структур, ROUT – слой рутин, MES – слой сообщений.

Слой структур представляет собой совокупность объектов, взаимодействующих друг с другом посредством посылки сообщений. Основа представления слоя структур – графы. Дуги графа определяют связи между объектами. Имитационная модель имеет иерархическое представление. Отдельные объекты, представляющие вершины графа, могут быть расшифрованы подграфом более низкого уровня и т.д. Для представления социальной сети вершины графа – это пользователи, сообщества; дуги – связи между пользователями, сообществами и т.д.

Для описания структуры имитационной модели используют графовые константы. Графовая константа – это граф некоторого стандартного вида. Например, кольцо, путь, дерево, полный граф и т.д. Количество вершин графа задают с помощью параметров. В Triad.Net реализованы несколько моделей случайных графов, такие как случайный граф Эрдеша–Реньи, граф Боллобаши–Риордана, модель П. Бакли и Д. Остгуса и модель копирования. Известно, что случайные графы позволяют исследовать виртуальные социальные сети.

Слой структур позволяет выполнять операции над графами. Для исследования характеристик построенного графа в слое структур реализованы процедуры, которые позволяют определять количество вершин в графе, диаметр, степень вершины, диаметр графа и т.д. Кроме того, реализованы процедуры для исследования социальных сетей.



Объекты структуры действуют по определённом сценарию, который описывают с помощью рутины. Рутинa представляет собой последовательность событий  $e_i$ , планирующих друг друга. Выполнение события сопровождается изменением состояния объекта. Состояние объекта определяется значениями переменных рутины. Таким образом, система имитации является событийно ориентированной. Совокупность рутин определяет слой рутин ROUT.

Слой сообщений (MES) предназначен для описания сообщений сложной структуры. Система моделирования Triad реализована таким образом, что пользователю необязательно описывать все слои. Так, если возникает необходимость в исследовании структурных особенностей модели, то можно описать только слой структур.

Одним из ключевых преимуществ Triad.Net является наличие графического интерфейса, позволяющего исследователю формировать модель наглядно — в виде графа, где узлы и связи отображаются с учётом заданных параметров. Пользователь может интерактивно задавать правила поведения агентов, настраивать свойства элементов системы и наблюдать за динамикой процессов в реальном времени.

Особенно стоит уделить внимание поддержке интеграции с онтологиями — формальными описаниями предметной области, содержащими информацию о сущностях (например, пользователях, сообществах, событиях) и их отношениях. Использование онтологий позволяет автоматически загружать и структурировать данные из реальной социальной сети [15]. Онтологическая модель обеспечивает согласованное представление информации и упрощает анализ семантических связей в социальной сети.

Таким образом, Triad.Net представляет собой мощный инструмент для комплексного моделирования, анализа и визуализации социальных сетей, обеспечивая удобство работы и гибкость при настройке модели. Интерфейс Triad.Net представлен на рисунке 4.

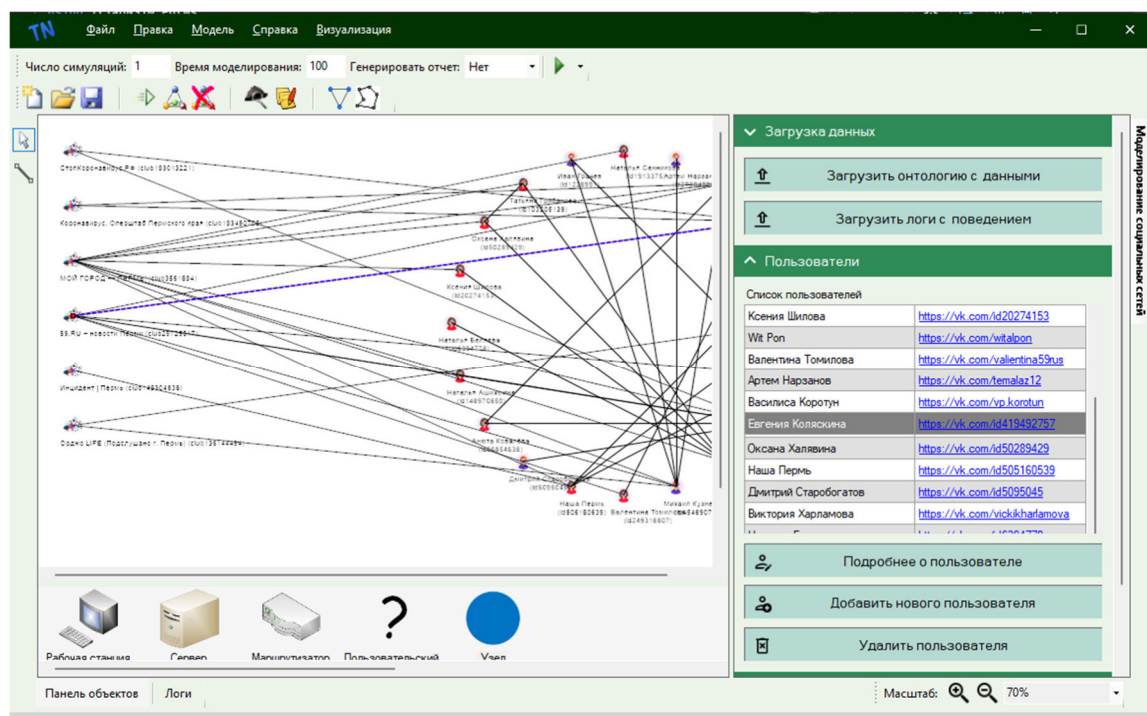


Рисунок 4. Интерфейс Triad.Net

На основе анализа работ [16, 17] был реализован модуль графического представления социальных сетей и визуализации результатов данных по распространению информации в сети. Интерфейс модуля для работы с социальными сетями представлен на рисунке 5.



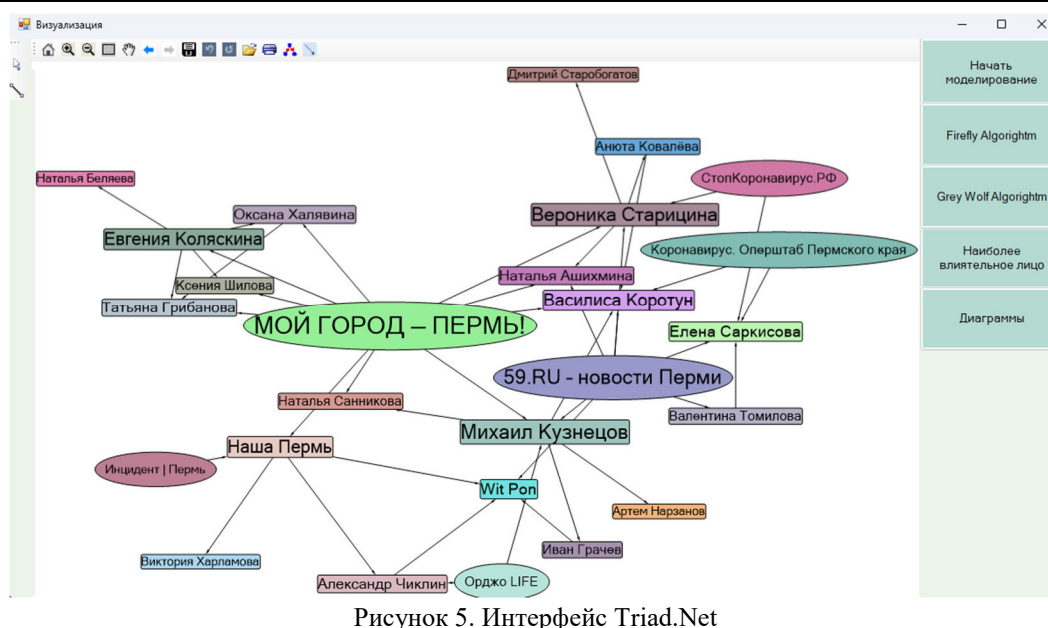


Рисунок 5. Интерфейс Triad.Net

### Результаты экспериментов

Для практической оценки методов рассмотрен следующий эксперимент. Модельная социальная сеть содержит 20 пользователей, среди которых заданы группы с разными связями (участники групп «Мой город – Пермь!», «59.RU – новости Перми» и др.). На рисунке 5 показано графическое представление такой сети.

С помощью SIR-модели эмулируется однократная вспышка новости. Сначала определяются три наиболее влиятельных узла сети средствами алгоритмов: наибольшая метрика выдается сообществу «МОЙ ГОРОД – ПЕРМЬ!», а также узлам «59.RU – новости Перми» и «Михаил Кузнецов» (по совокупности мер). Затем два самых активных лидера (за исключением стартового «Мой город – Пермь!», который изначально является источником новости) искусственно удаляются из сети. После этого модель повторно запускается на модифицированной сети без этих узлов. На рисунках 6, 7 показана сравнительная диаграмма времени распространения информации до и после удаления лидеров. Видно, что время затухания новости заметно сократилось, а часть узлов вовсе не получила информацию из-за разрыва ключевых связей. Это свидетельствует о существенном влиянии выявленных лидеров на динамику распространения.

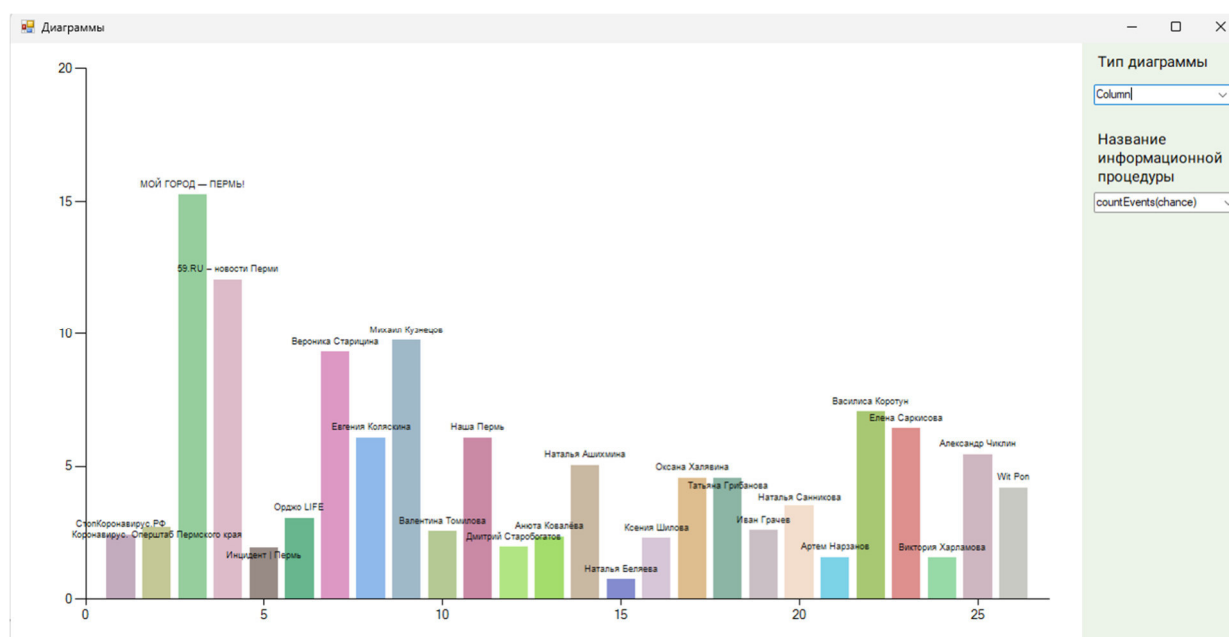


Рисунок 6. Столбчатая диаграмма времени распространения информации

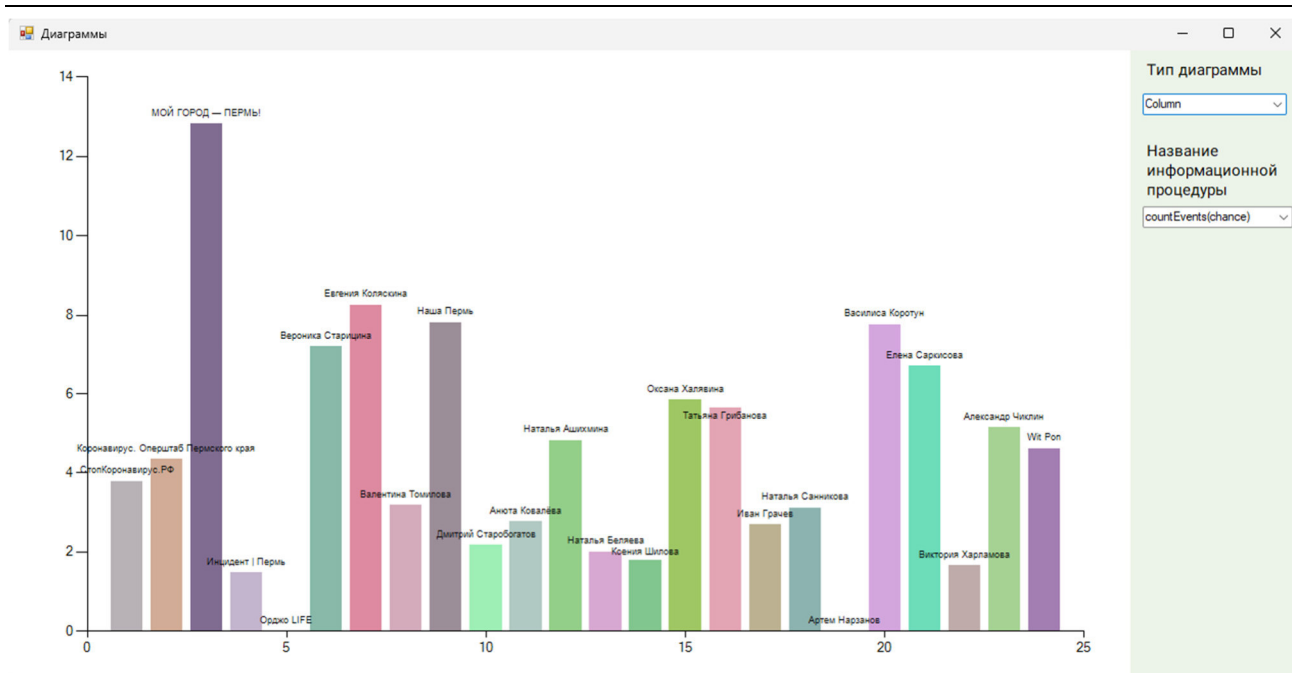


Рисунок 7. Столбчатая диаграмма времени распространения информации после удаления влиятельных пользователей

### Обсуждение результатов

Результаты эксперимента подтверждают эффективность реализованных методов. После удаления наиболее влиятельных узлов ключевые связи в сети нарушаются, что приводит к замедлению и фрагментации информационного потока. Выявление таких узлов позволяет моделировать сценарии управления распространением, например, влияние на маркетинговые кампании. Отметим также, что алгоритмы Firefly и GWO в рамках тестовой сети дали схожий набор лидеров, но GWO обеспечивает более стабильную сходимость при увеличении размеров сети. Наглядное отображение социальной сети и диаграмм существенно упрощает анализ: цветовое кодирование влиятельных узлов и диаграммы времени распространения позволяют быстро оценить ключевые результаты моделирования.

### Заключение

Были разработаны прикладные средства для вычисления метрик центральности и визуализации результатов моделирования социальных сетей. В частности, в платформу TriadNSim добавлены инструменты:

- Редактирование сети – добавление/удаление вершин и рёбер, интеграция с онтологическими данными.
- Методы поиска лидеров – внедрены алгоритмы Firefly и GWO для нахождения наиболее влиятельных узлов.
- Средства визуализации – автоматическая раскладка графа и построение диаграмм (включая цветовые метки узлов и графики времени распространения).

Реализованная система предоставляет уникальные возможности для контролируемых экспериментов:

- интерактивное вмешательство в структуру сети (удаление/изоляция узлов) с синхронизированной визуализацией последствий в реальном времени;
- глубокая интеграция с онтологиями, обогащающая узлы атрибутами (тип пользователя, тематика сообщества) для контекстной интерпретации;
- адаптивная визуализация, где цвет и размер узлов кодируют метрики влияния (PageRank, Firefly, GWO), а диаграммы отображают динамику распространения информации;
- цикл анализа «метрики → вмешательство → визуализация → интерпретация», недоступный в статических инструментах.

Применение этих инструментов позволяет проводить быстрый анализ структуры сети и динамики информационных процессов. Например, можно «отключить» ключевых пользователей в эксперименте и наблюдать последствия для распространения новости, что важно для оптимизации маркетинговых стратегий или исследования моделирования пандемий в социальных сетях. Разработанный модуль визуализации объединяет вычисление метрик и наглядное представление данных, что повышает эффективность исследования социальных сетей.

Проведённые исследования показали, что поставленные цели достигнуты. В дальнейшем планируется расширить функциональные возможности разрабатываемого модуля: добавить новые алгоритмы анализа, улучшить интерактивность визуализации и исследовать применение метода в реальных больших сетях.

### Список литературы

1. Грибанова Е. Б., Логвин И. Н., Ширенков И. В. Алгоритм оценки маркетинговых мероприятий онлайн-социальной сети "ВКонтакте" на основе каскадной модели распространения информации. Доклады Томского государственного университета систем управления и радиоэлектроники. 2018. Т. 21, вып. 3. С. 69-74.
2. Методика выявления каналов распространения информации в социальных сетях / А. А. Проноза, Л. А. Виткова, А. А. Чечулин, И. В. Котенко, Д. В. Сахаров // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2018. Т. 14, вып. 4. С. 362–377.
3. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Модели влияния в социальных сетях // УБС. 2009. № 27. С. 205–281.
4. Абрамов М.В., Тулупьев А.Л., Сулейманов А.А. Задачи анализа защищенности пользователей от соционинженерных атак: построение социального графа по сведениям из социальных сетей // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18, № 2. С. 313–321.
5. Дмитриев И.П., Замятина Е.Б. Опыт моделирования действий по предотвращению распространению запрещенной информации в социальной сети // Прикладная информатика. 2019. Т. 14, № 1 (79). С. 82–95.
6. Zimin I., Zamyatina E. Strategies of the Social Network Immunization: An Experience of an Investigation by Simulation Tools // 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT). IEEE, 2020. P. 1–6.
7. Давыдов А. А. Системная социология: Social Networks Mining. М.: ИС РАН, 2009.
8. Mikov A., Zamyatina E., Germanova D. Program Tools for Dynamic Investigation of Social Networks // Supplementary Proceedings of the 5th International Conference on Analysis of Images, Social Networks and Texts (AIST-SUP 2016), Yekaterinburg, Russia, April 7-9, 2016 / ed. by D. I. Ignatov. Vol. 1710. Aachen : CEUR Workshop Proceedings, 2016. P. 224-236.
9. Райгородский А.М. Модели случайных графов и их применение // Труды Московского физико-технического института (национальный исследовательский университет). 2010. С. 130–140.
10. A simulation-based approach to analyze the information diffusion in Microblogging Online Social Network / M. Gatti, A. P. Appel, C. Nogueira dos Santos, C. Z. Pinhanez // Proceedings of Winter Simulation Conf. Piscataway, New Jersey, 2013. P. 1685–1696. Usanin A., Zimin I., Zamyatina E. Study of Strategies for Disseminating Information in Social Networks Using Simulation Tools // Analysis of Images, Social Networks and Texts: 9th International Conference, AIST 2020, Skolkovo, Moscow, Russia, October 15–16, 2020, Revised Selected Papers Vol. 12602. Springer, 2021. P. 303–315.
11. Zamyatina E., Mikov A., Lanin V. Automation of Simulation Steps using Ontological Approach // Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018). Volume 2: KEOD / Ed. by D. Aveiro, J. Dietz, J. Filipe. Portugal: SCITEPRESS – Science and Technology Publications, Lda., 2018. P. 223-230.
12. Нуньес-Иглесиас Х. Элегантный SciPy / в. д. Уолт, Х. Дэшноу. Москва: ДМК Пресс, 2018.
13. Sebin J., P.V Paul. A survey on identification of influential users in social media networks using bio inspired algorithms. // International Conference on Machine Learning and Data Engineering. 2023.
14. Интеллектуальные системы: модели и методы метаэвристической оптимизации: монография / Л. А. Гладков, Ю. А. Кравченко, В. В. Курейчик, С. И. Родзин. Чебоксары, 2024.
15. An Approach to Developing Data Visualization Tools Based on Domain Specific Modeling / A. D. Dzheiranian, I. D. Ermakov, K. A. Proskuryakov, L. N. Lyadova // Scientific Vizualisation. 2024. Vol. 16, No. 4. P. 82–101.
16. Li Y., Chan J. Mixed emotion extraction analysis and visualization of social // Data & Knowledge Engineering. 2013. P. 82–92.
17. Chien-Po L., Chien Y., Yu-His L. Progress and future directions for research on social media // Telematics and Informatics. 2023.

УДК 659.442

DOI: 10.25686/978-5-8158-2474-4-2025-424-433

## Исследование и разработка алгоритма создания инклюзивного сайта в сфере государственных услуг

Т. М. Терещенко, С. Ю. Щур

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

**Аннотация.** Статья посвящена разработке алгоритма для создания инклюзивного веб-сайта в сфере государственных услуг. В работе рассматриваются ключевые моменты, способствовавшие формированию инклюзивного подхода в дизайне, а также проводится анализ существующих алгоритмов создания сайта.

**Ключевые слова:** инклюзивность, веб-дизайн, алгоритм

## Research and development of an algorithm for creating an inclusive website in the field of public services

T. M. Tereshchenko, S. Yu. Shchur

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

**Abstract.** The article is devoted to the development of an algorithm for creating an inclusive website in the field of public services. The work examines the key points that contributed to the formation of an inclusive approach to design, and also analyzes existing algorithms for creating a website.

**Keywords:** inclusivity, web design, algorithm

### Введение

В современном цифровизированном мире толерантного общества применение инструментов инклюзивного дизайна при создании сайтов набирает актуальность, способствуя решению проблемы доступности и удобства использования для всех пользователей, независимо от физических возможностей, возраста или технических навыков.

Литературная база статьи включает в себя научные труды отечественных авторов, таких как Е. Альтовский, И. Мальцев, Н. Инькова, М. Захарова и др. Зарубежные исследования и научные работы в области инклюзивного дизайна начали появляться за десятилетие до отечественных. Существующие исследования и практики создания веб-сайтов акцентируют внимание на реализации с точки зрения программирования. Однако многие из них не интегрируют качественные исследования и инклюзивные решения в актуальные проблемы общества.

Отвечая требованиям жизни в современном толерантном обществе, практическая значимость работы заключается в возможности внедрения инклюзивных решений в условиях неравной доступности интернет-ресурсов для людей с ограниченными возможностями.

### Постановка задачи

Цель данной статьи – проанализировать существующие и разработать собственный алгоритм создания инклюзивного веб-сайта, а также его реализация в сфере государственных услуг.

Поставленная цель определила следующие задачи:

- 1) определить понятие и принципы инклюзивного дизайна;
- 2) проанализировать существующие алгоритмы создания веб-сайтов;
- 3) разработать собственный алгоритм создания инклюзивного сайта;
- 4) реализовать разработанный алгоритм в сфере государственных услуг;
- 5) сформулировать выводы.

Методы исследования включили в себя анализ данных, качественные опросы респондентов, обобщение и структурирование информации. Важным аспектом станет также прототипирование платформы.

### Теория

Инклюзивный дизайн представляет собой концепцию, направленную на создание продуктов и услуг, доступных для всех пользователей, включая людей с ограниченными возможностями [1].

Если раньше обществом и государством было принято, что людям с ограниченными возможностями необходимы специальные учреждения, отдельная среда, то сегодня тенденции демократичного социального государства стремятся приобщить таких людей в обычную повседневную жизнь, не отделяясь от остального общества.

Обратимся к определению понятия инклюзивного дизайна. Инклюзивный дизайн — это подход к проектированию, который учитывает потребности и возможности как можно большего числа людей. Вместо того чтобы ориентироваться на гипотетического пользователя, инклюзивный дизайн заботится о разнообразии людей, создавая интерфейсы, которые могут использоваться всеми. Этот подход не нацелен на определённую группу, а учитывает индивидуальные предпочтения и общие ситуации [2].

Получается, инклюзивный дизайн не должен исключать стандартных пользователей, как и ограниченных. Он не работает на какую-то определённую группу, а ориентируется под различные ситуации и особые предпочтения. Инклюзивный дизайн ставит людей в центр процесса проектирования, как и любой другой дизайн, но обращает внимание на возможные ограничения в первую очередь.

Существует три категории ограничений: постоянные, временные, ситуационные.

В каждой из этих категорий есть несколько нарушений или ограничений: визуальные, слуховые, когнитивные, речевые [3].

Важно понимать разницу между инклюзивностью и универсальностью. Универсальность входит в понятие инклюзивности и устраняет постоянные ограничения или инвалидность, в то время как инклюзивные подходы к проектированию учитывают временные и ситуационные факторы [4].

Универсальность предполагает создание единого паттерна для решения ограничений, не рассматривая адаптацию и специализированность дизайна. Универсальный дизайн (UD) был первоначально определен Центром универсального дизайна при Университете штата Северная Каролина (CUD) в 1997 году как создание продуктов и сред, доступных всем людям, без необходимости специализированного дизайна или адаптации [5].

Алгоритм обеспечивает структурированный подход к разработке сайта, позволяя эффективно управлять процессом, минимизировать ошибки и затраты, а также гарантировать, что конечный продукт будет соответствовать изначальным целям и ожиданиям пользователей.

Алгоритмов создания веб-сайтов огромное множество, каждый автор выделяет тот или иной аспект проектирования исходя из направленности работы. Превалирующее большинство алгоритмов подробно освещает сторону разработки, предназначенную для программистов, когда о дизайнерской части может упоминаться лишь вскользь. Для анализа были подобраны наиболее общие алгоритмы. Например, в совместной работе Н.А. Иньковой, Е.А. Зайцевой, Н.В. Кузьминой и С.Г. Толстых, выделяются следующие этапы [6]:

- 1) планирование;
- 2) реализация;
- 3) тестирование;
- 4) публикация;
- 5) рекламирование;
- 6) сопровождение.

Планирование авторы выделяют как самый важный этап, в который входит определение целей, задач, аудитории проекта. Сюда же относится определение структуры и особенностей оформления. В этап реализации входит работа по созданию сайта, подготовка текстового и графического материала, а также написание кода программистами. В книге тестирование проводится на фокус-группах на работоспособность и удобство пользования интерфейсом. На последних трёх этапах сайт размещается на сервере и продвигается.

В данном алгоритме основной акцент сделан на разработке структуры и функционала сайта, а также реализации с помощью кодирования. Приведём в пример алгоритм из руководства по проектированию сайтов И. П. Мальцева. Автор выделяет такие этапы [7]:

- 1) анализ ситуации – проведение анализа существующих сайтов-примеров для подражания или отстройки от конкурентов;

- 2) функционал сайта – определение цели создания, типа сайта, его вида и компонентов. Включает в себя составление пути пользователя, определение количества страниц и глубины сайта;
- 3) дизайн сайта – цветовая гамма, графическое оформление, создание шаблона-прототипа, затем разработка полного концептуального макета с анимацией и переходами;
- 4) управление сайтом – хостинг, доменное имя, выбор системы управления, перевод дизайн-концепции в код и размещение в браузере, создание обслуживающего сервиса сайта;
- 5) наполнение сайта – составление карты сайта, сбор текстового и графического контента, иконок, написание ключ-слов. Также сюда автор отнёс самопроверку и правки;
- 6) решающий – полная сборка, запуск в открытый доступ и финансирование.

Автор выделил наполнение сайта контент в отдельный этап от создания самого дизайна, что довольно нетипично относительно других алгоритмов. Также отсутствует этап разработки самой идеи, интервью с возможным заказчиком, алгоритм начинается сразу с анализа конкурентов. Приведём в пример статью преподавателя Кубанского государственного университета М. В. Захаровой [8].

В работе автора выделены такие этапы, как:

- 1) структурирование информации (юзабилити). В этот этап входят вопросы по содержанию и структуре сайта, о форме подачи информации;
- 2) веб-дизайн: проработка концепции графического оформления, результатом чего является готовый файл;
- 3) верстка. С применением технологий HTML и CSS дизайн-макет трансформируется в код, который может функционировать в браузере.

В описанном алгоритме этапы достаточно обобщенные, отсутствует подробное описание действий, а также исключен исследовательский этап. Проанализировав несколько популярных алгоритмов, можно понять, что в каждом подчёркивается разная составляющая: дизайн, продвижение и реализация. В работе основной уклон следует сделать на создание дизайн-концепции, её составляющих в контексте разработки инклюзивного веб-сайта.

Был разработан и предложен собственный алгоритм для создания инклюзивного веб-сайта:

1. *Концептуальный этап.* На первом этапе происходит встреча с заказчиком (если он есть), составление брифа, технического задания и обозначение сроков работы. Важно определить концептуальное ядро, что создаётся, зачем и для какой целевой аудитории. Также на первом этапе выбирается тип разрабатываемого сайта: сайт-визитка, интернет-магазин, информационный сайт, игровой, корпоративный, функциональный, сайт-форум, блог и т.д.

2. *Исследовательский этап.* Самый важный этап для разработки инклюзивного веб-сайта. Перед началом разработки для более глубокого понимания целевой аудитории, а также для обоснования выбранного инклюзивного функционала, чтобы снизить риск ошибок при разработке, проводятся разного рода исследования. Следует проанализировать конкурентов и аналоги, определить сильные и слабые стороны с точки зрения дизайна и удобства пользования сайтом.

На этом этапе также проводится глубинное интервью, о котором в 2001 году написал в своей работе С. А. Белановский, которое подразумевает под собой личную беседу с заранее подготовленными вопросами [9]. По ходу ответов интервьюер должен подстраиваться и «копать» интересующую тему глубже для выявления проблем и «болей» респондента. После проведения интервью составляется список выводов для дальнейшей работы и определения инклюзивных решений.

Исследований, направленных на целевую аудиторию, огромное множество, каждое подбирается индивидуально под проект и может принести массу ценной информации. Для работы были выбраны глубинное интервью, метод наблюдения и метод опроса.

Основными популярными способами также являются анализ отзывов конкурентов, использование метрик, проведение опросов, исследования с применением сервисов статистики (Яндекс.Вордстат, Google Trends и др.)

3. *Создание дизайна веб-сайта.* В этот этап входит составление структуры сайта, пути пользователя, определение основных разделов. После определяется цветовая гамма, подбираются шрифты. Далее разрабатывается дизайн концепция веб-сайта, куда интегрируются выбранные инклюзивные инструменты и графические элементы в виде иконок, инфографики, изображений. В рамках

продвижения фирменного стиля также разрабатывается оформление сопутствующих материалов, социальных сетей, маскота.

4. *Тестирование*. Готовый макет следует повторно протестировать с помощью уже знакомых исследований, сделать работу над ошибками и окончательно доработать сайт по необходимости.

5. *Реализация сайта*. Это заключительный этап. Им занимаются специализированные люди в области backend и frontend разработки, а также маркетологи. Сюда входит определение домена, написание кода, финансовая поддержка и seo-продвижение.

Если говорить о прикладном характере инклюзивного дизайна, то стоит обратить внимание на сферу государственных услуг. Ведь это именно те сервисы, которыми пользуются граждане различных аудиторий и социальных групп, включая людей с ограниченными возможностями, пожилых пользователей и тех, кто сталкивается с трудностями в цифровой среде. Приведём в пример реализацию первых пунктов алгоритма и проведения глубинного интервью для создания государственного портала. Самый популярный сервис в России «Госуслуги» обладает некой спецификой, в связи с чем перегружен информацией и функционалом. Было предложено создать более узконаправленный портал, предоставляющий пенсионные услуги под названием «Забота».

Целевая аудитория данного проекта будет обширной, так как заинтересоваться пенсионными услугами может любой совершеннолетний гражданин, однако для удобства взаимодействия и определения вектора разработки следует разбить её на сегменты по возрастному критерию, выделить ядерную целевую аудиторию и прописать сценарий и риски для отдельного пользователя.

- Молодёжь (18-24 года): студенты, молодые специалисты, которые только начинают работать официально, и проявляют интерес к пенсионным накоплениям. Легко интуитивно адаптируются в непривычных интернет-пространствах.

- Взрослые (25-54 года) – люди, создающие или уже создавшие семью, интересующиеся пенсионными вопросами своих близких (пожилых родственников), а также информацией о пенсионных программах для себя. Составляют основную долю от всех пользователей сервиса, обладают средними навыками в пользовании веб-сайтами.

- Пожилые (55 лет и старше) – люди, заинтересованные (или вынужденные) люди в информации по пенсионным услугам, обладающие, как правило, базовыми навыками пользования интернетом. Как правило, люди этого сегмента предпочитают личный визит или телефонный звонок.

- Пользователи с ограничениями (от 18 лет и старше). В проекте будут рассматриваться дееспособные люди 3-й группы инвалидности с возможностью самостоятельно без посторонней помощи пользоваться интернетом, компьютером и телефоном. Данную группу можно разбить на подгруппы и описать ключевые особенности:

- люди с физическими ограничениями: со слабой моторикой, недомоганием, с восприимчивостью к свету, мигающей графике, громкой музыке и резким сменам картинки;

- люди с ограничениями органов чувств: Дальтоники, слабослышащие и слабослышащие (временно или на постоянной основе);

- люди с ограничениями в восприятии: с синдромом Дауна, аутизмом, умственными расстройствами, рассеиванием внимания, переутомлением.

Для дальнейшей работы с проектом необходимо провести анализ различных сервисов предоставления государственных услуг, чтобы определить необходимые инклюзивные решения для разработки.

В качестве ядерной целевой аудитории принято выделить группу граждан, стоящие на пороге пенсионного возраста или только что получившие право на получение пенсии (40-65 лет), и ориентировать удобство пользования веб-сайтом на них.

*Глубинное интервью* — это метод качественного исследования, заключающийся в личной беседе с респондентом. Интервьюер задаёт заранее заготовленные вопросы и по ходу ответов может задавать ситуативные, углубляясь в суть проблемной темы. В результате такой моментальной обратной связи можно получить:

- 1) глубокое понимание проблемы, возможность оценить ситуацию на данный момент;

- 2) индивидуальную информацию о мнении человека, опыте взаимодействия, переживаниях и чувствах;



3) раскрытие неочевидных мотивов и «болей», которые могут оказаться ключевыми при решении проблемы.

Для проведения глубинного интервью была выбрана ядерная целевая аудитория проекта (40-60 лет). Это люди предпенсионного возраста, которые, вероятнее всего, будут чаще вынуждены взаимодействовать с сервисом и решать пенсионные вопросы. Необходимо прописать стартовые вопросы, благодаря которым можно будет понять, с какими неудобствами и трудностями сталкиваются пользователи при взаимодействии с основным сайтом предоставления государственных услуг.

Приведём пример одного из интервью, героем которого стал Александр. Ему 40 лет, есть двое детей, родом из города Балаково, но много лет проживает в Санкт-Петербурге. Имеет проблемы со зрением. Грамматика и пунктуация в транскрибации интервью отредактированы автором с целью сокращения, интервью представлено в сжатом смысловом пересказе.

Место проведения – личная встреча в общественном месте.

– *Вам часто приходится сталкиваться с решением вопросов, где нужно взаимодействовать с государством, ведомствами? Расскажите о последнем опыте.*

– Как и всем россиянам периодически приходится сталкиваться. Недавно меняли загранпаспорт с супругой. Множество бумажной волокиты было, определённых проблем с сайтами. Но по итогу всё получилось, подали успешно. Сейчас ещё в процессе, оформляется.

– *Использовали ли вы для подачи документов веб-сайт или приложение?*

– Так получилось на самом деле, что скорее пришлось воспользоваться Госуслугами, потому что, когда ходили подавать лично, возникли проблемы в МФЦ и сотрудники порекомендовали подать заявление через сайт.

– *Если бы не было необходимости, то вы предпочли бы личный визит, нежели решать вопрос электронно?*

– Наверное, зависит от того, какой вопрос. Абсолютный приоритет одному из решений не отдавал бы.

– *Как вы оцените уровень безопасности, доверяете ли свои данные государственным ресурсам?*

– Госуслугам, понятное дело, доверяю, да и прочим сайтам. Если они официальные, то спокойно оставляю свои паспортные данные. У меня нет поводов этого не делать, по-другому вопросы не решишь.

– *Испытывали трудности со входом в личный кабинет или при регистрации?*

– Со входом время от времени возникает проблема. Не понимаю, почему у меня постоянно происходит автоматический выход из личного кабинета и данные не сохраняются. Приходится искать, вводить и заходить заново. Это не сильно проблематично, скорее некомфортно и времязатратно.

– *Легко ли в принципе адаптируетесь на государственных сайтах, находите необходимые разделы?*

– Опять же от раза к разу. Если возникают затруднения, то я пишу роботу, который там есть, либо в поддержку, и через какое-то время в любом случае подскажут и вопрос решается. Ну а в целом обычно довольно понятно, со всем разбираюсь и желаемую информацию получаю. Если вопрос какой-то срочный и не может решиться благодаря роботу, то я лучше позвоню.

– *В целом как вы оцените визуальное приложение «Госуслуги»?*

– В целом устраивает, каких-то конкретных недочётов так сразу не назову. Мелко бывает написано и ничего не разберёшь. У меня последние несколько лет садится зрение, думаю, что тоже сказывается. Просто бывает, что банально не вижу и приблизить никак нельзя, вот в этом плане бывает тяжело, но это опять же из-за моих субъективных проблем.

– *При решении вопросов какие эмоции чаще всего испытываете?*

– Из-за трудностей бывают негативные эмоции, когда что-то долго не получается. В целом мне не нравится система сообщений в приложении. Часто бывает, что там 20 уведомлений висит, я их перечитываю, а там спам, информация какая-то вместе с одним нужным сообщением. Очень неудобно разбираться в той информации, которая приходит, часто она бессмысленная. Никак не систематизирована.

- Как вы оцените количество графики на сайте, картинок, иконок?
- Я бы не стал жаловаться, робот и другие элементы все понятно сделаны. Бывает, после рабочего дня, когда заходишь, яркие цвета режут глаз. Тогда на телефоне яркость понижаю.
- Пользуетесь ли вы функциями ручной настройки сайтов, например размер шрифта, версия для дальтоников и т.д.?
- На самом деле первый раз про это слышу, не видел и никогда не пользовался. Если бы мне показали, где это и как делается, то я с радостью увеличивал бы шрифт.
- Можете подвести итог того, что вы бы поменяли на сайте?
- Я бы увеличил шрифт, поменял цвета на более приглушённые. И самое главное – настроил бы раздел с уведомлениями, чтоб там не было ничего лишнего. Я считаю, что пастельные цвета не то, что стоит выбирать, но убрать яркость можно.
- Было бы удобно, если все сервисы государственные были в одной стилистике?
- Многие сервисы, которыми я пользуюсь от «Госуслуг», и так в одной стилистике, в одном цвете. Но у меня даже разная стилистика не вызывает отрицания.
- Было бы легче пользоваться сервисами, если были бы пошаговые инструкции в виде иллюстраций?
- Скорее нет, потому что я хорошо ориентируюсь на сайтах или пользуюсь ботом-помощником сразу, а вот моей маме бы это очень помогло. Время от времени она обращается ко мне с какими-то вопросами, потому что не понимает, куда нажимать. Для неё это было бы однозначно плюсом. Людям в возрасте тяжело. Им нужно объяснять.
- Задумывались ли вы о своей будущей пенсии, знаете ли свою ставку, программы?
- Я задумывался, но не могу сказать, что владею какой-то конкретной информацией. Просто руки не доходили, но мысли уже были.

### Общие выводы исследования

1. Было проведено 30 интервью и выяснилось, что только 7 человек периодически пользуются государственными сайтами и приложениями. Остальные стараются как можно реже с ними взаимодействовать, а при необходимости едут сразу в офис по живой очереди и консультируются лично, даже если на это приходится потратить много времени. Поэтому было решено переосмыслить ядерную целевую аудиторию и ориентироваться больше на взрослых семейных людей в возрасте около 40 лет. У них достаточно времени до пенсии, большинство из них активные пользователи интернета, и именно они могли бы подписаться на новостные группы по теме.

2. Пошаговые инструкции, иллюстрации и инфографика смогли бы помочь пожилым людям или тем, кто испытывает трудности с интерфейсом.

3. Отсутствие функции запоминания пароля или быстрого входа в целом понижает уровень пользования сервисом. Часто проблема входа является причиной откладывания решения вопроса.

4. Функции ручной регулировки сайта не пользуются популярностью в связи с неосведомлённостью пользователей, хотя такая потребность есть. Часто граждане не замечают инклюзивный функционал из-за скрытого меню.

5. Отсутствие осведомлённости граждан о пенсионных программах и ставках является главной проблемой. Несмотря на интерес, многие не знают, где получить необходимую информацию, из-за чего откладывают этот вопрос вплоть до выхода на пенсию.

Также после интервьюирования можно выделить практические решения для внедрения в платформу:

- 1) подобрать цветовую палитру, оптимальную для всех пользователей;
- 2) сделать корреляцию по размерам шрифта;
- 3) сделать хорошую адаптивность под различные устройства;
- 4) создать полноценный фирменный стиль платформы для идентификации на рынке;
- 5) развивать социальные сети для осведомлённости граждан;
- 6) добавить функцию быстрого поиска;
- 7) добавить функцию выбора темы сайта для ситуативных неудобств;

- 8) добавить функцию быстрой записи в офис или консультации по телефону;
- 9) разработать чёткую структуру и разделы, пошаговые материалы.

После проведения концептуального и исследовательского этапов можно переходить к созданию дизайн-прототипа. Это будет государственный веб-сайт под названием «Забота», специализирующийся на вопросах пенсионного спектра. Разработку дизайн-прототипа следует начать с определения разделов сайта, функционала и составления пути пользователя – системы навигации, карты перехода пользователя по интерфейсу от экрана к экрану (рис. 1).

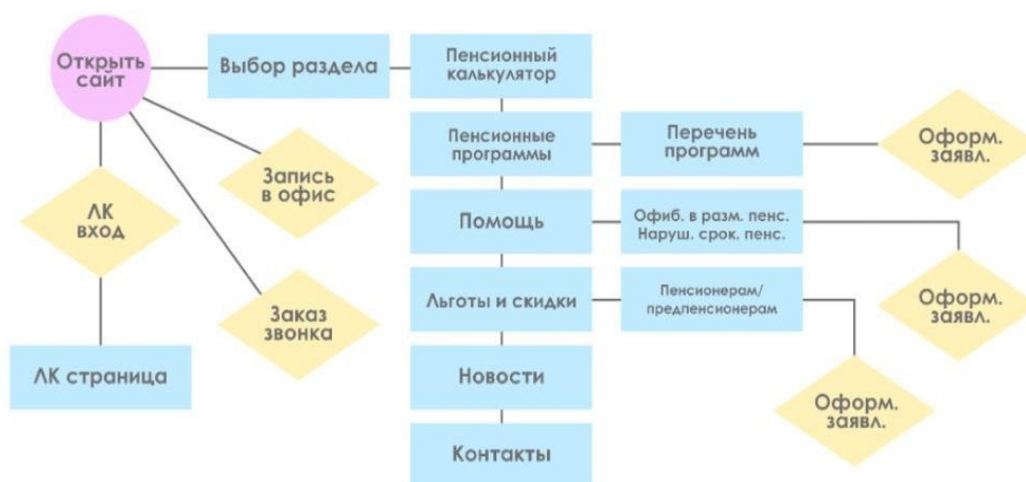


Рисунок 1. Составление пути пользователя

Каждый раздел, предполагающий обращение в ведомство, заканчивается оформлением заявления, что и является конечной точкой посещения портала. Альтернативной целью может являться очная запись в ведомство или же поиск контактов, что можно сделать сразу на главной странице сайта. Благодаря составлению пути пользователя обозначается тип сайта и общее видение итогового портала. Сайт представляет собой многостраничный портал с разделами.

Основная задумка – создать функционал для быстрого решения различных задач уже на главной странице сайта, без скрытых меню и лишних переходов «вглубь» сайта, а также сделать возможность, не изучая сайт, записаться на приём в офис или заказать звонок оператора для ситуаций, когда пользователю так удобнее.

Следующим шагом, исходя из ранее разработанного алгоритма, является разработка фирменного стиля для собственной платформы.

Поскольку платформа нацелена на обширную целевую аудиторию, принято решение использовать цвета, воспринимаемые большинством граждан, в том числе людьми с дальтонизмом. Люди с дальтонизмом (особенно протанопией и дейтеранопией) могут испытывать трудности при различении некоторых оттенков синего и зелёного. Однако многие дальтоники всё равно воспринимают голубой как отдельный цвет благодаря его яркости и контрасту с другими цветами [10]. В качестве основных цветов были выбраны голубой и лимонно-желтый как контрастирующий цвет для взаимодействия, кнопок (рис. 2).

Также при проведении исследований выяснилось, что пользователи испытывают дискомфорт от сильных отличий в дизайне разных государственных платформ, поэтому за основу-референс был выбран основной государственный портал на сегодняшний день — Госуслуги. Цвета играют важную роль в восприятии, эмоциональном состоянии и взаимодействии пользователей с платформой. Согласно семантике цвета по П.В. Яньшину, синие оттенки ассоциируются с мудростью, интуицией, а также создают спокойную атмосферу, что способствует лучшему усвоению информации [11].

Далее необходимо выбрать фирменные шрифты, используемые для сайта, и фирменный стиль (рис. 2).



Рисунок 2. Цветовая палитра и шрифтовая пара

Intro — современный, геометрический шрифт с минималистичным дизайном. Он отличается строгой структурой, равномерной толщиной линий и аккуратными формами букв. Был создан в 2017 году дизайнером Колином Тирни. Имеет легкую читаемость, простоту восприятия, универсальность. В проекте шрифт решено было использовать для заголовков, основного текста и рекламных материалов. Также шрифт был взят за основу логотипа.

Comfortaa — это округлый шрифт без засечек, был разработан дизайнером и типографом Юрием Костенко в 2008 году. Шрифт имеет характерный плавный и округлый стиль, что придает ему дружелюбный и доступный вид. Comfortaa был разработан с учетом возможности использования в веб-дизайне и печатных материалах и идеально подходит для текстового блока. Этот шрифт часто используется в проектах, требующих эмоциональной связи с пользователями.

Следующим этапом следует разработка логотипа (рис. 3).



Рисунок 3. Логотип платформы

Логотип проекта «Забота» представляет собой шрифтовое написание нейма с видоизмененной буквой в форме сердца, символизируя тепло и заботу о гражданах. Также была разработана серия иконок для сайта с плавными линиями, скругленными краями и монолитными элементами для большей читаемости (рис. 4).



Рисунок 4. Серия иконок для платформы

После определения основных элементов фирменного стиля были разработаны основные страницы сайта (рис. 5-7).

На каждой странице в шапке расположена ручная регулировка портала для людей с ограничениями и поисковая строка для поиска по ключевым словам. Исследования показали, что многие граждане не замечали ручную регулировку сайтов, хотя имели потребность в её использовании, поэтому на нашем портале инклюзивные инструменты располагаются не в скрытом меню, а выделены контрастирующим блоком в самом начале взаимодействия с порталом.

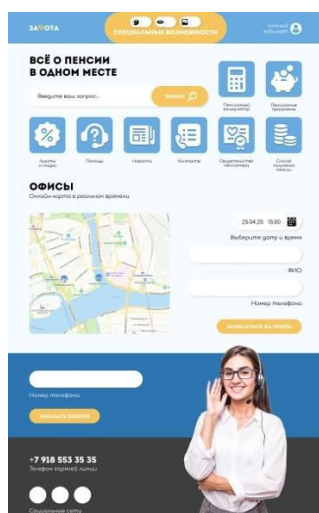


Рисунок 5. Главная страница сайта



Рисунок 6. Раздел «Пенсионный калькулятор»

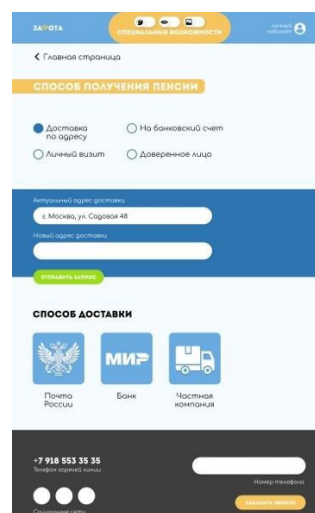


Рисунок 7. Раздел «Способ получения пенсии»

Инклюзивность портала заключается не только в функции подстраиваться под различных пользователей, но также в адаптивности портала под различные устройства: ноутбук, телефон, планшет и т.д. После создания основной версии для компьютеров и ноутбуков была разработана мобильная версия сайта, а также отдельное мобильное приложение (рис.7).

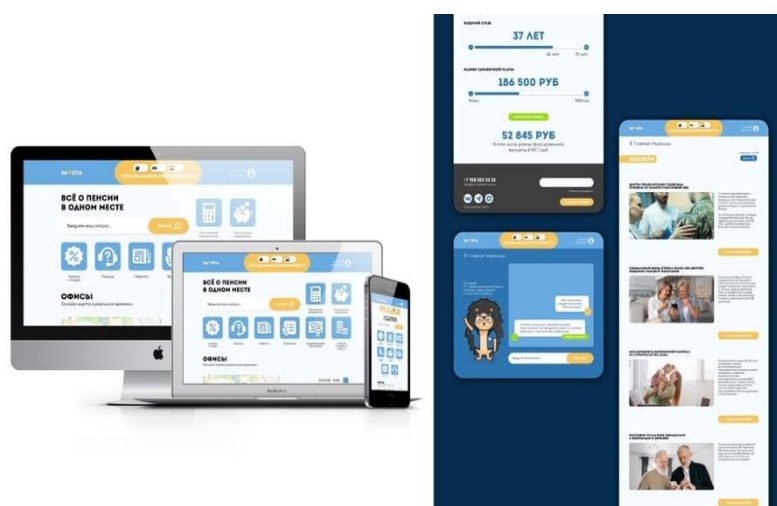


Рисунок 7. Адаптивные версии сайта

Пользовательский интерфейс (UI) относится к визуальным и интерактивным компонентам, которые позволяют пользователям взаимодействовать с продуктом или услугой. Он служит средством навигации и доступа к функциям.

Пользовательский опыт (UX), с другой стороны, относится к тому, как пользователи взаимодействуют с продуктом или услугой через UI. В совокупности UI/UX представляет собой интерфейс, который облегчает взаимодействие между пользователем и приложением, охватывая программные и аппаратные элементы [12].

## Заключение

Анализируя результаты исследования, можно сделать соответствующие **выводы**:

1. В современном мире создание доступной среды государственных интернет-ресурсов является важной задачей для обеспечения равных прав граждан с ограниченными возможностями и других социальных групп.

2. Общепринятые алгоритмы направлены больше на часть реализации с точки зрения программирования, нежели на дизайн и удобство интерфейса в целом.

3. Многие существующие подходы недостаточно учитывают особенности целевой аудитории, что может приводить к созданию менее доступных и удобных сайтов.

4. Разработанный алгоритм делает акцент на тщательном исследовании потребностей пользователей, что способствует созданию более инклюзивных и эффективных веб-ресурсов, а также помогает снизить риск ошибок.

5. Разработанный алгоритм также можно применять в разработке инклюзивных сайтов в других сферах как основу последовательных действий, способствующих правильному определению набора инклюзивных инструментов.

Сегодня важно разрабатывать не только эстетичные веб-ресурсы, но и доступные людям с ограничениями для вовлечения их в общество и предоставления равных возможностей со стороны государства. Представленная статья демонстрирует необходимость системного подхода, включающего анализ потребностей целевой аудитории, применение стандартов доступности и использование инструментов оценки качества. Разработанный алгоритм позволяет повысить эффективность процесса создания инклюзивных ресурсов, что в конечном итоге способствует повышению уровня цифровой грамотности и социальной интеграции различных групп населения. В будущем важно продолжать совершенствовать методики и внедрять инновационные решения, чтобы обеспечить максимально широкий и удобный доступ к государственным информационным системам для всех граждан.

#### Список литературы

1. Альтовский Е. В., Демидов А. А., Краснова Г. А. Доступность интернет-ресурсов инвалидам и лицам с ограниченными возможностями : информационно-аналитический обзор. Москва : Дело, 2021. 114 с.
2. Официальный сайт Центра инклюзивного дизайна Великобритании (CAE). Режим доступа: <http://www.cae.org.uk> (дата обращения: 20.01.2025).
3. Показатели инклюзии / науч. ред. Н. Борисова; пер. И. Аникеев. Москва : Перспектива, 2013.
4. Causey A. The Importance of Accessibility in Retail Websites // Journal of Business & Retail Management Research, 2019. P. 10-22.
5. Erdtman E., Rassmus-Gröhn K., Hedvall P. O. Scandinavian Journal of Disability Research. 2021. Vol. 23. P. 158–168.
6. Создание Web-сайтов. Ч. 5 / Н.А. Инькова, Е.А. Зайцева, Н.В. Кузьмина, С.Г. Толстых. Тамбов: Изд-во Тамб. гос. техн. ун-та, 2002. 56 с.
7. Мальцев И. П. Проектирование сайтов. 2017. 178 с.
8. Захарова М. В. Основные этапы и принципы процесса создания веб-сайта // Наука, образование и инновации: сборник статей международной научно-практической конференции. Часть 3. Екатеринбург: Аэтерна, 2016.
9. Белановский С. А. Глубокое интервью. Москва : Никколо-Медиа, 2001. 320 с.
10. Color discrimination in color vision deficiency. Mollon & Reffin, 1989.
11. Яньшин П. В. Психосемантика цвета. 2-е изд., перераб. и доп. Москва: Юрайт, 2024. 417 с.
12. Sharma V., Tiwari A. K. World Journal of Research and Review 2021. Vol. 12. P. 41–43.
13. Терскова С. Г. Механизм формирования доступной среды для инвалидов // Гуманитарные научные исследования. 2015. № 7, Ч. 2.

## **Автоматизированный поиск в видеопотоке: практический инструмент для исследований и контент-анализа видеофрагментов**

**Н. А. Бондарева**

ИПМ им. М.В. Келдыша РАН, Москва, Россия

### **Введение**

Нарастающая роль видеоконтента во всех областях исследований, от социальных до научных, все чаще вызывает необходимость обращаться к количественным методам анализа видеоконтента. Исследователи в различных областях, например в области социологии медиа и цифровых гуманитарных наук, сталкиваются с необходимостью анализировать огромные объемы видеоматериалов для выявления паттернов репрезентации того или иного явления, изучения экранного времени персонажей, анализа гендерного баланса и других аспектов медиаконтента [1, 2].

Однако традиционные методы поиска и анализа появлений конкретных объектов в видеоматериалах остаются крайне трудозатратными. Исследователь вынужден просматривать часы видеоматериала, вручную фиксируя временные интервалы появления интересующих персонажей. Такой подход не только требует значительных временных ресурсов, но и подвержен субъективным ошибкам, особенно при работе с длительными видеоматериалами или большими медиаколлекциями [3].

Существующие коммерческие решения для видеоанализа либо ориентированы на профессиональное видеопроизводство (Adobe Premiere Pro [4], Final Cut Pro [5]) и не нацелены на задачи видеоанализа, либо требуют значительных финансовых вложений и технической экспертизы (облачные сервисы AWS Rekognition [5], Google Video Intelligence API [6]). На сегодняшний день существует очень ограниченный диапазон инструментов, которые способны предоставить исследователю возможность анализировать значительный объем материала согласно его потребностям. Часто задачи исследования имеют столь узконаправленную специфику, что широкодоступные программные комплексы функционально не соответствуют требованиям поставленных задач. Это ставит перед исследователем дополнительную задачу непосредственно перед научным исследованием сначала подобрать подходящий инструмент для поиска или анализа, а в случае отсутствия такового создать его самому.

В данной работе рассматривается задача целенаправленного поиска в видеопотоке определенных объектов по запросу и фиксирование временных меток их появления. Поскольку представленные на современном рынке решения не соответствовали поставленным задачам, то было принято решение реализовать такой инструмент самостоятельно в рамках текущего проекта исследований, проводимых в ИПМ им. М.В. Келдыша РАН.

Целью исследования является разработка и описание практического инструмента для автоматизированного поиска объектов в видеоконтенте, доступного для использования исследователями без специальных технических навыков.

В число поставленных задач входит:

1. Анализ существующих решений для поиска персонажей в видео и выявление их ограничений для исследовательского применения;
2. Описание архитектуры разработанного инструмента;
3. Разработка программной реализации;
4. Выявление перспектив развития и возможности интеграции инструмента в различные области исследования.



### Обзор существующих решений

Современный рынок программного обеспечения предлагает ряд решений для работы с видеоконтентом, однако их применимость для исследовательских задач ограничена рядом нюансов, имеющих узконаправленную специфику применения.

Профессиональные видеоредакторы, такие как Adobe Premiere Pro и Final Cut Pro, включают функции автоматического распознавания лиц, однако они ориентированы преимущественно на задачи видеопроизводства. Основные ограничения этих решений для исследовательского применения включают высокую стоимость лицензий, сложность освоения для пользователей без технического бэкграунда, отсутствие возможности экспорта структурированных данных для последующего статистического анализа [8].

Облачные сервисы (AWS Rekognition Video, Google Video Intelligence API, Microsoft Azure Video Indexer) предоставляют мощные возможности для анализа видеоконтента, включая распознавание и отслеживание лиц. Однако их применение в академических исследованиях может быть затруднено по ряду причин, начиная от высокой стоимости обработки больших объемов видеоматериалов и заканчивая сложностью политической ситуации, продуцирующей ряд ограничений для российских ученых. Также следует учитывать тот факт, что облачные технологии подходят не для всех научных задач, так как некоторые данные требуют исключительно локальной обработки в силу своей специфики и требований к конфиденциальности при загрузке исследовательских материалов на внешние серверы [9, 10].

В академической среде существует ряд специализированных инструментов для анализа видеоконтента. К примеру, ELAN (EUDICO Linguistic Annotator) широко используется лингвистами для аннотации видеоматериалов, однако требует ручной разметки и не включает автоматическое распознавание лиц или объектов [11]. Решения OpenCV-based предоставляют мощную техническую базу для создания систем видеоанализа, однако их использование требует значительных программистских навыков. Большинство готовых реализаций ориентировано на технических специалистов и не предоставляет простого интерфейса для исследователей-гуманитариев [12].

Проект VideoANT (Video Annotation Tool) от Университета Миннесоты предлагает веб-интерфейс для аннотации видео, но также требует ручной работы и не включает функции автоматического распознавания [13]. В области цифровых гуманитарных наук разработаны отдельные инструменты для анализа медиаконтента. Cinemetrics-проект фокусируется на анализе монтажных характеристик фильмов, но не включает функции поиска объектов или определенных персонажей [14]. CLARIAN Media Suite предоставляет исследователям доступ к большим коллекциям медиаконтента с возможностями поиска и анализа, однако функции автоматического распознавания объектов ограничены и у него [15].

Среди существующих программных решений, представленных в открытых интернет-источниках, можно выявить несколько ключевых пробелов, в частности:

- 1) недостаток простых решений для решения задачи поиска конкретных объектов/персонажей в видеопотоке;
- 2) отсутствие технических средств, позволяющих запустить поиск по образцу по определенному видеоматериалу, предоставленному пользователем;
- 3) недоступность инструментов для исследователей без технического бэкграунда (большинство решений требует программистских навыков или значительных финансовых ресурсов);
- 4) ограниченные возможности экспорта структурированных данных для последующего статистического анализа и визуализации;

Эти аспекты подтверждают необходимость и актуальность разработки простого, доступного инструмента для автоматизированного поиска определенных объектов в видеоконтенте, ориентированного на потребности исследователей в различных областях – от экспериментальной физики до медиаисследований и цифровых гуманитарных наук.

### Техническая реализация

Разработанная система представляет собой инструмент для автоматизированного поиска персонажей в видеоконтенте, состоящий из пяти основных этапов: предобработка эталонного изображения, декомпозиция видеоматериала, обнаружение объекта (в нашем случае лица) в кадрах, сопоставление с эталоном и формирование выходных данных в следующем виде:

- директория с вырезанными видеофрагментами;
- файл с временными отметками появления искомого объекта в кадре.

Для работы программы создается организованная структура директорий, где группируются необходимые материалы.

После запуска исполняющего файла программа автоматически загружает эталонные лица, добавленные пользователем, затем проводит поиск по загруженным пользователем видеофайлам, анализируя кадры на наличие искомого объекта и сравнивая обнаруженные лица с эталонным образцом, а затем формирует директорию с результатами работы. В ней находятся:

- видеофайлы с найденными фрагментами, содержащими заданные лица;
- JSON-файл с таймкодами сцен, описывающий время начала и окончания каждого фрагмента;
- лог-файл с информацией о процессе обработки;
- файл настроек (для возможной корректировки параметров).

Работа пользователя сводится к подготовке данных (фото и видео) и запуску программы. После этого программа автоматически выполняет поиск лиц и создает результаты в указанной папке.

### Архитектура системы

Система представляет собой модульное приложение, разработанное для автоматического извлечения фрагментов видео, содержащих определенные лица. Основной принцип работы заключается в сравнении лиц, обнаруженных в видеопотоке, с эталонными изображениями, представленными пользователем в виде фотографий.

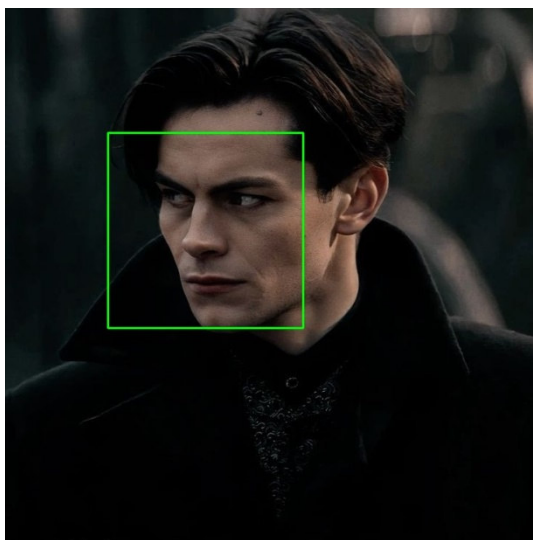


Рисунок 1. Пример распознанного лица

Алгоритм работы программы представлен на рисунке 2.

Основным компонентом системы является класс FaceExtractor, который инкапсулирует всю логику обработки видео и распознавания лиц.

Ключевые методы класса FaceExtractor:

- `load_reference_faces()`: загружает эталонные изображения лиц из указанной директории и кодирует их в векторы признаков;
- `process_video()`: обрабатывает видеофайл покадрово, обнаруживает лица на каждом кадре, кодирует их и сравнивает с эталонными векторами;

- `_merge_scenes()`: объединяет близкие по времени сцены с обнаруженными лицами в единые фрагменты;
- `_save_results()`: сохраняет видеофрагменты с обнаруженными лицами и метаданные в указанную директорию.

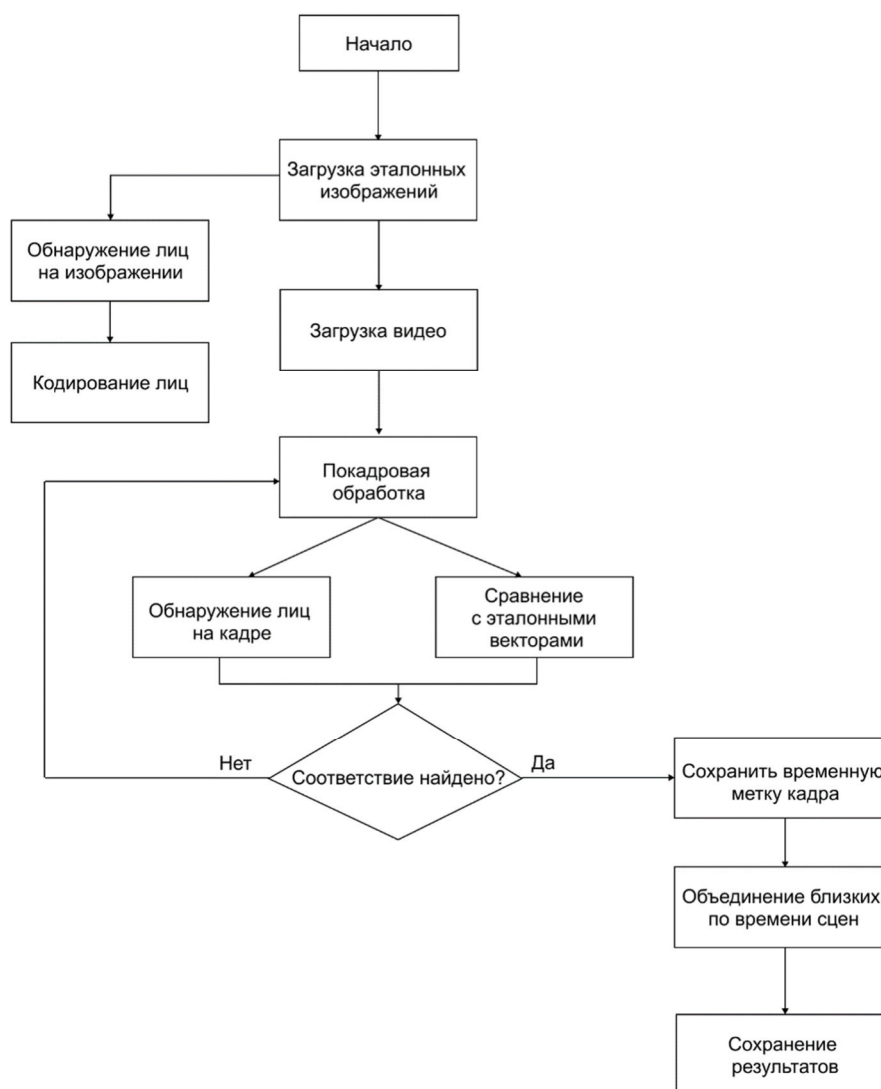


Рисунок 2. Схема алгоритма

### Технические параметры

Система разработана на языке программирования Python 3.9. Для распознавания и сопоставления лиц используется библиотека `face_recognition` [16], представляющая собой Python-обертку над библиотекой `dlib` [17]. В основе алгоритма лежит подход FaceNet [18], который создает 128-мерные векторы признаков (embeddings) для каждого обнаруженного лица с использованием глубокой нейронной сети на основе архитектуры ResNet.

Для реализации основных функций программы использовались следующие библиотеки:

- `opencv-python (cv2)` (версия 4.8.1.78) для работы с видео и изображениями, используемая для загрузки, декодирования и покадровой обработки видео;
- `moviepy` (версия 1.0.3) для редактирования видео, используемая для объединения сцен и сохранения видеофрагментов;
- `numpy` (версия 1.24.3) для работы с массивами, используемая для хранения и обработки векторов признаков лиц.

Выбор данных программных средств обусловлен простотой использования, широкой доступностью библиотек для машинного обучения и обработки видео, а также хорошей производительностью.

### Предобработка данных

Для обеспечения высокой точности и скорости распознавания лиц к входным данным применяется ряд шагов предобработки.

Эталонные изображения лиц должны быть представлены в формате JPG, JPEG или PNG с разрешением, достаточным для четкого отображения лица. Процесс загрузки эталонных изображений включает в себя чтение файлов, обнаружение лиц на изображениях и кодирование лиц в 128-мерные векторы признаков с использованием библиотеки `face_recognition`.



Рисунок 3. Эталонные образцы лица

Система поддерживает видеофайлы в формате MP4, AVI и MKV. Процесс обработки видео включает в себя загрузку видеофайла, декодирование кадров и покадровую обработку. Для уменьшения вычислительной нагрузки размер кадров изменяется до `max_image_size` пикселей (параметр настраивается в файле конфигурации). Для ускорения обработки используется параметр `frame_interval`, который определяет, как часто будут обрабатываться кадры (например, `frame_interval = 5` означает, что будет обрабатываться каждый пятый кадр).

### Формат выходных данных

Результаты работы системы представлены в виде видеофрагментов, содержащих сцены с обнаруженными лицами, и метаданных, описывающих эти фрагменты.

Видеофрагменты сохраняются в формате MP4 с использованием кодека H.264. Метаданные сохраняются в формате JSON и содержат следующую информацию: `video_name` (имя исходного видеофайла); `fragments` (список фрагментов с обнаруженными лицами с временными отметками); `settings` (параметры обработки) и т.д. Также ведется запись лог-файлов (рис. 4).

```
!025-09-03 00:45:03,882 - INFO - Начало загрузки референсных изображений
!025-09-03 00:45:04,488 - INFO - Успешно загружено изображение: im1.jpg
!025-09-03 00:45:05,124 - INFO - Успешно загружено изображение: im2.jpg
!025-09-03 00:45:05,606 - INFO - Успешно загружено изображение: im3.jpg
!025-09-03 00:45:05,608 - INFO - Начало обработки видео: c:\Users\NiciBond\face_detection_project\input_videos\sara.mp4
!025-09-03 00:45:23,477 - INFO - Начало новой сцены: 3.30 сек
!025-09-03 00:45:34,524 - INFO - Конец сцены: 5.30 сек
!025-09-03 00:45:40,445 - INFO - Начало новой сцены: 6.30 сек
!025-09-03 00:45:45,913 - INFO - Конец сцены: 7.30 сек
```

Рисунок 4. Пример отчета программы

Тестирование системы проводилось на следующем оборудовании:

- CPU: Intel Core i7-8700K (6 ядер, 12 потоков);
- RAM: 16 GB DDR4;
- GPU: NVIDIA GeForce GTX 1070 (8 GB);
- тип накопителя: SSD;
- программное обеспечение;
- операционная система: Windows 10 (64-bit);
- Python: 3.9.

### Описание тестов

Для оценки производительности системы были проведены следующие тесты.

1. *Тест размера изображения.* Целью данного теста является определение влияния размера изображения на скорость обработки. В ходе теста изменялся размер изображения (от 500 до 2000 пикселей) и измерялось время обработки одного кадра.

2. *Тест параллельной обработки.* Целью данного теста является определение оптимального размера пакета для параллельной обработки. В ходе теста изменялся размер пакета (от 1 до 16) и измерялось время обработки одного кадра.

3. *Тест обработки видео.* Целью данного теста является определение влияния интервала между кадрами на скорость обработки. В ходе теста изменялся интервал между кадрами (от 1 до 30) и измерялась скорость обработки видео (кадров в секунду).

Для оценки результатов тестов использовались следующие метрики:

- время обработки (в секундах);
- скорость обработки (кадров в секунду);
- использование памяти (в MB);
- загрузка CPU (в процентах).

Результаты тестов показали, что система успешно обнаруживает заданных персонажей в видеопотоке, формирует видеофрагменты с их участием и сохраняет метаданные, описывающие эти фрагменты. В частности, тесты подтвердили:

- эффективность использования библиотеки `face_recognition` для распознавания лиц;
- положительное влияние предобработки данных (изменение размера изображений, пропуск кадров) на скорость обработки видео;
- возможность настройки параметров системы (например, `max_image_size`, `frame_interval`, `recognition_threshold`) для адаптации к различным требованиям к производительности и точности.

Таким образом, можно заключить, что первая версия программного обеспечения успешно реализована и выполняет поставленные задачи. Система демонстрирует стабильную работу и позволяет эффективно находить заданных персонажей в видеопотоке.

Полученные результаты открывают широкие перспективы для дальнейшего развития и применения системы в различных областях, таких как:

- автоматическая разметка видеоархивов;
- системы безопасности и видеонаблюдения;
- анализ медиаконтента и выявление ключевых персонажей;
- создание персонализированных видеоподборок.

### Обсуждение

Разработанная система обладает рядом *достоинств*, которые делают ее эффективной для использования в различных задачах:

- простота использования: система основана на широко известных и доступных библиотеках Python, что упрощает ее настройку;
- эффективность: благодаря использованию предобученной нейронной сети и оптимизированным алгоритмам обработки видео, система обеспечивает высокую скорость и точность распознавания лиц;

- гибкость: система позволяет настраивать параметры обработки видео (например, размер изображения, интервал между кадрами, порог распознавания) для адаптации к различным требованиям к производительности и точности;
- модульность: модульная архитектура системы упрощает ее расширение и модификацию для решения новых задач.

Вместе с тем стоит отметить, что система в первой версии имеет узконаправленный функционал и ориентирована только на поиск лиц в видеопотоке, а также имеет и некоторые *ограничения*, которые необходимо учитывать при ее использовании:

- зависимость от качества эталонных изображений: точность распознавания лиц напрямую зависит от качества эталонных фотографий; нечеткие, плохо освещенные или повернутые фотографии могут привести к снижению точности распознавания;
- ограниченность предобученной модели: система использует предобученную нейронную сеть, которая может быть неоптимальной для распознавания лиц в определенных условиях (например, при плохом освещении, сильных изменениях ракурса или наличии окклюзий);
- вычислительная сложность: обработка видео требует значительных вычислительных ресурсов, что может быть проблемой при использовании системы на маломощных устройствах.

По мере работы над улучшениями характеристик системы и расширением ее функциональности можно выделить следующие *направления развития*:

- использование более современных архитектур нейронных сетей: замена предобученной модели на более современную архитектуру (например, ResNet, EfficientNet) может повысить точность распознавания, расширить области её применения и распознавать не только лица, но и более специфические объекты, к примеру определенные явления в физических экспериментах;
- разработка адаптивных алгоритмов: разработка алгоритмов, которые автоматически настраивают параметры системы (например, размер изображения, интервал между кадрами, порог распознавания) в зависимости от характеристик видеопотока, позволит повысить ее эффективность и гибкость;
- оптимизация для работы на GPU: использование GPU для ускорения вычислений может значительно повысить скорость обработки видео, особенно при использовании сложных нейронных сетей;
- интеграция с другими системами: интеграция системы с другими системами (например, системами видеонаблюдения, базами данных лиц) позволит расширить ее функциональность и область применения;
- разработка графического интерфейса пользователя (GUI): создание удобного графического интерфейса упростит использование системы для пользователей, не имеющих опыта работы с командной строкой.

Разработанная система имеет потенциал для применения как для прикладных задач в сфере анализа медиа, так и в различных областях науки – от экспериментальных установок до социологических исследований.

## Заключение

В данной работе представлена система для автоматического поиска лиц в видеопотоке, основанная на использовании библиотеки `face_recognition` и оптимизированная для достижения высокой производительности.

Разработанная система представляет собой инструмент для поиска лиц в видеопотоке, который может быть использован в различных областях науки и техники. Предложенные направления развития системы, такие как использование более современных архитектур нейронных сетей, разработка адаптивных алгоритмов и оптимизация для работы на GPU, позволят расширить ее функциональность и повысить эффективность, что откроет новые возможности для применения разработанной системы в решении широкого круга задач в различных областях науки и техники.

Вклад данной работы заключается в демонстрации практической применимости существующих библиотек и методов оптимизации для решения реальной задачи поиска лиц в видеопотоке.

Разработанная система может быть использована в качестве основы для создания более сложных и функциональных приложений в области анализа видеоданных.

### Список литературы

1. Moretti F. Distant Reading. London: Verso, 2013. 298 p.
2. Manovich L. Cultural Analytics: Analysing Cultural Patterns in the Era of "More Media" // Domus. 2009. No. 923. P. 1-7.
3. Salt B. Film Style and Technology: History and Analysis. London: Starword, 2009. 398 p.
4. Adobe Premiere Pro. URL: <https://www.adobe.com/products/premiere.html>
5. Final Cut Pro. URL: <https://www.apple.com/final-cut-pro/>
6. Amazon Rekognition. URL: <https://aws.amazon.com/rekognition/>
7. Video Intelligence API documentation. URL: <https://cloud.google.com/video-intelligence/docs>
8. Adobe Systems Inc. Adobe Premiere Pro User Guide // Adobe Documentation. 2023. URL: <https://helpx.adobe.com/premiere-pro/user-guide.html>
9. Amazon Web Services. Amazon Rekognition Video Developer Guide. 2023. URL: <https://docs.aws.amazon.com/rekognition/>
10. Google Cloud. Video Intelligence API Documentation. 2023. URL: <https://cloud.google.com/video-intelligence/docs>
11. Wittenburg P., Brugman H., Russel A. ELAN: a Professional Framework for Multimodality Research // Proceedings of the 5th International Conference on Language Resources and Evaluation. 2006. P. 1556-1559.
12. Bradski G., Kaehler A. Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library. Sebastopol: O'Reilly Media, 2016. 1024 p.
13. VideoANT Project. Video Annotation Tool / University of Minnesota. URL: <https://ant.umn.edu/>
14. Cinemetrics. Film Measurement and Analysis // Cinemetrics Database. URL: <http://www.cinemetrics.lv/>
15. CLARIAH Media Suite. Research Platform for Media Studies // CLARIAH Consortium. URL: <https://mediasuite.clariah.nl/>
16. Geitgey A. Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning // Medium. 2016. URL: <https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffe121d78>
17. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren, J. Sun // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770-778.
18. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 815-823.



## Синергетическая интеграция методов роевого интеллекта в задачах визуализации и аналитики

С. А.Третьяков, И. С. Стародубцев  
ИММ УрО РАН, Екатеринбург, Россия

**Аннотация.** Адаптивные мультиагентные роевые системы (АМРС) представляют перспективное направление для решения сложных задач визуализации и аналитики в условиях неопределенности и динамически изменяющейся среды. Данная работа представляет обзор современных подходов к построению АМРС, основанных на синергетической интеграции четырех ключевых технологий: мультиагентного обучения с подкреплением (MARL), графовых нейронных сетей (GNN), методов вычислительного интеллекта (CI) и трансферного обучения (TL). Особое внимание уделено применению синергетического подхода к задачам роевой разведки местности и одновременной локализации и картографирования (SLAM), где визуализация процесса обучения и анализа траекторий агентов играют критическую роль в оценке эффективности системы. Проведен сравнительный анализ классических и современных методов роевого интеллекта, выявлены ограничения существующих подходов и показаны преимущества их интеграции. Результаты анализа демонстрируют, что синергетический подход MARL+GNN+CI+TL обеспечивает повышенную масштабируемость, устойчивость к отказам и адаптивность по сравнению с классическими методами, открывая новые возможности для интерактивной визуализации и аналитики роевых систем.

**Ключевые слова:** адаптивные роевые системы, мультиагентное обучение с подкреплением, графовые нейронные сети, вычислительный интеллект, трансферное обучение, научная визуализация, SLAM

### Введение

Задачи координации множества автономных агентов в условиях ограниченной информации и децентрализованного управления возникают в широком спектре прикладных областей, включая робототехнику, системы наблюдения и мониторинга окружающей среды. Особый интерес представляют системы, способные адаптироваться к изменениям в структуре задач и условиях функционирования без перепроектирования базовых алгоритмов управления.

Традиционные подходы к построению мультиагентных систем основаны на заранее определенных правилах взаимодействия (алгоритм Boids [1], потенциальные поля [2]) или иерархических схемах координации [3]. Данные методы обеспечивают предсказуемое поведение в стационарных условиях, однако демонстрируют ограниченную способность к адаптации при изменении параметров среды или отказах отдельных агентов. Альтернативным направлением является применение методов машинного обучения, в частности, мультиагентного обучения с подкреплением (Multi-Agent Reinforcement Learning, MARL), позволяющего агентам вырабатывать стратегии поведения на основе взаимодействия со средой [4].

Одной из ключевых проблем MARL является экспоненциальный рост пространства состояний при увеличении числа агентов, что затрудняет масштабирование алгоритмов на большие роевые системы. Частичным решением данной проблемы служит использование графовых нейронных сетей (Graph Neural Networks, GNN) для структурированного представления взаимодействий между агентами [5, 6]. GNN позволяют учитывать локальную топологию связей и обрабатывать системы переменного размера, сохраняя при этом возможность обобщения на новые конфигурации.

Дополнительные возможности для повышения эффективности обучения предоставляют методы вычислительного интеллекта (Computational Intelligence, CI), включающие эволюционные алгоритмы и роевую оптимизацию. Данные подходы эффективны в задачах глобального поиска и настройки гиперпараметров нейросетевых архитектур [7–9]. Трансферное обучение (Transfer Learning, TL) позволяет переносить выученные стратегии между родственными задачами, значительно сокращая время адаптации к новым условиям [10, 11].

Особую значимость методы координации мультиагентных систем приобретают в контексте задач одновременной локализации и картографирования (Simultaneous Localization and Mapping, SLAM), где

группа агентов должна совместно исследовать неизвестную территорию, строя карту местности и определяя собственные координаты. Роевой подход к решению SLAM обеспечивает повышенную отказоустойчивость и возможность распараллеливания вычислений по сравнению с централизованными методами [12].

Анализ и верификация поведения мультиагентных роевых систем требует развитых средств визуализации и аналитики. Визуальное представление траекторий агентов, динамики обучения и топологии взаимодействий позволяет исследователям выявлять закономерности группового поведения и оценивать эффективность различных алгоритмов координации [13–15].

Цель данной работы состоит в систематическом анализе современных подходов к построению адаптивных мультиагентных роевых систем с акцентом на синергетическое использование методов MARL, GNN, CI и TL. Рассматриваются возможности применения интегративного подхода к задачам роевого SLAM, а также вопросы визуализации и аналитики поведения роевых систем.

### **Методы построения адаптивных роевых систем**

Современные подходы к построению роевых систем можно разделить на несколько основных категорий: классические методы, основанные на предопределенных правилах взаимодействия, и адаптивные подходы, использующие машинное обучение и методы искусственного интеллекта. Рассмотрим каждую из этих категорий более подробно.

#### *Классические подходы*

Основу традиционных методов координации роевых систем составляют алгоритмы, основанные на локальных правилах взаимодействия между агентами. Алгоритм Boids [1] формализует групповое поведение через три базовых правила: разделение (избегание столкновений с соседями), выравнивание (согласование направления движения) и сплочение (стремление к центру локальной группы). Данный подход обеспечивает эмерджентное формирование стайного поведения без централизованного управления.

Методы потенциальных полей [2] представляют среду в виде скалярного поля, где каждая точка характеризуется потенциальной энергией. Агенты движутся в направлении уменьшения потенциала, что позволяет реализовать притяжение к целям и отталкивание от препятствий. Консенсусные алгоритмы [16] обеспечивают достижение согласованного состояния группы агентов через итеративный обмен информацией с соседями.

Указанные подходы характеризуются вычислительной эффективностью и предсказуемостью поведения в стационарных условиях. Однако их применимость ограничена при необходимости адаптации к изменяющимся условиям среды или выполнения сложных многоэтапных задач, требующих долгосрочного планирования. Для преодоления этих ограничений в последние годы активно развиваются методы машинного обучения для роевых систем.

#### *Мультиагентное обучение с подкреплением*

Мультиагентное обучение с подкреплением представляет естественное расширение классического обучения с подкреплением на случай множественных взаимодействующих агентов. В отличие от одноагентных формулировок, MARL оперирует в нестационарной среде, где действия каждого агента влияют на функции вознаграждения остальных участников системы.

Основные подходы к MARL можно классифицировать на value-based и policygradient методы. К первой категории относятся алгоритмы VDN (Value Decomposition Networks) [17], QMIX [18] и QTRAN [19, 20], которые аппроксимируют совместную функцию ценности через композицию индивидуальных Q-функций агентов. Данные методы обеспечивают монотонность между локальными и глобальными оценками, что критично для корректности децентрализованного исполнения политик.

Policy-gradient подходы, включая MADDPG [18, 21, 22], COMA [18, 23] и MAPPO [18], оптимизируют стратегии агентов напрямую в пространстве параметров политик. MADDPG использует централизованных критиков для каждого агента при сохранении децентрализованных акторов, что позволяет эффективно обучать непрерывные стратегии. COMA вводит контрфактуальные оценки для решения проблемы распределения вознаграждения между агентами.

Парадигма CTDE (Centralized Training with Decentralized Execution) получила широкое распространение в кооперативных задачах. Агенты имеют доступ к расширенной информации о состоянии системы во время обучения, но полагаются исключительно на локальные наблюдения при исполнении, что обеспечивает практическую применимость в условиях ограниченной коммуникации.

Ключевыми ограничениями MARL являются экспоненциальный рост сложности с увеличением числа агентов, нестационарность среды обучения и сложность обеспечения стабильной сходимости. Эти факторы затрудняют применение чистых MARL-подходов к крупномасштабным роевым системам. Частичным решением указанных проблем является использование графовых представлений для структурирования взаимодействий между агентами.

#### *Графовые нейронные сети*

Графовые нейронные сети предоставляют эффективный механизм для обработки структурированных данных, где объекты связаны отношениями произвольной топологии. В контексте мультиагентных систем GNN естественным образом моделируют сетевую структуру взаимодействий между агентами.

Базовая архитектура GNN основана на итеративном процессе передачи сообщений (message passing), где каждый узел агрегирует информацию от своих соседей и обновляет собственное представление. Классические варианты включают Graph Convolutional Networks (GCN) [5, 24], GraphSAGE [6] и Graph Attention Networks (GAT) [6, 25], различающиеся способами агрегации и взвешивания соседних узлов.

Применение GNN в мультиагентных системах позволяет решить несколько фундаментальных проблем. Во-первых, GNN инвариантны к перестановкам узлов, что обеспечивает корректную обработку систем с изменяющимся составом агентов. Во-вторых, параметры сети разделяются между всеми узлами, что позволяет масштабировать алгоритмы на произвольное число агентов без переобучения.

Интеграция GNN с MARL реализуется на различных уровнях архитектуры. GNN могут использоваться для построения эмбедингов состояний, обработки коммуникационных сообщений или формирования централизованных критиков с учетом графовой структуры системы [5, 6, 26, 27]. Такая интеграция особенно эффективна в задачах с явно выраженной пространственной или логической топологией взаимодействий. Дополнительные возможности для повышения эффективности обучения предоставляют методы, не требующие градиентной информации.

#### *Методы вычислительного интеллекта*

Вычислительный интеллект включает семейство алгоритмов оптимизации, не требующих градиентной информации и эффективных в задачах с мультимодальными функциями цели. Основные направления включают эволюционные алгоритмы, роевую оптимизацию и нейроэволюцию.

Эволюционные стратегии (Evolution Strategies, ES) [28, 29] и генетические алгоритмы (Genetic Algorithms, GA) [30] используют принципы естественного отбора для поиска оптимальных решений в пространстве параметров. Алгоритм роя частиц (Particle Swarm Optimization, PSO) [31, 32] моделирует социальное поведение стай для глобальной оптимизации.

В контексте мультиагентных систем CI-методы применяются для настройки гиперпараметров нейросетевых архитектур, оптимизации топологии коммуникационных графов и эволюции стратегий поведения агентов. Нейроэволюция [28] позволяет одновременно оптимизировать архитектуру и веса нейронных сетей, что особенно полезно в задачах с неявно определенной структурой вознаграждения.

Гибридные подходы, сочетающие CI с градиентными методами, демонстрируют перспективные результаты. Например, эволюционные алгоритмы могут использоваться для глобального поиска архитектуры сети с последующей градиентной донастройкой весов [33]. Однако высокая вычислительная стоимость обучения ограничивает применимость CI-методов, что делает актуальными подходы к переносу знаний между задачами.

#### *Трансферное обучение*

Трансферное обучение направлено на повторное использование знаний, полученных при решении одной задачи, для ускорения обучения в родственных задачах. В мультиагентных системах TL особенно актуально ввиду высокой вычислительной стоимости обучения с нуля.

Основные подходы к TL в MARL включают дистилляцию политик [21], где знания экспертных агентов передаются новым участникам системы, и метаобучение [34, 35], позволяющее быстро адаптироваться к новым задачам на основе ограниченного опыта.

Федеративное обучение (Federated Learning) [36] обеспечивает распределенное обучение без централизованного обмена данными, что критично для систем с ограничениями на приватность. Sim2Real-подходы [37] решают проблему переноса политик из симуляции в реальные условия функционирования.

Ключевым вызовом TL является предотвращение негативного переноса, когда знания из исходной задачи ухудшают производительность в целевой области. Для этого применяются методы регуляризации, оценки релевантности задач и архитектурного разделения специфичных и общих компонентов знаний.

#### *Синергетический подход MARL+GNN+CI+TL*

Анализ отдельных методов построения роевых систем показывает, что каждый из подходов обладает специфическими преимуществами и ограничениями. В данной работе предлагается синергетический подход, основанный на интеграции методов мультиагентного обучения с подкреплением, графовых нейронных сетей, вычислительного интеллекта и трансферного обучения. Ключевая идея заключается в том, что недостатки отдельных технологий могут быть компенсированы за счет их комплементарного взаимодействия, что позволяет получить систему с качественно новыми свойствами.

#### **Архитектура интегрированной системы**

Предлагаемая архитектура основана на модульном принципе, где каждый компонент выполняет специализированную функцию в общей системе координации агентов. В основу архитектурного решения положена парадигма CTDE (Centralized Training with Decentralized Execution), обеспечивающая эффективное обучение при сохранении автономности агентов на этапе исполнения.

MARL обеспечивает основной механизм выработки стратегий поведения через взаимодействие со средой. GNN отвечает за структурированное представление взаимодействий между агентами и эффективную коммуникацию в рамках локальных топологий. CI-модули осуществляют глобальную оптимизацию гиперпараметров системы и архитектурных решений. TL-компоненты обеспечивают быструю адаптацию к новым задачам и условиям функционирования.

Центральным элементом архитектуры является GNN-слой, который формирует графовые эмбединги на основе локальных наблюдений агентов и структуры их взаимодействий. Эти эмбединги служат входными данными для MARL-политик, что позволяет агентам принимать решения с учетом состояния ближайшего окружения. Параллельно CI-модули оптимизируют параметры GNN-архитектуры (глубина сети, размерность эмбедингов, механизмы внимания) и MARL-алгоритмов (размер буферов опыта, коэффициенты обучения).

Общая схема взаимодействия компонентов представлена на рисунке 1.

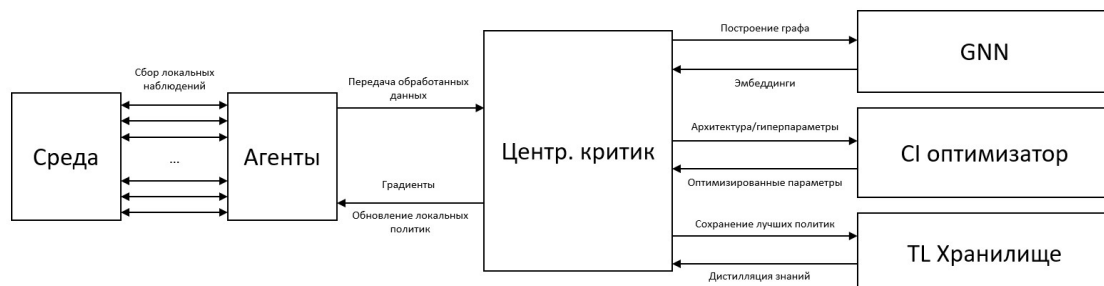


Рисунок 1. Архитектура синергетического подхода MARL+GNN+CI+TL

Особенностью предлагаемой архитектуры является гибкий механизм коммуникации при нарушении прямой связи с централизованным критиком. В случае фрагментации сети агенты используют многошаговую маршрутизацию сообщений через GNN-граф, где промежуточные узлы

выполняют роль ретрансляторов градиентной информации. Это обеспечивает поддержание координации даже при значительных нарушениях топологии сети.

TL-компоненты интегрированы на нескольких уровнях системы. Дистилляция политик позволяет инициализировать новых агентов на основе знаний экспертов.

Метаобучение обеспечивает быструю адаптацию к изменениям в составе роя или параметрах среды. Федеративные схемы позволяют агентам обмениваться обобщенными знаниями без раскрытия локальных данных.

### **Компенсация ограничений методов**

Синергетическое взаимодействие компонентов системы позволяет преодолеть ключевые ограничения отдельных подходов. Проблема масштабируемости MARL решается через применение GNN, которые обеспечивают линейный рост вычислительной сложности с увеличением числа агентов благодаря локальности обработки сообщений. Нестационарность среды обучения компенсируется CI-модулями, которые адаптивно подстраивают параметры системы при обнаружении изменений в динамике среды.

Высокая вычислительная стоимость CI-методов снижается за счет применения TL для инициализации популяций эволюционных алгоритмов на основе предыдущего опыта. Это значительно сокращает число поколений, необходимых для достижения приемлемого качества решений. Проблема негативного переноса в TL минимизируется через использование GNN-эмбедингов, которые предоставляют структурированные представления для оценки релевантности между задачами.

Фиксированная структура классических MARL-архитектур преодолевается через нейроэволюцию топологий GNN, что позволяет системе адаптироваться к изменениям в составе роя или типе решаемых задач. CI-модули обеспечивают поиск оптимальных коммуникационных протоколов и схем агрегации информации в GNN-слоях.

### **Применение к задачам роевого SLAM**

Задача одновременной локализации и картографирования для группы агентов представляет комплексную проблему, требующую координации исследовательского поведения, обмена информацией о наблюдениях и совместного построения карты местности. Синергетический подход естественным образом адаптируется к специфике SLAM-задач.

MARL-компонент обучает агентов стратегиям исследования, балансирующим между детальным изучением известных областей и поиском новых территорий. Функция вознаграждения формируется на основе метрик покрытия территории, качества локализации и эффективности коммуникации между агентами. Централизованный критик в фазе обучения позволяет координировать глобальную стратегию исследования, учитывая состояние всего роя и прогресс картографирования.

GNN-слой моделирует пространственные и информационные связи между агентами. Узлы графа соответствуют агентам и ключевым точкам карты, ребра отражают возможности коммуникации и пространственную близость. Механизм передачи сообщений позволяет агентам обмениваться наблюдениями и координировать планы исследования без централизованной координации.

При фрагментации роя на изолированные подгруппы (что характерно для SLAM в сложных средах с препятствиями) каждая группа способна поддерживать локальное согласование стратегий через реер-to-реер-дистилляцию политик. Агенты используют ранее синхронизированные параметры для построения согласованных локальных карт. При восстановлении связи между подгруппами происходят синхронизация и слияние карт на основе общих ориентиров и доверительных метрик качества наблюдений.

CI-модули оптимизируют параметры алгоритмов SLAM (пороги обнаружения петель, стратегии слияния карт) и конфигурацию роя (количество агентов, распределение ролей). Эволюционные алгоритмы позволяют найти компромисс между точностью картографирования и вычислительными ресурсами.

TL-компоненты обеспечивают адаптацию к новым типам местности и условиям наблюдения. Политики, обученные в симуляции, адаптируются к реальным условиям через механизмы доменной адаптации. Метаобучение позволяет быстро настраивать параметры системы при переходе между помещениями и открытыми пространствами.

### Сравнительный анализ подходов

Для систематического сравнения различных подходов к построению роевых систем рассмотрим ключевые характеристики по критериям масштабируемости, устойчивости, скорости обучения и адаптивности (табл. 1).

Таблица 1. Сравнительный анализ подходов к построению роевых систем

Подход	Масштабируемость	Устойчивость	Скорость обучения	Адаптивность
MARL	Средняя: экспоненциальный рост сложности ограничивает большие рои	Низкая: критик снижает неустойчивость, но сбои агентов требуют перезапуска	Низкая: требуется много эпизодов, особенно для off-policy методов	Средняя: обучаются новые стратегии, но без TL требуется долгое дообучение
GNN	Хорошая: инвариантность к числу агентов, теоретически масштабируемы	Высокая при структурной связи: работает при отказе отдельных узлов	Средняя: добавляет коммуникацию, замедляя шаги обучения	Хорошая: обрабатывает изменяющиеся связи и новых агентов
CI	Высокая для параллелизма: хорошо масштабируется на кластерах	Умеренная: стойки к шуму, но при сбоях требуется перезагрузка	Низкая: требуют большого количества оценок	Низкая: переобучение при изменениях среды
Transfer Learning	Высокая: переносит опыт, снижая нагрузку при масштабах	Высокая: улучшает устойчивость, агенты быстро переобучаются	Очень высокая: использование предобученных компонентов ускоряет обучение	Очень высокая: перенос политик обеспечивает адаптацию «на лету»
MARL + GNN + CI + TL	Очень высокая: комбинированное решение масштабируется на большие роевые структуры	Очень высокая: устойчивость за счет GNN-коммуникации и CI-модулей	Высокая: TL ускоряет инициализацию, CI оптимизирует архитектуру	Очень высокая: адаптивная коммуникация и совместное обучение политик

Представленное в таблице сравнение демонстрирует преимущества предлагаемого синергетического подхода по всем рассматриваемым критериям. Новым является подход, при котором четыре различные технологии интегрированы в единую архитектуру, где каждый компонент не только выполняет свою функцию, но и активно компенсирует ограничения других методов. Ключевым преимуществом является способность системы адаптироваться к изменениям масштаба, состава роя и условий функционирования без существенной деградации производительности, что принципиально отличает предлагаемое решение от существующих подходов, использующих максимум два-три из рассматриваемых компонентов.

Практическая реализация синергетического подхода требует тщательной балансировки между компонентами системы. Чрезмерная сложность архитектуры может приводить к увеличению времени обучения и снижению интерпретируемости поведения агентов. Необходимо развитие методов автоматической настройки весов различных компонентов в зависимости от специфики решаемых задач.

### Заключение

В данной работе представлен систематический анализ современных подходов к построению адаптивных мультиагентных роевых систем с акцентом на их применение в задачах визуализации и

аналитики. Рассмотрены ключевые ограничения классических методов координации роев и показаны возможности их преодоления за счет интеграции методов искусственного интеллекта.

Предложенный синергетический подход, объединяющий мультиагентное обучение с подкреплением, графовые нейронные сети, методы вычислительного интеллекта и трансферное обучение, демонстрирует качественные преимущества по сравнению с использованием отдельных технологий. Ключевые достоинства интегративного решения включают повышенную масштабируемость за счет локальности GNN-обработки, улучшенную устойчивость к отказам через адаптивную реконфигурацию системы и сокращение времени адаптации к новым задачам благодаря механизмам трансферного обучения.

Особое внимание уделено применению предлагаемого подхода к задачам роевой разведки местности и одновременной локализации и картографирования. Рассмотрена адаптация синергетической архитектуры к специфике SLAM-задач, включая координацию исследовательского поведения агентов и распределенное построение карт местности. Кроме того, предлагаемый подход открывает новые возможности для моделирования и визуализации сложных динамических систем, включая физические процессы, биологические популяции и социальные сети, где поведение множества взаимодействующих элементов определяет макроскопические свойства системы.

Вместе с тем практическая реализация комплексных систем, интегрирующих четыре различные технологии, сопряжена с рядом технических вызовов. К основным ограничениям относятся увеличение архитектурной сложности, необходимость тщательной балансировки параметров различных компонентов и снижение интерпретируемости поведения системы. Данные аспекты требуют дальнейших исследований в направлении разработки методов автоматической настройки и анализа гибридных архитектур.

Перспективными направлениями развития являются исследование применимости синергетического подхода к другим классам задач коллективного интеллекта, разработка специализированных метрик для оценки эффективности интегрированных роевых систем, а также создание инструментов интерактивной визуализации для анализа динамики обучения и поведения больших групп агентов. В контексте научной визуализации особый потенциал представляет использование адаптивных роевых систем для моделирования и визуального анализа многочастичных систем, турбулентных течений, эволюционных процессов и других явлений, характеризующихся сложной пространственно-временной динамикой. Важным направлением является интеграция методов объяснимого искусственного интеллекта для повышения прозрачности принятия решений в мультиагентных системах и разработка визуальных интерфейсов для интерактивного анализа коллективного поведения.

### Список литературы

1. C. W. Reynolds, Flocks, herds and schools: A distributed behavioral model, in: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '87, Association for Computing Machinery, New York, NY, USA, 1987, p. 25–34. URL: <https://doi.org/10.1145/37401.37406>. doi:10.1145/37401.37406.
2. O. Khatib, Real-time obstacle avoidance for manipulators and mobile robots, in: Proceedings. 1985 IEEE international conference on robotics and automation, volume 2, IEEE, 1985, pp. 500–505.
3. I. N. Dubina, N. M. Oskorbin, D. S. Khvalynskiy, Decision-making coordination in hierarchical systems, World of Economics and Management 19 (2019) 5–18. URL: <https://doi.org/10.25205/2542-0429-2019-19-2-5-18>. doi:10.25205/2542-0429-2019-19-2-5-18, in Russian.
4. S. V. Albrecht, F. Christianos, L. Schöafer, Multi-Agent Reinforcement Learning: Foundations and Modern Approaches, MIT Press, 2024. URL: <https://www.marl-book.com/>.
5. A. Goeckner, Y. Sui, N. Martinet, X. Li, Q. Zhu, Graph neural network-based multiagent reinforcement learning for resilient distributed coordination of multi-robot systems, in: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2024, pp. 5732–5739.
6. Z. Liu, J. Zhang, E. Shi, Z. Liu, D. Niyato, B. Ai, X. S. Shen, Graph neural network meets multi-agent reinforcement learning: Fundamentals, applications, and future directions, IEEE Wireless Communications (2024).
7. Z. Huang, S. Batra, T. Chen, R. Krupani, T. Kumar, A. Molchanov, A. Petrenko, J. A. Preiss, Z. Yang, G. S. Sukhatme, Quadswarm: A modular multi-quadrotor simulator for deep reinforcement learning with direct thrust control, 2023. URL: <https://arxiv.org/abs/2306.09537>. arXiv:2306.09537.



8. Охрименко, Методы вычислительной математики в нейронных сетях, in: Международная научно-техническая конференция молодых ученых БГТУ им. ВГ Шухова, посвященная 170-летию со дня рождения ВГ Шухова, 2023, pp. 354–357.
9. Родзин, Вычислительный интеллект: немонотонные логики и графическое представление знаний, Программные продукты и системы (2002) 21–23.
10. M. Zhao, G. Wang, Q. Fu, X. Guo, Y. Chen, T. Li, X. Liu, Mw-maddpg: a metalearning based decision-making method for collaborative uav swarm, *Frontiers in Neurorobotics* 17 (2023) 1243174.
11. A. W. Mamond, M. Kundroo, S.-e. Yoo, S. Kim, T. Kim, Fldqn: Cooperative multi-agent federated reinforcement learning for solving travel time minimization problems in dynamic environments using sumo simulation, *Sensors* 25 (2025) 911.
12. Касаткин, Приходько, Роевые алгоритмы одновременной локализации и построения карты, in: Прикладная математика: современные проблемы математики, информатики и моделирования, ФГБУ "Российское энергетическое агентство" Минэнерго России Краснодарский ЦНТИ – филиал ФГБУ "РЭА" Минэнерго России, Краснодар, 2022, pp. 124–128. In Russian.
13. A. Dorri, S. S. Kanhere, R. Jurdak, Multi-agent systems: A survey, *Ieee Access* 6 (2018) 28573–28593.
14. Бронников, Адаптивное визуальное управление производственным агентом, Технология приборостроения (2014) 3–6.
15. Фролов, Применение многоагентных технологий при построении интеллектуальных компьютерных обучающих систем, Гаудеамус (2013) 223–225.
16. R. Olfati-Saber, J. A. Fax, R. M. Murray, Consensus and cooperation in networked multi-agent systems, *Proceedings of the IEEE* 95 (2007) 215–233.
17. P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls, T. Graepel, Value-decomposition networks for cooperative multi-agent learning, 2017. URL: <https://arxiv.org/abs/1706.05296>. arXiv:1706.05296.
18. G. Papoudakis, F. Christianos, L. Schöfer, S. V. Albrecht, Benchmarking multiagent deep reinforcement learning algorithms in cooperative tasks, *arXiv preprint arXiv:2006.07869* (2020).
19. T. Rashid, M. Samvelyan, C. S. De Witt, G. Farquhar, J. Foerster, S. Whiteson, Monotonic value function factorisation for deep multi-agent reinforcement learning, *Journal of Machine Learning Research* 21 (2020) 1–51.
20. K. Son, D. Kim, W. J. Kang, D. E. Hostallero, Y. Yi, Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning, in: *International conference on machine learning*, PMLR, 2019, pp. 5887–5896.
21. S. Wadhwan, D.-K. Kim, S. Omidshafiei, J. P. How, Policy distillation and value matching in multiagent reinforcement learning, in: *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 2019, pp. 8193–8200.
22. R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, I. Mordatch, Multi-agent actor-critic for mixed cooperative-competitive environments, *Advances in neural information processing systems* 30 (2017).
23. J. Foerster, G. Farquhar, T. Afouras, N. Nardelli, S. Whiteson, Counterfactual multi-agent policy gradients, in: *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
24. T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2017. URL: <https://arxiv.org/abs/1609.02907>. arXiv:1609.02907.
25. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, 2018. URL: <https://arxiv.org/abs/1710.10903>. arXiv:1710.10903.
26. L. Ratnabala, A. Fedoseev, R. Peter, D. Tsetserukou, Magnnet: Multi-agent graph neural network-based efficient task allocation for autonomous vehicles with deep reinforcement learning, *arXiv preprint arXiv:2502.02311* (2025).
27. H. Park, B. Seong, S.-K. Ko, Spectra: Scalable multi-agent reinforcement learning with permutation-free networks, *arXiv preprint arXiv:2503.11726* (2025).
28. N. Vaughan, Multi-agent reinforcement learning for swarm retrieval with evolving neural network, in: *Biomimetic and Biohybrid Systems: 7th International Conference, Living Machines 2018, Paris, France, July 17–20, 2018, Proceedings 7*, Springer, 2018, pp. 522–526.
29. T. Salimans, J. Ho, X. Chen, S. Sidor, I. Sutskever, Evolution strategies as a scalable alternative to reinforcement learning, 2017. URL: <https://arxiv.org/abs/1703.03864>. arXiv:1703.03864.
30. Z. Zhu, C. Yu, J. Wang, A hybrid genetic algorithm and proximal policy optimization system for efficient multi-agent task allocation, *Systems* 13 (2025) 453.
31. L. Leuzzi, S. Jones, S. Hauert, D. Bacciu, A. Cossu, Lifelong evolution of swarms, 2025. URL: <https://arxiv.org/abs/2503.17763>. arXiv:2503.17763.

- 
32. S. Lin, J. Wang, B. Huang, X. Kong, H. Yang, Bio particle swarm optimization and reinforcement learning algorithm for path planning of automated guided vehicles in dynamic industrial environments, *Scientific Reports* 15 (2025) 463.
  33. E. Real, A. Aggarwal, Y. Huang, Q. V. Le, Regularized evolution for image classifier architecture search, in: *Proceedings of the aaai conference on artificial intelligence*, volume 33, 2019, pp. 4780–4789.
  34. A. S. Nipu, S. Liu, A. Harris, Enabling multi-agent transfer reinforcement learning via scenario independent representation, in: *2023 IEEE Conference on Games (CoG)*, 2023, pp. 1–8. doi:10.1109/CoG57401.2023.10333236.
  35. C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: *International conference on machine learning*, PMLR, 2017, pp. 1126–1135.
  36. B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, Communicationefficient learning of deep networks from decentralized data, in: *Artificial intelligence and statistics*, PMLR, 2017, pp. 1273–1282.
  37. J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel, Domain randomization for transferring deep neural networks from simulation to the real world, in: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 2017, pp. 23–30.

УДК 004.9, 666.3.032, 620.1, 532.13, 621.74  
DOI: 10.25686/978-5-8158-2474-4-2025-451-458

## Исследование процесса литья керамического шликера под низким давлением

М. А. Петров<sup>1</sup>, М. В. Корнюшин<sup>2</sup>, А. В. Смирнов<sup>2</sup>, В. В. Рыбальченко<sup>3</sup>

<sup>1</sup>Московский политехнический университет, Москва, Россия

<sup>2</sup>МИРЭА – Российский технологический университет, Москва, Россия

<sup>3</sup>Российский университет транспорта, Москва, Россия

**Аннотация.** Современные технологии изготовления керамических деталей развиваются вместе с развитием и самих инженерных материалов. Интересным является способ литья керамической суспензии под низким давлением. В статье рассматриваются эксперименты по изучению реологических характеристик корундового шликера, который при температурах литья представляет собой густую суспензию, и численное моделирование простых и сложных прототипов деталей. Получено хорошее совпадение результатов по экспериментальному и численному изучению литейной способности шликера.

**Ключевые слова:** алумооксидная керамика, вязкость керамической суспензии, течение суспензии, литейная способность суспензии, литьё под низким давлением, численное моделирование, CFD.

## Research into the low pressure casting of ceramic slurry

M. A. Petrov<sup>1</sup>, M. V. Korniyushin<sup>2</sup>, A. V. Smirnov<sup>2</sup>, V. V. Rybalchenko<sup>3</sup>

<sup>1</sup>Moscow Polytechnic University, Moscow, Russia

<sup>2</sup>MIREA – Russian Technological University, Moscow, Russia

<sup>3</sup>Russian University of Transport (MIIT), Moscow, Russia

**Abstract.** Modern technologies for ceramic parts manufacturing are developing parallel with the development of engineering materials. A low-pressure casting process of slurry is a rather new one process, applied to engineering ceramics. To master it, a series of experiments are conducted to study the rheological characteristics of corundum-based slurry, which represents thick slurry at casting temperatures, and numerical simulation of simple and complex prototypes manufacturing. Good agreement was obtained between the results of experimental and numerical study of the slurry's flowability.

**Keywords:** alumina ceramics, viscosity of ceramic slurry, slurry flow, flowability of ceramic slurry, low-pressure casting, numerical simulation, CFD.

### Введение

Изготовление керамических деталей является актуальной задачей. Технологии обработки керамических материалов до готовых изделий включают следующие способы формообразования: сухое, мокрое и пластическое. При сухом варианте керамические порошки спрессовывают или компактируют под высокими давлениями (одноосное или всестороннее сжатие). При мокром варианте керамический материал доводится до вида суспензии, которой далее одним из методов литья (шликерное, на ленту, намораживание, под давлением и т.д.) придаётся требуемая форма. В третьем варианте керамический материал представляет собой пластическую массу, которой методами обработки давлением придают нужную форму. После выполнения формообразующей операции получаемые «зелёные», хрупкие изделия подвергают термообработке, в процессе которой удаляется связующее и достигается консолидация частиц, повышаются прочностные характеристики детали.

Технология горячего шликерного литья (ГШЛ) [1, 2] расширила группу технологических процессов по работе с керамическими суспензиями. На рисунке 1 представлена УГШЛ, которая по своей схеме напоминает установку для литья под низким давлением или ЛНД (данный способ также известен, как литьё под регулируемым давлением, литьё на компрессорных машинах), применяемую в литейном производстве [3, 4], для которого избыточное давление, оказывающее воздействие на поверхность расплава, составляет, как правило, не более 1 атм. Преимуществом данного процесса является плавное заполнение полости формы за счёт создания ламинарного потока шликера, поднимающегося из тигля по литнику (металлопроводу) к питателю и в форму. Время заполнения и значение прикладываемого внешнего давления ( $P$ ) регулируется и влияет на качество заполнения формы, согласно уравнению (1). Ламинарное течение позволяет избежать турбулентно-дисперсионного течения материала в форме. Соблюдение ламинарного течения является условием выполнения условия, согласно уравнению (2). В

упрощённом виде скорость потока и/или заполнения формы может быть определена по уравнению (3) [5, 6].

$$P = P_{\phi} + P_m = k\rho g(h_{\phi} + h_m), \quad (1)$$

$$Re_{\text{лам}} = \frac{vd}{\nu} = \frac{vd\rho}{\eta} < 2300, \quad (2)$$

$$v = k \sqrt{\frac{2P}{\rho}}, \quad (3)$$

где  $P_{\phi}$  и  $P_m$  – давление, необходимое для заполнения формы и для поднятия суспензии по металлопроводу соответственно, [Па];  $\rho$  – плотность суспензии, [кг/м<sup>3</sup>];  $g$  – ускорение свободного падения, [м/с<sup>2</sup>];  $h_{\phi}$  и  $h_m$  – высота формы отливки и металлопровода соответственно, [м];  $k$  – коэффициент, учитывающий гидравлические сопротивления, [-];  $\nu$  – скорость потока суспензии, [м/с];  $d$  – средний диаметр потока, [м];  $\nu$  – кинематическая вязкость, [м<sup>2</sup>/с];  $\eta$  – динамическая вязкость, [Па с].

### Постановка задачи

Исследуемый материал – парафиновый шликер на основе корундовой шихты с содержанием оксида алюминия ( $\alpha\text{-Al}_2\text{O}_3$ ) не менее 94 %, марки ВК-94-1 или 22ХС. Химический состав вакуумплотной керамики представлен в таблице 1. Шликер находится в твёрдом состоянии при комнатной температуре и плавится/кристаллизуется в температурном интервале 70 °С...100 °С.

Таблица 1. Химический состав ВК-94-1 (согласно ГОСТ Р 70659-2023)

Al <sub>2</sub> O <sub>3</sub>	SiO <sub>2</sub>	Cr <sub>2</sub> O <sub>3</sub>	MgO <sub>2</sub>
94,4	2,8	0,5	2,3

На рисунке 1 представлена упрощённая схема установки горячего шликерного литья (УГШЛ) от компании «ЭКОН» [2], являющаяся также расчётной схемой для численного моделирования процесса ЛНД. Задачи для численного моделирования разделены на две группы. Первая группа задач направлена на определение текучести материала. Задачи второй группы подразумевают изготовление нескольких отливок-демонстраторов. Детали размещаются внутри формы по одной, а также в виде гроздей (рис. 2).

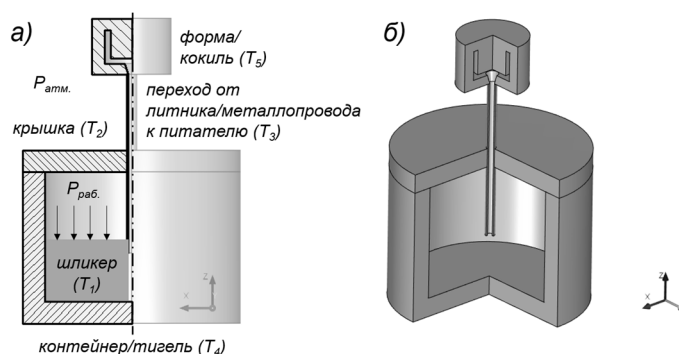


Рисунок 1. Исходная геометрия в 2D-постановке (а), с указанием пяти контролируемых областей нагрева, и 3D-постановке (б)

При литье керамического шликера необходимо учитывать процессы теплообмена с тиглем и формой, так как при температуре свыше 160 °С парафин начинает кипеть и испаряться. Поскольку перед заливкой формы подвергаются вакуумированию, то растворимость выделяющегося газа в жидкой фазе снижается. Этот этап необходим для устранения внутренней пористости керамической отливки. В процессе движения шликера по каналам его жидкотекучесть может изменяться. Так, с падением температуры увеличивается внутреннее трение или динамическая/кинематическая вязкость, и жидкотекучесть снижается.

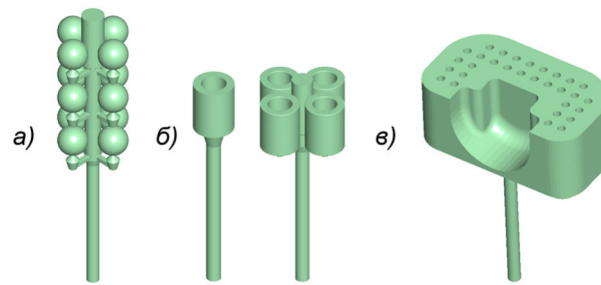


Рисунок 2. Группа отливок, рассматриваемых в рамках задачи 2: шарик (а), стаканчик (б) и смеситель (в)

Описание изменения вязкости материала или его реологические свойства с изменением температуры могут быть описаны уравнением (4) для густой суспензии, которое состоит из двух слагаемых: первое слагаемое отвечает за характер течения суспензии внутри ёмкости по всему объёму, а второе – за пристеночное течение, а также уравнением (2), закон Гершеля-Балкли, согласно уравнению (5) [7, 8, 9].

$$P = 2(\sigma_0 + \alpha_1 V^m) \ln\left(\frac{D_0}{D}\right) + 4(\tau_0 + \beta_1 V^n) \frac{L}{D}, \quad \sigma_0 = \frac{F}{\pi L^2 \tan^2 \theta}, \quad (4)$$

$$\sigma = \sigma_0 + k\dot{\gamma}^n \quad (5)$$

где  $\sigma_0$  – предел текучести материала, измеренный при помощи конусного пенетromетра, [Па];  $F$  – прикладываемая сила, [Н];  $\theta$  – половина угла конуса при вершине, [град];  $L$  – глубина внедрения, [м];  $\dot{\gamma}$  – скорость сдвига, [ $\text{с}^{-1}$ ]).

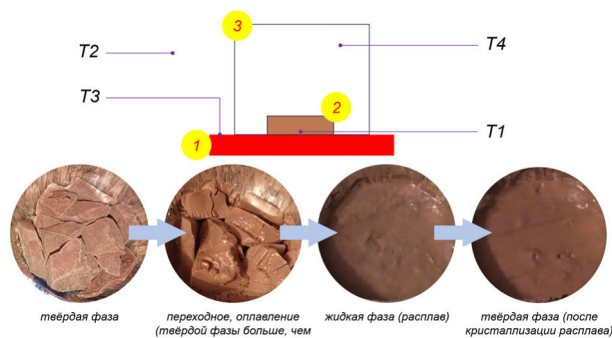
Целью исследования является численное моделирование процесса ЛНД для получения отливок сложной конфигурации из корундовой керамики. Обозначенная цель достигается выполнением следующих задач: 1) выполнение экспериментов для оценки реологических свойств керамического шликера; 2) геометрическое проектирование области для выполнения численного моделирования; 3) проверка качества геометрии при помощи технологии быстрого прототипирования – настольная фотополимерная 3D-печать (технология LCD); 4) задание граничных условий и свойств шликера; 5) анализ полученных результатов.

### Экспериментальные исследования

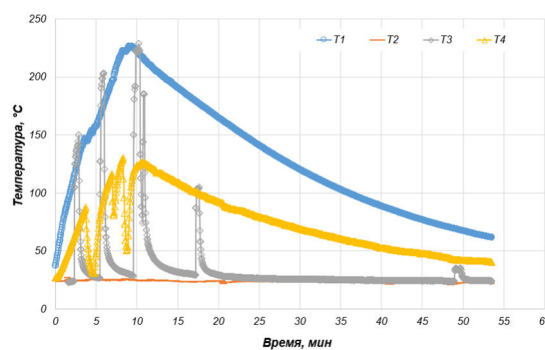
На рисунке 3 показана схема выполнения эксперимента по характеристике материала с определением контрольных интервалов температур стабильного агрегатного состояния и его визуальной фиксацией. На нагреваемое основание (1) устанавливается тигель с материалом (2) и закрывается защитным экраном для уменьшения конвективного теплообмена с окружающей средой. При помощи четырёх термопар типа К выполняется измерение температур: в донной части тигля (Т1), окружающей среды (Т2), периодическое измерение температуры основания (Т3) и на расстоянии 100 [мм] от основания тигля внутри защитного экрана (Т4). Литейная способность начинает резко снижаться при температуре 80 °С и ниже. В жидком состоянии шликер является весьма густым.

Определение вязкости выполнено на реометре AntonPaar QC по схеме «цилиндр–цилиндр» ( $\varnothing 27$  мм), оборудованном нагревательной камерой C-PTD 180/AIR/QC. На рисунке 4 представлены графики изменения динамической вязкости в интервале температур 70...110 °С, для максимальной скорости сдвига/среза  $\dot{\gamma} = 100$  [ $\text{с}^{-1}$ ].

Оценка текучести шликера выполнялась в специальной оснастке ПЛС-1 (рис. 5), устанавливаемой вертикально на машине ЭКОН-УГШЛ. Исходными условиями проведения эксперимента являлись давление на входе 2...3 [атм], температура шликера  $t_{\text{ш}} = 75$  °С, температура формы  $t_{\text{ф}} = 20$  °С. В зависимости от процента наполнителя ( $\omega = 10...12$  %<sub>мас.</sub>) и давления высота подъёма шликера составляла 65...122 [мм]. Следует отметить, что при одном и том же проценте наполнения литейные способности увеличиваются с повышением прикладываемого давления.



а) схема постановки эксперимента и изменение агрегатного состояния шликера с повышением температуры



б) результаты измерений температур

Рисунок 3. Изучение консистенции шликера

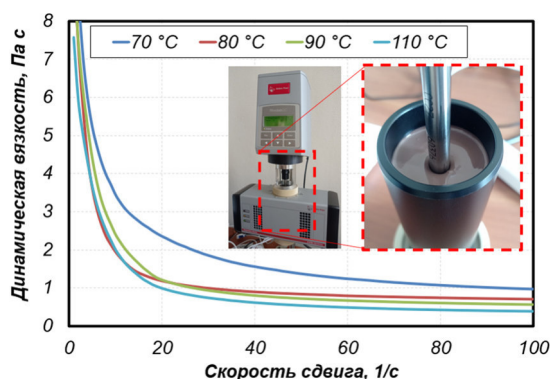


Рисунок 4. Графики изменения динамической вязкости от скорости сдвига

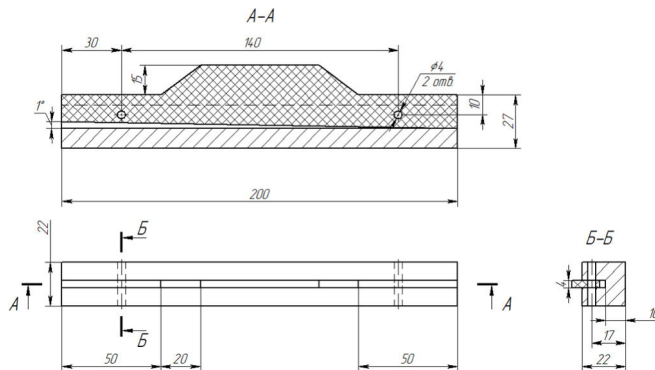


Рисунок 5. Чертёж оснастки для ПЛС-1

### Численное моделирование

Численное моделирование в трёхмерной постановке выполнялось по методу многофазного течения материалов, находящихся в разных агрегатных состояниях. Volume of Fluid (VOF) [10, 11], позволяющий определить долю вещества в определённый момент времени, для которого решается уравнение переноса фаз, представлено уравнением (6). В качестве веществ выбраны воздух (газовая фаза) и парафин (жидкая фаза).

$$\frac{\partial f}{\partial t} + V \nabla f = 0, \quad (6)$$

где  $f$  – объёмная доля фазы в ячейке.

Абсолютные значения температуры и давления ( $T_{\text{abs}}$ ,  $P_{\text{abs}}$ ) рассматриваются как суммы опорных значений ( $T_{\text{ref}}$ ,  $P_{\text{ref}}$ ) и избыточных, прикладываемых величин ( $T$ ,  $P$ ) [11]. За опорные значения приняты нормальные атмосферные условия:  $T_{\text{ref}} = 293,15$  [K],  $P_{\text{ref}} = 101000$  [Па]. Рассматривались две задачи на основании которых были сделаны выводы о наиболее благоприятных значениях граничных условий проведения процесса ЛНД. В первой задаче проверялись литейные способности шликера (ПЛС-1), а во второй задаче рассматривался собственно процесс изготовления прототипов отливок деталей.

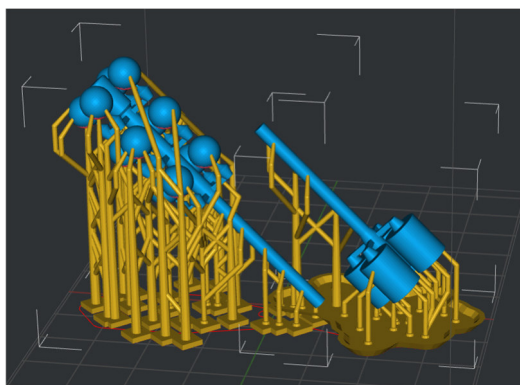
Массовая скорость шликера определялась как отношение массового расхода к площади поперечного сечения входа (inlet). Для задачи проверки литейной способности шликера (задача 1) он составил  $715$  [кг/м<sup>2</sup> с], а для серии задач изготовления прототипов деталей (задача 2) –  $2105$  [кг/м<sup>2</sup> с]. Теоретическое время заполнения при заданном общем массовом расходе  $0,01$  [кг/с] для задачи 1 составило  $0,126$  [с], а для задачи 2 время заполнения  $0,0341$  [с]. Шаг по времени составил  $0,001$  [с].

### Проверка корректности геометрии модели

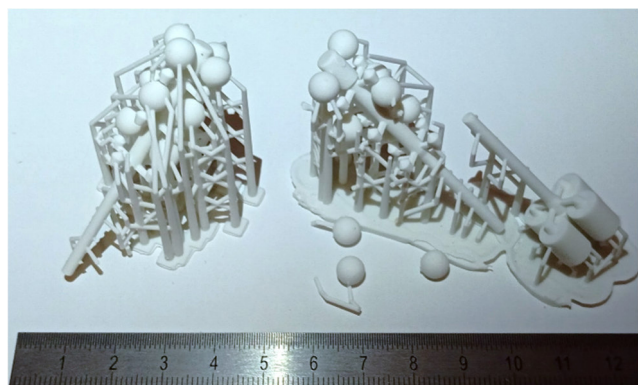
Для численного моделирования процессов вычислительной гидродинамики (CFD) применяются специальные форматы представления геометрии. Для этого в программах компьютерного

геометрического моделирования (CAD) инструментами твердотельного моделирования проектируют интересующий объект, который затем экспортируют в формат полигональной или фасеточной 3D-модели (\*.stl, \*.vrm, \*.obj, \*.ply и другие). Качество полигональной модели (ПМ) может не всегда удовлетворять требованиям, которые необходимы для выполнения численного моделирования. В современных CFD-программах, в том числе и в FlowVision, встроены инструменты контроля и исправления геометрии, помогающие сократить время на подготовку геометрии. Ещё одним способом проверки как качества геометрии, так и самой разработанной в виртуальной среде инструментами цифрового проектирования геометрии, является её изготовление методами быстрого прототипирования, например, при помощи настольной фотополимерной 3D-печати (метод Liquid Cristal Display (LCD)).

В качестве материала был выбран керамоподобный фотополимерный материал с массовой долей наполнителя не более  $\omega = 10...15 \text{ \%}_{\text{мас.}}$ . Для 3D-печати выбрано оборудование Creality HalotOne. Подготовка управляющей программы выполнена в программе Halot Box. На рисунке 6 показаны этапы подготовки и результаты 3D-печати. Установлено, что диаметр канальцев, связывающих шарики с питателями, достаточно хрупки, что может привести к их разрушению при извлечении прототипов из формы УГШЛ в случае реализации технологии ЛНД. Также оказалось, что для изучаемых прототипов важны правильный выбор угла наклона к основной платформе и наличие поддерживающих структур.



а) 3D-модели и поддерживающие структуры



б) прототипы отливок деталей (вид сверху-сбоку), после 3D-печати, промывки и отверждения

Рисунок 6. Подготовка управляющей программы (gcode, а) и готовые прототипы отливок (б)

Это обстоятельство может быть также важно и при ЛНД, так как канальцы могут играть роль местного сопротивления, что будет менять характер течения шликера. Полезным будет отметить и то, что из нескольких исследованных коммерческих керамоподобных шликеров при общем времени отверждения в УФ-печи 30 минут лишь при  $\omega = 15 \text{ \%}_{\text{мас.}}$  тактильно чувствуется выступание наполнителя на поверхности прототипов. Отделение от поддержек целесообразнее производить до проведения отверждения в УФ-печи, так как поддерживающие структуру более пластичны до обработки, чем после. Отсутствие поддерживающих структур не позволяет напечатать шарики. Вышеизложенных проблем в случае изготовления грозди стаканчиков не наблюдается.

### Обсуждение результатов

Установлено, что при разной температуре стенки (60 °С против 80 °С) вид полей фазового распределения (VOF) и температуры практически не изменяется, но изменяются численные значения, например 83,625 °С против 87,886 °С. Поле же динамической вязкости отличается, так как в случае более высокой температуры стенки вязкость шликера ниже и его жидкотекучесть выше (рис. 7). При увеличении температуры наблюдается и рост проходимого шликером расстояния или его литейная способность, которое составляет 87,4 [мм] при температуре стенки 80 °С. Так, в эксперименте при температуре стенки 20 °С, давлении подачи 0,3 [МПа], или 3 [атм], и  $\omega = 10 \text{ \%}_{\text{мас.}}$  расстояние составляет 85 [мм].



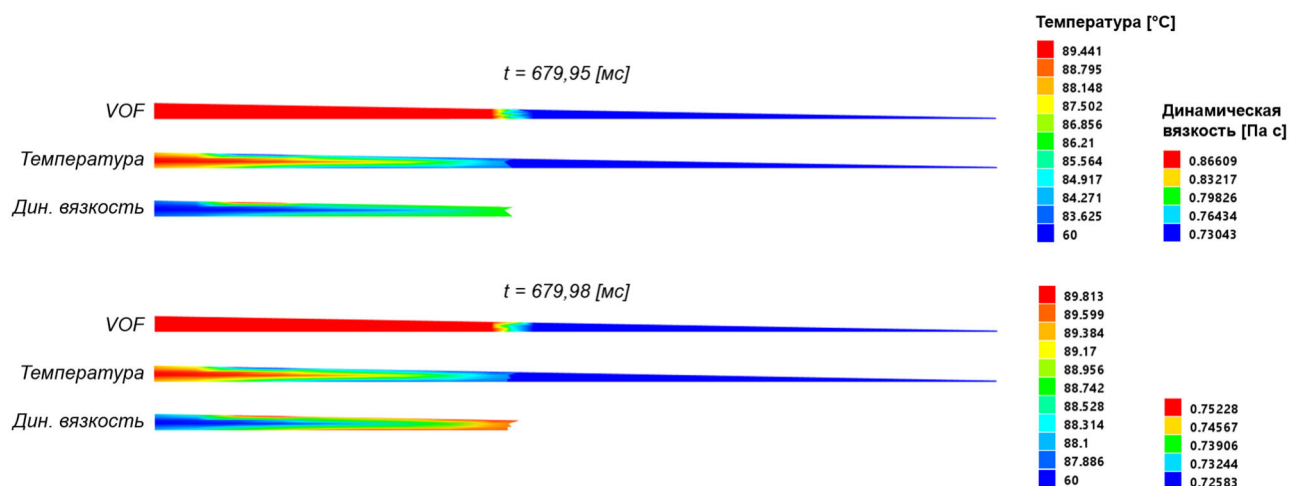


Рисунок 7. Расчётные поля для задачи 1, при разной температуре стенки: 60 °C (вверху) и 80 °C (внизу)

На рисунке 8 представлены полученные результаты по моделированию заливки грозди из 12 шариков. Начиная со среднего уровня, заполнение происходит неравномерно. К итоговому расчётному моменту времени 144,31 [мс] полного заполнения формы шликером не наблюдается, что указывает на необходимость увеличения входного давления. Задача о заполнении одного шарика показала, что, наоборот, давление на входе весьма высоко, так как вместо плавного заполнения формы наблюдается образование струи в полости шарика (эффект «фонтана»), её удар о верхнюю часть стенки и последующее турбулентно-дисперсное заполнение полости.

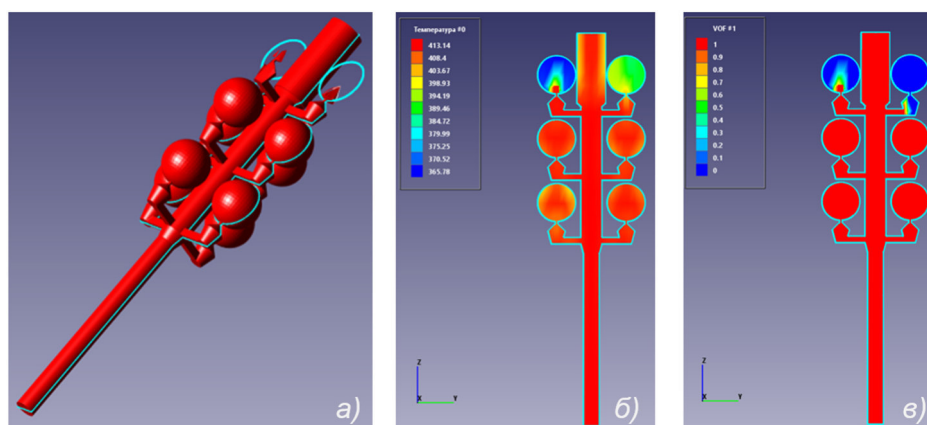


Рисунок 8. Поля распределения: доли шликера к моменту времени 173,57 [мс] (а) и 172,02 [мс] (б), и температуры к моменту времени 171,82 [мс] (в)

На рисунке 9 представлено распространение жидкой фазы (красный цвет) в рассматриваемом объёме для двухмерной и трёхмерной визуализации результатов по плоскости и по объёму соответственно для разных моментов времени (рис. 10). На основании полученной информации можно сделать вывод о том, что в области металлопровода происходит нерегулируемое заполнение по высоте, резкий удар в нижнюю часть формы, что может привести, как и в случае с отливкой шарика, к образованию турбулентно-дисперсионных потоков при смене направления движения суспензии и сильном её перегреве на этапе заполнения донной части стаканчика.

Также было установлено, что ближе к окончанию процесса заполнения формы в её верхней части наблюдается волнение жидкости, которое может нарушать процесс ламинарного течения, с образованием пор в верхних областях. Общее время заполнения формы составило 32,77 [мс] (рис. 10).

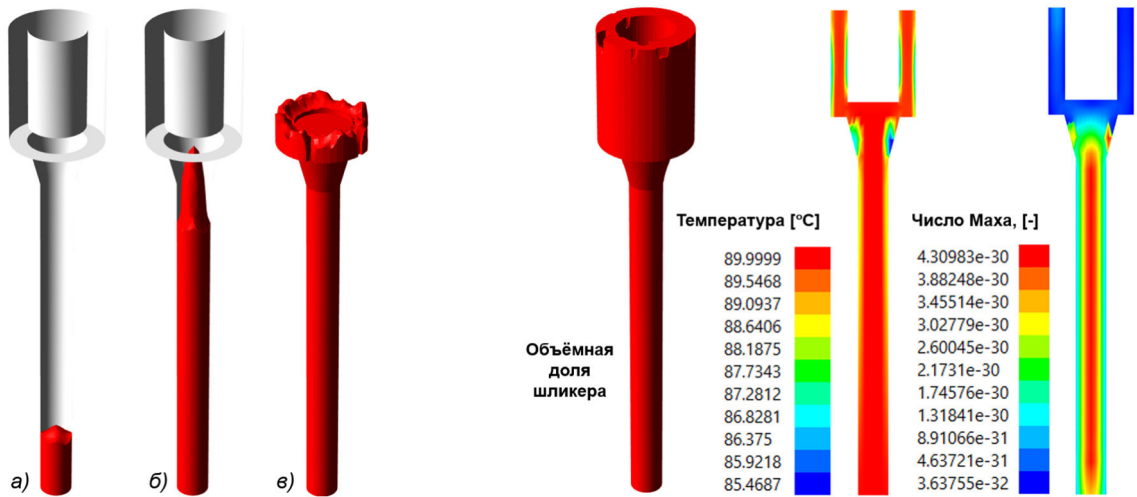


Рисунок 9. Распределение жидкой фазы в процессе её подачи в разные моменты времени: (а)  $t=1,35$  [мс]; (б)  $t=10,40$  [мс]; (в)  $t=18,35$  [мс]

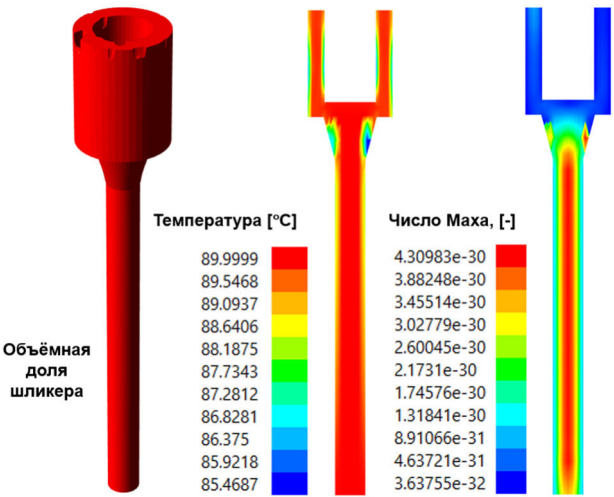


Рисунок 10. Расчётные поля к моменту времени  $t = 32,77$  [мс]

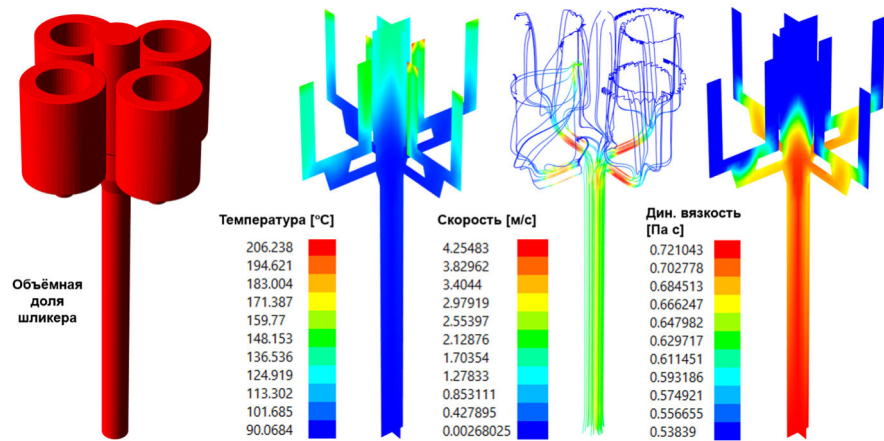


Рисунок 11. Расчётные поля к моменту времени  $t = 170,08$  [мс]

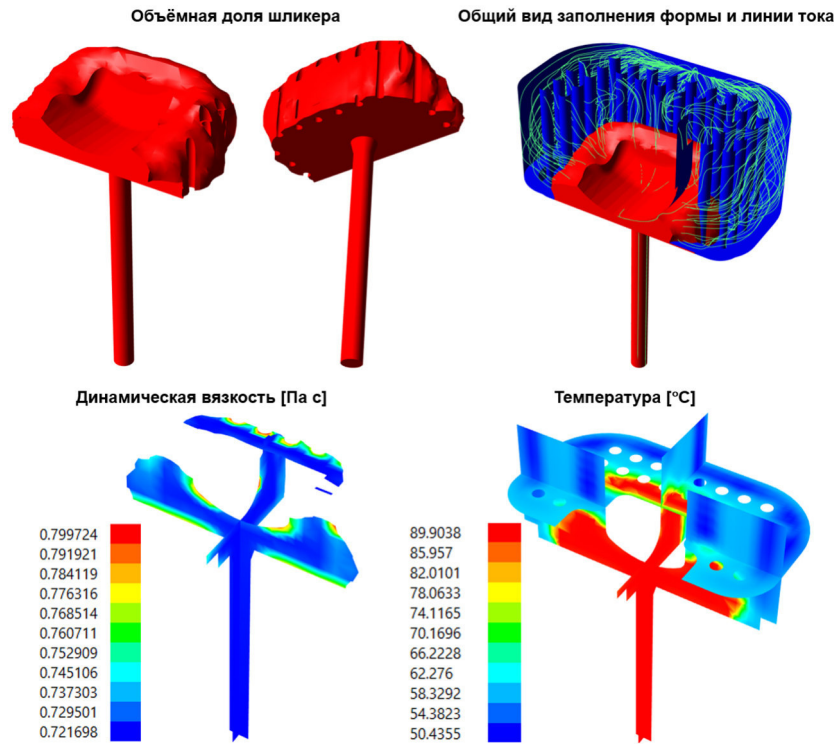


Рисунок 12. Расчётные поля к моменту времени  $t = 169,06$  [мс]

Изготовление грозди из четырёх стаканчиков длится около 170 [мс] (рис. 11). Наблюдаются повышение скорости течения в питателе, что обусловлено уменьшением проходного сечения, хорошая литейная способность в области формы. И, наконец, наиболее сложная деталь – смеситель, показана 1/3 от общей высоты заливки, требует столько же времени заполнения, что и гроздь из четырёх стаканчиков (рис. 12). Наличие большого количества каналов с уменьшающимся в два раза диаметром по высоте может приводить к захлаживанию области и понижению литейной способности шликера.

### Выводы

В статье представлены результаты экспериментального и численного исследования процесса литья под низким давлением керамической суспензии. Показано, что в зависимости от сложности геометрии деталей и температурных условий изменяется жидкотекучесть корундового шликера. Турбулентно-дисперсное течение вызвано высокими скоростями заполнения. Для их устранения требуется более точный контроль давления на входе.

Было выполнено определение характера течения материала корундового шликера и его консистенции в интересующем интервале температур обработки с графическим определением изменения динамической вязкости, которая использовалась для корректного описания течения шликера в процессе численного моделирования ЛНД, и окна температур, гарантирующего жидкотекучесть шликера. Изготовленные по технологии быстрого прототипирования (LCD) прототипы отливок из керамоподобного материала дали общее представление о правильности построения геометрии литейной грозди, в которой были отмечены слабые места. Моделирование литейной способности шликера по схеме оснастки ПЛС-1 показало корректность задания граничных условий и свойств шликера и хорошее совпадение с результатами экспериментов.

### Список литературы

1. Грибовский П.О. Горячее литье керамических изделий: учеб. пособие. М.: Госэнергоиздат, 1961. 400 с.
2. ЭКОН. Установка горячего шликерного литья / НПП АО "Экон". URL: <https://econobninsk.ru/ru/produktsiya/litevye-ustanovki/ekon-ugshl> (дата обращения: 14.08.2025).
3. Степанов Ю.А., Баландин Г.Ф., Рыбкин В.А. Технология литейного производства. Специальные виды литья. М.: Машиностроение, 1983. 287 с.
4. Литьё под давлением / М.Б. Беккер, М.Л. Заславский, Ю.Ф. Игнатенко, Р.А. Коротков, В.Я. Невзоров. Изд. 3-е, перераб. и доп. М.: Машиностроение, 1990. 400 с.
5. Гини Э.Ч., Зарубин А.М., Рыбкин В.А. Технология литейного производства. Специальные виды литья, М.: Академия, 2005. 352 с.
6. Белопухов А.К. Технологические режимы литья под давлением. Изд. 2-е перераб. и доп. М.: Машиностроение, 1985. 272 с.
7. Рейнер М. Реология. М.: Наука, 1965. 224 с.
8. Barnes H.A. A handbook of elementary rheology, Wales: Cambrian Printers, 2000. 200 p.
9. Открытый отчёт компании Anton Paar, Viscosity Measurement of Ceramic Slurries with the Rotational Rheometer RheolabQC. URL: <https://www.anton-paar.com/corp-en/services-support/document-finder/application-reports/ceramics-viscosity-measurement-of-ceramic-slurries-with-rheolabqc/> (дата обращения: 14.08.2025).
10. Sardarian M., Mirzaee O., Habibolahzadeh A. Numerical simulation and experimental investigation on jetting phenomenon in low pressure injection molding (LPIM) of alumina // Journal of Materials Processing Technology DOI: 10.1016/j.jmatprotec.2016.12.014
11. FlowVision. Руководство пользователя FlowVision 3.13.04 / ООО «Тесис». 2023.

## Применение метода корреляции цифровых изображений для оценки деформации листовой заготовки

М. А. Петров<sup>1</sup>, В. В. Исаков<sup>2</sup>

<sup>1</sup>Московский политехнический университет, Москва, Россия

<sup>2</sup>Центральный институт авиационного моторостроения имени П.И. Баранова, Москва, Россия

**Аннотация.** В статье рассматривается применение метода корреляции цифровых изображений (КЦИ) при выполнении технологического испытания на штампуемость по Эриксену и одноосному растяжению алюминиевых и медных листовых заготовок соответственно. Особенностью оценки результата тестов является применение поверхностного рисунка или паттерна без дополнительного покрытия поверхности исследуемой заготовки, а также с дополнительным покрытием и хаотичным паттерном, создаваемого краской, в сочетании с параметрами настройки программы DICe, в которой выполняется сравнение снимков. Установлено, что паттерны с частым рисунком способны поддерживать непрерывность результатов по полю деформации.

**Ключевые слова:** АМг2, МКЭ, DICe, паттерн, спекл-интерферометрия, QForm, ParaView.

## Application of digital image correlation method for strain field assessment of a sheet workpiece

M. A. Petrov<sup>1</sup>, V. V. Isakov<sup>2</sup>

<sup>1</sup>Moscow Polytechnic University, Moscow, Russia

<sup>2</sup>Central Institute of Aviation Motors, Moscow, Russia

**Abstract.** The paper considers the method of the digital image correlation (DIC) for practical application especially for technological test of formability according to Erichsen and uniaxial tension for aluminium and copper sheet workpieces, correspondingly. The peculiarity of test results' assessment is the application of a surface sketch or pattern without additional coating of the surface of the workpiece under study, as well as with an additional coating and a chaotic pattern created by paint, in combination with the parameters of the DICe program settings, in which the comparison of images is performed. It was found out that patterns with a frequent pattern are able to maintain the strain field discontinuity.

**Keywords:** AA5051, FEM, DICe, pattern, speckle interferometry, QForm, ParaView.

### Введение

В современной практике бесконтактные технологии оценки различных параметров неуклонно вытесняют контактные методы. Этот процесс обусловлен тем, что бесконтактные методы хорошо укладываются в концепцию «умного» производства, становление которого взаимосвязано с процессом цифровой трансформации производства, в рамках четвёртой промышленной революции (Industry 4.0). Можно выделить группу бесконтактных методов, позволяющих оценить качество материала, обрабатываемого методами обработки давлением на последующих участках. Наиболее востребованной данная технология является в производстве оболочковых деталей сложной формы из листового металла. Взятый курс на снижение веса конструкции без потери жёсткости требует либо применения прочных металлов и сплавов с пониженной плотностью, либо композиционных материалов, либо деталей с изменённой геометрией конструкции, полученных, например, при помощи топологической оптимизации.

В основе упомянутых бесконтактных методов лежат технологии машинного зрения. В этом смысле хорошо известна такая библиотека машинного зрения, как OpenCV, широко применяемая на практике. Реализацию бесконтактных методов можно встретить в системах 3D-сканирования (3D-scanning), теневой фотографии (shadow projection), корреляции цифровых изображений (КЦИ, Digital Image Correlation (DIC) и т. д. Сначала происходит сбор информации аппаратными средствами, а затем следует её обработка программными средствами с выводом интересующих результатов [1].

В методике корреляции цифровых изображений задача может выполняться как в двухмерном (моно), так и в трёхмерном (стерео) режимах. Не всегда очевидно, что может получиться на выходе, какой режим выбрать для того или иного испытания. Так, например, при отсутствии большой

пластической деформации можно стереорежим заменить на двухмерный в случае, если материал заготовки не разрушается. В основе метода лежит сопоставление полученных изображений друг с другом в разный период времени на основе сравнения особенностей рисунка на поверхности, который может являться результатом обработки поверхности, либо искусственно создан на ней. Аппаратной частью системы КЦИ являются камеры, как правило, с разрешением не менее Full HD (1920×1080) или HD (1280×720), ведущие непрерывную съёмку в процессе эксперимента. Программной частью в специализированном программном обеспечении (ПО) выполняется обработка полученных результатов либо во время съёмки (в режиме онлайн), либо после эксперимента (в режиме оффлайн). Программы делятся на свободные, или open source, (например, DICe, py2DIC, ADIC2D/3D, OpenCorr) и проприетарные (например, GOM Correlate/Zeiss Inspect, VIC-2D/3D, Istra4D). Исследуемый образец рекомендуют подготовить, уменьшить отражающую способность поверхности и нанести уникальный неповторяющийся паттерн, например хаотично распределённые точки разного размера. Как правило, подготовка поверхности выполняется в два этапа: сначала наносят антибликовое покрытие, а затем аэрозольным методом наносят точки. Также требуется создать контраст между цветом поверхности и цветом точек. Самой простой парой являются чёрно-белые шаблоны (белая поверхность и чёрные точки). При использовании технологии подмешивания цвета или создания низкочастотного шума (dithering) в зависимости от восприимчивости камерой оттенков цветов при определённом освещении можно выбрать и другие цветовые сочетания паттерна и поверхности.

Методика КЦИ применяется:

- для получения значения деформации при проведении механических испытаний (ZwickRoell 2D DIC (videoXtens, laserXtens), VDA-3D, VIC-2D/3D);
- характеристики колебаний и оценки дефектов в условиях динамических испытаний (Dantec DIC, GOM Aramis, EikoTwin DIC).

Встречаются исследования, в которых КЦИ применяется для изучения биомеханики растений [2], изучения разрушения образцов почвы [3] и других областей науки и техники [4]. С определёнными сложностями сталкиваются исследователи при выполнении механических испытаний в условиях повышенных температур [5, 6]. Развитие данной методики ведётся зарубежными сообществами, например Международным сообществом по КЦИ [7], целью которого является стандартизация методики, повышение точности и достоверности результатов, которые достигаются совместной работой специалистов в этой области, в том числе и через организацию бенчмарков и соревнований [8, 9]. Сообществом германских инженеров в 2019 году разработан стандарт VDI 2626, в котором закреплены основные определения, методики проведения КЦИ и оценка точности результатов.

**Целью исследования** является оценка работоспособности программы с открытым кодом для разных испытаний и типов подготовки поверхности заготовки. К **задачам** исследования относятся: 1) получение неразрывных полей для образцов до их разрушения для случаев одноосного нагружения и выдавливания полусферическим пуансоном; 2) выполнение валидации или сопоставления результатов численного моделирования и расчёта деформаций методом КЦИ для образцов со спекл-структурой (тип III); 3) определение степени влияния различных настроек в программе DICe на результат расчёта поля деформации.

### Экспериментальная часть

В обработке материалов давлением (ОМД) листовых материалов выполняются тесты на проверку технологических и механических свойств деформируемого материала. Это необходимо для того, чтобы определить технологические параметры, при которых разрушения материала наблюдаться не будет, а также будут гарантированы эксплуатационные свойства изготавливаемой детали. Тест на одноосное растяжение выполнялся на образцах, размеры которых соответствовали ГОСТ 1497-84, из тонколистового проката (фольги) меди марки М1 толщиной  $s = 100$  мкм (ГОСТ 859-2001), без специальной подготовки поверхности, на испытательной машине Мелитэк МЭС-Л-0.5. Тест по Эриксену (ГОСТ 10510-80 и ISO 20482) выполнялся на установке МТЛ 10Г-1 для заготовок из стали 10 толщиной  $s = 0,8$  мм; для поверхности образца без антибликового покрытия и с отсутствующим паттерном (тип I) и из алюминиевого сплава марки АМг2 толщиной  $s = 1,2$  мм; для поверхности без

антибликового покрытия с нанесенным паттерном (тип II); для поверхности с антибликовым покрытием и паттерном в виде спекл-структуры (тип III).

На рисунке 1 представлены изучаемые образцы из АМг2, разделённые по типам. Паттерны для заготовок типа II наносились по методу лазерной гравировки на установке, оборудованной лазерным модулем Raylase 1504 (Nd:YAG,  $\lambda=1064$  нм,  $W=20$  Вт/см<sup>2</sup>). Маркировка выполнялась за один проход, применялись образцы-карточки размером 40×40 мм (типоразмер 1) и 85×95 мм (типоразмер 2). Время маркировки для образцов типоразмера 1 составило 45 секунд, а для типоразмера 2 – максимально 700 секунд, в зависимости от плотности заполнения паттерна.

Всего рассматривалось 10 паттернов с разной конфигурацией повторяющихся элементов, а также процентом заполнения площади деформируемой поверхности. Выбор типа паттерна осуществлялся исходя из следующих соображений: невысокая плотность покрытия поверхности паттерном затруднит процедуру анализа по методу КЦИ из-за большого удаления соседних элементов, а слишком высокая, наоборот, даст избыточную информацию, которая также приведёт к неправильным результатам. Экспериментально установлено, что разрушение образца наступает при перемещении пуансона на 11 мм, силе деформирования 0,0059 МН и максимальном утонении в месте разрушения материала  $\delta = 0,4$  мм. Эти исходные данные являются опорными.

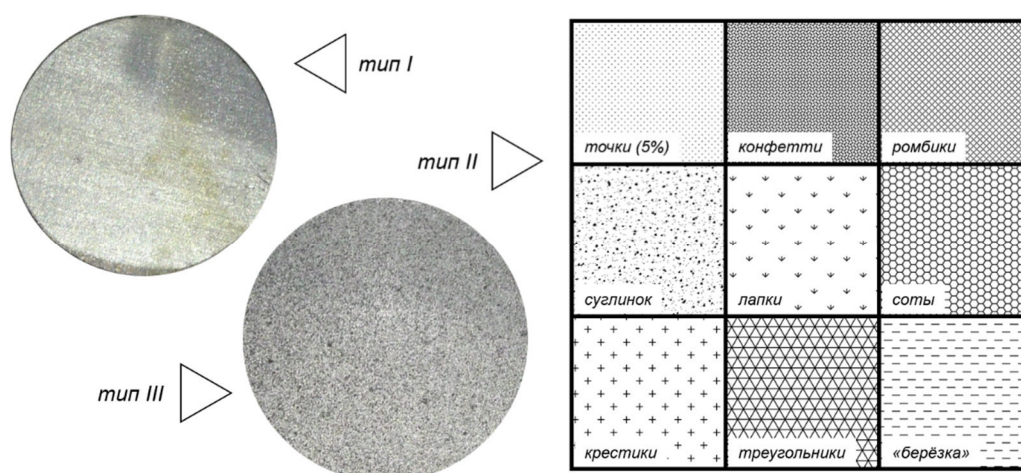


Рисунок 1. Изучаемые паттерны поверхности листовой заготовки

### КЦИ и численное моделирование

В качестве программного средства для выполнения КЦИ была выбрана программа с открытым кодом DICe (Digital Image Correlation Engine), разработанная в Sandia Laboratory (США) и, согласно данным GitHub, в январе 2016 года впервые представленная на платформе репозитория [10]. Программа DICe имеет интуитивный интерфейс для загрузки данных, но в силу наличия специальных функций требует дополнительных пояснений. Наиболее важными параметрами из представленных в интерфейсе программы являются initialization method, sssig threshold (пороговая граница контрастности), subset size (размер реперной подобласти), gauge size (размер датчика) и Gauss filter (фильтр Гаусса). SSIG измеряет контрастность в каждой подгруппе и не позволяет DICe выделить подгруппу с низким значением SSSIG, то есть данный параметр отфильтровывает области изображения с низким градиентом контрастности. Если у подгруппы значение SSSIG ниже указанного значения, то она исключается. Для случая, когда имеются бликующие или затемнённые поверхности, значение этого параметра выставляют равным нулю. Subset size выполняет проверку того, попадает ли точка на границу конформного определения и не находится ли она в исключённой области, то есть размер данной области определяет выборку, по которой будет вестись сверка в процессе поиска решения. Gauss filter – нижнечастотный фильтр, целью которого является изменение значений уровня серого в пикселях. Чаще всего с целью сглаживания перехода применяется Гауссова весовая функция (квадратичная, либо линейная). Контроль точности решения выполняется по четырём параметрам и визуальной оценке пользователя. Переменная SIGMA показывает прогнозируемую вариацию



смещения, оценивает прогнозируемую вариацию решения по смещению с учётом вариаций данных, вызванных шумом и смещением интерполяции. Переменная GAMMA отражает качество соответствия шаблону; измеряет, насколько хорошо шаблон или подмножество эталонного изображения соответствует деформированному изображению; является критерием корреляции (или модуля целевого функционала). Переменная BETTA является функцией потерь и измеряет чувствительность функции потери к малым возмущениям в решении по смещению. Четвёртый параметр – экземпляр пост-процессора UNCERTAINTY, который вычисляет неопределенность в оценке смещения в ходе выполнения КЦИ.

Валидируемым выходным полем является поле напряжений виртуального датчика (Virtual Strain Gauge (VSG\_Strain) для двух главных направлений (XX, YY) и в плоскости (XY), при этом деформация для третьего направления ZZ вычисляется в программе ParaView (рис. 2).

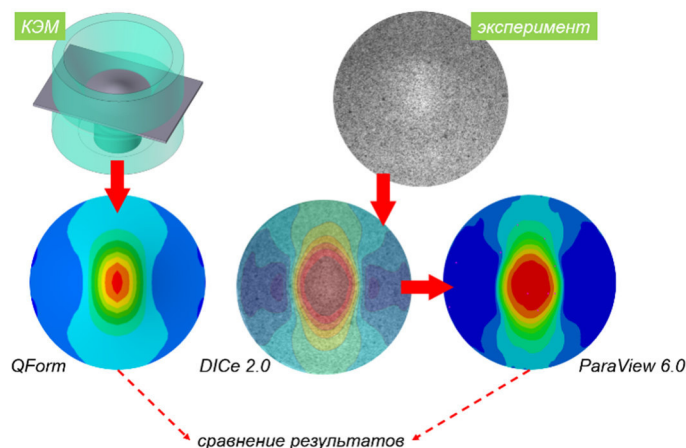


Рисунок 2. Схема выполнения валидации результатов

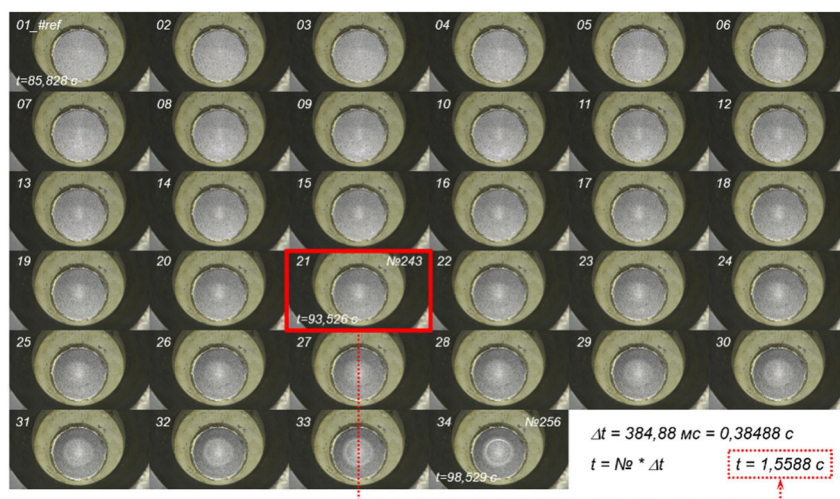


Рисунок 3. Схема выполнения валидации результатов

Численная оценка значений деформаций в программе DICe является непростой задачей, несмотря на наличие легенды, так как простановка меток со значениями деформаций имеющимися в программе инструментами невозможна. Оценку величины деформации можно выполнить визуально, а также путём сравнения цветов по полю с цветами на шкале легенды, например, по цветовой схеме RGB при помощи инструмента «пипетка» в любом редакторе изображений. Более продвинутым способом является экспорт результатов расчётов, представляемых в формате Exodus (\*.e), и их импорт в программу ParaView, разработанную также на принципе открытого кода. Численное моделирование теста по Эриксену по методу конечных элементов было выполнено в программе QForm и более подробно рассматривается в работе [11]. Для одноосного растяжения сравнительное численное



моделирование не выполнялось. Анализировались только результаты КЦИ. Для сопряжения решений численной задачи и результата КЦИ определялось время расчёта и вычислялось время по порядковому номеру кадра раскадровки, выполненной в программе VLC в стандартных фильтрах сцены, с периодичностью каждый седьмой и пятнадцатый кадры (рис. 3).

### Обсуждение результатов

На рисунке 4 представлены виды разрушений образцов. Несмотря на то что изготовление образцов выполнялось из одного листа, геометрия трещины указывает на то, что паттерн выступил в случае, например, ромбиков, шестигранников, конфетти и треугольников в роли концентраторов напряжений, по которым развивалась трещина.

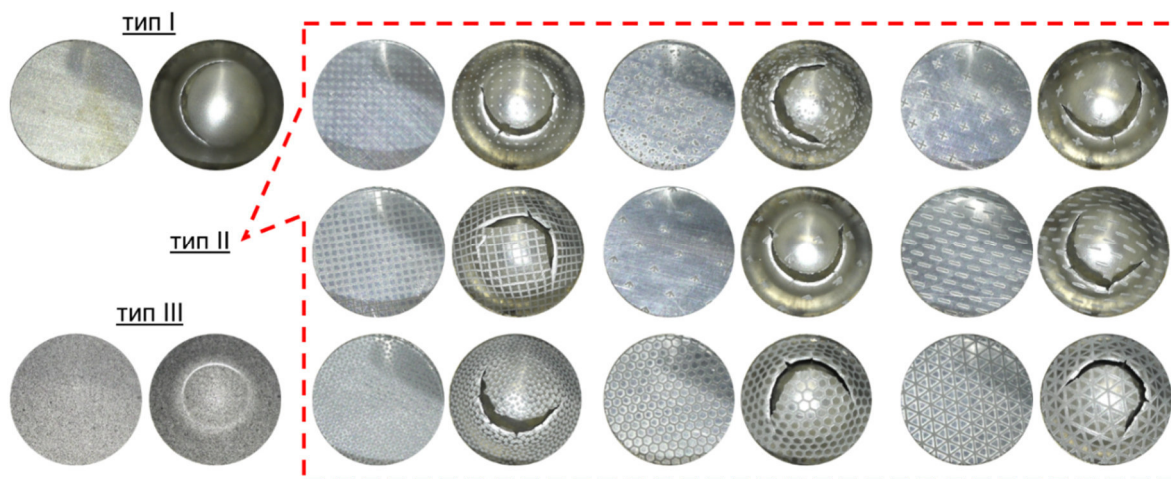


Рисунок 4. Вид образцов до (слева) и после (справа) разрушения, тест по Эриксену

На рисунке 5 показано сравнение значений напряжений для случаев численного моделирования и КЦИ образца третьего типа. Расхождение значений получено во втором и третьем знаках после запятой, что указывает на достаточно высокое совпадение результатов. На рисунке 6 показаны результаты одноосного растяжения. Отметим, что наилучший результат был получен для случая максимизации значений gauge size и gauss filter image window, уменьшения шага между контрольными точками с 14 до 9 пикселей, а также с минимизацией значения sssig. Это позволило получить поле, показывающее уровень критического значения напряжения, при котором наблюдается разрушение материала образца из меди. Размер области subset size менялся в ходе реализации КЦИ, но не сильно оказывал влияние на результат.

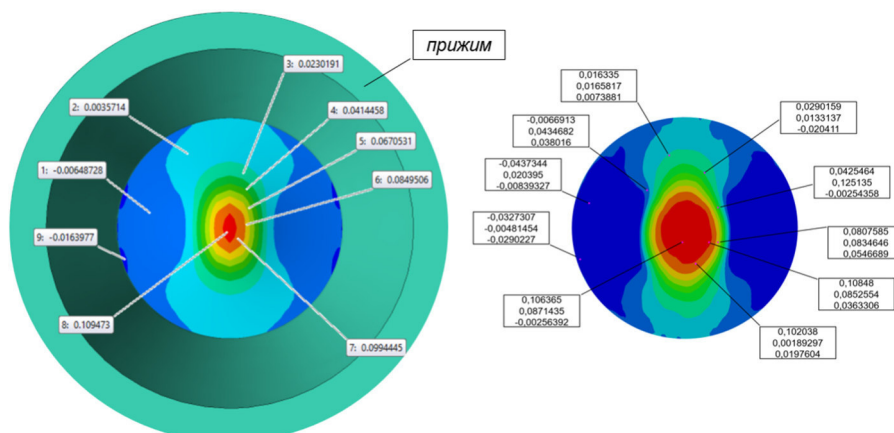


Рисунок 4. Сравнение значений деформаций: QForm, деформация\_XX (слева) и ParaView, деформации\_XX, YY и ZZ (справа, надписи сверху-вниз в выноске)

На рисунке 7 представлены наиболее удачные результаты теста по Эриксену для образцов второго типа. Для сравнения приводится результат, полученный на спекл-структуре. Стоит отметить, что наличие бликующей поверхности образцов второго типа не всегда является проблемой. Так, паттерн в виде точек, конфетти, шестигранников и суглинка хорошо зарекомендовал себя на первых этапах расчётов (до пятого-шестого шага). Однако в дальнейшем в виду бликов и перераспределения отражённого света от паттерна поле значений разваливается. Добавление промежуточных фотографий не улучшает восприятие картинки программой.

Сравнивая характеристику поля деформаций по главному направлению XX, отметим, что в случае спекл-структуры поле деформации лучше описывает физический эксперимент и тот результат, который наблюдался в численном моделировании. И наоборот, на втором шаге расчёта у образцов второго типа с точечным паттерном наблюдаются некоторая хаотичность и асимметрия в полях после его образования и развития.

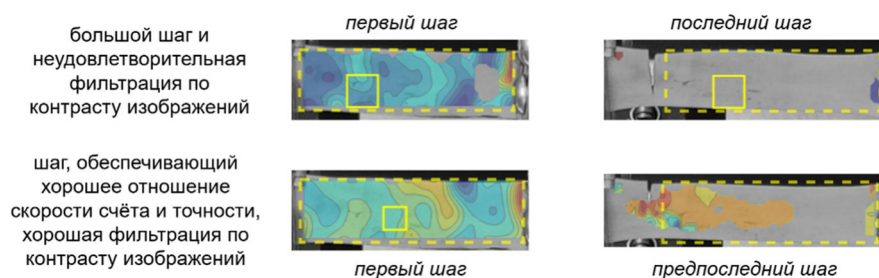


Рисунок 6. Результаты одноосного растяжения по КЦИ

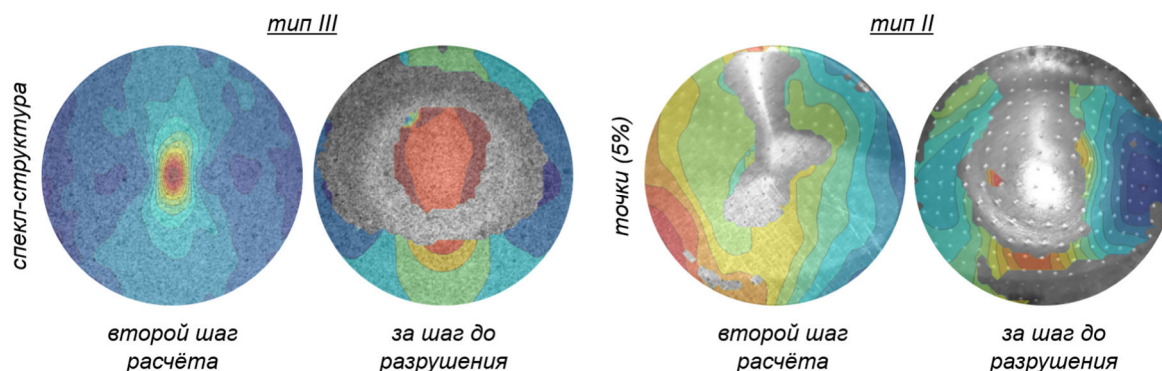


Рисунок 7. Результаты теста по Эриксену после анализа по методу КЦИ

## Выводы

В результате выполненных исследований установлено следующее: 1) получить неразрывные поля для образцов второго типа на растяжение и выдавливание полусферическим пуансоном возможно лишь на первых 5...6 шагах расчётов, так как в дальнейшем блики мешают построению поля значений; 2) выполненная валидация численного моделирования и КЦИ для образца со спекл-структурой является хорошей, так как несоответствие значений выявлено во второй-третьем знаке после запятой; 3) такие настройки, как фильтрация данных и размер датчика деформации, в сильной степени влияют на результат расчёта.

Вычисление напряжения в третьем направлении осуществляется в программе ParaView, что позволяет говорить об оценке третьей координаты в режиме двухмерного КЦИ-анализа. Поля показывают разрывность по мере приближения момента образования трещины. В связи с этим появляется возможность выполнить гибридный вариант КЦИ: в монорежиме до разрушения получить поле интересующих значений, а при подходе к моменту разрушения материала перейти в стереорежим, что позволит при работе в онлайн-режиме сократить время на обработку меньшего количества получаемой информации без потери качества оценки величин.

## Заключение

Рассмотренный метод бесконтактной оценки деформации КЦИ является весьма перспективным для области ОМД, для оценки деформации материалов при комнатных температурах. При повышенных температурах данная методика интересна лишь для условий неинтенсивного конвективного теплообмена и отсутствия свечения объекта исследования. Открытые библиотеки машинного зрения, например OpenCV, позволяют самостоятельно разработать приложение для реализации рассмотренных выше задач. Однако в Российской Федерации нет открытых программ, подобно DICe или OpenCorr, для применения и апробации техники КЦИ, которая весьма полезна для развития как национальной базовой и фундаментальной науки, так и в инженерной практике. Интерес и перспективы развития метода постепенно поспособствуют вовлечению ИТ-специалистов и, несомненно, приведут к появлению доступных программ и оборудования для КЦИ, разработанных в России.

## Список литературы

1. Осинцев А.В., Очков К.Ю. Визуализация перемещений и деформаций методом цифровой корреляции изображений // Научная визуализация. 2016. № 8:2. С. 15–23.
2. Mylo M.D., Poppinga S. Digital image correlation techniques for motion analysis and biomechanical characterization of plants. *Front. Plant Sci.* 14:1335445 (2024). DOI: 10.3389/fpls.2023.1335445
3. Arza-García M., Núñez-Temes C., Lorenzana J.A., *et al.* Evaluation of a low-cost approach to 2-D digital image correlation vs. a commercial stereo-DIC system in Brazilian testing of soil specimens. *Archiv.Civ.Mech.Eng* 22, 4 (2022). DOI: 10.1007/s43452-021-00325-0
4. Périé J.-N., Passieux J.-C. Advances in Digital Image Correlation (DIC). *Applied Sciences* (2020). DOI: 10.3390/books978-3-03928-515-0
5. Lyons J.S., Liu J., Sutton M.A. High-temperature deformation measurements using digital-image correlation. *Experimental Mechanics* 36, 64–70 (1996). DOI: 10.1007/BF02328699
6. Kempny M. Digital image correlation - method development, scope, principle of functioning, and future goals. *Journal of Metallic Materials*, 3–4. 30–41. (2022). DOI: 10.32730/imz.2657-747.22.3-4.4
7. International Digital Image Correlation Society. Jones E.M.C. and Iadicola M.A. (Eds.). A Good Practices Guide for Digital Image Correlation (2018). DOI: 10.32720/idics/gpg.ed1/print.format
8. Reu P.L., Toussaint E., Jones E., *et al.* DIC Challenge: Developing Images and Guidelines for Evaluating Accuracy and Resolution of 2D Analyses. *Exp Mech* 58, 1067–1099 (2018). <https://doi.org/10.1007/s11340-017-0349-0>
9. Reu P.L., Blaysat B., Andó E., *et al.* DIC Challenge 2.0: Developing Images and Guidelines for Evaluating Accuracy and Resolution of 2D Analyses. *Exp Mech* 62, 639–654 (2022). <https://doi.org/10.1007/s11340-021-00806-6>
10. Репозиторий GitHub “Digital Image Correlation Engine (DICe)”. URL: <https://github.com/dicengine> (дата обращения: 29.08.2025).
11. Петров М.А., Ромашов Д.А., Исаков В.В. Применение бесконтактных способов оценки деформации при испытаниях на выдавливание алюминиевой листовой заготовки по Эриксену // GraphiCon 2023: труды 33-й Междунар. конф. по компьютерной графике и машинному зрению (Москва, 19–21 сент. 2023 г.). М.: Институт прикладной математики им. М.В. Келдыша РАН, 2023. С. 274–284.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-466-471

## Применение искусственного интеллекта для визуализации контуров природных объектов на спутниковых снимках

Е. В. Попов, П. В. Юрченко

Нижегородский государственный архитектурно-строительный университет, Нижний Новгород, Россия

**Аннотация.** Авторами изучено применение методов машинного обучения для классификации изображений на спутниковых снимках. Для визуализации контуров природных объектов был применен нейросетевой алгоритм. При подготовке обучающего набора данных использовались методы машинного обучения. Проведён анализ полученных результатов классификации и дана оценка точности алгоритма, предложены способы его усовершенствования.

**Ключевые слова:** геометрическое моделирование, геоинформационное моделирование, нейросеть, природные ресурсы, визуализация границ объектов, классификация изображений

## Using artificial intelligence to visualize the contours of natural objects in satellite images

E. V. Popov, P. V. Yurchenko

Nizhny Novgorod State University of Architecture and Civil Engineering, Nizhny Novgorod, Russia

**Abstract.** The authors studied the application of machine learning methods for image classification in satellite images. A neural network algorithm was used to visualize the contours of natural objects. Machine learning methods were used to prepare the training dataset. The classification results were analyzed, and the accuracy of the algorithm was evaluated. Suggestions were made for improving the algorithm.

**Keywords:** geometric modeling, geoinformation modeling, neural network, natural resources, visualization of object boundaries, image classification.

### Введение

Спутниковые снимки сегодня являются одним из часто применяемых материалов для изучения природных объектов. Анализ спутниковых снимков проводится с целью распознавания на них природных объектов или объектов антропогенного происхождения и использования полученных данных для аналитики и прогноза. Инструментами в этих исследованиях, как правило, выступают технологии искусственного интеллекта, основанные на методах машинного обучения и компьютерного зрения.

Применение возможностей искусственного интеллекта в совокупности с геоинформационными технологиями позволяет решить несколько важных взаимосвязанных задач. Во-первых, визуализировать контуры изучаемых объектов на спутниковом снимке, а во-вторых, при регулярных наблюдениях накапливать материал для аналитики и прогноза возможных изменений в контурах и площадях целевых объектов, используя только рабочее место за компьютером. Так можно наблюдать за изменениями в контурах целевых объектов в течение определенного периода времени, и при накоплении необходимого количества данных прогнозировать дальнейшее развитие территории. Качественный прогноз является важным для принятия решений в управлении как конкретными ресурсами, так и всей территорией исследования.

Вопросам исследования возможностей искусственного интеллекта в процессе распознавания природных и антропогенных объектов на спутниковых снимках, при котором визуально выделяются их контуры и занимаемые ими площади, в последнее десятилетие посвящают свои исследования многие российские и зарубежные ученые. Ряд из них сосредоточены на максимально достоверном распознавании и визуализации изучаемых объектов и их автоматизированной категоризации и типизации. Применяемые методы машинного обучения исследователи, как правило, сочетают между собой, часто используя как основной метод для комбинирования – нейронную сеть. Так, использование нейросети глубокого обучения U-Net в сочетании с семантической нейросетью ES-Net при изучении состояния лесных массивов описано в [1]; использование временных рядов и многомасштабной географически взвешенной регрессионной модели (MGWR), данные для которой распознаются с

помощью метода случайного леса на спутниковых снимках в исследовании сельскохозяйственных культур рассмотрено в [2]. Иные исследования посвящены прогнозированию развития ситуации с изменениями контуров изучаемых объектов. Основными предметами изучения являются природные объекты – лесные массивы, земли сельскохозяйственного назначения и культивируемые на них насаждения, водотоки и водоемы, прибрежные зоны рек, морей и океанов. Так, исследование [3] посвящено прогнозу развития сельскохозяйственных культур, в частности, пшеницы, и использует при этом пять различных алгоритмов машинного обучения (Random Forest, Gradient Boosting, AdaBoost, LightGBM и XGBoost); делая вывод об их производительности, наиболее успешной считает модель XGBoost. Вместе с тем использование данных из нескольких источников в сочетании с алгоритмами машинного обучения повышает точность прогнозирования и добавляет новые возможности в системы поддержки принятия решений в сельском хозяйстве. Подобный вывод делается в исследовании [4], где применяются методы оценки малых площадей, статистические регрессионные методы в совокупности с данными, полученными из открытых источников – спутник Landset 8, платформа Google Earth Engine при исследовании качества лесных массивов.

Для представляемого исследования были выбраны природные объекты, расположенные на территории Большеболдинского округа Нижегородской области и представляющие существенную ценность для данной территории – лесные массивы. Значимость данных природных объектов и применяемые методы делают исследование актуальным и востребованным.

### **Постановка задачи**

Известно, что при изучении объектов, занимающих на поверхности Земли значительные по площади территории, используют данные дистанционного зондирования. К таким объектам, безусловно, относятся и лесные массивы [5]. Чаще всего их можно получить из материалов спутниковых снимков, данных аэрофотосъемки, а также снимков, сделанных с помощью беспилотных летательных аппаратов. Поскольку последние два источника съемки требуют специфических инструментов и не всегда возможны для свободного доступа, было решено использовать материалы хранилищ спутниковых снимков, доступные широкому кругу пользователей. Из существующих открытых источников были отобраны снимки исследуемой территории за период с 2014 по 2024 год, сделанные спутниками Landset 8 и 9 в период вегетации (с мая по сентябрь) в хорошую ясную и безоблачную погоду, так как таким снимкам требовалось меньше предварительных процедур обработки, и их можно было использовать для создания геоинформационной модели, позволяющей визуализировать на растровом снимке контуры изучаемых объектов.

Обработанные в геоинформационных системах данные дистанционного зондирования земной поверхности, полученные со спутниковых снимков, позволяют работать как со всеми природными ресурсами региона, так и сосредоточиться на отдельных, наиболее важных объектах. Для этой работы требовался достаточно эффективный и относительно несложный алгоритм, позволяющий проводить распознавание изображений исследуемых природных объектов на спутниковых снимках, визуализировать их контуры и использовать полученные результаты для дальнейших решений.

### **Теория**

#### *1. Особенности изучаемых объектов*

Лесные массивы относятся к категории объектов, не имеющих типичных контуров, что подтверждается визуальным анализом выбранных для исследования фрагментов спутниковых снимков (рис. 1). Это затрудняет распознавание и визуализацию их контуров на спутниковых снимках методами широко распространенных в последнее время сверточных нейронных сетей (CNN). По нашему мнению, использование классического нейросетевого алгоритма «с учителем», имеющего в базе метод обратного распространения ошибки, для распознавания изображений лесных массивов является более приемлемым. Здесь мы учитывали также и требование к экономии ресурсов при разработке и внедрении алгоритма.



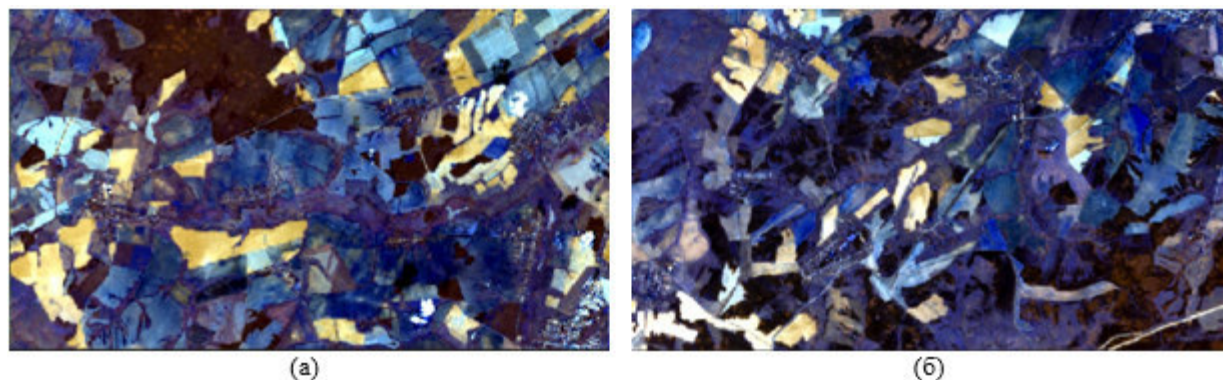


Рис. 1. Фрагменты векторных слоев снимка с лесными массивами: а) – фрагмент 1; б) – фрагмент 2

## 2. Основные этапы работы по визуализации лесных массивов

Следующие этапы работы с выбранными снимками в рамках рассматриваемого подхода предполагали применение геоинформационных технологий и проводились в геоинформационной системе QGIS:

- сборка растрового снимка с помощью модуля GDAL Tools;
- выделение сцен снимка, содержащих исследуемую территорию;
- отделение фрагментов сцен, содержащих лесные массивы;
- моделирование лесных массивов с помощью плагинов QGIS [6];
- получение размеченного одноканального растра с изображениями лесных массивов.

При сборке снимка были взяты данные каналов с 1 по 7 и объединены в виртуальной среде GDAL Tools QGIS. Затем на выделенных фрагментах снимка, содержащих лесные массивы, были добавлены векторные слои для осуществления разметки и классификации изображений лесных массивов, используя эталонные полигоны.

В рамках проводимого исследования визуализация изображений лесных массивов означает проведение дихотомической классификации выбранного фрагмента сцены спутникового снимка, посредством которой были получены одноканальные растры с пикселями двух типов: относящиеся к лесным массивам и не относящиеся к таковым. Эта классификация осуществлялась с помощью основного метода исследования – нейросетевого алгоритма. Обучающие наборы данных для него были подготовлены с помощью четырех известных алгоритмов машинного обучения – Gaussian Mixture Model, Random Forest, Support Vector Machines, K-Nearest Neighbors, реализованных в QGIS посредством встроенных плагинов [6]. После подготовки наборов данных и обучения нейросети были произведены распознавание и визуализация контуров лесных массивов на выбранных фрагментах спутниковых снимков. Результаты визуализации лесных массивов нейросетевым алгоритмом приведены для фрагмента 1 (рис. 2) и фрагмента 2 (рис. 3) с обучающими наборами, подготовленными каждым из упомянутых методов.

Необходимо заметить, что на предыдущих этапах исследования, описанного в [7], в нейросетевом алгоритме использовалось 14 входных узлов и две функции активации ReLu и Softmax, так как требовалось реализовать дихотомическую классификацию. Сеть была реализована на языке программирования Python, были применены функции библиотек numpy, pirsgis, tensorflow. Однако при попытке внедрить алгоритм в открытые геоинформационные системы QGIS (ГИС QGIS) было обнаружено, что не все версии QGIS позволяют корректно его запустить. Поэтому алгоритм был изменен для устранения этих проблем следующим образом: для повышения точности взяты 64 узла и те же функции активации, применяемые библиотеки Python – numpy, pirsgis, scikit-learn. А это позволит использовать его в перспективе как плагин для открытой ГИС QGIS, что расширит возможности его применения и может представлять определенную новизну в проводимом исследовании.

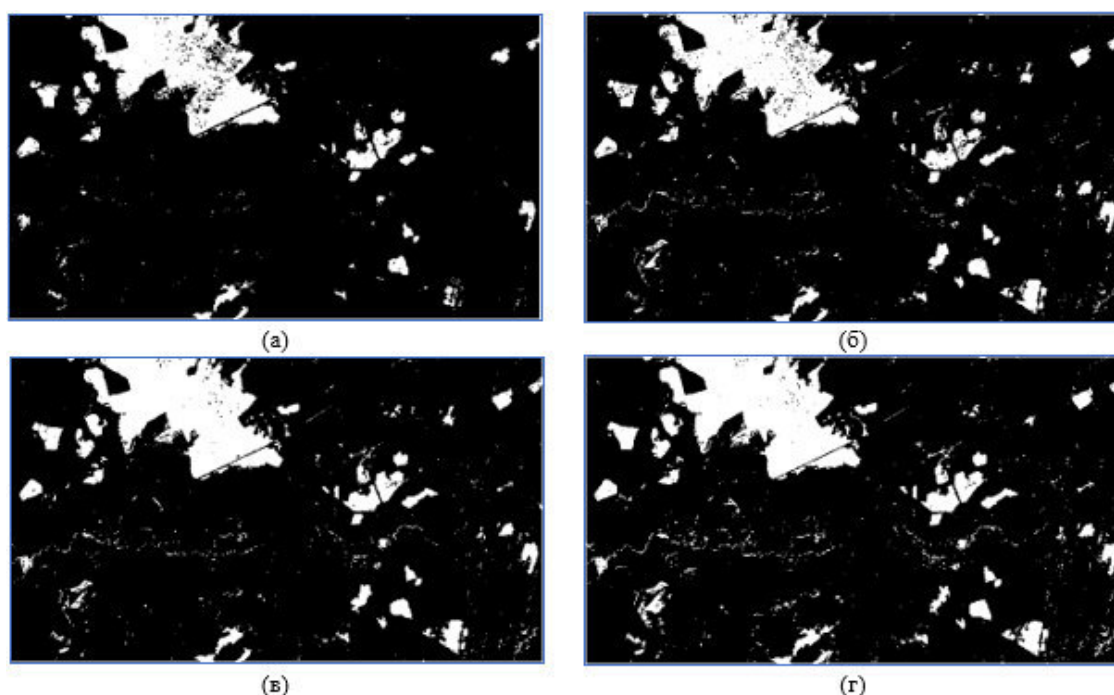


Рис. 2. Визуализация лесных массивов на фрагменте 1 с обучающими наборами, подготовленными методами: а) Gaussian Mixture Model; б) K-Nearest Neighbors; в) Random Forest; г) Support Vector Machines

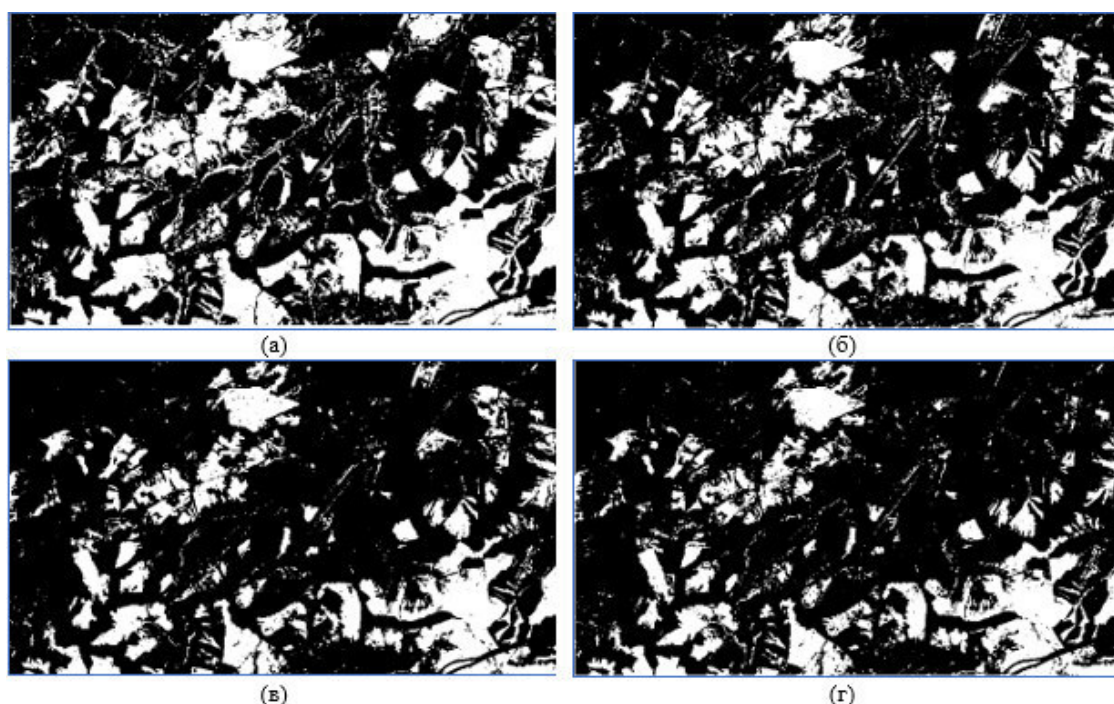


Рис. 3. Визуализация лесных массивов на фрагменте 2 с обучающими наборами, подготовленными методами: а) Gaussian Mixture Model; б) K-Nearest Neighbors; в) Random Forest; г) Support Vector Machines

### Результаты экспериментов

После обучения с применением наборов данных, подготовленных с помощью указанных методов машинного обучения, нейросеть показала в целом приемлемые результаты матрицы ошибок и соответствующих метрик – точности (precision), полноты (recall) и общей метрики (accuracy). Это позволяет считать предсказанные значения визуализации удовлетворительными и свидетельствующими о достижении поставленных целей.

Данные матриц ошибок для визуализации изображений лесных массивов на фрагменте 1 приведены в таблице 1, на фрагменте 2 – в таблице 2. Из данных таблиц следует, что фрагмент с большим



количеством лесных массивов распознается так же эффективно и достоверно, как и фрагмент, на котором лесных массивов не так много, что позволяет сделать вывод о независимости алгоритма от количества пикселей распознаваемых объектов.

Таблица 1. Данные матрицы ошибок для фрагмента 1

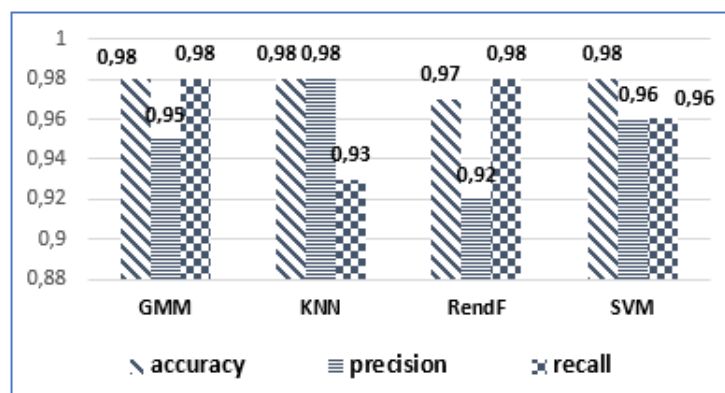
Метод подготовки обучающего набора	TN	FN	FP	TP
GMM	17657	115	299	5629
KNN	17747	427	104	5422
RendF	17371	122	516	5691
SVM	17243	268	263	5926

Таблица 2. Данные матрицы ошибок для фрагмента 2

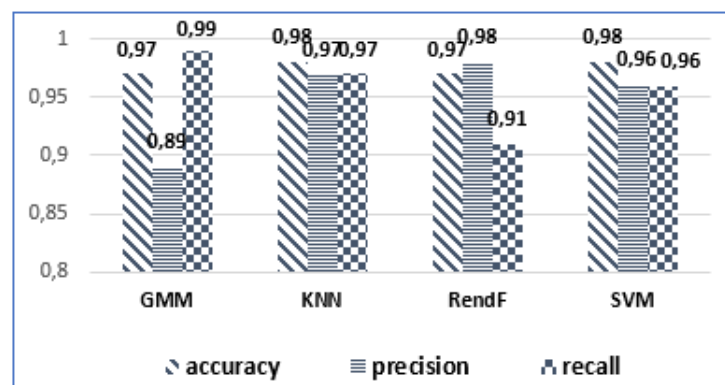
Метод подготовки обучающего набора	TN	FN	FP	TP
GMM	17276	33	680	5711
KNN	17567	180	180	5669
RendF	17792	523	95	5290
SVM	17264	268	242	5926

### Обсуждение результатов

Рассчитанные на базе приведенных в таблицах 1, 2 показателей матриц ошибок метрики позволяют говорить о достаточной точности и успешности нейросетевого алгоритма с каждым из подготовленных наборов данных. Вместе с тем можно говорить о более или менее успешных запусках алгоритма с некоторыми из обучающих наборов и также о наборе, который наиболее стабильно работает в сочетании с разработанным нейросетевым алгоритмом. Как видно из рисунка 4, наиболее стабильно работает алгоритм с обучающим набором данных, подготовленным с применением метода опорных векторов (SVM).



(a)



(b)

Рис. 4. Визуализация значений метрик нейросетевого алгоритма: а) фрагмент 1; б) фрагмент 2

Показатели метрик нейросети более высокие по сравнению с описанными в [7], они варьируются в диапазонах: accuracy 0,97-0,98, precision 0,91-0,98, recall 0,93-0,98.

### Заключение

На данном этапе исследования сделан вывод о том, что автоматизированные способы распознавания объектов на спутниковых снимках требуют более тщательной подготовки исходных данных. В частности, обучающий набор данных может быть подготовлен с помощью алгоритмов машинного обучения. Тогда в результате будет получена более точная геоинформационная модель исследуемой территории, визуализирующая контуры и площади изучаемых объектов.

### Список литературы

1. Wang, Y., Zhang, Z., Xia, J., Zhao, F., & Dong, P. (2025). ES-Net Empowers Forest Disturbance Monitoring: Edge–Semantic Collaborative Network for Canopy Gap Mapping. *Remote Sensing*, 17(14), 2427. <https://doi.org/10.3390/rs17142427>.
2. Xiao, P., Zhou, Y., Qian, J., Liu, Y., & Li, X. (2025). Spatiotemporal Mapping and Driving Mechanism of Crop Planting Patterns on the Jiangnan Plain Based on Multisource Remote Sensing Fusion and Sample Migration. *Remote Sensing*, 17(14), 2417. <https://doi.org/10.3390/rs17142417>.
3. Rufaioğlu, S. B., Bilgili, A. V., Savaşlı, E., Özberk, İ., Aydemir, S., Ismael, A. M., Kaya, Y., & Matos-Carvalho, J. P. (2025). Sensor-Based Yield Prediction in Durum Wheat Under Semi-Arid Conditions Using Machine Learning Across Zadoks Growth Stages. *Remote Sensing*, 17(14), 2416. <https://doi.org/10.3390/rs17142416>.
4. Alegbeleye, O. M., Poudel, K. P., VanderSchaaf, C., & Yang, Y. (2025). Improving the Estimates of County-Level Forest Attributes Using GEDI and Landsat-Derived Auxiliary Information in Fay–Herriot Models. *Remote Sensing*, 17(14), 2407. <https://doi.org/10.3390/rs17142407>.
5. Габитова А.А. Дистанционное зондирование и ГИС в оценке лесозаращения неиспользуемых сельскохозяйственных земель // Успехи современного естествознания. 2022. № 11. С. 42-46. Режим доступа: <https://natural-sciences.ru/ru/article/view?id=37926>.
6. Репозиторий плагинов QGIS Python. Плагин dzetsaka. Режим доступа: <https://plugins.qgis.org/plugins/dzetsaka/#plugin-about>.
7. Попов Е.В., Юрченко П.В. Использование нейросети для геометрического моделирования природных и антропогенных объектов // Приволжский научный журнал. 2024. № 4. С.336-341.

УДК 004.8, 519.178

DOI: 10.25686/978-5-8158-2474-4-2025-472-486

## Метод идентификации неявных связей в решении аналитических задач

И. Д. Соколов, К. В. Ионкина, М. С. Улизко, Е. В. Антонов, Е. Н. Бажанова, А. А. Артамонов  
Национальный исследовательский ядерный университет «МИФИ», Москва, Россия

**Аннотация.** Инструменты по визуализации позволяют преобразовать большие массивы данных в понятные графики для пользователя. В данной работе рассматривается разработка инструмента визуализации графов для выявления неявных связей между информационными объектами. В статье представлена реализация метода для выявления неявных связей между объектами и разработка инструмента для построения графовой визуализации, позволяющая пользователю взаимодействовать с графом посредством фильтрации. Реализованы функциональные возможности и предложен специализированный язык запросов для изменения вида узлов и ребер графа.

Разработанный инструмент и предложенный метод апробированы на двух реальных наборах данных: обнаружение потенциальных нарушений обязательств по ядерному нераспространению, выявление перспективных направлений научного сотрудничества организаций, подтвердив практическую значимость данного исследования.

## Method of identification of implicit connections in solving analytical problems

I. D. Sokolov, K. V. Ionkina, M. S. Ulizko, E. V. Antonov, E. N. Bazhanova, A. A. Artamonov  
National Research Nuclear University “MEPhI”, Moscow, Russia

### Введение

В настоящий момент существует множество инструментов для визуализации данных. Однако каждый из них имеет определенные ограничения. Коммерческие программные решения могут быть недоступны из-за высокой стоимости, другие решения могут не обладать достаточной функциональностью или сложны в освоении. Например, приложение Gephi, широко используемое для визуализации графовых данных, требует от пользователя высокой степени подготовки, что обусловлено сложностью его интерфейса и многоступенчатым процессом построения графов. VOSviewer требует структуру и выполняет одну узконаправленную задачу (задача библиометрического анализа). В связи с этим возникает потребность в создании специализированного программного обеспечения, обладающего возможностями интеграции интерактивных механизмов фильтрации и методов выделения ключевых вершин и ребер для анализа больших объемов данных и выявления закономерностей, при этом не привязанного к определенной структуре данных.

Таким образом, необходимость в гибком и функциональном инструменте для анализа сложных данных становится особенно актуальной при исследовании неоднородных объектов, где важно выявлять скрытые связи и зависимости.

Для определения связи между объектами  $A$  и  $B$  необходимо выделить характеристики этих объектов и выявить общие признаки, которые и будут указывать на взаимосвязь между исходными объектами. Пусть  $C$  — объект, такой что  $C$  имеет общие значения характеристик с объектом  $A$  и общие значения характеристик с объектом  $B$ . В этом случае можно предположить, что объекты  $A$  и  $B$  связаны через объект  $C$ . Таким образом, объект  $C$  становится общим признаком для объектов  $A$  и  $B$ , что позволяет связать эти два объекта между собой.

В статье «Нахождение скрытых зависимостей между объектами на основе анализа больших массивов библиографических данных» [1] рассматривается задача определения тематической близости журналов и конференций, используя несколько подходов определения близости автора к научной области. Автор работы рассматривает подход построения графов соавторства статей, в котором рассчитывается вес ребра, соединяющего несколько журналов. Предполагается, что если один и тот же автор имеет публикацию в разных журналах, то это может свидетельствовать о тематической близости журналов.

В статье «Методы анализа гетерогенных данных для построения социального профиля» [2] рассматривается выявление неявных связей для построения социального профиля. В качестве метода предлагается использовать графовый подход для нахождения неявных связей.

Указанные работы доказывают целесообразность использования графовой визуализации при выявлении взаимосвязей между объектами. Для определения взаимосвязи объектов аналитик должен выбрать интересующие его характеристики в рамках рассматриваемого объекта и указать, как связаны сущности внутри, определяя правило построения графа.

Рассмотрим существующие методы выявления связей между объектами. В работе [1] рассматривался метод графа соавторства. Данный метод применяется в основном для определения тематической близости журналов и конференций за счет построения графа. Метод имеет ограничение, связанное с масштабируемостью, так как применяется только для одного вида задач.

Рассмотрим метод корреляции (определения рассмотрены в книге [3]). Данный метод определяет взаимосвязь между двумя переменными, что позволяет выявить зависимость изменения одной переменной от другой. Заметим, что хотя метод и не предполагает наличие структуры, но может работать только с одним типом данных.

В статье [4] представлено описание методов «Выявление причинности» и «Оценка причинности», позволяющих выявить причинно-следственную связь между рассматриваемыми объектами. Данные методы сконцентрированы на оценке воздействия одного элемента на другой и предполагают наличие причинности этой взаимосвязи. Однако в рамках системного анализа применение их нецелесообразно, так как они нацелены не на структуру объекта, а на причину и следствие влияния одного объекта на другой.

В статьях [5] и [6] рассмотрены использования подходов семантического анализа и построения онтологий. Эти два метода позволяют построить модель, которая описывает структуру рассматриваемой области, ее свойства и внутренние связи, а также представляет данную область в виде графа. Указанные методы решают задачу взаимосвязи объектов, но они привязаны к заранее известной структуре данных.

### Постановка проблемы

Рассмотренные выше методы позволяют решать поставленные задачи на выявление связей между объектами, но требуют либо определенной структуры входных данных, либо определенного типа таких данных. В рамках настоящей работы необходимо решить проблему построения графа на входных данных различной структуры и типов, а также выявления явных и неявных связей между объектами, поскольку представленные выше методы такую задачу решить не позволяют. В ходе проведенного обзора решено разработать собственный метод для выявления явных и неявных связей независимо от структуры и типа входных данных и разработать инструмент для визуализации данных при различной структуре данных.

### Теория

#### Метод идентификации связей между объектами

Пусть  $N = \{obj = (a_1, a_2, \dots, a_M) | a_i - \text{характеристика объекта } obj, i = 1, 2, \dots, M\}$  – множество объектов с характеристиками  $a_1, a_2, \dots, a_M$ . Каждому объекту  $obj = (a_1, a_2, \dots, a_M) \in N$  поставим в соответствие граф  $G = \langle V, E \rangle$ , где множество вершин  $V = \{v_1, v_2, \dots, v_T\}, T \geq M$ , составляют значения характеристик  $a_1, a_2, \dots, a_M$  объекта  $obj$ , и множество ребер  $E = \{(v_i, v_j) | \exists obj \in N : obj(a_i) = v_i \text{ и } obj(a_j) = v_j\}$ , то есть две вершины  $v_i, v_j \in V$  соединены ребром, если найдется объект  $obj = (a_1, a_2, \dots, a_M) \in N$ , такой что значение характеристики  $a_i$  объекта  $obj$  равно  $v_i$  и значение характеристики  $a_j$  объекта  $obj$  равно  $v_j$ , где характеристики  $a_i$  и  $a_j$  задаются аналитиком.

Каждой вершине из множества  $V$  поставим в соответствие ее тип – характеристику из множества  $\{a_1, a_2, \dots, a_M\}$ . Тогда множество вершин  $V = K_1 \cup K_2 \cup \dots \cup K_M$ , где  $K_i = \{v_j \in V | obj(a_i) = v_j\}$  – множество вершин типа  $a_i, i = 1, 2, \dots, M$ . Если вершина  $v_j \in K_i$ , то для удобства будем обозначать ее  $v_j^{K_i}, j \in \{1, 2, \dots, T\}, i \in \{1, 2, \dots, M\}$ . Таким образом, граф  $G = \langle V, E \rangle$  можно представить в виде

гетерогенного графа  $G = \langle K_1 \cup \dots \cup K_L, E \rangle$ , где  $L \in \{1, 2, \dots, M\}$ , при этом если  $v_i, v_j \in K_W$ , то  $(v_i, v_j) \notin E$  (характеристики и их количество  $L$  задает аналитик).

Пусть  $K_L$  и  $K_P$  – множества вершин типа  $a_L$  и  $a_P$  соответственно. Приведем следующие определения:

**Явная связь** между вершинами  $v_i \in K_L$  и  $v_j \in K_P$  существует тогда и только тогда, когда есть ребро  $(v_i, v_j) \in E$ .

При построении графа, согласно правилу построения, между вершинами разных типов образуется явная связь. В таком случае между вершинами одного типа можно определить неявную связь.

**Неявная связь** между вершинами  $v_i \in K_L$  и  $v_j \in K_L$  существует тогда и только тогда, когда имеется вершина  $v_x \in K_P$ , такая что имеется ребро  $(v_i, v_x) \in E$  и ребро  $(v_j, v_x) \in E$ , где  $L \neq P$ .

**Явная связь порядка  $n$**  между вершинами  $v_i \in K_L$  и  $v_j \in K_P$  существует тогда и только тогда, когда есть путь  $v_i \rightarrow v_j$  длины  $n$ , не содержащий вершин одинаковых типов, причем  $n \in \mathbb{N}$  – наименьшее натуральное число с таким свойством.

**Неявная связь порядка  $n$**  между вершинами  $v_i \in K_L$  и  $v_j \in K_L$  существует тогда и только тогда, когда имеется вершина  $v_x \in K_P$ , такая что существует явная связь  $(v_i, v_x) \in E$  и  $(v_j, v_x) \in E$ , порядок хотя бы одной из которой равен  $n$ .

**Очевидная неявная связь** – неявная связь между вершинами  $v_i \in V$  и  $v_j \in V$ , такая что существует  $obj \in N$ , значение характеристики  $a_i$  которого содержит  $v_i$  и  $v_j$ . В случае, если рассматриваются связи между объектами одного типа, очевидная неявная связь превращается в явную связь.

Стоит отметить, что данные определения справедливы для неориентированного графа, так как в ориентированном графе присутствует направление от одной вершины к другой и отсутствует потребность в типах вершин.

#### Алгоритм построения графа на различной структуре входных данных

В зависимости от поставленной задачи структура входных данных может быть различной. Каждый объект может отличаться по наименованию полей, типу полей и разной структурой вложенности. В таком случае рассмотрены четыре вида связи полей между собой внутри любой структуры:

- Вид 1: Связывание полей, не имеющих общих предков;
- Вид 2: Связывание полей, находящихся на одном уровне, имеющих общего предка;
- Вид 3: Связывание полей, находящихся на разном уровне, имеющих общего предка;
- Вид 4: Связывание полей одного поля.

```
{
  "Title": "article 1",
  "author": {
    "full_name": "Author's name",
    "affiliation": {
      "name": "National Resaerch Nuclear University MEPhI",
      "country": "Russian Federation"
    },
    "published_year": 2021,
    "keyword": "React"
  }
}
```

Рис. 1. Пример файла с обычной вложенностью

Рассмотрим файл со следующей структурой (рис. 1). Для получения полей из такой структуры данных и их последующего связывания будет достаточно обратиться напрямую к полю. Если интересующее поле находится на  $n$ -ом уровне вложенности, то данное поле получает предка, который описывает путь получения данных по полю. Предком считается предшествующее поле в данных. Если структура данных представляет сильную вложенность, то вводится понятие ближайшего предка. Ближайший предок – поле в данных, при обращении к которому выводится интересующее поле. Для одного поля может существовать только один ближайший предок, таким образом каждое поле будет

иметь путь, состоящий из предков и одного ближайшего предка. Ближайший предок может относиться к нескольким полям, например поле `affiliation` является ближайшим предком для `name` и `country`, по которым можно получить название аффилиации и страну.

Общим предком считается поле, по которому необходимо обратиться для получения искомым данных. Стоит отметить, что ближайший предок может быть и общим, при условии, что уровень вложенности равен одному. В противном случае общим предком будет считаться поле, находящиеся на верхнем уровне вложенности, а ближайший предок – на предпоследнем уровне вложенности.

Рассмотрим файл с вложенной структурой. На рисунке 2 представлен пример данных, в котором по одному полю можно получить список данных с одинаковыми полями. В этом случае обратиться напрямую к полю и связать их не получится, так как при обращении к предку выводится список значений. Для такого случая необходимо получить данные из списка и работать с элементами списка по отдельности, используя один и тот же алгоритм связи. В таком случае реализовывается декартово произведение элементов.

```
{
  "Title": "article_1",
  "author": [
    {
      "full_name": "Author's name",
      "affiliation": [
        {
          "name": "National Research Nuclear University MEPhI",
          "country": "Russian Federation"
        }
      ]
    },
    {
      "full_name": "Another Authors's name",
      "affiliation": [
        {
          "name": "National Research Nuclear University MEPhI",
          "country": "Russian Federation"
        },
        {
          "name": "Moscow State University",
          "country": "Russian Federation"
        }
      ]
    }
  ],
  "published_year": 2021,
  "keyword": [
    "React",
    "Python"
  ]
}
```

Рис. 2. Пример файла со вложенностью данных

Для таких структур ниже рассмотрены различные примеры связи полей.

#### Связывание полей, не имеющих общих предков

Из-за того, что поля не имеют общих предков, для связи таких полей будет достаточно связать их, используя путь до интересующего поля. При структуре, изображенной на рисунке 1, для связывания доступны следующие пары полей:

- Title – `author.full_name`;
- Title – `author.affiliation.name`;
- Title – `author.affiliation.country`;
- Title – `published_year`;
- и другие варианты, при которых узлы без общих предков будут связываться.

Если значением одного из выбранных полей будет список, то для связывания такого рода необходимо произвести декартово произведение. Например, для связи авторов с ключевыми словами следует сначала пройти по списку, получаемому по ключу `author`, получить данные по полю `full_name`, после пройти по полю `keyword`, то есть произвести декартово произведение выбранных полей. Таким образом, получатся следующие пары узлов:

- Author's name – React;
- Author's name – Python;
- Another Author's name – React;
- Another Author's name – Python;
- и другие, при которых узлы без общих предков будут связываться.

### Связывание полей, находящихся на одном уровне, имея общего предка

Данный вид позволяет связать поля, находящиеся на одном уровне вложенности, например название аффилиации и страна аффилиации. Возьмем структуру, представленную на рисунке 2.

Для связывания полей, находящихся на одном уровне, необходимо получить данные общего ближайшего предка и соединить их.

Данная связь необходима для того, чтобы связать поля, относящиеся к определенному классу данных, например: полная информация об авторе, аффилиации и т. д. При этом обеспечивается связь только внутренних данных, что позволяет сделать поэлементное соединение, то есть соединение, основанное на позиции элементов внутри данных, где соединение происходит по правилу  $[[a], [b]], [[x], [y]] \rightarrow [[a, b], [x, y]]$ .

При структуре данных (см. рис. 2) возможно связать следующую пару полей:

- author.affiliation.name – author.affiliation.country.

### Связывание полей, находящихся на разном уровне, но имея общего предка

Данный вид связи представляет собой комбинацию первого и второго вида, при котором необходимо получить данные в рамках одного предка, как при использовании второго вида связи. После получить информацию по второму полю и произвести декартово умножение как для первого вида.

Предположим, пользователь хочет связать следующие поля: author.full\_name и author.affiliation.country. Для этого нужно привести данные к одному уровню и связать поля. При структуре данных (рис 2) возможно связать следующие пары полей:

- author.full\_name – author.affiliation.country,
- author.full\_name – author.affiliation.name.

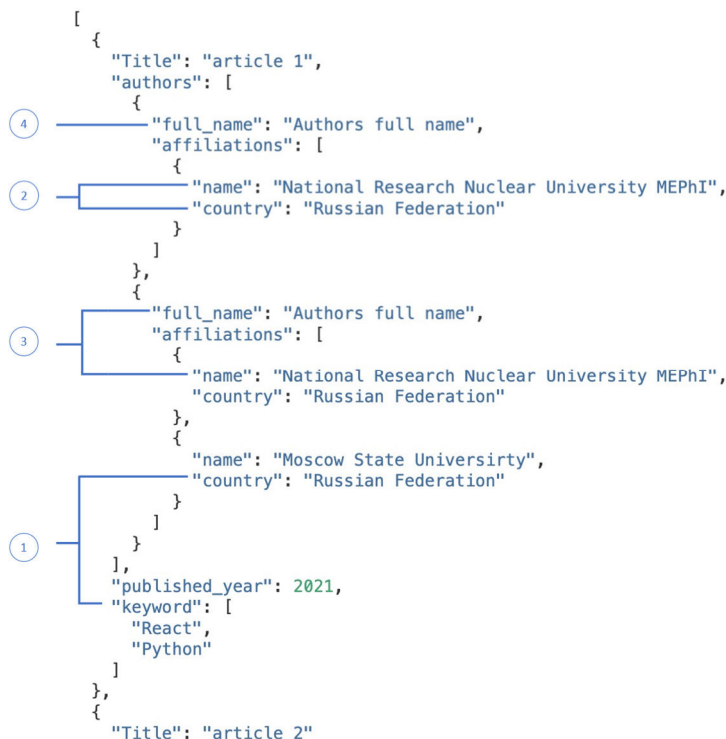


Рис. 3. Пример структуры данных



### Связывание элементов одного поля

Этот вид представляет собой частный случай первого вида связи, при котором выбраны два одинаковых поля. Такой вид может использоваться при связывании элементов в рамках одной записи. Например, рассмотрим поле “full\_name”. В рамках одного объекта (в нашем случае статьи Title 1) все авторы будут связаны между собой. Таким образом получаются узлы со следующими связями:

- Author 1 – Author 1 (для первой статьи);
- Author 1 – Author 2 (для второй статьи);
- Author 2 – Author 3 (для третьей статьи).

При рассмотренных видах связи можно сформировать следующие варианты связывания полей внутри одной структуры данных. Каждый вид обозначен своим номером: 1 – первый вид связи, 2 – второй вид связи, 3 – третий вид связи, 4 – четвертый вид связи (рис. 3).

В зависимости от вида связи будет изменяться алгоритм связывания (рис. 4). Для каждого вида реализован программный код на основе предложенного алгоритма, который позволяет проходить по всей структуре и получать необходимые данные по полям, заранее приводя их в структуру для построения графа.

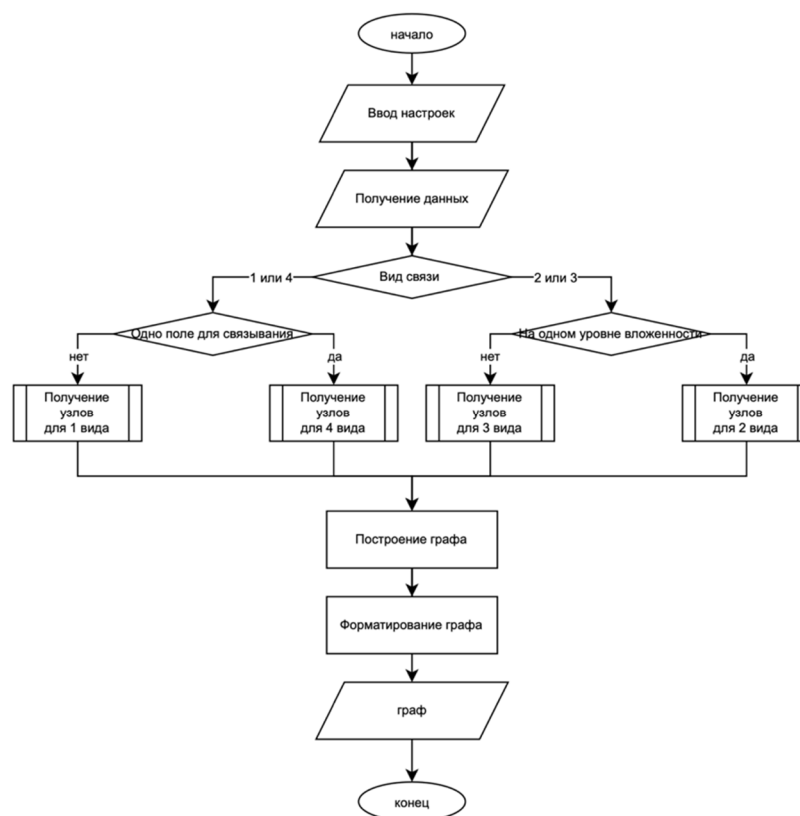


Рис. 4. Алгоритм построения графа

### 3. Проектирование и разработка инструмента для визуализации графа

Разработанный инструмент представляет собой клиент-серверную архитектуру. В данном случае клиент задает настройку графа, а сервер выполняет запрос в базу данных и преобразовывает данные в структуру для построения графа.

Инструмент состоит из трех компонентов (рис. 5): графического интерфейса, программного интерфейса и базы данных. Графический интерфейс позволяет пользователю взаимодействовать с программным интерфейсом посредством использования определенных фильтров. В свою очередь программный интерфейс взаимодействует с базой данных. В качестве хранения выбран Elasticsearch. При выгрузке данных программный интерфейс преобразует их в формат для построения графа.



При выборе узла в блоке дополнительной информации отобразится сводка по узлу, то есть выведется название узла, его вес и группа (рис. 7).



Рис. 7. Пример отображения дополнительной информации по узлу

## Сброс выделения

Необходимость сбрасывания выделения вершины обусловлена возможностью выбора других элементов графа без остаточных эффектов от предыдущего выделения. При нажатии кнопки «сброс выделения» все вершины перекрашиваются в свой первоначальный цвет. Также блок дополнительной информации уберет сводку по выбранному узлу.

## Подсвечивание вершин

Для акцентирования внимания аналитика на определенных узлах добавляется подсвечивание вершин графа (рис. 8). Данная функциональная возможность позволит привлечь внимание аналитика к определенной вершине графа по заданному условию для прослеживания связи с объектами. Подсветка реализуется с использованием мигающей сферы.

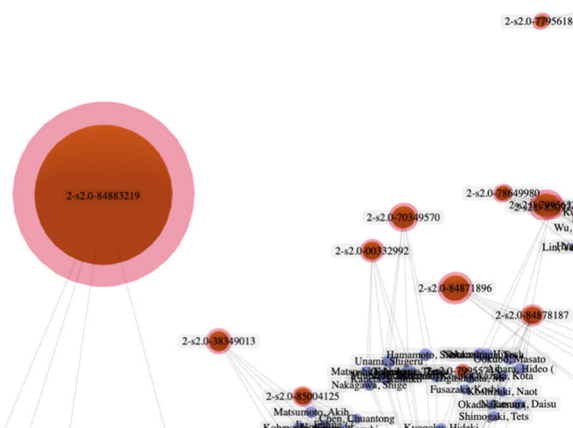


Рис. 8. Сфера для акцентирования внимания пользователя

## Кастомная настройка узлов

Пользователь для интересующих его узлов может задать цвет и группу, к которой узлы будут относиться. Цвет для той или иной группы узлов генерируется автоматически (рис. 9). Также

пользователю необходимо задать способ связи групп в отведенном поле «Отношения между полями». Выбранные поля будут отмечаться цветом той группы, к которой они принадлежат. При необходимости пользователь может настроить для группы узлов наличие подписи и мигающей сферы.

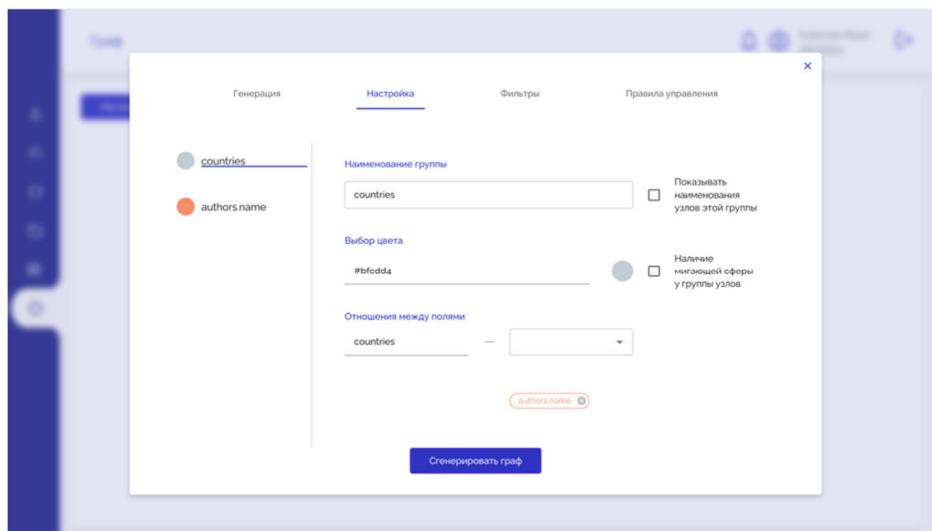


Рис. 9. Настройка узлов графа по выбранным полям

### Фильтрация данных

Данный механизм позволяет осуществлять фильтрацию графа для уменьшения количества данных. Фильтрация производится с использованием языка запросов Domain Specific Language (DSL) на основе формата JSON для составления запросов. Для поиска определенного документа, загруженного в базу данных, необходимо обратиться к созданному полю посредством определенного синтаксиса языка запросов {query: {<тело запроса>}}.

Для эксплуатации необходимо наличие два способа фильтрации: фильтрация посредством выбора соответствующих полей, операторов и типа фильтрации; расширенная фильтрация, позволяющая писать запросы, используя язык запросов Domain Specific Language (DSL) (рис. 10).

Первый метод позволяет соединять выбранные фильтры логическими операторами (must, should, must not). Между собой фильтры соединяются логическим и, для учитывания всех введенных фильтров по интересующим полям. Например, фильтрация документов по странам и фильтрация документов по определенным ключевым словам будут соединяться между собой логическим «И», то есть в документе должны учитываться условия как первого фильтра, так и второго.

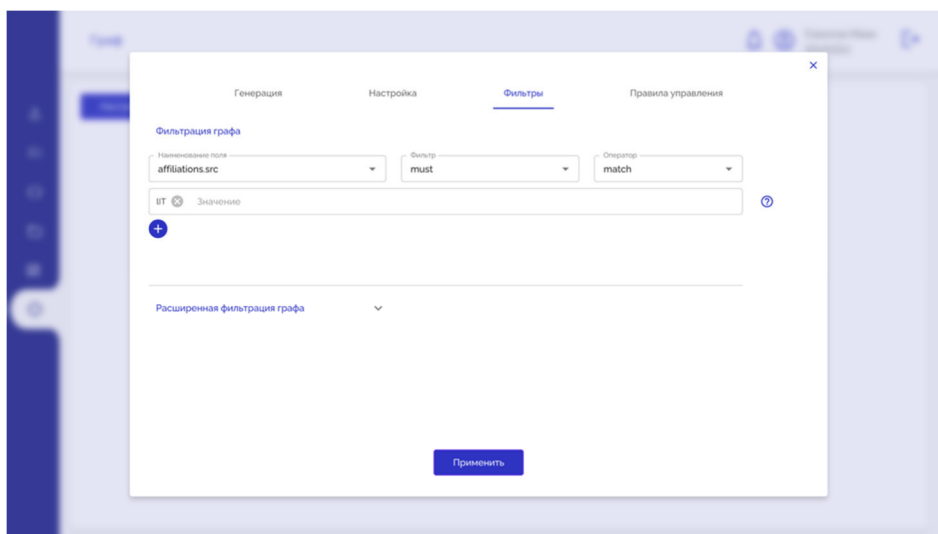


Рис. 10. Фильтрация графа по запросу в Elasticsearch

### Правила для изменения выбранных узлов

Для решения аналитических задач предложен язык формирования запросов (правил). Рассмотрим случай, при котором пользователю необходимо выделить из определенной группы узлов узел или узлы, удовлетворяющие определенным условиям. Для этого реализованы правила, на основе которых пользователь может изменить значения узла (например, изменить вес узла, цвет, группу, наименование и подсвечивание). Для этого реализован специализированный язык запросов:

Базовая структура правил

```
field_type : <values> {conditions = conditions} [change_pattern = change_pattern]
```

Здесь

- field\_type – синтересующее поле, узлы которого нужно рассматривать;
- <values> – список узлов, которые необходимо рассмотреть. Для этого списка узлов будет применяться изменение. Узлы разделяются запятыми;
- {conditions = conditions} – список условий, при удовлетворении которых будет применяться изменение. Новое условие разделяется запятой и связывается логическим И (документ должен выполнять все условия);
- [change\_pattern = change\_pattern] – список изменений.

Изменению подлежат следующие значения:

- val – вес узла (int);
- name – наименование узла (str);
- label – лэйбл узла (str);
- label\_vis – наличие лэйбла у узла (boolean);
- color – цвет узла (str);
- highlited – наличие мигающей сферы (boolean);
- group – группа узла (str).

Для присвоения одному из значений определенного типа данных необходимо перед этим значением указать сам тип данных. Реализованы следующие типы данных:

int() – числовые;

float() – десятичные или числа с плавающей точкой (0.5/1.0);

bool() – логические (True/False);

str() – текстовые. Если в текстовых данных присутствуют запятые, то нужно заключить их в слэш (/, /).

Если в поле field\_type указано значение all – {conditions = conditions} и [change\_pattern = change\_pattern] будут применяться ко всем полям и узлам.

Если в поле <values> указано <\*> – {conditions = conditions} и [change\_pattern = change\_pattern] будут применяться к узлам по выбранному полю field\_type.

Для указания {conditions = conditions} необходимо точно указать поле, по которому нужно провести сравнение (доступны следующие виды сравнения: =, ==, >, <=, <). Каждое условие связано с предыдущим логическим «И». Пример написания условия:

```
{condition 1 = value 1, condition 2 <= value 2}
```

Здесь condition 1/ condition 2 – поля, по которым нужно провести условие; value 1/ value 2 – значение для сравнения.

Для указания [change\_pattern = change\_pattern] необходимо указать изменение. Пример написания изменения:

```
[color = user_color, val = int(user_val)]
```

Здесь user\_color – цвет, который задал пользователь; user\_val – вес, который задал пользователь.

Пример реализованных правил:

```
author.gender: <female> [color = red];
countries: <US, RU> [color = blue, group = countries];
eid: <*> {probability >= float(0.5)} [val = int(10), highlited = bool(True)]
```

В графическом интерфейсе для того, чтобы использовать правила, реализована отдельная вкладка, в которую необходимо вводить правило (рис. 11).

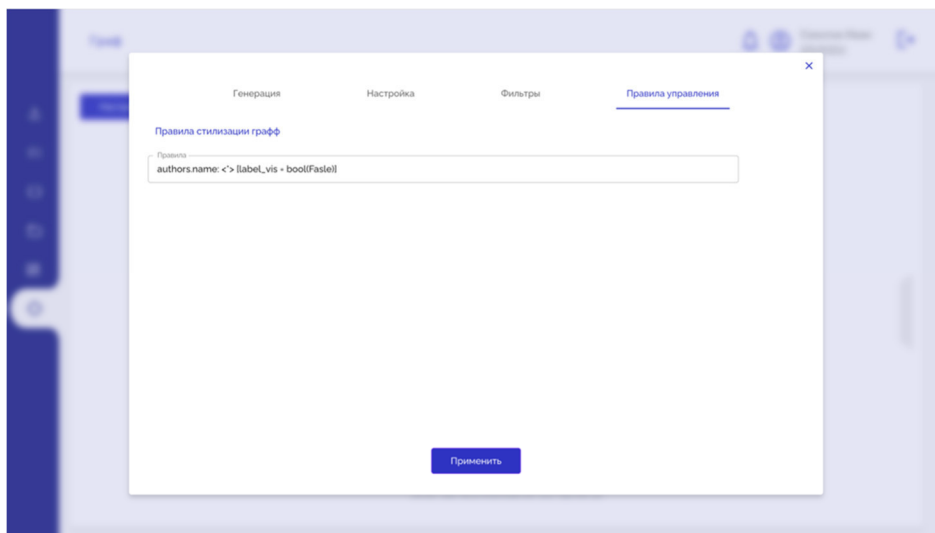


Рис. 11. Пример реализации правил для настройки узлов графа

### Результаты экспериментов

Разработанный инструмент используется для решения научно-технических задач. В рамках апробации инструмента рассмотрены две задачи: обнаружение нарушений обязательств по ядерному нераспространению, визуальный анализ публикационной активности научной организации.

#### 5. Обнаружение нарушений обязательств по ядерному нераспространению

Апробация системы в данной задаче проходит на 5 280 публикациях по ядерным технологиям. Для построения графа выбраны следующие поля: countries – страны (зеленые узлы), authors.full\_name – полное имя автора (голубые узлы), eid – идентификатор статьи (коричневые узлы). Полученный граф отображен на рисунке 12.

Для акцентирования внимания на ядерных державах узлы, представляющие страны «ядерной пятёрки», окрашены в зелёный цвет, тогда как все остальные страны отображаются красным.

При исследовании графа обнаружено, что у США есть совместные публикации с двумя странами не из ядерной пятёрки (Австралия и Норвегия), и по ним производилась фильтрация. Для уменьшения количества публикаций производилась фильтрация для статей США, в которых должны были встречаться Австралия или Норвегия.



Рис. 12. Граф взаимосвязи авторов, стран и статей

После применения фильтрации графа обнаружено, что с США писали еще Германия и Дания (рис. 13). Таким образом, можно выделить коллаборационные статьи, представляющие интерес для эксперта.

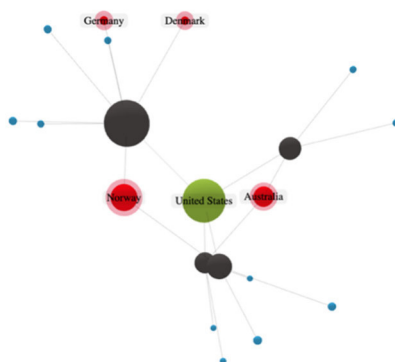


Рис. 13. Фильтрация графа по США

Также построен граф тематик исследований по странам (рис. 14). Из графа можно выделить пересечение стран, которые пишут по нескольким тематикам (красное пересечение) и группы стран, которые работают в основном только по одной тематике (синее пересечение).

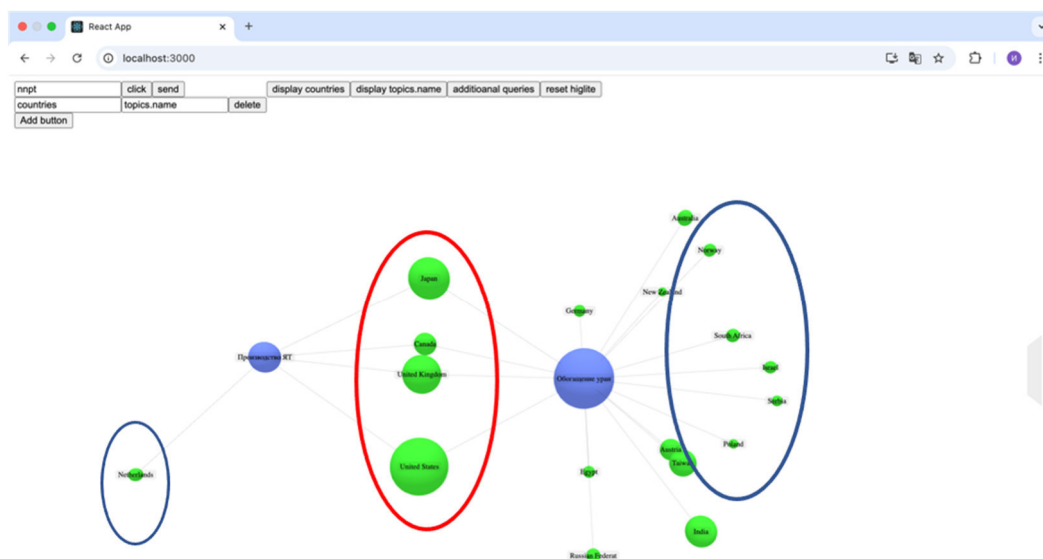


Рис. 14. Граф по тематикам исследований стран

## 6. Визуальный анализ публикационной активности научной организации

Для анализа структуры и тематической направленности публикационной активности Объединённого института ядерных исследований (ОИЯИ) использован метод визуализации на основе графовых моделей. В качестве данных применены метаданные 36 008 публикаций. На этих данных строится граф по следующим полям: *affiliations.name* (название аффилиации) – красные узлы и *keyword* (ключевые слова) – розовые узлы. В итоге реализован граф с 1470 узлами и 5291 ребром (рис. 15).

Полученный граф разделяется на несколько кластеров, которые описывают взаимосвязи организаций по определенным тематикам. На рисунке 16 представлены направления тематик ОИЯИ. По выделенным узлам можно выявить, с какими организациями ОИЯИ взаимодействует и по каким тематикам.

Для уменьшения количества данных производилась фильтрация по ОИЯИ. В итоге реализован граф из 1406 узлов и 5504 ребер (рис. 17).





Рис. 15. Граф публикационной активности ОИЯИ

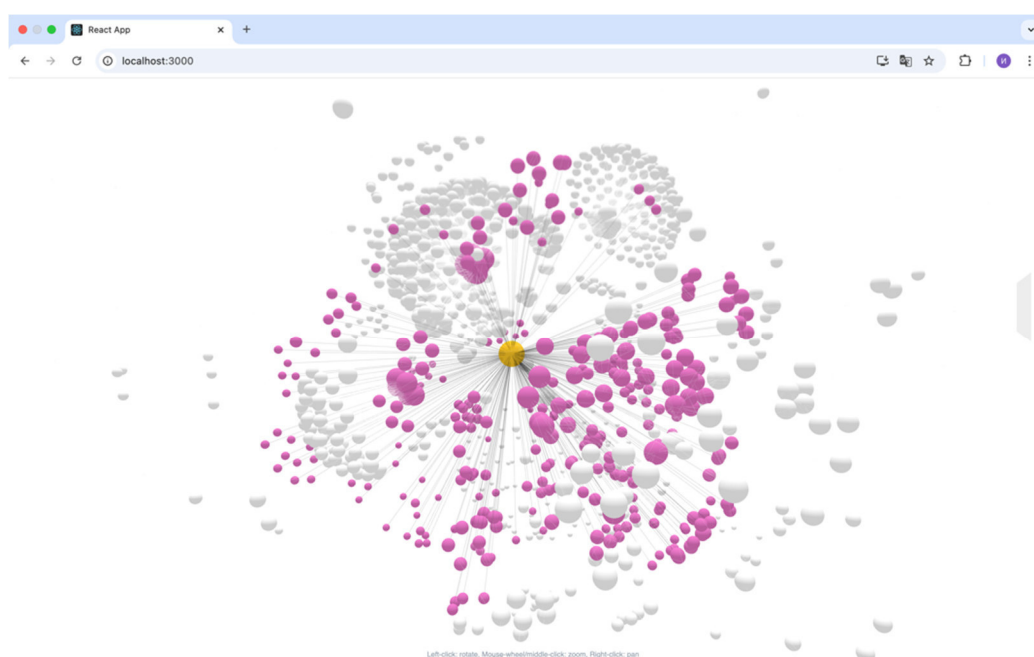


Рис. 16. Подсвечивание связей узла ОИЯИ

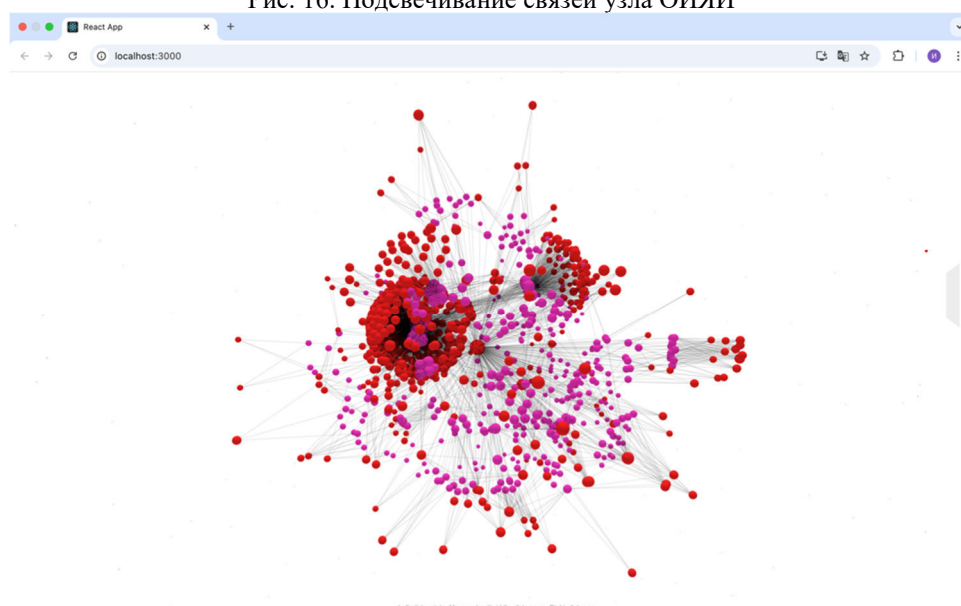


Рис. 17. Отфильтрованный граф по ОИЯИ

На полученном графе чётко прослеживаются связи ОИЯИ с внешними организациями. В частности, кластер, представленный на рисунке 18, посвящен тематике физики частиц и демонстрирует основные взаимодействия института в этой области.

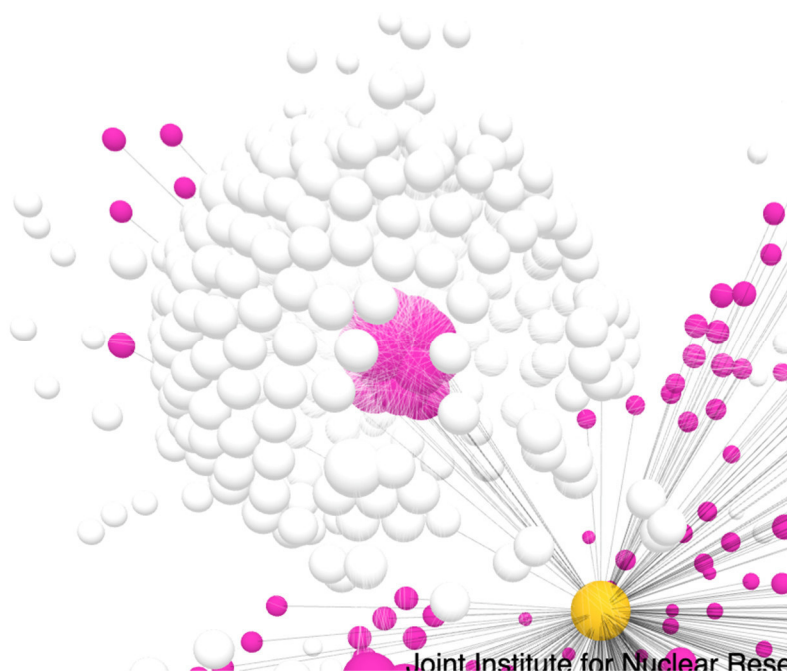


Рис. 18. Кластер тематики ОИЯИ, посвященный направлению физики частиц

Применение фильтрации по ключевым словам «jets» и «quark gluon plasma» позволяет детализировать эти связи и выделить конкретные организации-партнёры ОИЯИ в рамках данного научного направления (рис. 19).

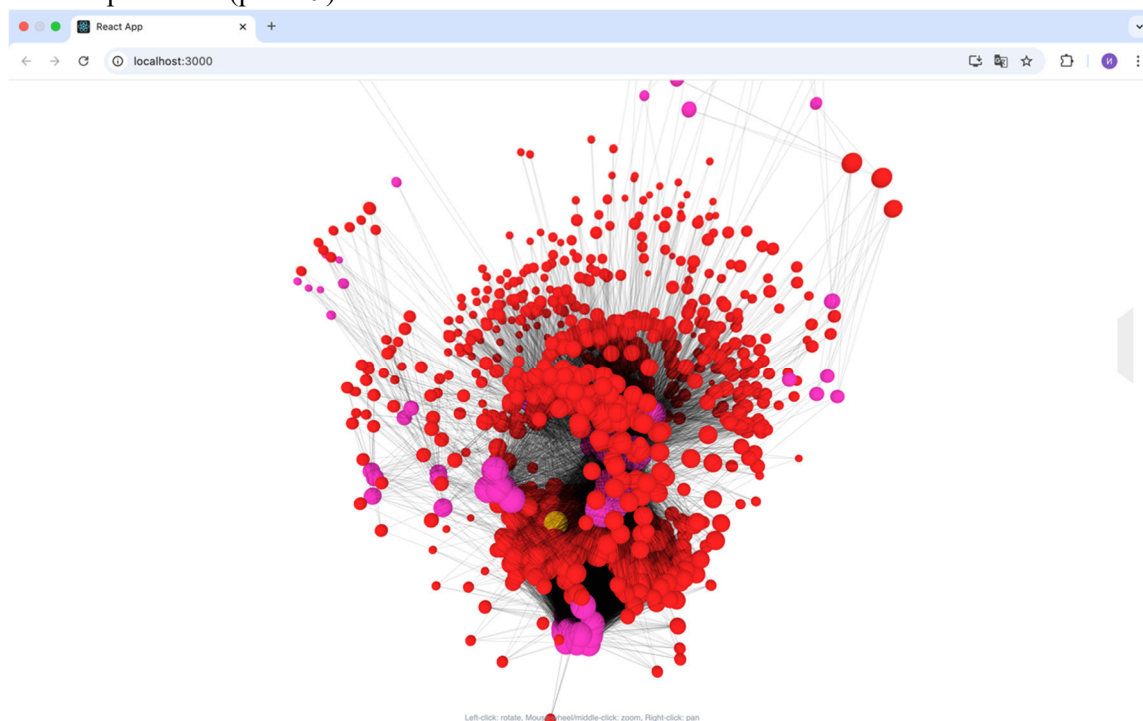


Рис. 19. Граф взаимосвязей между организациями по выбранной тематике (ОИЯИ – желтый узел)

### Обсуждение результатов

Разработанный метод идентификации явных и неявных связей между объектами, а также реализованный инструмент построения графовой модели продемонстрировали высокую

эффективность при решении аналитических задач различной сложности. Основным преимуществом предложенного подхода является его независимость от структуры входных данных, что позволяет обрабатывать как простые табличные данные, так и сложные вложенные форматы. Это делает инструмент универсальным и применимым к широкому кругу задач системного анализа.

Для реализации пользовательского интерфейса использовался ReactJS, а логика обработки и хранения данных была построена на основе Python и Elasticsearch. Поддержка различных типов данных, возможность динамического изменения графа и наличие интерактивных механизмов фильтрации обеспечивают аналитикам широкие возможности контроля над процессом анализа. Это позволяет эффективно использовать систему как опытным специалистам, так и пользователям с минимальной технической подготовкой.

Таким образом, предложенный метод и инструмент построения графовых моделей представляют собой мощное средство для выявления как явных, так и неявных связей между объектами, независимо от структуры и типа данных. Они могут быть использованы в различных областях: от научного анализа и библиометрии до политического и экономического мониторинга, обеспечивая аналитикам качественно новый уровень понимания сложных систем и взаимосвязей внутри них.

### **Выводы и заключение**

В современных условиях, когда объемы научно-технических данных стремительно растут, важными становятся инструменты, способные преобразовать исходные массивы в наглядные и удобные для анализа графовые структуры. В рамках данного исследования разработан метод построения графа на входных данных различной структуры и выявления явной и неявной связи между объектами, а также инструмент графовой визуализации, который позволяет решить проблему обнаружения взаимосвязей между информационными объектами с помощью механизмов выбора узлов и способа их связи по интересующим данным, настройкой графа и его визуального отображения.

Результаты апробации предложенного метода и инструмента позволяет анализировать два прикладных сценария, связанных с построением отношений между объектами, на примере выявления тематик публикаций по странам и организациям: для этого использованы два набора данных: 5 280 документов по тематике ядерных материалов, 36 008 научных публикаций ОИЯИ.

### **Список литературы**

1. Козицын А. С. Нахождение скрытых зависимостей между объектами на основе анализа больших массивов библиографических данных // VI Международная конференция «Актуальные проблемы системной и программной инженерии» (АПСПИ 2019). Москва, 2019.
2. Тимонин А. Ю., Бождай А. С. Методы анализа гетерогенных данных для построения социального профиля // Russian journal of management. 2017. Т. 5, № 3. С. 481-489.
3. Lindley D. V. Regression and Correlation Analysis. London: Palgrave Macmillan UK, 1990.
4. Georgi M. Methods for Mining Causality from Observations in Artificial Intelligence // Izvestiya SFedU. Engineering Sciences. 2023. Vol. 3, no. 192. Pp. 125-134.
5. Батищев С. В., Искварина Т. В., Скобелев П. О. Методы и средства построения онтологий для интеллектуализации сети интернет // Известия Самарского научного центра РАН. 2002. Т. 4, № 1. С. 91-103.
6. Батура Т. В. Методы и системы семантического анализа текстов // Программные продукты и системы: Международный журнал. 2016. Т. 12.

## Визуализация вихревой системы треугольного крыла

Т. В. Константиновская, В. Е. Борисов, А. Е. Луцкий

ИПМ им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В работе рассматривается задача образования вихревых структур при сверхзвуковом обтекании дельтовидного крыла. Показаны способы визуализации этих структур. Представлены результаты применения методов научной идентификации и визуализации вихревых течений к анализу и обработке полученных в результате расчетов численных данных. Дельта-крыло расположено под углом атаки  $14^\circ$  градусов к набегающему потоку. Моделирование результатов основывалось на использовании URANS уравнений. Численные расчеты проведены на гибридной суперкомпьютерной системе К-60 в ЦКП ИПМ им. М.В. Келдыша РАН. Использованы различные подходы и критерии для идентификации вихревых структур и визуализации результатов численных расчетов.

**Ключевые слова:** обтекание дельта-крыла, вихрь, идентификация вихревых структур, научная визуализация.

## Visualization of delta wing vortex system

T. V. Konstantinovskaya, V. E. Borisov, A. E. Lutsky

Keldysh Institute of Applied Mathematics, Moscow, Russia

**Abstract.** The problem of vortex structures formation in supersonic flow around a delta wing is considered in the paper. The methods of visualization of these structures are shown. The results of application of vortex flows scientific identification and visualization methods to the analysis and processing of numerical data obtained as a result of simulations are presented. The delta wing is located at an attack angle of 14 degrees to the incoming flow. URANS equations were used for the results modeling. Numerical simulations were carried out on a hybrid supercomputer system K-60 at the CKP of Keldysh Institute of Applied Mathematics of the RAS. Various approaches and criteria have been used for vortex structures identification and visualization the numerical simulations results.

**Keywords:** flow around delta wing, vortex, vortex structures identification, scientific visualization

## Введение

Дельтовидные, или треугольные, крылья начали активно использоваться в авиации с 1950-х годов с развитием высокоскоростных летательных аппаратов (ЛА), где получили широкое применение. Такие крылья используются также в области аэрокосмических технологий, например, для таких аппаратов, как Буран и Space Shuttle.

Сегодня интерес к треугольным крыльям возрастает не только в авиации, так как скорости передвижения имеют большое значение в современном мире, но и в аэрокосмической сфере в контексте разработки многоразовых космических систем. В частности, следует упомянуть космический корабль Starship, у которого вторая ступень оснащена треугольными аэродинамическими поверхностями как в носовой, так и в хвостовой частях [1].

Широкое применение треугольных крыльев способствовало активному исследованию их аэродинамических характеристик как в нашей стране, так и за рубежом.

При проектировании и создании высокоскоростных летательных аппаратов ключевое значение имеет тщательное изучение аэродинамических характеристик их полета. Одной из главных задач является исследование вихревых структур, неизбежно сопровождающих сверхзвуковое обтекание треугольного крыла.

Для обработки и анализа данных, полученных в ходе численного и экспериментального моделирования, могут быть эффективно использованы методы научной идентификации и визуализации вихревых структур. Эти методы предоставляют инструменты не только для визуализации потоков, но и для их глубокого анализа [2–5].

В данной работе представлены результаты визуализации численного исследования сверхзвукового обтекания дельтовидного крыла. Показано образование вихревых структур на кромке и поверхности крыла. Авторами используются различные методы идентификации и визуализации вихревых течений, в частности, Лютекс (Liutex) критерий [5–8], относящийся к третьему поколению подобных методов.

Для проведения численных расчетов используется подход URANS с моделью турбулентности SA. Расчеты проводились на многопроцессорной гибридной системе К-60 в ЦКП ИПМ им. М.В. Келдыша РАН [9].

### Постановка задачи

Исследовалось сверхзвуковое обтекание треугольного крыла, которое имело острые кромки, стреловидность  $78^\circ$ , полуразмах 0.1118 м, корневую хорду 0.526 м и угол атаки  $\alpha = 14^\circ$ . Число Маха набегающего потока принималось  $M_\infty = 3$ . Число Рейнольдса задавалось  $Re_L = 1 \times 10^7$  ( $L$  – характерная длина обезразмеривания вычислительной модели, здесь  $L = 1$  м). На рисунке 1 представлена схема расчетной области. Использовалась неструктурированная сетка, содержащая 7315200 ячеек – криволинейных гексаэдров.

Течение рассматривалось на расстоянии вплоть до 2.8 корневой хорды крыла вниз по потоку от задней кромки крыла.

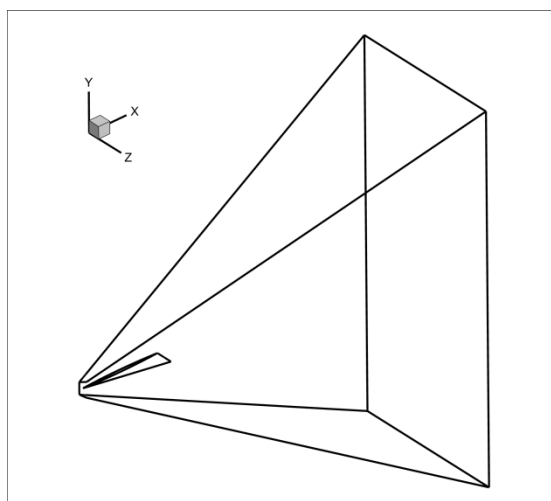


Рисунок 1. Расчетная область, установка треугольного крыла

### Численные расчеты

В расчетной области задавалась связанная декартова система координат, начало которой совпадает с вершиной треугольного крыла; ось  $O_x$  направлена вдоль корневой хорды крыла,  $Oy$  совпадает с плоскостью симметрии,  $O_z$  перпендикулярна плоскости симметрии. Срединная поверхность крыла лежит в плоскости  $z = 0$ .

Численные данные были получены с использованием авторского программного комплекса ARES [10] расчета трехмерных турбулентных течений вязкого сжимаемого газа, который был разработан и программно реализован в ИПМ им. М.В. Келдыша РАН.

Для расчетов трехмерного турбулентного течения сжимаемого газа использовалась система нестационарных осредненных по Рейнольдсу и Фавру уравнений Навье–Стокса (URANS). Использовалась однопараметрическая модель турбулентности Спаларта–Алмараса (SA) в модификации для сжимаемых течений [11]. Начальные и граничные условия ставились стандартным образом.

Аппроксимация уравнений модели осуществлялась в пространственном направлении с применением метода конечных объемов со схемой реконструкции второго порядка точности TVD. Метод конечных объемов в предположении, что расчетная область разбита на неперекрывающиеся многогранные ячейки, реализуется путем интегрирования системы уравнений модели для каждой ячейки, после чего объемные интегралы от потоков преобразуются в поверхностные интегралы по граням ячейки. Для расчета невязких потоков на гранях ячеек был применен обобщенный метод С. К. Годунова с точным римановским солвером. Для временной аппроксимации уравнений использовались как явная, так и неявная схемы. Подробное описание используемого численного метода можно найти в [12].

Расчеты проводились на гибридной суперкомпьютерной системе К-60 [13] в ЦКП ИПМ им. М.В. Келдыша РАН.

### Визуализация вихревой системы

Для визуального представления и анализа вихревых структур необходимо их определить и выделить из остального потока. На данный момент нет единого определения вихря, что приводит к разнообразию подходов и методов для его идентификации. В связи с этим продолжаются дальнейшие работы по поиску и созданию новых и оптимальных методов идентификации и визуализации вихревых структур для их изучения [14–16].

Для этих целей был разработан специальный отдельный модуль внутри авторского программного комплекса ARES, позволяющий в постпроцессинговом режиме обработки данных определять и анализировать вихревые структуры на гексагональных сетках. В его рамках реализованы некоторые классические методы научной идентификации и визуализации, такие как  $\lambda_2$ ,  $Q$ -критерий и др [17, 18]. Также модуль содержит Лютекс (Liutex) метод научной визуализации – один из последних и наиболее современных критериев идентификации вихревых структур, относящийся к третьему поколению таких методов. Упомянутый постпроцессинговый модуль формирует выходные данные в формате пакета программ Tecplot.

В представленной работе визуализация результатов численных вычислений и их анализ проводятся с помощью различных методов идентификации и визуализации вихревых структур. В том числе применяются подходы с прямым использованием градиентов основных и производных газодинамических свойств течения.

### Результаты расчетов и визуализация вихревых структур

На рисунке 2 представлен общий вид полученного течения при сверхзвуковом обтекании треугольного крыла. Для визуализации результатов течения на рисунке 2 используются изоповерхности модуля завихренности  $Vort$  (значение 200). Видно, что полученная вихревая система достаточно сложносоставная. На рисунке 2 различим основной вихрь, а также можно наблюдать влияние хвостового скачка на вихревую систему.

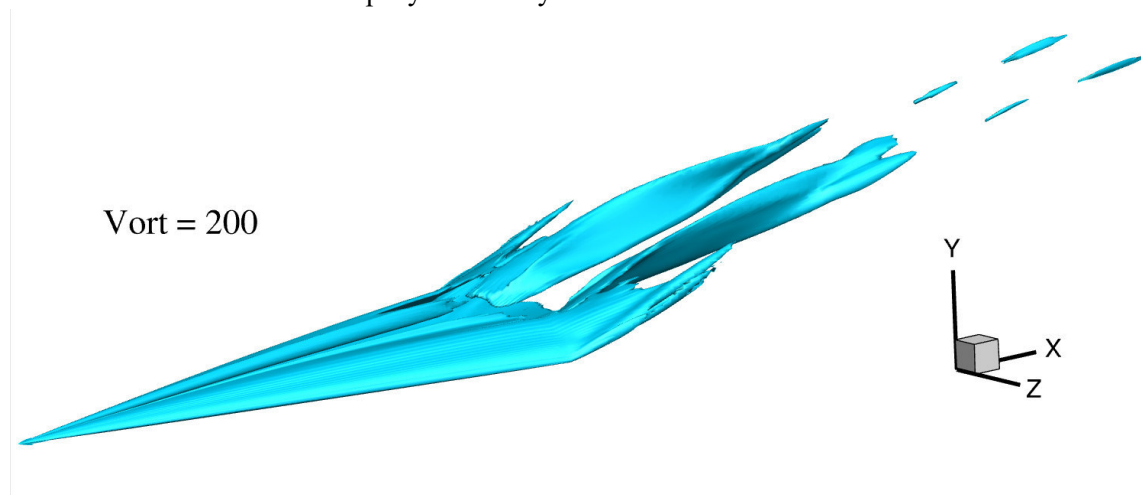


Рисунок 2. Общий вид полученного численно течения

В рассматриваемой постановке вихревая система образуется на подветренной стороне треугольного крыла и состоит из следующих основных элементов. Формируется вихрь на боковой кромке (первичный вихрь), основной вихрь и вторичный вихрь (рис. 3). На этом рисунке показаны распределение  $x$ -ой составляющей ротора скорости (завихренности)  $XVorticity$  и линии тока в поперечном сечении  $x = 0.4$ , пересекающем крыло ближе к задней кромке. Вторичный вихрь вызван отрывом пограничного слоя (на рисунке 3 обозначено красным цветом), который возникает из-за неблагоприятного градиента давления, а именно: роста давления по направлению к боковой кромке. Вторичный вихрь имеет направление вращения, противоположное направлению вращения основного вихря.



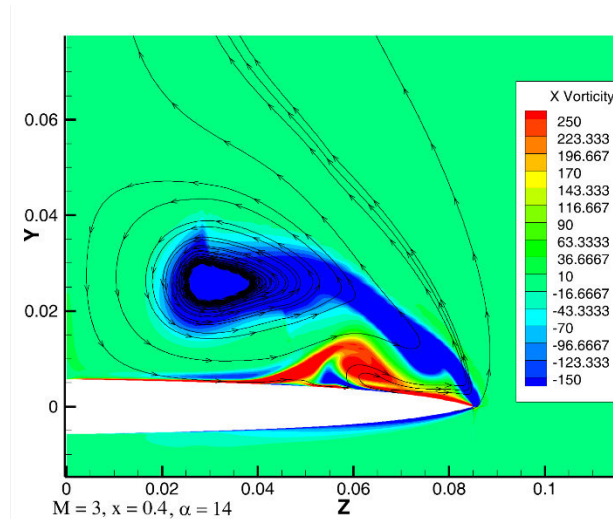


Рисунок 3. Распределение  $XVorticity$  и линии тока в проекции на поперечное сечение  $x = 0.4$

Использование продольной завихренности  $XVorticity$  позволяет показать направление вращения выявленных вихревых структур (рис. 3 и 4). На рисунке 4 показано применение  $\lambda_2$ -критерия, изображены изолинии  $\lambda_2 = -3000$ . При этом они раскрашены продольной составляющей вектора завихренности, что действительно дает в рассмотренной конфигурации представление не только о наличии и положении вихревых структур, но и о направлении их вращения. На рисунке 4 отчетливо различимы в пространстве основные вихри (на боковой кромке, вторичный, основной).

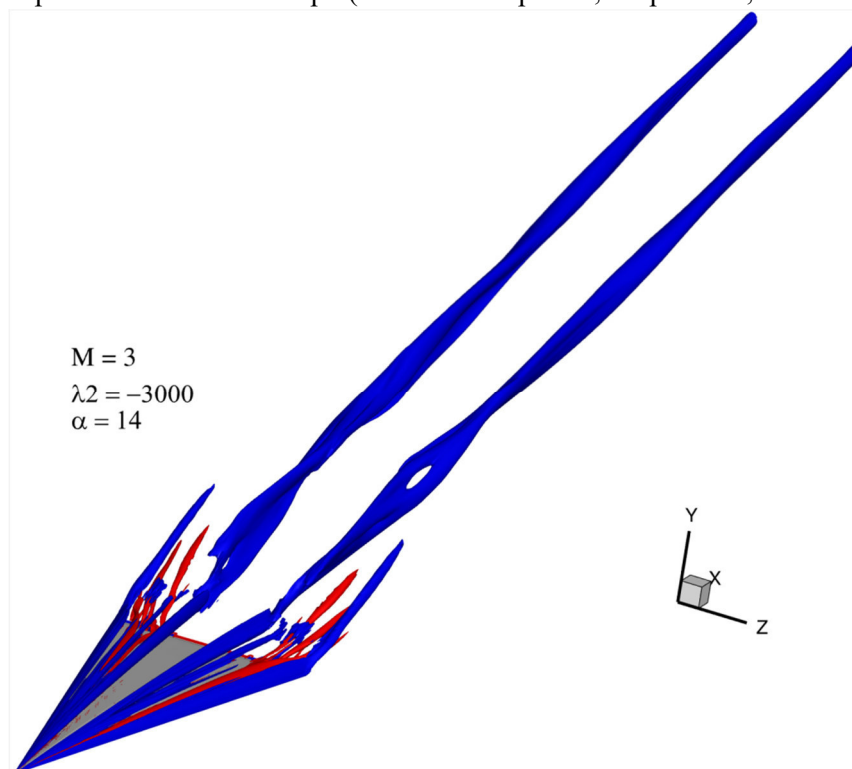


Рисунок 4. Визуализация вихревых структур при помощи  $\lambda_2$ -критерия: изоповерхности  $\lambda_2 = -3000$

На рисунке 5 показаны оси структурообразующих вихрей: на боковой кромке, основного и вторичного, получены они с помощью анализа различных свойств течения.

Комбинация применения модуля завихренности и  $\lambda_2$ -критерия для визуального представления свойств течения и положения вихревых структур представлена на рисунке 6, где показаны поперечные сечения для трех значений  $x$  ( $x = 0.45$ ,  $x = 0.75$  и  $x = 1.0$ ) с соответствующими параметрами в них. Можно видеть, что методы дают согласованные результаты.



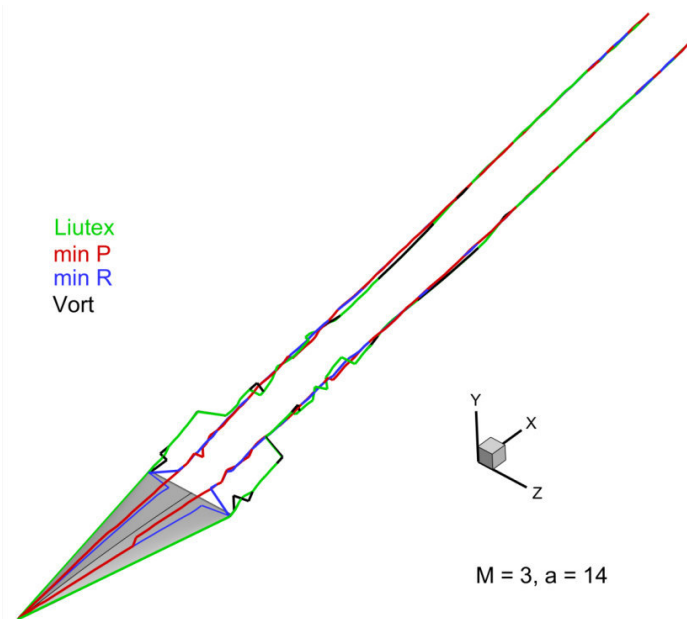


Рисунок 5. Определение осей вихрей: первичного (на боковой кромке), основного и вторичного:  $M = 3, \alpha = 14^\circ$

Получено, что на некотором расстоянии от крыла его вихревая система сливается в одну вихревую структуру, которая хорошо различима и простирается вплоть до границ рассматриваемой области (см. рис. 4–6).

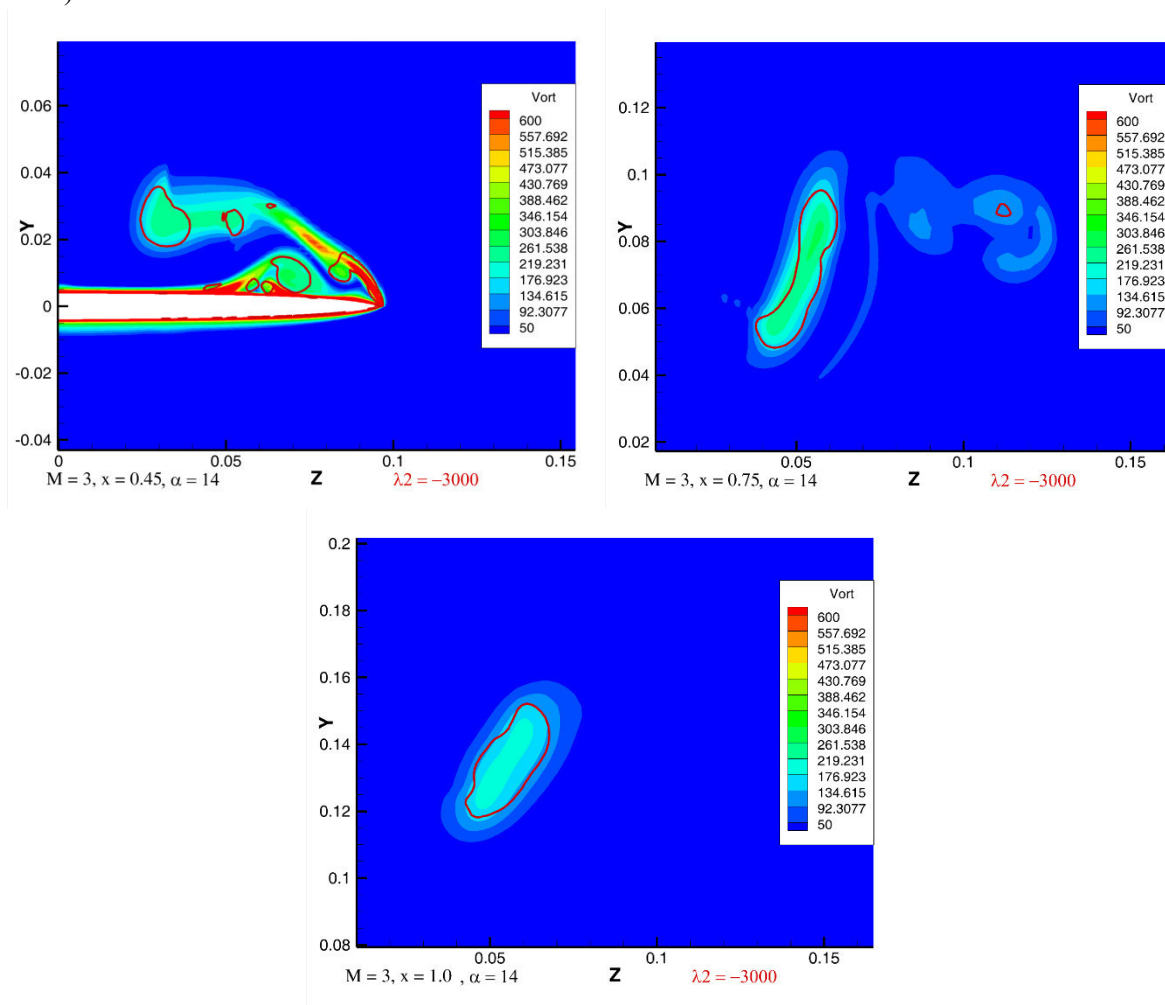


Рисунок 6. Визуализация вихревых структур при помощи модуля завихренности  $Vort$  и  $\lambda_2$ -критерия: показаны поперечные сечения при  $x = \text{const}$ , изоповерхности  $\lambda_2 = -3000$  показаны красной линией

На рисунке 7 представлены изоповерхности завихренности ( $Vort = 200$ ) с изображением коэффициента давления  $C_p$  поверх, показаны разные ракурсы: вид спереди-сверху (а), половина крыла до плоскости симметрии (б). Добавление коэффициента давления позволяет увидеть нюансы влияния хвостового скачка и изменения давления в вихревой системе в трехмерном пространстве.

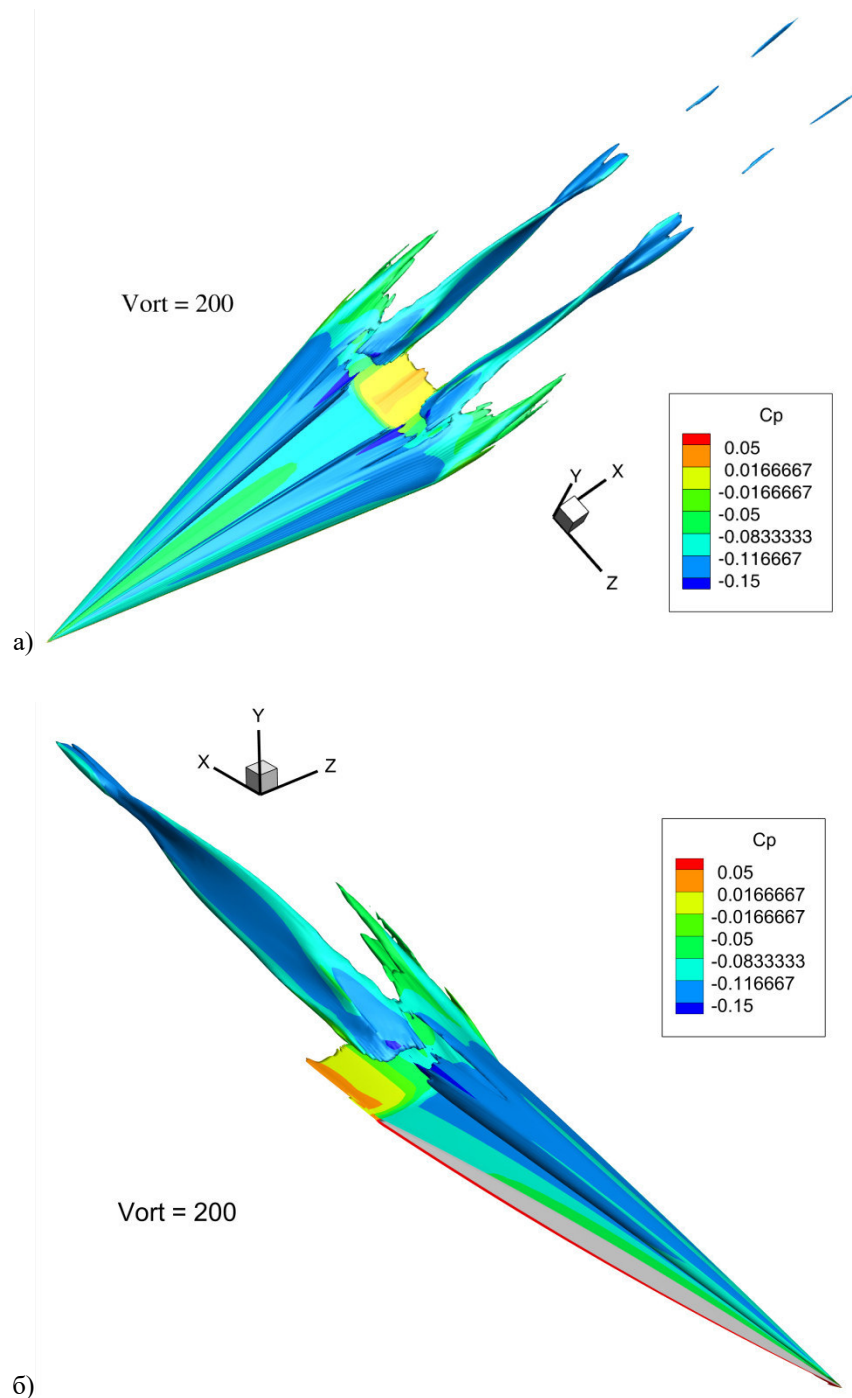


Рисунок 7. Визуализация вихревых структур при помощи изоповерхности завихренности ( $Vort = 200$ ) с изображением коэффициента давления  $C_p$  поверх, представлены разные ракурсы: вид спереди-сверху (а), половина крыла до плоскости симметрии (б)

Таким образом, различные методы идентификации и визуализации вихревых структур могут использоваться как отдельно, так и в комбинации для достижения более точной и информативной идентификации и визуализации вихревых структур в различных средах, позволяя при этом добиться взаимодополняющих результатов и более глубокого анализа данных.

## Заключение

В работе рассматриваются численные результаты расчетов сверхзвукового обтекания треугольного крыла. Представлены результаты анализа и научной визуализации полученных данных. Продемонстрировано применение различных методов и подходов к визуализации вихревых структур. Численные расчеты были выполнены с помощью авторского программного комплекса ARES на суперкомпьютере К-60 в ЦКП ИПМ им. М.В. Келдыша РАН.

Получено, что при сверхзвуковом обтекании треугольного крыла в рассмотренной конфигурации его вихревая система состоит из трех структур: первичного вихря (на боковой кромке), вторичного и основного. На некотором расстоянии за концевой кромкой крыла эти структуры сливаются в одну, составляя один вихрь, который простирается вплоть до конца интересующей нас области.

На примере рассмотренной задачи показано, как можно с помощью комбинирования различных способов и подходов к идентификации и визуализации вихревых структур реализовать разносторонний анализ численных данных.

## Источник финансирования

Работа выполнена при финансовой поддержке РФФИ, проект № 24-21-00230.

## Список литературы

1. <https://www.spacex.com/vehicles/starship>
2. Yang Wen-Jei. Handbook of flow visualization. 2nd edition. CRC Press, 2001. 724 p.
3. Волков К.Н. Методы визуализации вихревых течений в вычислительной газовой динамике и их применение при решении прикладных задач // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 3 (91).
4. Epps B.P. Review of Vortex Identification Methods // 55<sup>th</sup> AIAA Aerospace Sciences Meeting. 2017. AIAA 2017-0989. DOI: 10.2514/6.2017-0989
5. Third generation of vortex identification methods: Omega and Liutex/Rortex based systems / C. Liu, Y. Gao, X. Dong, Y. Wang, J. Liu, Y. Zhang, X. Cai, N. Gui // J. Hydrodyn. 2019. **31** 2, pp. 205–223.
6. Stretching and shearing contamination analysis for Liutex and other vortex identification methods / P. Shrestha, C. Nottage, Y. Yu, O. Alvarez, C. Liu // Advances in Aerodynamics. 2021. **3** 8.
7. Liu J., Liu C. Modified normalized Rortex/vortex identification method // Phys. Fluids. 2019. 31:061704, 6 p.
8. Kolář V., Šístek J. Stretching response of Rortex and other vortex-identification schemes // AIP Advances. 2019. **9**, 105025. DOI: 10.1063/1.5127178
9. ЦКП (центр коллективного пользования) ИПМ им. М.В. Келдыша РАН. URL: <https://ckp.kiam.ru/?home>
10. Программный комплекс ARES для расчета трехмерных турбулентных течений вязкого сжимаемого газа на высокопроизводительных вычислительных системах. Свидетельство о регистрации программы для ЭВМ RU 2019667338, 23.12.2019 / В.Е. Борисов, А.А. Давыдов, И.Ю. Кудряшов, А.Е. Луцкий.
11. Параллельная реализация неявной схемы на основе метода LU-SGS для моделирования трехмерных турбулентных течений / В.Е. Борисов, А.А. Давыдов, И.Ю. Кудряшов, А.Е. Луцкий, И.С. Меньшов // Математическое моделирование. 2014. № 26 (10). С. 64–78; Parallel Implementation of an Implicit Scheme Based on the LU-SGS Method for 3D Turbulent Flows / V.E. Borisov, A.A. Davydov, I.Yu. Kudryashov, A.E. Lutsky, I.S. Men'shov // Mathematical Models and Computer Simulations. 2015. No. 7 (3). P. 222–232.
12. Allmaras S.R., Johnson F.T., Spalart P.R. Modifications and Clarifications for the Implementation of the Spalart-Allmaras Turbulence Model // Seventh International Conference on CFD (ICCFD7), Big Island, Hawaii, 9-13 July 2012.
13. Вычислительный комплекс К-60. URL: <https://www.kiam.ru/MVS/resources/k60.html>.
14. Rortex—a new vortex vector definition and vorticity tensor and vector decompositions / C. Liu, Y. Gao, S. Tian, X. Dong // Phys. Fluids. 2018. 30:035103.
15. Surface techniques for vortex visualization / Ch. Garth, X. Tricoche, T. Salzbrunn, T. Bobach, G. Scheuermann // VISSYM'04: Proceedings of the Sixth Joint Eurographics – IEEE TCVG conference on Visualization, Konstanz Germany, May 19-21, 2004. P. 155-164.
16. Canivete Cuissa J.R., Steiner O. Innovative and automated method for vortex identification – I. Description of the SWIRL algorithm // A&A 668. 2022. A118. doi: 10.1051/0004-6361/202243740.
17. Jeong J., Hussain F. On the identification of a vortex // Journal of Fluid Mechanics. 1995. Vol. 285. P. 69–94.
18. Hunt J.C.R., Wray A.A., Moin P. Eddies, stream, and convergence zones in turbulent flows // Technical Report № CTR-S88. Palo Alto: Center for Turbulent Research. 1988. P. 193–208.

УДК 004.93, 531.715.2, 531.717.1, 67.05, 62.5  
DOI: 10.25686/978-5-8158-2474-4-2025-494-497

## Разработка методики позиционирования изделия при многокоординатной лазерной обработке

А. А. Молотков<sup>1</sup>, Д. Л. Сапрыкин<sup>2</sup>, О. Н. Третьякова<sup>1</sup>, Д. Н. Тужилин<sup>2</sup>, А. А. Шамордин<sup>2</sup>

<sup>1</sup> Московский авиационный институт (национальный исследовательский университет) МАИ, Москва, Россия

<sup>2</sup> Лаборатория промышленных исследований группы компаний НПЦ  
«Лазеры и аппаратура ТМ», Москва, Россия

**Аннотация.** В статье рассматривается методика видеопозиционирования изделий с криволинейными поверхностями при многокоординатной лазерной обработке. Разработанный функционал позволяет определить позицию нулевой точки чертежа по изображению с промышленной камеры. Основу алгоритма распознавания составляет модифицированный метод поиска в ширину BFS. Введение конвейерной обработки видеопотока с настройкой фильтров и детекторов обеспечивает адаптацию под различные материалы и условия съёмки. Метод позволяет динамически рассчитывать коэффициент пересчёта пикселей в миллиметры, что критически важно при работе с пространственно-сложными поверхностями. Представленное решение повышает точность позиционирования, снижает требования к оснастке и сокращает подготовительное время.

**Ключевые слова:** многокоординатное позиционирование, видеопозиционирование, обработка изображений, конвейер фильтрации, машинное зрение.

## Development of a positioning method for a workpiece in multi-axis laser processing

A. A. Molotkov<sup>1</sup>, D. L. Saprykin<sup>2</sup>, O. N. Tretyakova<sup>1</sup>, D. N. Tuzhilin<sup>2</sup>, A. A. Shamordin<sup>1</sup>

<sup>1</sup> Moscow aviation Institute (national research university) MAI, Moscow, Russia

<sup>2</sup> PROMIS LAB of the group of companies "Lasers and Equipment TM", Moscow, Russia

**Abstract.** The article presents a method for video-based positioning of parts with curved surfaces during multi-axis laser processing. The developed functionality enables the determination of the drawing's zero point position based on images from an industrial camera. The core of the recognition algorithm is a modified breadth-first search (BFS) method. The integration of a configurable video stream processing pipeline with filter and detector tuning ensures adaptability to various materials and imaging conditions. The method allows for dynamic calculation of the pixel-to-millimeter conversion factor, which is critical when working with spatially complex surfaces. The proposed solution improves positioning accuracy, reduces fixture requirements, and shortens production preparation time.

**Keyword:** multi-axis positioning, video positioning, image processing, filter pipeline, machine vision.

### Введение

Данная работа является продолжением и развитием наших исследований [1-7] по разработке программного обеспечения для автоматизации технологических процессов на отечественном лазерном оборудовании с числовым программным управлением. Текущее состояние и перспективы развития технического прогресса во многом определяются применением в различных машинах и механизмах деталей, имеющих криволинейные и пространственно сложные поверхности. К таким деталям относятся, например, лопатки газовых и паровых турбин, крыльчатки, импеллеры и другие. Использование таких сложных по форме рабочих поверхностей расширяет функциональные, технологические и эксплуатационные возможности оборудования.

Производство изделий с пространственно криволинейными поверхностями зачастую состоит из нескольких этапов, часть из которых выполняется на разном технологическом оборудовании. Перемещения заготовки непременно влекут за собой задачу многокоординатного позиционирования, для привязки чертежа, по которому осуществляется обработка к реальному положению изделия. Традиционно задачу позиционирования относительно кинематики станка решают через ощупывания датчиком касания или применения специальных высокоточных оснасток для однозначной установки. Однако не всегда ввиду сложной геометрии эти способы позволяют соотнести чертеж с координатами изделия. Для решения задачи сопоставления программного чертежа при многокоординатной лазерной обработки был разработан функционал по определению позиции нулевой точки чертежа на изделии с криволинейной поверхностью по видеоканалу цифровой камеры.

### Общая информация

Функционал реализован на базе уже существующего программного модуля FlexMV, интегрированного в систему управления. Программный код написан на языке C++ с использованием стандарта C++17 и программной платформы Qt 5.15. Взаимодействие с промышленной камерой осуществляется посредством библиотеки производителя – MvCameraControl. Для упрощения реализации алгоритмов обработки изображений и распознавания применяется библиотека OpenCV, обеспечивающая широкие возможности работы с кадрами в реальном времени. Визуализация видеопотока, а также отображение распознанных объектов выполняются с использованием встроенных средств Qt, предоставляющих интерфейс для работы с графической сценой на основе OpenGL. Благодаря применению кроссплатформенных технологий, система может быть эффективно использована как в операционной среде Windows, так и в Linux.

### Конвейер обработки видеопотока

Разнообразие условий эксплуатации лазерной установки, а также широкий спектр обрабатываемых материалов диктуют необходимость предоставления оператору возможности тонкой настройки не только параметров детектирующих алгоритмов, но и выбора типов фильтров и порядка их применения при обработке входного видеопотока. Для реализации данного функционала в программный модуль FlexMV была интегрирована система конвейерной обработки изображений, разработанная на основе методики, впервые представленной в работе [4] и получившей развитие в [5-6].

Рассматриваемая система оперирует двумя основными типами примитивов. Первый тип — детекторы, представляющие собой алгоритмы, предназначенные для распознавания образов на входном изображении. Второй тип — фильтры, выполняющие предварительную обработку кадров с целью улучшения условий для последующего анализа.

FlexMV предоставляет пользователю средства интерактивной настройки как параметров фильтров и детекторов, так и их последовательности. Это реализовано за счёт визуализации промежуточных результатов обработки, что позволяет оценить влияние каждого этапа конвейера на итоговое изображение. Особое значение имеет порядок применения фильтров, поскольку каждый последующий получает изображение, уже модифицированное предыдущим.

Такая архитектура обработки позволяет эффективно компенсировать вариации физических свойств материалов, которые влияют на визуальное отображение изделия на видеопотоке.

### Общая методика поиска нулевой точки чертежа

В решаемой задаче предполагается, что в качестве нулевой точки чертежа выступает граница на изделии, образованная перепадом высот, или центр отверстия (рис. 1).

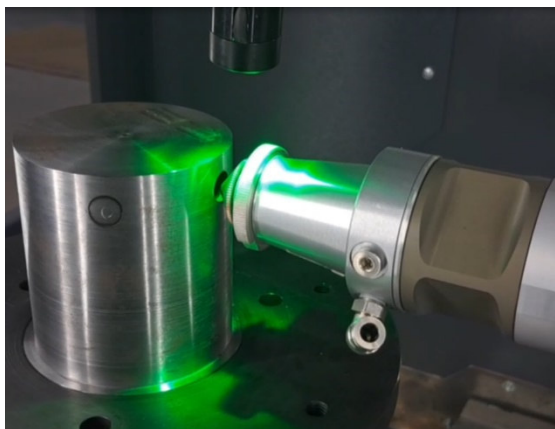


Рис. 1. Тестовый образец – цилиндр с отверстием

Поиск центра состоит из поиска четырёх точек на границы окружности, две точки служат для определения центра по вертикали и две по горизонтали, опустим тривиальный момент поиска центра отрезков и остановимся подробнее на нахождении самой границы. Для осуществления поиска в

предполагаемом месте нахождения границы запускается детектор распознавания (BorderDetector) и начинается движение по заданной оси в заданном направлении, цель этого – первичное (грубое) определение положения края. При обнаружении детектором границы необходимо переместиться таким образом, чтобы сопло лазерной установки оказалось ровно над ней, для этого нужно уточнить координаты распознанного объекта. Для того чтобы сопоставить координаты в пикселях изображения с координатами кинематической системы станка, необходимо рассчитать коэффициент пересчёта (pxToMm). При этом коэффициент pxToMm в задаче многокоординатного позиционирования при обработке деталей с пространственно-криволинейными поверхностями не получится рассчитать один раз заранее ввиду разного сечения, образованного пересечением фокальной плоскости с поверхностью изделия при движении. Таким образом, pxToMm необходимо рассчитывать по месту, для каждой обнаруженной границы. Для расчёта необходимо путём перемещения получить координаты распознанной границы на краях области видимости (рис. 2), в пикселях на изображении и миллиметрах от кинематики. После получения необходимых данных коэффициент рассчитывается как изменение позиции в миллиметрах, делённое на изменение позиции в пикселях:

$$pxToMm = \frac{secondPosMm - firstPosMm}{secondPosPx - firstPosPx}.$$

Таким образом, для успешного сопоставления чертежа с изделием складывается методика, при которой оператору необходимо определить конвейер обработки видеопотока и параметры перемещений поиска края.

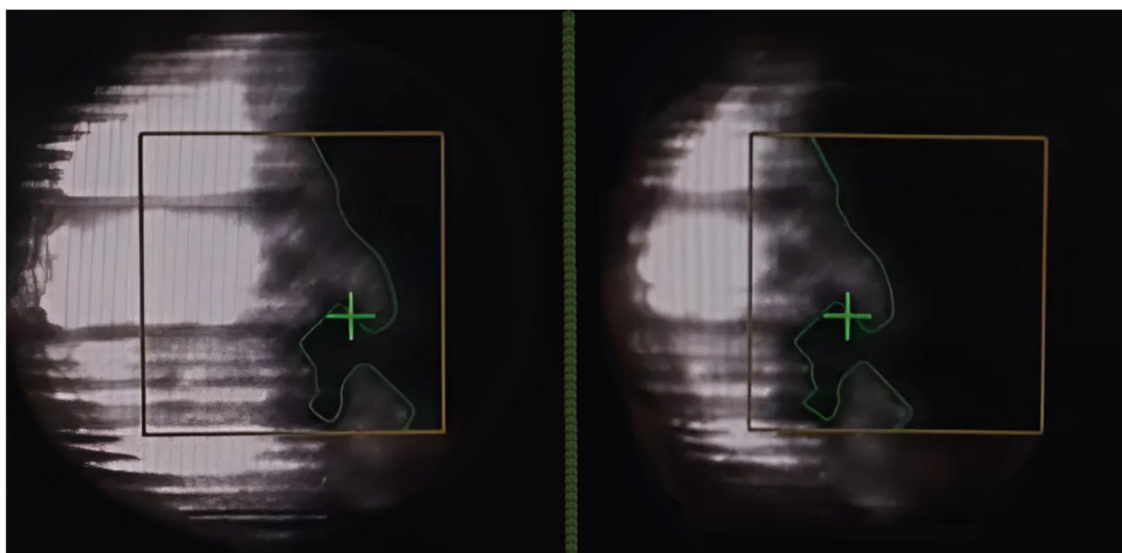


Рис. 2. Распознавание границы на краях области видимости для расчёта pxToMm

### Распознавание края BorderDetector

Представленный алгоритм реализует метод поиска границ объектов на изображении с использованием модифицированного поиска в ширину (BFS) с последующим восстановлением пути до границы. Алгоритм состоит из двух основных этапов:

1. *Построение карты расстояний.* Исходное изображение преобразуется в матрицу расстояний, где каждому пикселю присваивается значение, соответствующее его удалённости от начальной точки. Для этого применяется следующий подход: инициализируется матрица размером с исходное изображение, заполненная значением -1, обозначающим «непосещённые» пиксели. Пиксели, яркость которых ниже заданного порога, маркируются как -2, исключая их из дальнейшего анализа. Из начальной точки запускается волновой алгоритм: на каждом шаге соседние пиксели (окрестность 3×3) помечаются значением, равным текущей глубине поиска. Процесс продолжается до достижения границы изображения или исчерпания допустимого числа итераций.

2. *Восстановление пути до границы.* После обнаружения граничного пикселя выполняется обратный проход по карте расстояний: алгоритм последовательно переходит к соседним пикселям с

меньшим значением глубины поиска, формируя траекторию до исходной точки. Координаты всех промежуточных точек сохраняются в массив, который в дальнейшем используется для визуализации рисунка 1 или расчёта положения границы.

Для повышения устойчивости алгоритма граничные пиксели изображения проверяются в первую очередь, что позволяет быстро обнаруживать контуры объектов, пересекающие рамку кадра. Предварительное исключение тёмных пикселей сокращает область поиска и уменьшает влияние шумов.

Коэффициент пересчёта  $pxToMm$  корректируется для каждой обнаруженной границы, что обеспечивает точность позиционирования даже при изменении масштаба изображения из-за криволинейности поверхности.

### Преимущества и недостатки

Разработанный метод позиционирования на основе обработки изображений с промышленной камеры позволяет существенно повысить точность сопоставления чертежа с реальным положением детали на станке в ситуации, когда применение классических методов не представляется возможным.

Адаптивность к сложным поверхностям – динамический пересчёт коэффициента  $pxToMm$  обеспечивает точность даже при работе с криволинейными и пространственно неоднородными поверхностями. Гибкость настройки обработки изображений – возможность выбора и настройки фильтров и детекторов позволяет адаптировать алгоритм под различные материалы и условия освещения.

Однако метод имеет и ряд ограничений: зависимость от условий съёмки – неравномерное освещение, блики или низкая контрастность могут снижать точность распознавания границ. Необходимость предварительной настройки – оператор должен вручную подбирать параметры фильтрации и детектирования для каждого типа поверхности.

### Заключение

Внедрение алгоритма видеопозиционирования позволило решить задачу сопоставления чертежа с деталью при многокоординатной лазерной обработке в ситуации, где реперами выступают границы, образованные перепадом высот. Применение модифицированного BFS-алгоритма BorderDetector для распознавания границ обеспечивает высокую точность даже при работе с криволинейными поверхностями. Таким образом, предложенное решение не только повышает точность позиционирования, но и сокращает время подготовки производства, что особенно важно при обработке сложных деталей в отечественных авиационной, энергетической и других высокотехнологичных отраслях.

### Список литературы

1. Molotkov A.A., Tretiayakova O.N. On possible approaches to visualizing the process of selective laser melting // Scientific Visualization. 2019. Vol. 11, no. 4. Pp. 1-12.
2. Молотков А.А., Третьякова О.Н. Применение машинного зрения в лазерных технологиях // Труды МАИ: электронный журнал. 2022. № 127.
3. Molotkov A.A., Tretiayakova O.N., Tuzhilin D. N. About development and application of a software platform for machine vision for various laser technologies // Scientific Visualization. 2022. No. 5. Pp. 108-118.
4. Молотков А.А., Третьякова О.Н. Применение методов машинного зрения и математического моделирования для разработки технологий создания электронных приборов // Приборы. 2022. № 4. С. 55-58.
5. Разработка комплекса программ для создания промышленного лазерного технологического оборудования / А.А. Молотков, Д.Л. Сапрыкин, О.Н. Третьякова, Д.Н. Тужилин // Приборы. 2022. № 5. С. 15-22.
6. Молотков А.А., Третьякова О.Н. Отработка технологических режимов при создании SLM технологии // Приборы. 2023. № 8. С. 44-47.
7. Tretiayakova O.N., Tuzhilin D.N., Shamordin A.A. Research and Application of Machine Vision Algorithms for Defect Detection in Additive Technologies // Scientific Visualization. 2025. Vol. 17, no. 1. Pp. 114-121.



## A unified approach for EEG visualization on a mobile device

Ivan A. Labutin

Perm State University, Perm, Russia

**Abstract.** In this paper we propose a concept of applying unified methods for EEG visualization on a mobile devices based on the ontologically driven software generation pipeline. The importance of this work comes from the rapid development of wearable hobbyist-class neurocomputer interfaces and the request for the unified approach for bringing such devices and mobile gadgets together for the purpose of EEG visualization and analysis. This paper reviews existing solutions for visualizing EEG on mobile devices, touches the properties of Aurora OS and its graphics stack in application to visualization. The software solution based on an ontology-driven architecture presented in our previous work is employed to solve a problem of automatic generation of application code. This application is deployed and tested on a real hardware device running Aurora OS 4.0, further validating our approach.

**Keywords:** neurointerface, BCI, ontology engineering, IoT

### Introduction

There is a growing tendency of moving EEG research and experiments from the walls of medical institutions into more personalized and lax environment. Modern mobile devices are more than capable of collecting and processing EEG data. [1] notes that EEG pattern discovery can be beneficial in a wide spectrum of possible applications in natural conditions, such as sports, education, entertainment, rehabilitation, enriching traditional approaches. In last decade a few modern mobile EEG systems have emerged to fight the limitations imposed by a traditional clinical equipment, but their widespread employment still remains restricted due to interoperability issues. [2] identified more than 900 studies that employed numerous hobbyist-grade EEG devices (Emotiv Epoc, NeuroSky, Muse etc.) that shows a growing trend for interest in mobile EEG research. Due to this fact, it may be concluded that the task of developing unified methods for visualizing EEG signals on smartphones that became our everyday companion devices is of utmost importance.

In our previous work [3] we proposed a unified ontology-driven approach to describe all aspects of an interaction with BCI, generate a code of a smart mediator that can be used to query data from an EEG headset and preprocess it. Such a smart mediator may manifest itself as a mobile application, allowing for quick integration of the neurocomputer interface with different experimental settings and devices. This paper examines the relevance of mobile EEG visualization, analyzes existing developments on this topic and features the implementation of a mobile EEG real-time visualization application employing described approach.

### Relevant Studies

In recent years, mobile devices with powerful built-in sensors have opened up new possibilities for neuroscientific experiments outside of rigid medical environment. Some researchers note [4] that modern smartphones put computing power and data collection tools in your pocket, which is why there is a demand for portable EEG recording systems that can function outside of a laboratory. Indeed, commercial mobile and hobbyist-grade EEG devices have appeared (for example, Emotiv EPOC, Neurosky MindWave, Muse and others), combining compact headcaps with wireless data transmission to a smartphone or tablet [5]. For example, [6] describes a development of an 24channel mobile EEG system DreamMachine with a companion Android application EEGDroid,

which provides visualization of the signal in real time. The application offers scaling, enabling or disabling individual channels, choosing the gain and filters, and provides a live monitoring of the recording quality.

A number of studies have been focusing on describing the transfer of classical EEG analysis methods to mobile platforms. For example, in the work [4], the Smartphone Brain Scanner project was introduced. It is a system that combines a wireless 14-channel EEG headcap and a smartphone for the purpose of a 3D visualization of brain activity in real time. The authors used Qt framework, suitable for cross-platform development. Despite the lower signal quality compared to stationary high-density systems, the authors point out that mobile approach opens up tremendous new possibilities for neuroscientific research in non-stationary environment. However, the provided implementation had some disadvantages: the system required a rooted

smartphone and a specially prepared kernel, and the issues of delays and synchronization of stimuli required additional investigation.

In the work [7], a modular approach to creating a BCI applications on Android was proposed. They implemented the SCALA (Signal ProCessing and CLassification on Android) application within the framework of a “multi-app” architecture, where separate modules for stimuli presentation, EEG capture, online processing and classification are different apps that simultaneously launched on the smartphone, and the data is transmitted between them via TCP/UDP network connections. The system turned out to be quite flexible: according to the authors, it can be extended with different EEG devices. Their experiments confirmed that timing is not a problem - a system can achieve near-laboratory accuracy, but there are some issues related to OS task scheduling and device performance that can affect it.

Another approach, training neural network models to control a smartphone interface, is demonstrated in [8]. It presents a prototypical iPhone application that recognizes P300 potentials and blinks from EEG signals, allowing a person to control a phone dial interface.

There are also specialized mobile applications for educational or auxiliary purposes. For example, [9] created iBrain2 and iBrainEEG2 applications for Android and iOS, which can visualize brains’ functional connectivity networks and the arrangement of electrodes according to the 10–20 system on a 3D transparent model of the cortex. These applications were created in C# using the Unity game engine. The authors emphasize that their solutions are the first embedded mobile applications for interactively displaying functional networks of the brain. The main goal of iBrain2/ iBrainEEG2 is educational: they allow doctors and students to study the anatomy and connections in the brain in an interactive mode. However, these programs work with pre-calculated atlas and network data and do not connect to real EEG devices; that is, they do not perform signal analysis by themselves, but only demonstrate prerecorded information.

An important aspect in developing mobile EEG applications is the employed graphical user interface framework. For example, as was already mentioned, Smartphone Brain Scanner used Qt for cross-platform implementation and was able to provide 3D model display at 30~ frames per second and smooth touch control [4]. The subsequent version of this system (SBS2) is also implemented with Qt/C++ and runs on Windows, Linux, Android and other systems [5]. At the same time, some other applications are built using relatively high-level technologies – for example, the aforementioned iBrain/iBrainEEG is based on the Unity engine and C# [9]. The choice of lightweight GUI components in regards to a mobile device seems logical due to the limited resources of the device: interfaces should be as simple as possible and not take away computing power. For example, the EEGDroid application provides only a few basic controls and a graph view to ensure a lightweight and comprehensible visualization of EEG without unnecessary bloat [6].

There’s a relatively distinct area of research that attempts to standardize the description of neuroscientific experiments, their methods, tools and data properties. This embodies projects with the aim to create ontologies for EEG, for example NEMO (Neural ElectroMagnetic Ontologies) [10] or BCI-O (Brain-Computer Interface Ontology) [11]. Both of them have the goal of formalizing the concepts of signals, episodes, stimuli, etc. Ontology-driven software generation could facilitate the creation of specialized brain-computer interfaces (e.g., by automatically linking UI elements to signal semantics), and this was demonstrated in our previous work [12].

To summarize, the field of EEG visualization on mobile devices demonstrates the possibility of transferring many functions of neurocomputer interfaces to portable applications [13, 9]. At the same time, current mobile solutions are still limited in signal quality and versatility: low-density recording and consumer sensors produce more noise [13], and smartphone operating systems have delays that affect the accuracy of timestamps [7]. Most existing applications perform auxiliary, demonstration or educational functions and do not claim to be competitive with the classic laboratory-grade systems [9]. In particular, there are still no signs of widespread use of unified, adaptive systems capable of automatically generating an interface for various types of experiments. The proposed ontology-driven approach is intended to close this gap by providing a single conveyor from the description of the experiment to the final mobile application.

### **Suggested Solution**

The proposed solution is based on the ideas from our previous works [14, 15, 12, 16], which formulated an approach of ontology-driven integration of neural interfaces into an IoT infrastructure. The main essence of

this approach is the introduction of a smart mediator between the neurocomputer interface and other IoT components, with its firmware being automatically generated based on a control ontology compiled from a set of ontological descriptions of the environment, data transformation modules, and the neural interface itself [14, 15]. This ontology describes the characteristics of all data transfer and transformation parameters, which allows for automatic construction of interaction pipeline. We also proposed [12] a mechanism for constructing a pipeline that combines ontological descriptions of smart mediator components through the interface of the SciVi scientific visualization and visual analytics platform [17, 16, 18, 19], where the researcher describes the desired components and their interconnection, and the ontologically controlled generator automatically creates the mediators' firmware [15].

In this work, we expanded the available toolkit of our system with the „EEG Visualize” operator, which can be included in the smart mediator to perform visualization of raw EEG data on the screen of a mobile device in real time. This operator is a separate part of the data processing pipeline, similar to the „Sliding Window” or „Montage Provider” we previously introduced [16]. At the stage of creating the firmware of the smart mediator, this visualization module is included in the code generation process in order to embed into the mediator with the purpose of receiving and displaying „raw” EEG signals.

The key advantages of the proposed concept are that it is based on proven principles of ontology-driven integration [14, 15], thus automating the inclusion of the visualization module into the system without a need for manual code creation, providing unification (the EEG device driver and other modules can be easily swapped without manual modification of the mediators' code [16]) and allowing to employ a mobile device as a platform for EEG data visualization.

As a result of the code generating, the final firmware of the smart mediator contains a built-in EegVisualize module, which draws an EEG graph on the display in real time. This approach is consistent with the ideas of earlier publications, where it was noted that the approach for BCI-IoT interoperability should support the integration of neurocomputer interfaces without deep knowledge of neuroscience or coding skills [14, 15]. With our method, the researcher employs high-level tools to solve his specific research problem („visualize EEG signals”), and the system itself turns the pipeline he built into a firmware code.

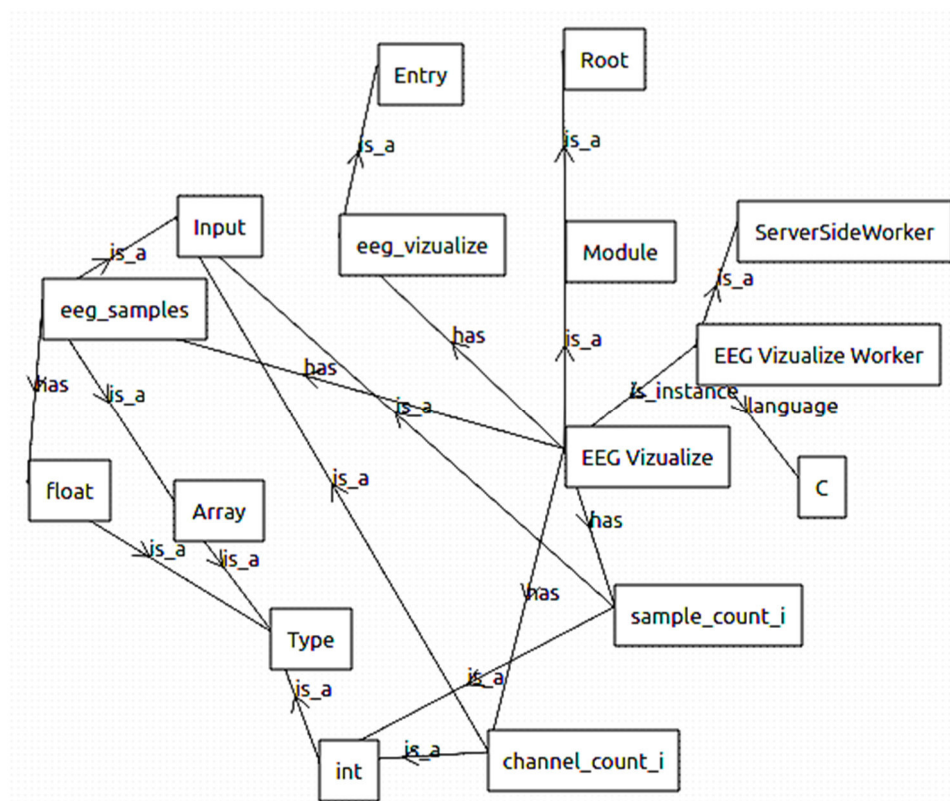


Figure 1. „EEG Visualize” operator ontology

### Testing and Validation

Д To verify the viability of the concept, the new „EEG Visualize” (Fig. 1) operator was incorporated into a data processing pipeline (Fig. 2) employed to run raw EEG visualization on the Aquarius NS220 tablet device running Aurora OS. This device is equipped with a high-quality 10.1-inch IPS screen with a resolution of 1920×1200, which ensures a clear view of fine details of EEG graphs. In the experimental setup, an external neurocomputer interface EBNeuro BE Plus LTM was used, connected to the tablet via Wi-Fi network. Signal acquisition rate was set to 512 Hz, and we used 19-channel configuration (21-electrode headcap). The assembled mediators’ code including the implementation of visualization operator is loaded onto the tablet in form of a mobile application, and during its operation, raw EEG data are received by it in real time via the constructed pipeline.

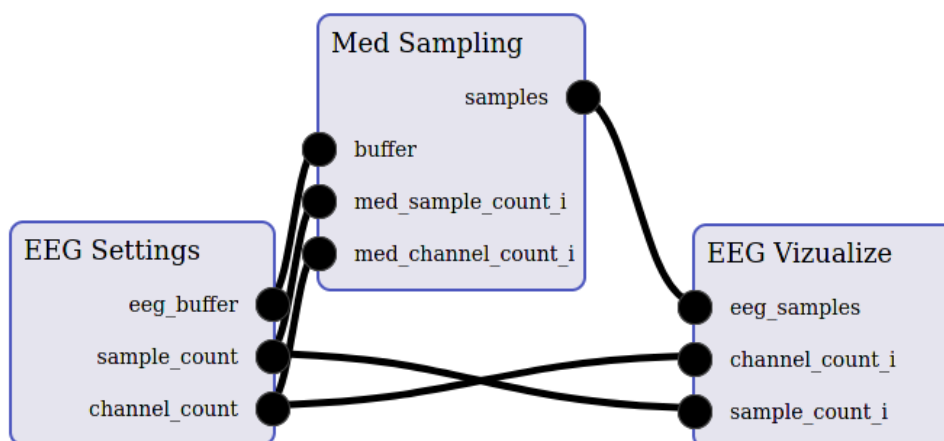


Figure 2. An overview of a pipeline

Aurora OS is a Russian mobile OS based on the Linux kernel and the libhybris compatibility layer, which allows using Android drivers in GNU/Linux environment. It employs the Wayland/Qt stack for graphics and Silica’s own QML components for the user interface. Applications for Aurora OS are developed using Qt Quick (QML/JavaScript) and, if necessary, C++. The user interface is created declaratively in QML, and the application logic can be implemented in C++ or JavaScript. This architecture allows for a nice balance between a hardware-accelerated rendering and smooth operation of the interface on mobile devices and an ease of programming.

The visualization operator introduced in this paper consists of both C++ and QML/JavaScript code. It uses the basic Canvas class for displaying EEG signals as polylines without any smoothing. Screen update is performed at 30 frames per second regardless of the signal acquisition rate. Each channel is drawn using a distinct color from a predefined palette. The testing process included the following steps:

1. Creation of firmware for the smart mediator. The SciVi editor was used to describe the sequence of operators (Fig. 2). The „EEG Visualize” operator is added to the diagram at the end of the flow. After that, the glue source code combining the modules of the smart mediator is generated based on the ontological description of modules and the created pipeline (Fig. 3).
2. Deployment to the device. The compiled mediator application was installed on the Aquarius NS220 tablet. When launched, the application establishes a connection with the neurocomputer interface and begin signal acquisition.
3. Visualization examination. The tablet screen displays the raw EEG signal graph (Fig. 4) in real time.
4. Performance evaluation and estimation. During operation, graph update delays were visually estimated. No noticeable delays were spotted, which confirms the suitability of the solution for real-time applications.

During testing, the application was launched on the real device, receiving an EEG stream from EBNeuro BE Plus LTM EEG amplifier. After the launch, the application initializes the EEG amplifier driver, starts to receive the data and displays signal curves as stripes on the screen, updating them as the new data arrive. The interface is really simplistic and offers no governance over the process, but it may be extended to incorporate

control widgets: channel selection, scaling, stopping the stream, etc. In practice, it was observed that the QML/Qt graphics engine is capable of drawing several dozen EEG channels without delay, providing smooth animation of graphs. It may be beneficial to add a way for the user to interact with widgets via the touch screen: for example, provide the ability to scale the time window. The application correctly responds to background events (minimizes, continuing recording) thanks to the built-in Aurora OS life cycle management mechanisms.

```
float* eeg_buffer;
int channel_count;
int sample_count;
float* samples;

void eeg_settings(float** eeg_buffer, int* channel_count, int* sample_count);
void med_sampling(float* buffer, int med_sample_count_i, int med_channel_count_i,
float** samples);
void eeg_vizualize(float* eeg_samples, int channel_count_i, int sample_count_i);

eeg_settings(&eeg_buffer, &channel_count, &sample_count);
med_sampling(eeg_buffer, sample_count, channel_count, &samples);
eeg_vizualize(samples, channel_count, sample_count);
```

Figure 3. Listing of generated glue code

During the tests, the Aquarius NS220 tablet demonstrated the convenience of the platform: a stable, uninterrupted connection to the neurocomputer interface via Wi-Fi, robust operation of the OS, and excellent high-quality graphical output to the screen. Thus, it was confirmed that the „EEG Visualize” operator functions correctly as part of a smart intermediary, and the selected equipment is a suitable platform for such experiments. Most importantly, our solution [12, 16] proved to be applicable for the task of real-time visualization of EEG data collected in the IoT environment on the client device.

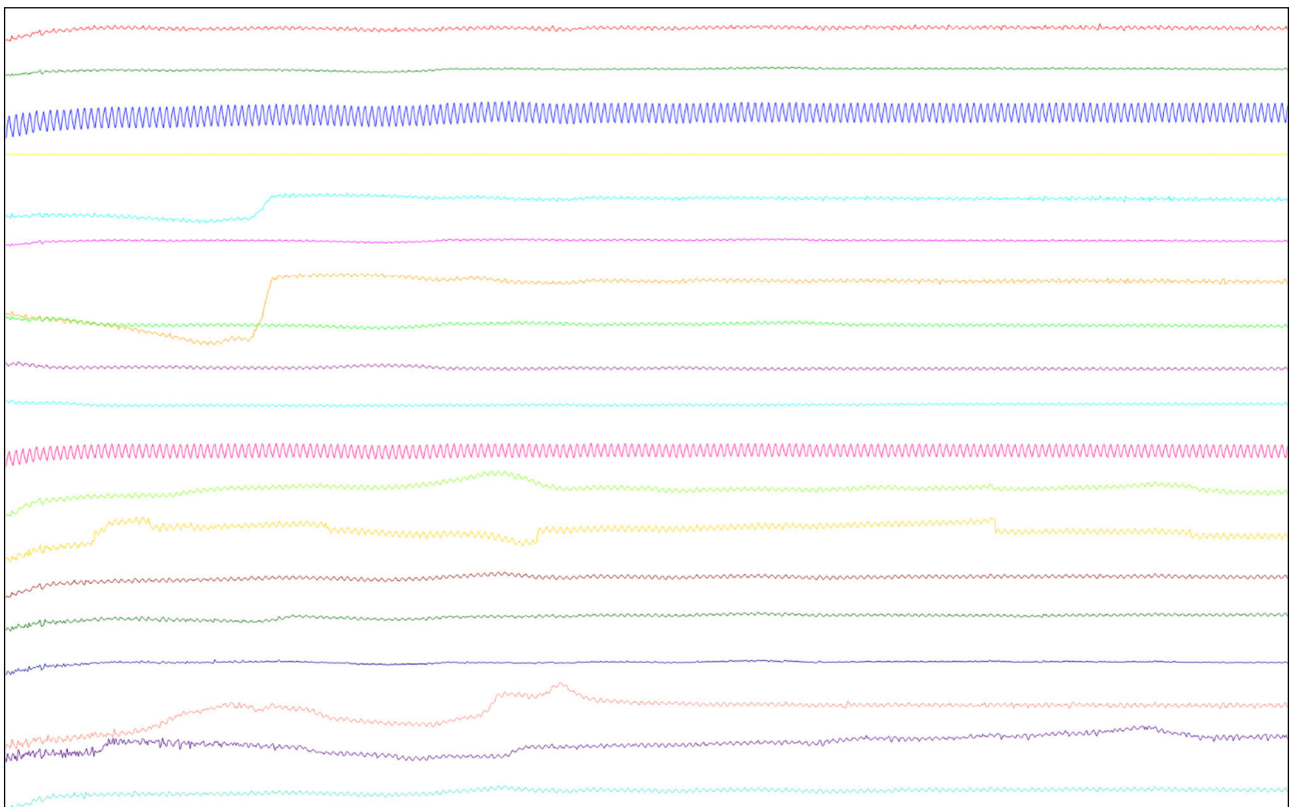


Figure 4. Application screenshot displaying visualization of an EEG signal.  
Different EEG channels are represented with different colors

## Conclusion

The paper reviews modern approaches to visualization of EEG data on mobile devices. A review of existing mobile applications and tools for working with EEG data is conducted. An ontology-driven solution is employed that simplifies the creation of an EEG data visualization application. Aurora OS-powered Aquarius NS220 tablet is used as a platform. Testing showed the applications' performance and responsiveness under normal loads. In the future, it may be feasible to expand the functionality of the pipeline, including support for new types of visualization and EEG preprocessing methods.

The conducted study verified the applicability of an ontology-driven solution for mobile EEG visualization task on a new device, and also allowed us to expand the toolset of our platform with new visualization operator.

## Acknowledgments

The authors express their deep gratitude to the Educational and Scientific Laboratory of Sociocognitive and Computer Linguistics, Faculty of Philology, Perm State University, for the provided equipment and support for the research.

## References

1. K. E. Mathewson, J. P. Kuziek, J. E. M. Scanlon, D. Robles, The moving wave: Applications of the mobile EEG approach to study human attention, *Psychophysiology* 61 (2024). doi:10.1111/psyp.14603.
2. J. Sabio, N. S. Williams, G. M. McArthur, N. A. Badcock, A scoping review on the use of consumer-grade EEG devices for research, *PLOS ONE* 19 (2024) e0291186. doi:10.1371/journal.pone.0291186.
3. K. Ryabinin, S. Chuprina, I. Labutin, Ontology-driven tools for EEG-based neurophysiological research automation, *Scientific Visualization* 13 (2021). doi:10.26583/sv.13.4.08.
4. A. Stopczynski, C. Stahlhut, J. E. Larsen, M. K. Petersen, L. K. Hansen, The smartphone brain scanner: A portable real-time neuroimaging system, *PLoS ONE* 9 (2014) e86733. doi:10.1371/journal.pone.0086733.
5. E. D. McKenzie, A. S. P. Lim, E. C. W. Leung, A. J. Cole, A. D. Lam, A. Eloyan, D. K. Nirola, L. Tshering, R. Thibert, R. Z. Garcia, E. Bui, S. Deki, L. Lee, S. J. Clark, J. M. Cohen, J. Mantia, K. T. Brizzi, T. R. Sorets, S. Wahlster, M. Borzello, A. Stopczynski, S. S. Cash, F. J. Mateen, Validation of a smartphone-based eeg among people with epilepsy: A prospective study, *Scientific Reports* 7 (2017). doi:10.1038/srep45567.
6. P. Samimisabet, L. Krieger, T. Nethar, G. Pipa, Introducing a new mobile electroencephalography system and evaluating its quality in comparison to clinical electroencephalography, *Sensors* 23 (2023) 7440. doi:10.3390/s23177440.
7. S. Blum, S. Debener, R. Emkes, N. Volkening, S. Fudickar, M. G. Bleichner, Eeg recording and online signal processing on android: A multiapp framework for braincomputer interfaces on smartphone, *BioMed Research International* 2017 (2017) 1–12. doi:10.1155/2017/3072870.
8. A. Campbell, T. Choudhury, S. Hu, H. Lu, M. K. Mukerjee, M. Rabbi, R. D. Raizada, Neurophone: brain-mobile phone interface using a wireless eeg headset, in: *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds, SIGCOMM '10, ACM, 2010*, pp. 3–8. doi:10.1145/1851322.1851326.
9. G. M. Rojas, J. A. Fuentes, M. Gálvez, Mobile device applications for the visualization of functional connectivity networks and eeg electrodes: ibrain and ibraineeg, *Frontiers in Neuroinformatics* 10 (2016). doi:10.3389/fninf.2016.00040.
10. G. Frishkoff, P. LePendou, R. Frank, H. Liu, D. Dou, Development of neural electromagnetic ontologies (nemo): Ontology-based tools for representation and integration of event-related brain potentials, *Nature Precedings* (2009). doi:10.1038/npre.2009.3458.1.
11. S. J. R. Méndez, Modeling actuations in bci-o: a context-based integration of sosa and iot-o, in: *Proceedings of the 8th International Conference on the Internet of Things, IOT '18, ACM, 2018*, pp. 1–6. doi:10.1145/3277593.3277914.
12. S. Chuprina, I. Labutin, A high-level adaptation toolkit for unified integration of software systems with neural interfaces, *Scientific Visualization* 16 (2024) 11–24. doi:10.26583/sv.16.4.02.
13. A. Stopczynski, C. Stahlhut, M. K. Petersen, J. E. Larsen, C. F. Jensen, M. G. Ivanova, T. S. Andersen, L. K. Hansen, Smartphones as pocketable labs: Visions for mobile brain imaging and neurofeedback, *International Journal of Psychophysiology* 91 (2014) 54–66. doi:10.1016/j.ijpsycho.2013.08.007.
14. I. A. Labutin, S. I. Chuprina, Ontology-based neurointerface iot integration approach, *Proceedings of the Institute for System Programming of the RAS* 36 (2024) 91–108. doi:10.15514/ispras-2024-36(2)-8.
15. K. Ryabinin, S. Chuprina, I. Labutin, Tackling IoT Interoperability Problems with OntologyDriven Smart Approach, *Springer International Publishing*, 2021, pp. 77–91. doi:10.1007/978-3-030-89477-1\_9.
16. S. I. Chuprina, I. A. Labutin, Scientific visualization tools to improve utilizing neural interface, in: *Proceedings of the 32nd International Conference on Computer Graphics and Vision, GraphiCon-2022, Keldysh Institute of Applied Mathematics*, 2022, pp. 391–402. doi:10.20948/graphicon-2022-391-402.

- 
17. K. Ryabinin, S. Chuprina, Development of ontology-based multiplatform adaptive scientific visualization system, *Journal of Computational Science* 10 (2015) 370–381. URL: <http://www.sciencedirect.com/science/article/pii/S1877750315000307>. doi:<https://doi.org/10.1016/j.jocs.2015.03.003>.
18. S. I. Chuprina, Using data fabric architecture to create personalized visual analytics systems in the field of digital medicine, *Scientific Visualization* 15(5) (2023) 50–63. URL: <https://api.semanticscholar.org/CorpusID:266356696>. doi:10.26583/sv.15.5.05.
19. S. I. Chuprina, Adaptatsiya tekhnologii fabrik dannykh k razrabotke sistem vizual'noi analitiki v oblasti tsifrovoy meditsiny, in: *Trudy 33 Mezhdunarodnoi konferentsii po komp'yuternoi grafike i mashinnomu zreniyu "Grafikon-2023"*, Institut problem upravleniya im. V.A. Trapeznikova RAN, 2023, pp. 405–416. doi:10.20948/graphicon-2023-405-416.



## Машинное обучение для анализа доступности коротковолнового канала связи

Н. А. Конкин, М. В. Вершинин

Поволжский государственный технологический университет, Йошкар-Ола, Россия

**Аннотация.** Применение методов машинного обучения для прогнозирования максимально применимой частоты коротковолновых (КВ) каналов связи позволяет повысить надежность передачи информации в условиях помех, замираний сигналов и многолучевого распространения. Разработана методика оценки доступности КВ-канала на трассе о. Кипр — г. Йошкар-Ола с использованием модели Comb\_Voting\_Huber\_p-10\_mae, обеспечивающей прогноз МПЧ с отклонением менее 0.5 % в среднем за весь период исследования. На основе прогнозных и реальных значений МПЧ рассчитываются параметры сигнала, включая рабочую частоту, скорость передачи и длительность защитного интервала, для адаптации к динамической сигнально-помеховой обстановке. Анализ сезонных и суточных вариаций показал высокую точность модели: среднегодовая прогнозная доступность составила 93.46 % против реальной 93.23 %, с минимальной разницей SNR (0.06 дБ) и BER ( $2.76 \cdot 10^{-4}$ ). Несмотря на редкие отклонения, вызванные ионосферными возмущениями, модель демонстрирует стабильность и пригодность для практического применения в адаптивных системах радиосвязи.

**Ключевые слова:** машинное обучение (МО); максимально применимая частота (МПЧ); оптимальная радиочастота (ОРЧ)

## Machine learning for analyzing the availability of a shortwave communication channel

N. A. Konkin, M. V. Vershinin

Volga State University of Technology, Yoshkar-Ola, Russia

**Abstract.** The application of machine learning methods for predicting the Maximum Usable Frequency (MUF) of shortwave (HF) communication channels enhances the reliability of information transmission under conditions of interference, signal fading, and multipath propagation. A methodology was developed to assess the availability of an HF channel on the Cyprus–Yoshkar-Ola path using the Comb\_Voting\_Huber\_p-10\_mae model, achieving MUF prediction with a deviation of less than 0.5 %. Based on predicted and actual MUF values, signal parameters such as operating frequency, transmission rate, and guard interval duration are calculated to adapt to the dynamic signal-to-noise environment. Analysis of seasonal and diurnal variations demonstrated high model accuracy: the annual predicted channel availability was 93.46 % compared to the actual 93.23 %, with a minimal difference in SNR (0.06 dB) and BER ( $2.76 \cdot 10^{-4}$ ). Despite occasional deviations due to ionospheric disturbances, the model exhibits stability and suitability for practical use in adaptive radio communication systems.

**Keywords:** machine learning (MO), Maximum Usable Frequency (MUF), Optimal Radio Frequency (ORF)

### Введение

Доступность коротковолнового (КВ) радиоканала может снижаться из-за высокой загруженности помехами, затуханий сигналов, многолучевого распространения и других факторов, что ухудшает вероятность передачи сообщения в целом.

При передаче сообщений по КВ-радиоканалам на фиксированной частоте с постоянной скоростью передачи и неизменяющимся помехоустойчивым кодом можно передавать файлы небольших объемов (не более 10 Кбайт) со средней информационной скоростью порядка 100 бит/с и с вероятностью ошибки на бит  $10^{-3} - 10^{-5}$  при использовании *BPSK* модуляции без мультиплексирования.

Для повышения достоверности передачи информации по КВ-каналу необходимо использовать методы адаптации к динамически изменяющейся сигнально-помеховой обстановке. В частности, в зависимости от обстановки следует изменять:

- 1) рабочую частоту;
- 2) техническую скорость передачи за счет изменения кратности модуляции;
- 3) избыточность и параметры помехоустойчивого кода;
- 4) длительность защитного интервала и элементарного сигнала;
- 5) количество и расстановку используемых субчастот;
- 6) адаптивное перераспределение информационного потока и мощности передатчикам между субчастотами сигналов.

Также для обеспечения безошибочного доведения сообщения по КВ-радиоканалу необходимо использовать помехоустойчивые коды с контролем достоверности и осуществлять переспрос неприятых или искаженных информационных блоков.

Применение методов машинного обучения позволит заблаговременно получать значения МПЧ — одной из основных характеристик коротковолновой радиосвязи — и использовать их в адаптивных системах передачи информации, тем самым увеличивая доступность коротковолновых радиоканалов.

### Постановка задачи

*Цель работы:* оценка доступности коротковолнового радиоканала связи между о. Кипр и г. Йошкар-Ола с использованием модели машинного обучения *Comb\_Voting\_Huber\_p-10\_mae* на основе реального и прогнозного параметра МПЧ и их сравнение.

*Задачи:*

1. Разработка методики оценки доступности коротковолнового радиоканала с использованием реальных и прогнозных значений максимально применимой частоты;
2. Программная реализация методики оценки доступности КВ-каналов на основе машинного обучения с использованием языка программирования Python и специализированных библиотек Pandas и Scipy;
3. Проведение эксперимента оценки доступности КВ-каналов связи с использованием реальных и прогнозных значений максимально применимой частоты;
4. Анализ результатов эксперимента доступности КВ-каналов связи путем сравнения доступности радиоканалов при использовании реальных и прогнозных данных;
5. Составление вывода о применимости машинного обучения для устранения или снижения негативного воздействия причин возникновения высокой загруженности.

### Теория

Одной из основных характеристик коротковолновых радиоканалов является максимально применимая частота (МПЧ, *MUF – Maximum Usable Frequency*), поскольку она определяет наивысшую частоту, которая может быть отражена ионосферой, для связи на дальние расстояния. Использование частоты ниже МПЧ гарантирует, что радиоволны будут отражены обратно к Земле. Если же частота выше МПЧ, то волны пройдут сквозь ионосферу и связь не состоится. Выбор наиболее близкой к МПЧ частоты обычно обеспечивает наилучшее распространение, но для стабильности часто используют частоту немного ниже, поскольку она непостоянна и меняется в зависимости от времени суток, года и солнечной активности, поэтому знание текущего значения МПЧ позволяет выбрать наиболее подходящую частоту для надежной связи в конкретный момент. Стабильная частота, вычисляемая как 85 % от МПЧ, называется оптимальной радиочастотой (ОРЧ, *ORF – Optimal Radio Frequency*) и является такой же важной по значимости характеристикой радиоканала.

Анализ ошибки определения доступности КВ-канала связи осуществляется на основе прогнозного МПЧ по сравнению с реальным. Максимально применимая частота является опорной для расчета оптимальной радиочастоты, на которой должен работать передатчик, чтобы обеспечить минимально необходимую напряженность электромагнитного поля для приемника. Вычисление ОРЧ выполнено следующим образом:

$$f_{orf} = 0,85 \times f_{muf}. \quad (1)$$

Рассмотрим доступность канала целевого датасета исследования (о. Кипр — г. Йошкар-Ола). Доступность оценим по метрике *BER* для *BPSK*-модуляции:

$$Avail_{BER} = \frac{(t \mid BER(t) \leq 10^{-5})}{N} \times 100, \quad (2)$$

где  $BER(t) \leq 10^{-5}$  — пороговое значение *BER* для *BPSK* модуляции КВ-канала связи; *t* — временная точка, соответствующая заданному условию; *N* — общее число точек.

Эффективный SNR с учетом кодирования

$$SNR_{eff} = SNR_{corr} + G_c, \quad G_c = 5 \text{ дБ}, \quad (3)$$

с переводом в линейную шкалу

$$SNR_{eff,lin} = 10^{SNR_{corr}/10}. \quad (4)$$

Рассчитаем BER как

$$BER = \frac{1}{2} \operatorname{erfc} \left( \sqrt{SNR_{eff,lin}} \right), \quad (5)$$

где  $\operatorname{erfc}$  — дополнительная функция ошибок.

$SNR_{corr}$  определяется с учетом затухания  $F$ , которое моделируется распределением Рэлея с параметром  $\sigma$ :

$$F \sim \text{Rayleigh}(\sigma), \quad \sigma = 0.7. \quad (6)$$

Таким образом, определяем  $SNR_{corr}$  как

$$SNR_{corr} = SNR - M \times F, \quad (7)$$

где  $M$  — параметра глубины замирования (8 дБ);  $F$  — случайная составляющая (по распределению Рэлея).

Выполним расчет базового SNR:

$$SNR = P_{rx} - P_{noise}, \quad (8)$$

где  $P_{rx}$  — принимаемая мощность,  $P_{noise}$  — мощность шума. Мощность шума определим как

$$P_{noise} = 10 \log_{10} (k \cdot T \cdot B) + NF + 30, \quad (9)$$

где  $k = 1.38 \times 10^{-23} \text{ Дж/К}$  — постоянная Больцмана;  $T = 290 \text{ К}$  — шумовая температура (в Кельвинах);  $B$  — полоса канала (Гц),  $NF$  — шумовая фигура (8 дБ). Принимаемая мощность

$$P_{rx} = 10 \log_{10} (P_{tx}) + G_{tx} + G_{rx} - L_{tot}, \quad (10)$$

где  $P_{tx}$  — мощность передатчика,  $G_{tx}$  — коэффициент усиления передающей антенны,  $G_{rx}$  — коэффициент усиления приемной антенны,  $L_{tot} = L_{fslp} + L_{abs}$  — итоговые потери ( $L_{fslp}$  — потери в свободном пространстве,  $L_{abs}$  — потери на поглощении ионосферой). Потери в свободном пространстве определим как

$$L_{fslp} = 20 \log_{10} (d) + 20 \log_{10} (f_{orf}) + 32.45, \quad (11)$$

где  $d$  — расстояние от передатчика до приемника,  $f_{orf}$  — оптимальная радиочастота. Потери на поглощении ионосферой определим следующим образом:

$$L_{abs} = \alpha \frac{d}{1000} \cdot \frac{1}{f_{orf}^2}, \quad \alpha = 0.02 \frac{\text{дБ}}{\text{МГц} \cdot 1000 \text{ км}}. \quad (12)$$

Расчет направлен на количественную оценку доступности коротковолнового радиоканала связи между двумя точками: о. Кипр — г. Йошкар-Ола. Основным исходным параметром служит максимальная применимая частота (МПЧ,  $MUF$ ), которая получена в результате прогноза и измерений. На ее основе рассчитывается оптимальная рабочая частота ( $ORF$ ) (1) как 85 % от  $MUF$ , что отражает устойчивую область передачи в ионосферном канале. Используя  $ORF$ , проводим оценку потерь сигнала в канале. В расчет включаются потери в свободном пространстве ( $F SPL$ ), которые зависят от расстояния и частоты, а также дополнительное поглощение в D-слое ионосферы (11, 12). Эти потери суммируются, чтобы определить общее затухание. Сигнал также подвержен замираниям, что моделируется с помощью распределения Рэлея, влияющего на итоговый уровень SNR (6). Затем рассчитывается мощность принимаемого сигнала с учетом мощности передатчика, усиления антенн и всех потерь (10). Также оценивается мощность шума приемника, основанная на температуре, ширине полосы и шумовой фигуре (9). Разность между мощностью сигнала и шумом дает SNR в децибелах (8). Полученное значение SNR используется для оценки битовой ошибки (BER) на основе выбранной

модуляции (5). В данной реализации используется *BPSK*, для которой *BER* выражается через дополнительную функцию ошибок и зависит от эффективного значения *SNR* с учетом кодирующего выигрыша. Итоговая доступность канала определяется как доля временных точек, в которых уровень *SNR* превышает минимальный допустимый порог (например, 10 дБ) (2) или  $BER \leq 10^{-5}$ .

В качестве основного метода прогнозирования применён *Voting Regressor*, который объединяет предсказания трёх моделей — *ElasticNet*, *Huber* и *XGBoost* — посредством усреднения. Такой ансамбль позволяет эффективно компенсировать недостатки отдельных моделей: линейные регрессоры (*ElasticNet* и *Huber*) обеспечивают устойчивость к выбросам и надёжное моделирование линейных зависимостей, в то время как *XGBoost*, обладая мощными механизмами градиентного бустинга, идентифицирует сложные нелинейные паттерны и взаимодействия между признаками. Основным мотивом применения *Voting Regressor* — достижение большей точности и стабильности прогноза за счёт синергии моделей с разной природой.

Специальное внимание уделено настройке *Huber Regressor*. Его гиперпараметры были оптимизированы, включая подбор параметра  $\delta$  (*epsilon*), разделяющего квадратичную и линейную зоны функции потерь, что позволяет значительно снизить влияние выбросов. Для предотвращения переобучения была реализована стратегия ранней остановки (*early stopping*), контролируемая по метрике *MAE* на валидационной выборке; обучение прекращается при отсутствии улучшения *MAE* в течение 10 итераций. Такой подход обеспечивает баланс между чувствительностью к малым отклонениям и устойчивостью к сильным выбросам.

Настройка оказалась оправданной выбором только *Huber Regressor* для тонкой настройки: *Voting Regressor* уже обеспечивает ансамблевое усиление, углублённая оптимизация всех моделей была бы избыточной и ресурсоёмкой. При этом ключевой эффект достигается за счёт регуляризации наиболее чувствительного к выбросам компонента — именно *Huber* — без необходимости усложнять весь ансамбль.

Программная реализация методики оценки доступности коротковолновых (КВ) каналов связи выполнена на языке Python с использованием библиотек *Pandas*, *Scipy* и *Plotly*. Разработанный код обеспечивает загрузку и обработку данных о максимально применимой частоте (МПЧ) из файлов Excel и CSV, расчет оптимальной радиочастоты (ОРЧ) как 85 % от МПЧ, а также вычисление ключевых параметров канала: отношения сигнал/шум (*SNR*), вероятности битовой ошибки (*BER*) и доступности канала на основе порогового значения *BER* ( $\leq 10^{-5}$ ). Для учета потерь сигнала реализованы модели свободного пространства и ионосферного поглощения, а также моделирование замираний с использованием распределения Рэлея. Библиотека *Pandas* применяется для управления временными рядами и их агрегации, *Scipy* — для вычислений, включая функцию ошибок (*erfc*) для расчета *BER*. Визуализация результатов осуществляется через тепловые карты и графики с помощью *Plotly*, что позволяет анализировать сезонные и суточные вариации отклонений. Сравнение реальных и прогнозных данных выполнено с использованием метрик *MAE* и *MSE*, обеспечивая количественную оценку точности модели *Comb\_Voting\_Huber\_p-10\_mae*. Программный код включает модульную структуру с функциями для обработки данных, расчета метрик и визуализации, что обеспечивает гибкость и воспроизводимость анализа.

### Результаты экспериментов

Для прогнозирования максимально применимой частоты (МПЧ) и оценки доступности коротковолнового радиоканала связи на трассе о. Кипр — г. Йошкар-Ола (протяженность около 2700 км) были сформирован датасет, охватывающий период с 22 января по 31 декабря 2023 года. Данные собирались с использованием метода наклонного зондирования и программно-определяемых радиоприемников (SDR) с дискретизацией 5 минут, что соответствует 288 замерам в сутки. Общее количество отсчетов составило 99072 (344 дня  $\times$  288 замеров) для полного года, из которых 98784 отсчета приходятся на указанный период.

Описанные расчеты применены для реального хода МПЧ и прогнозного для модели *Comb\_Voting\_Huber\_p-10\_mae*, сводный результат представлен в таблице 1.

Таблица 1. Сводные данные доступности по сезонам

Сезон	Реальный, %	Прогноз, %	Разница (прогн. – реал.)
Весна	92.47	92.03	–0.44
Лето	94.09	94.56	+0.47
Осень	92.45	92.63	+0.18
Зима	93.23	93.46	+0.23

Для каждого из сезонов была построена тепловая карта отклонения (разности) значений доступности канала связи по прогнозу МПЧ (модель Comb\_Voting\_Huber\_p-10\_mae) от доступности по реальному ходу МПЧ за различные сезоны.

Тепловая карта представляет собой графическое представление данных, где отдельные пороговые значения отклонения в матрице отображаются в виде цветов (ось OZ). На оси OX отображается дата, в момент которой были получены значения, на оси OY — время суток. Данный способ визуализации данных использует цвет для представления значения переменной. Использование данного метода особенно полезно для обнаружения закономерностей и тенденций в больших наборах данных, которые сложно увидеть, рассматривая просто массив чисел.

Тепловая карта доступности за весенний период (рис. 1) показывает, что по мере приближения к летнему периоду суточное окно, в котором сконцентрировано примерно 80 % периодов с отклонениями, отличными от 0, расширяется. Так, минимальное окно 01.03 начинается с 07:00, заканчивается в 15:00 и имеет длительность 8 часов. Максимальное окно в 24 часа наблюдается 31.05. Ночное время (с 22:00 до 6:00) характеризуется максимальным отклонением в 42 %, медианное значение отклонения — 0 %, а среднее — 4.2 %. Дневное время характеризуется максимальным отклонением в аналогичные 42 %, медианное значение отклонения — 8.3 %, а среднее — 8.7 %.

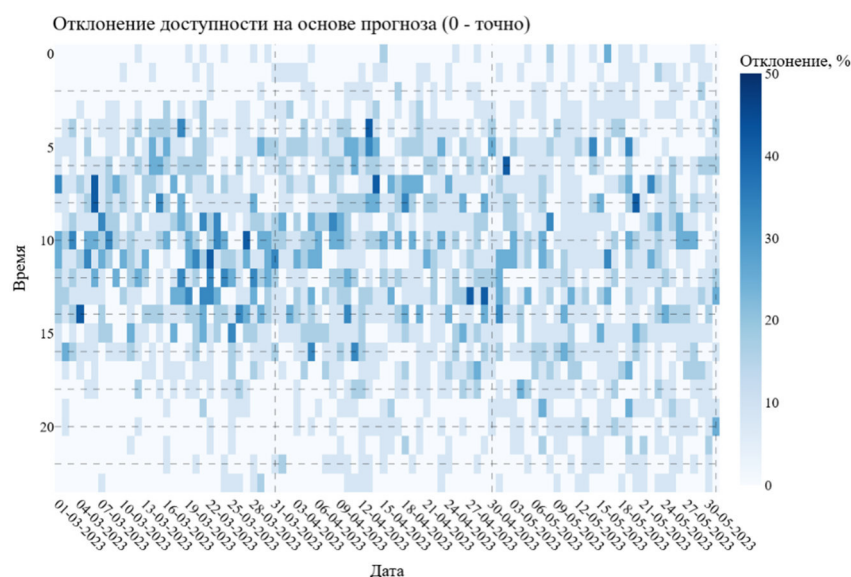


Рис. 1. Тепловая карта отклонения значений доступности канала связи по прогнозу МПЧ-модели Comb\_Voting\_Huber\_p-10\_mae от доступности по реальному ходу МПЧ за весенний период

Тепловая карта доступности за летний период (рис. 2) демонстрирует, что суточное окно отклонений идентично за весь сезон и составляет 24 часа. Ночное время (с 22:00 до 05:00) характеризуется максимальным отклонением в 41.67 %, медианным значением отклонения — 4.17 %, а средним — 5.09 %. Дневное время (с 06:00 до 21:00) характеризуется максимальным отклонением в 41.67 %, медианным значением отклонения — 8.33 %, а средним — 7.51 %.

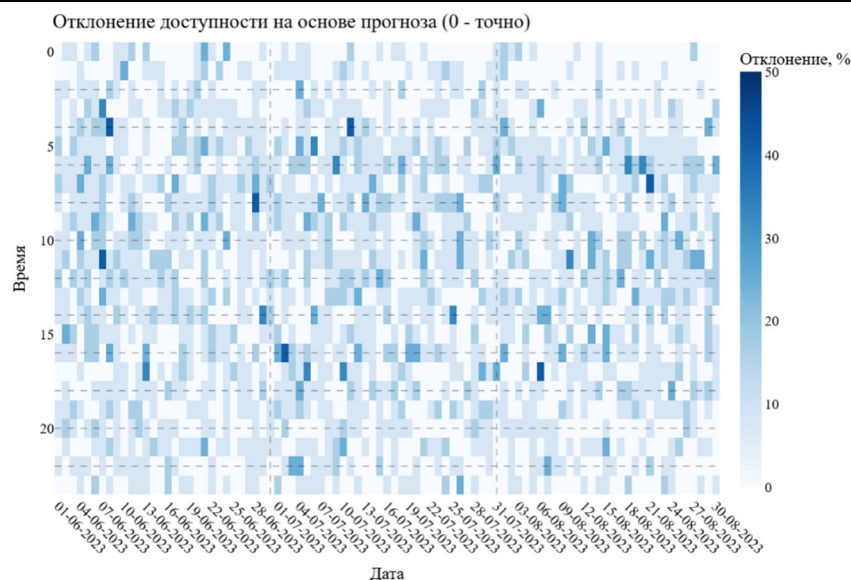


Рис. 2. Тепловая карта отклонения значений доступности канала связи по прогнозу МПЧ-модели Comb\_Voting\_Huber\_p-10\_mae от доступности по реальному ходу МПЧ за летний период

Тепловая карта доступности за осенний период (рис. 3) показывает, что суточное окно отклонений сужается с начала до конца сезона. Так, 01.09 – 05.09 начинается примерно с 01:00, заканчивается примерно в 19:00 и соответственно имеет длительность 18 часов. В конце сезона 28.11 – 30.11 начинается примерно с 05:00, заканчивается примерно в 15:00 и соответственно имеет длительность 10 часов. Ночное время характеризуется максимальным отклонением в 33.33 %, медианным значением отклонения — 0 %, а средним — 3.37 %. Дневное время характеризуется максимальным отклонением в 50 %, медианным значением отклонения — 8.33 %, а средним — 7.81 %.

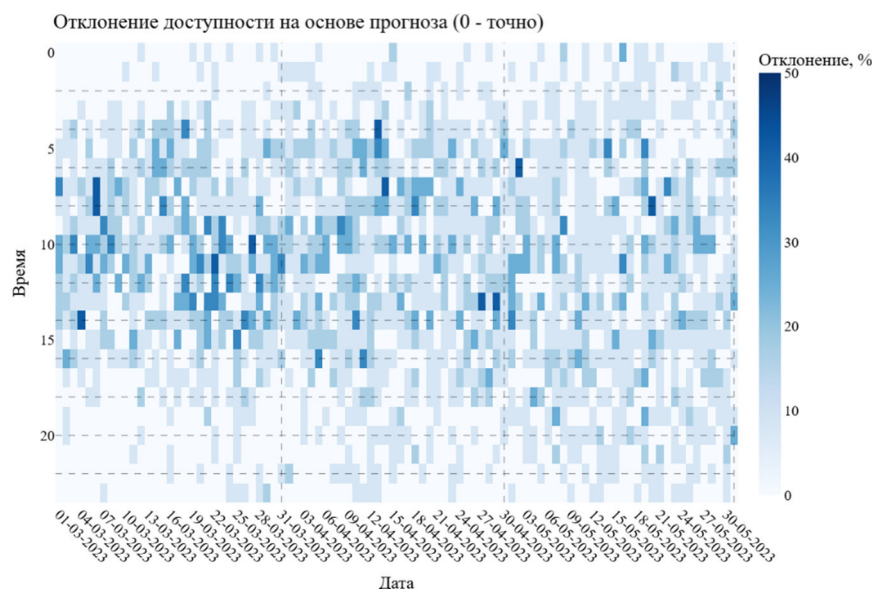


Рис. 3. Тепловая карта отклонения значений доступности канала связи по прогнозу МПЧ-модели Comb\_Voting\_Huber\_p-10\_mae от доступности по реальному ходу МПЧ за осенний период

Тепловая карта доступности за зимний период (рис. 4) демонстрирует самую стабильную ситуацию по сравнению с другими временами года. Суточное окно отклонений продолжает сужение примерно на 4 часа, с 15 в начале сезона до 7 в конце. Ночное время характеризуется максимальным отклонением в 25 %, медианным значением отклонения — 0 %, а средним — 2.43 %. Дневное время характеризуется максимальным отклонением в 50 %, медианным значением отклонения — 4.17 %, а средним — 7.03 %.

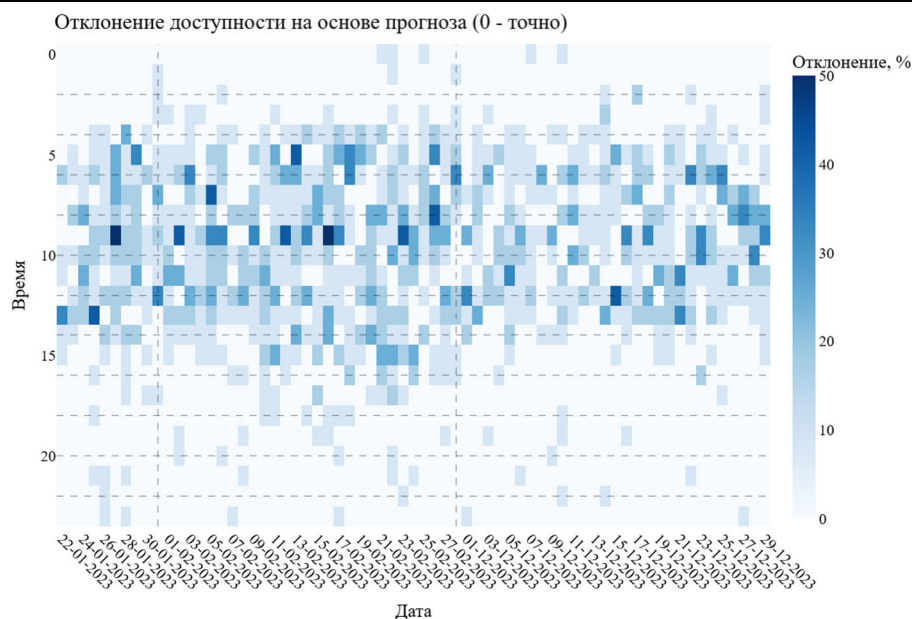


Рис. 4. Тепловая карта отклонения значений доступности канала связи по прогнозу МПЧ-модели Comb\_Voting\_Huber\_p-10\_mae от доступности по реальному ходу МПЧ за зимний период

Тепловая карта доступности за год (рис. 5) демонстрирует ярко выраженную годовую цикличность, где летний период характеризуется высокой изменчивостью с максимальным отклонением в 41.67 %, медианным значением 4.17 % в ночное и 8.33 % в дневное время, а средними значениями 5.09 % и 7.51 % соответственно. Зимний период демонстрирует наибольшую стабильность с максимальным отклонением в 25 % в ночное время и 50 % в дневное, медианным значением 0 % в ночное время и 4.17 % в дневное, а средними значениями 2.43 % и 7.03 % соответственно. Значения в ночные часы демонстрируют большую стабильность, чем значения в дневное время, что подтверждается более низкими медианными и средними значениями отклонений в ночной период по сравнению с дневным во всех сезонах.

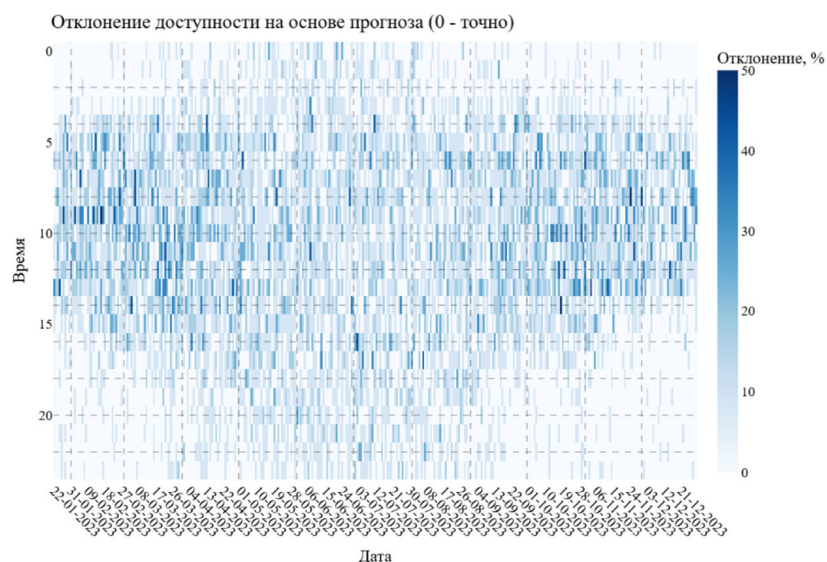


Рис. 5. Тепловая карта отклонения значений доступности канала связи по прогнозу МПЧ-модели Comb\_Voting\_Huber\_p-10\_mae от доступности по реальному ходу МПЧ за весь год

Сезонные различия в результатах обусловлены не только ионосферными процессами, но и внешними факторами, такими как геомагнитные возмущения и уровень солнечной активности. Летний период, с его высокой изменчивостью, отражает интенсивные ионосферные процессы, вызванные



максимальной солнечной активностью, что приводит к увеличению отклонений (максимум 41.67 % в ночное и дневное время) и расширению суточного окна. Зимний период, напротив, характеризуется низкой солнечной активностью, что снижает вариабельность ионосферных параметров и обеспечивает наибольшую стабильность прогнозов, особенно в ночное время (медиана – 0 %, среднее – 2.43 %). Осенний и весенний периоды представляют переходные состояния, где изменения солнечной активности вызывают сужение или расширение окна отклонений. Дневное время во всех сезонах демонстрирует более высокие отклонения (медиана – до 8.33 %, максимум до 50 % осенью и зимой), что связано с усилением ионосферных возмущений в дневные часы из-за солнечного излучения. Ночные часы, напротив, более стабильны, особенно зимой, когда минимальная ионизация снижает вероятность значительных отклонений.

### Обсуждение результатов

По результатам анализа доступности канала по сезонам и за весь год прогнозный набор *MUF* демонстрирует очень близкие показатели к реальным измерениям. Весной (с марта по май) прогноз несколько занижил доступность: если реальный канал был доступен в 92.47 % времени, то прогнозный — в 92.03 %, то есть отставание всего на 0.44 %. Летом ситуация обратная: реальная доступность составила 94.09 %, а прогноз поднялся до 94.56 % – превышение на 0.47 %. Осенью (сентябрь-ноябрь) прогноз оказался выше, чем реальные данные, на 0.18 % (92.63 % против 92.45 %). За год реальная доступность составила 93.23 %, а прогнозная — 93.46 %, что говорит о небольшом положительном смещении прогнозной модели на 0.23 %.

Средний скорректированный *SNR* по реальным данным составил 11.79 дБ, а по прогнозным — 11.73 дБ, разница 0.06 дБ, что находится в пределах погрешности измерений и моделирования. Тем не менее среднеквадратичная ошибка прогноза *SNR* достигает примерно 4.18 дБ, что указывает на редкие, но заметные отклонения на отдельных временных отрезках.

*BER* оказалась на одинаково низком уровне: в среднем  $2.76 \cdot 10^{-4}$  как для реального, так и для прогнозного каналов. Средняя абсолютная ошибка прогноза по *BER* не превышает  $5 \cdot 10^{-4}$ .

Прогноз МПЧ позволяет воспроизвести основные характеристики *HF*-канала — доступность, *SNR* и *BER* — с точностью, достаточной для практического применения. Прогнозная модель чуть завышает доступность летом и осенью, занижает ее весной, но в масштабах года наблюдается стабильность и среднегодовая доступность выше 93%.

### Выводы и заключение

Была проведена оценка доступности коротковолнового радиоканала связи между о. Кипр и г. Йошкар-Ола с использованием модели машинного обучения *Comb\_Voting\_Huber\_p-10\_mae*, основанной на прогнозировании параметра максимально применимой частоты (МПЧ). Для достижения поставленной цели были выполнены следующие задачи: разработана методика оценки доступности коротковолнового радиоканала с использованием реальных и прогнозных значений МПЧ, реализована программная платформа на языке Python с применением библиотек *Pandas* и *Scipy* для обработки данных и анализа, проведен эксперимент по оценке доступности канала связи, а также выполнен детальный анализ полученных результатов с учетом сезонных и суточных вариаций.

Анализ результатов показал, что прогноз МПЧ, выполненный с помощью модели *Comb\_Voting\_Huber\_p-10\_mae*, позволяет воспроизвести ключевые характеристики коротковолнового радиоканала — доступность, отношение сигнал/шум (*SNR*) и вероятность битовой ошибки (*BER*) — с высокой точностью, достаточной для практического применения в адаптивных системах радиосвязи. Для весеннего сезона прогнозная доступность радиоканала составила 92.03 % против реальной — 92.47 % (разница –0.44 %), для летнего сезона — 94.56 % против 94.09 % (разница +0.47 %), для осеннего сезона — 92.63 % против 92.45 % (разница +0.18 %), а для зимнего сезона — 93.46 % против 93.23 % (разница +0.23 %). Среднегодовая прогнозная доступность достигла 93.46 %, что лишь на 0.23 % выше реального значения 93.23 %, это свидетельствует о высокой стабильности модели в годовом масштабе.

Прогнозная модель демонстрирует тенденцию к небольшому завышению доступности в летний и осенний периоды, что связано с повышенной ионосферной активностью и сложностью учета всех факторов, таких как геомагнитные возмущения. Весной модель слегка занижает доступность, что обусловлено переходным характером ионосферных условий в этот период. Средний скорректированный SNR составил 11.73 дБ для прогнозных данных против 11.79 дБ для реальных, с разницей всего 0.06 дБ, что находится в пределах погрешности измерений. Среднеквадратичная ошибка прогноза SNR составила 4.18 дБ, указывая на редкие, но заметные отклонения в отдельных временных интервалах, вызванные случайными ионосферными возмущениями. Вероятность битовой ошибки (BER) для прогнозных и реальных данных оказалась идентичной на уровне  $2.76 \cdot 10^{-4}$ , а средняя абсолютная ошибка прогноза BER не превысила  $5 \cdot 10^{-4}$ , что подтверждает высокую точность модели в воспроизведении характеристик канала.

Тепловые карты отклонений, построенные для каждого сезона, выявили четкую сезонную динамику. Весной суточное окно отклонений расширяется с 8 до 24 часов, что связано с увеличением солнечной активности. Летом окно стабильно составляет 24 часа, отражая высокую изменчивость ионосферных условий. Осенью и зимой наблюдается сужение окна (с 18 до 10 часов осенью и с 15 до 7 часов зимой), что соответствует снижению солнечной активности и повышению стабильности ионосферы. Ночные часы во всех сезонах демонстрируют меньшие отклонения (медиана 0 % в весне, осени и зиме, 4.17 % летом), по сравнению с дневными (медиана до 8.33 %), что объясняется меньшей ионосферной турбулентностью в ночное время. Максимальные отклонения достигают 50 % в дневное время осенью и зимой, что указывает на влияние дневных ионосферных возмущений.

Полученные результаты подтверждают эффективность применения ансамблевой модели Comb\_Voting\_Huber\_p-10\_mae для прогнозирования МПЧ и оценки доступности коротковолнового радиоканала. Высокая точность прогнозов (отклонение менее 0.5 %) и способность модели адаптироваться к сезонным и суточным вариациям ионосферных условий делают ее перспективным инструментом для реализации адаптивных систем радиосвязи, способных минимизировать воздействие помех, замираний и многолучевого распространения. Дальнейшие исследования могут быть направлены на учет дополнительных геофизических параметров, таких как индексы солнечной активности, для дальнейшего повышения точности прогнозов в условиях высокой ионосферной нестабильности.

#### Источник финансирования

Работа выполнена при финансовой поддержке Российского научного фонда, грант (Теория, методы и средства сенсорной диагностики доступности каналов когнитивной КВ радиосвязи с учетом частотно-временной дисперсии, канальных помех и с применением технологий программно-конфигурируемых радиосистем и анализом больших данных, № 24-29-00442).

#### Список литературы

1. Рябова Н. В., Конкин Н. А., Елчанинова А. С. Разработка методики прогнозирования максимальной применимой частоты ионосферных КВ радиолиний на основе синтеза нейросети // Инфокоммуникационные и радиоэлектронные технологии. 2024. Т. 7, № 4. С. 691-705.
2. Елчанинова А. С., Конкин Н. А., Чернов А. А. Исследование нейронной сети типа LSTM с целью прогнозирования параметров трансionoсферного радиоканала // VI Научный форум «Телекоммуникации: теория и технологии» ТТТ-2023 : материалы XXV Международной научно-технической конференции, Казань, 22–24 ноября 2023 года. Казань: Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, 2023. С. 129-131.
3. Рябова Н. В., Конкин Н. А., Вершинин М. В. Восстановление временного хода ПЭС с использованием методов машинного обучения // Распространение радиоволн : сборник докладов I Всероссийской молодежной научной школы-конференции, посвященной памяти Д. С. Лукина, Йошкар-Ола, 11–14 ноября 2024 года. Йошкар-Ола: Поволжский государственный технологический университет, 2024. С. 113-117.
4. Шаптала В. С. Моделирование системы цифровой коротковолновой радиосвязи // Техника средств связи. 2019. № 2(146). С. 154-156.

## Visualization of unrealistic details in synthetic images using feature maps of a neural network

Victor Aleksandrov<sup>1</sup>, Vladimir V. Kniaz<sup>1,2</sup>, Pavel Gomozov<sup>1</sup> and Daniil Novikov<sup>1</sup>

<sup>1</sup>State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia

<sup>2</sup>Moscow Institute of Physics and Technology (MPT), Moscow, Russia

**Abstract.** The generation of synthetic datasets has become a pivotal tool in training neural models for object detection and recognition tasks. These datasets offer an efficient and scalable solution to the challenges of acquiring large volumes of labeled data. However, synthetic images often suffer from a lack of realistic details, which can significantly degrade the performance of models trained on them. The absence of nuanced features such as texture, lighting variations, and complex backgrounds leads to a gap between synthetic and real-world data, hindering model generalization.

In this paper, we propose a novel method for visualizing and identifying unrealistic details in synthetic images by leveraging feature maps extracted from neural networks. Our approach involves analyzing the activations within convolutional layers to highlight discrepancies between synthetic and real images. By comparing these feature maps, we can pinpoint areas where synthetic images diverge from expected patterns observed in real-world data. This visualization technique provides valuable insights into the specific elements that require enhancement, guiding improvements in synthetic image generation processes.

**Keywords:** generative modeling, image synthesis, image enhancement, object detection

### Introduction

The advent of deep learning has revolutionized the field of computer vision, particularly in tasks such as object detection and recognition. A critical component in the success of these models is the availability of large, diverse, and well-annotated datasets. However, collecting and annotating real-world images is a labor-intensive and costly process. To address this challenge, synthetic datasets have emerged as a viable alternative, offering an efficient means to generate vast amounts of labeled data. These datasets are created using computer graphics techniques that simulate various objects and scenes. Despite their advantages, synthetic images often lack the intricate details present in real-world images, which can adversely affect model performance when deployed in practical scenarios.

One of the primary issues with synthetic images is their inability to accurately replicate the complex textures, lighting conditions, and environmental variations found in natural settings. These shortcomings result in a domain gap between synthetic and real data, which neural networks struggle to bridge during training. This gap manifests as a degradation in model accuracy and robustness when applied to real-world tasks. Consequently, there is a pressing need for methods that can enhance the realism of synthetic images or at least identify where these images fall short. Example of visualization of unrealistic details in synthetic images is presented in Figure 1.

In recent years, feature maps within neural networks have been utilized to gain insights into how models perceive input data. These feature maps capture hierarchical representations of images at different levels of abstraction, from simple edges to complex shapes and textures. By examining these activations, researchers can infer which aspects of an image are emphasized by the network during processing. This capability presents an opportunity to analyze synthetic images critically and identify unrealistic details that may contribute to the observed domain gap.

In this paper, we introduce a method for visualizing unrealistic details in synthetic images using feature maps extracted from convolutional neural networks (CNNs). Our approach involves comparing feature map activations between synthetic and real images across various layers of the network. By highlighting discrepancies in these activations, we can pinpoint specific areas where synthetic images deviate from expected patterns seen in real-world data. This visualization technique not only aids in diagnosing deficiencies within synthetic datasets but also provides actionable insights for improving image generation processes. Ultimately, our method aims to facilitate the creation of more realistic synthetic datasets that enhance model performance on object detection and recognition tasks.



Figure 1. Example visualization of unrealistic details in synthetic images

## Related Work

The field of synthetic image generation and its application in training neural networks has garnered significant attention in recent years. Researchers have explored various methodologies to bridge the gap between synthetic and real-world data, aiming to improve model performance and generalization. This section reviews the key advancements in enhancing synthetic images and the use of such images for network training.

### Enhancement of Synthetic Images

Enhancing the realism of synthetic images is crucial for reducing the domain gap between synthetic and real-world datasets. Techniques such as style transfer, domain adaptation, and generative adversarial networks (GANs) have been employed to imbue synthetic images with more naturalistic features. Style transfer methods adjust the texture and color distribution of synthetic images to match those found in real images, while domain adaptation techniques focus on aligning feature distributions across domains. GANs, particularly CycleGANs, have shown promise by learning mappings from synthetic to real image domains without paired examples, thereby enhancing visual fidelity.

### Network Training using Synthetic Images

Training deep neural networks with synthetic data offers a cost-effective alternative to collecting large-scale annotated datasets. Several studies have demonstrated that models pre-trained on synthetic datasets can achieve competitive performance when fine-tuned with a smaller set of real images. Transfer learning techniques are often employed to adapt models trained on synthetic data to real-world tasks. However, challenges remain in ensuring that models do not overfit to unrealistic artifacts present in synthetic images. Recent research has focused on developing robust training strategies that mitigate these issues by incorporating domain adaptation layers or adversarial training schemes.

## Method

In this section, we introduce our proposed method for visualizing unrealistic details in synthetic images using feature maps from convolutional neural networks (CNNs). Our approach leverages a novel framework named FakeSegment, which systematically analyzes discrepancies between feature map activations of synthetic and real images.

### FakeSegment Framework Overview

The FakeSegment framework is designed to identify unrealistic details in synthetic images by comparing their feature map activations against those derived from real images. The framework consists of several stages: preprocessing, feature extraction, activation comparison, and visualization. By focusing on areas where activation patterns diverge significantly, FakeSegment highlights regions within synthetic images that may contribute to the domain gap.

### Network Architecture

Our method employs a SSD object detection architecture known for its effectiveness in capturing hierarchical image representations. The network comprises multiple convolutional layers followed by pooling layers, culminating in fully connected layers for classification tasks. We utilize intermediate feature maps from various convolutional layers to perform our analysis, as these maps provide insights into different levels of abstraction within the image.

**Loss Function**

To optimize our framework's performance, we define a custom loss function that penalizes discrepancies between feature map activations of synthetic and real images. This loss function incorporates both pixel-wise differences and structural similarity metrics to ensure comprehensive evaluation across different abstraction levels.

**Dataset Generation**

For our experiments, we generate a diverse set of synthetic images using state-of-the-art rendering software. These images encompass various object categories and environmental conditions to simulate realistic scenarios encountered in practical applications. Additionally, we curate a corresponding set of real-world images for comparative analysis within our framework.

**Evaluation**

Our evaluation strategy encompasses both qualitative and quantitative assessments to validate the effectiveness of our method in identifying unrealistic details in synthetic images.

**Evaluation Protocol**

We establish a rigorous evaluation protocol that involves comparing feature map activations across multiple CNN architectures trained on both synthetic and real datasets. This protocol ensures that our findings are consistent and generalizable across different network configurations.

**Qualitative Evaluation**

Qualitative evaluation involves visual inspection of highlighted regions within synthetic images where significant activation discrepancies occur. These visualizations provide intuitive insights into specific areas requiring enhancement for improved realism.

**Quantitative Evaluation**

For quantitative assessment, we measure the degree of alignment between feature map activations from synthetic versus real datasets using statistical metrics such as mean squared error (MSE) and structural similarity index (SSIM). These metrics offer objective measures of how closely aligned the representations are across domains.

**Conclusion**

In this paper, we presented a novel approach for visualizing unrealistic details in synthetic images using feature maps from neural networks. Our FakeSegment framework effectively identifies areas where synthetically generated content deviates from realistic patterns observed in natural imagery. By providing actionable insights into these discrepancies, our method facilitates improvements in image generation processes aimed at producing more realistic datasets. Future work will explore extending this approach to other modalities such as video sequences or 3D models while integrating additional machine learning techniques for enhanced analysis capabilities.

**Acknowledgments**

The research was carried out at the expense of a grant from the Russian Science Foundation No. 24-21-00314, <https://rscf.ru/project/24-21-00314/>

**Declaration on Generative AI**

The author(s) have not employed any Generative AI tools.

## Стабилизация взаимодействия или автоволны в параллельных вычислениях?

Д. В. Манаков, П. А. Васёв

Институт математики и механики им. Н.Н. Красовского УрО РАН, Екатеринбург, Россия

**Аннотация.** Цель данной работы состоит в том, чтобы отобразить (нарисовать) стабилизацию взаимодействия в виде волны, максимально абстрагируясь от значений вычисляемой функции (меняя только параметры распараллеливания) с целью выявления общих закономерностей. Для ускорения стабилизации рассматриваются сбалансированные вычисления (более общее требование, чем однотипность агентов). Таким образом, начальное распределение данных (равномерное с учетом гетерогенности вычислительной системы и графа задач) выбирается наиболее близким к «логистической кривой». Балансировка нагрузки была ранее рассмотрена как задача автоматического управления с очевидной стратегией – следующая задача из графа задач назначается тому процессору, у которого математическое ожидание нагрузки процессора меньше. Вообще говоря, это стандартная задача интерполяции (сглаживания). Стоит отметить, что рассматривалась не только задача Дирихле, но и частично задача Пуассона, поскольку на практике была замечена высокая чувствительность к коэффициенту, связанному с обработкой очереди сообщений.

Наибольший интерес в теоретическом и практическом плане представляет задача определения условий, при выполнении которых волна стабилизации взаимодействия будет близка к автоволне. Такая постановка задачи обоснована. Известно, что автоволны (а именно спиральные волны) возникают в химической реакции Белоусова-Жиботинского, а также что подобный процесс в некоторой степени описывается моделью Лотки-Вольтера. Получается, что, распараллеливая решение, например, системы двух линейных уравнений (модель хищник-жертва без учета нелинейной составляющей (взаимодействия)), автоматически получится численное решение стохастической модели Лотки-Вольтера, поскольку в параллельных вычислениях всегда имеются случайная составляющая и взаимодействие, а протяженные объемы соответствуют, например, линейке процессоров. На практике можно рассмотреть не только линейку процессоров, но и решетку процессоров (динамическая 3D-визуализация) и разные схемы взаимодействия, среди которых особый интерес представляют «недетерминированные» (без синхронизации, на бесконечном горизонте) стационарные процессы (предполагается, что схема распараллеливания явно не зависит от времени). Кроме того, в параллельных вычислениях можно ввести псевдослучайные помехи, например, запаздывание некоторого процессора по времени или по координате, в том числе через управление параметрами в режиме онлайн-визуализации.

В теоретическом плане важно, что оптимизационная задача максимизации энтропии (производительности параллельных вычислений) сводится к параметризованной модели белошумного случайного процесса. И вид отображения (волна) строится с учетом этой модели. Также хотелось обсудить задачу параллельной фильтрации данных (частный случай онлайн-визуализации, актуальный при работе с большими данными) в контексте задачи Канторовича (транспорт мер) и принципа динамического программирования.

**Ключевые слова:** параллельные вычисления, стабилизация взаимодействия, автоволны, стохастическая модель Лотки-Вольтера.

## Stabilization of interaction or autowaves in parallel computing?

D. V. Manakov, P. A. Vasev

Krasovsky Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russia

**Abstract.** The purpose of this work is to display the stabilization of interaction in the form of a wave abstracting as much as possible from the values of the calculated function, changing only the parallelization parameters in order to identify general patterns. In order to accelerate stabilization balanced calculations are considered. It is a more general requirement than the uniformity of agents so the initial data distribution uniform taking into account the heterogeneity of the computing system and the task graph is chosen as the closest to the "logistic curve". Previously load balancing was considered as an automatic control problem with an obvious strategy the next task from the task graph is assigned to the processor with the lower mathematical expectation of the processor load. This is a standard interpolation (smoothing) problem. It is worth noting that not only the Dirichlet problem was considered but also partially, the Poisson problem since in practice high sensitivity to the coefficient associated with the processing of the message queue was noted. Of greatest theoretical and practical interest is the problem of determining the conditions under which the interaction stabilization wave will be close to an autowave. This formulation of the problem is justified. It is known that autowaves namely spiral waves arise in the Belousov-Zhibotinsky chemical reaction, and that a similar process is to some extent described by the Lotka-Volterra model. It turns out that by parallelizing the solution, for example of a system of two linear equations, it is the predator-prey model without taking into account the nonlinear component, interaction we will

automatically obtain a numerical solution of the stochastic Lotka-Volterra model since parallel calculations always have a random component and interaction, and extended volumes correspond, for example to a line of processors. In practice, it is possible to consider not only a processor line, but also a processor grid and various interaction schemes, among which “non-deterministic” it is without synchronization, on an infinite horizon stationary processes are of particular interest it is assumed that the parallelization scheme is clearly independent of time. In addition, pseudo-random noise can be introduced into parallel computations, for example a delay of some processor in time or coordinate, including through parameter control in the online visualization mode.

In theoretical terms, it is important that the optimization problem of maximization of entropy, that is parallel computing performance is reduced to a parameterized model of a white-noise random process, and that the view (wave) is constructed taking this model into account.

I would also like to discuss the problem of parallel data filtering that is a special case of online visualization relevant when working with big data in the context of the Kantorovich problem or transport of measures and the principle of dynamic programming.

**Keywords:** parallel computing, interaction stabilization, autowaves, Lotka-Volterra stochastic model.

## Введение

Существует такая точка зрения, что численные методы в контексте параллельных вычислений морально устарели. Возникает вопрос: как правильно распараллеливать программы? Необходимо отталкиваться не только от практики, но и развивать математическую теорию. Известно, что хорошо работают интерполяционные алгоритмы, например, аналоги метода Ньютона, в которых определяется касательная в точке (точках) траектории (касательное расслоение). Таким образом, желательно для распараллеливания сразу выбирать асимптотически устойчивый алгоритм. С другой стороны, любая параллельная программа, если она масштабируется, является энтропийноустойчивой. Можно привести и другой пример. Так, значения решения последовательной программы методом Рунге-Кутты на тех же данных будут отличаться по точности порядка три вторых [1] от параллельного вращения, которое всегда имеет случайную составляющую. Получается так, что формализация параллельных вычислений хорошо ложится на стохастические дифференциальные уравнения (СДУ) [2]. Особый интерес для исследований представляют установившиеся возмущенные процессы (стабилизирующиеся). Так, в данной работе представлены практические и теоретические исследования по «автоволнам» в параллельных вычислениях.

В принципе, идея рассмотрения параллельных вычислений с точки зрения динамических систем не такая уж новая: аналогичный подход используется, например, для клеточных автоматов, кроме того актуальны исследования для мультиагентных систем, в которых требуется однотипность агентов, например, агентов программирования. Стоит подчеркнуть, что в данной работе декларируется более общее требование – балансировка нагрузки для графа задач.

С точки зрения визуализации задача заключается в том, чтобы нарисовать эти волны в зависимости, в первую очередь, от схемы распараллеливания, а также числовых параметров: количества процессов -  $p$  и количества данных -  $N$  для некоторых тестовых примеров задач. Формально такой перебор можно рассматривать как решение задачи методом Монте-Карло с выборкой по значимости, для которой выборка осуществляется на основе экспертного подхода. Причем геометрическое распараллеливание приводит к рассмотрению смещенного распределения (блочный подход, в котором с блоком ассоциируется процессор). Для начала надо разработать адекватный вид отображения, далее вид отображения, сопряженный с математической моделью (образ меры).

## Постановка задачи

Цель данной работы состоит в том, чтобы отобразить (нарисовать) стабилизацию взаимодействия в виде волны, максимально абстрагируясь от значений вычисляемой функции (меняя только параметры распараллеливания) с целью выявления общих закономерностей. Наибольший интерес в теоретическом и практическом плане представляет задача определения условий, при выполнении которых волна стабилизации взаимодействия будет близка к автоволне. Вероятно, что триггерные автоволны могут возникнуть только в случае недетерминированных процессов на решетке процессоров.



## Теория

### 1. Автоволны

Общее определение: автоволны представляют собой самоподдерживающиеся сигналы, которые индуцируют процессы локального высвобождения запасенной в среде энергии, затрачивающейся на выпуск аналогичных процессов в соседних областях. Параллельные (распределенные) вычисления можно интерпретировать как сигнальные графы (частный случай СДУ) [2], тогда передача информации (сообщений), рассматриваемой как кинетическая энергия [3], обеспечивает запуск аналогичных процессов на процессорах, которым предназначено это сообщение. Таким образом, стабилизацию взаимодействия можно рассматривать как триггерную автоволну, что и является основной целью данной работы. При изучении фазовых автоволн необходимо учитывать значения вычисляемой функции, хотя эта задача оставлена на будущее, но в разрабатываемый вид отображения заложена возможность изучения фазовых автоволн.

Еще одно важное определение. Ревербератор – это источник волн: он может существовать в ждущей среде, в которой нет автоколебательных элементов. Кроме того, поскольку он не связан с геометрией среды, то может мигрировать по ней. Начальное значение процессора  $-p_0$ , с которого начинается вычисление, так же как и номер процессора, на котором заданы псевдослучайные помехи, будем рассматривать как ревербератор.

### 2. Вид отображения

С целью масштабируемости изображения используются безразмерные переменные. Областью определения является распределение данных по процессорам (в случае линейки процессоров одна координата, в случае решетки процессоров – две). Областью значений является номер итерации при фиксированном астрономическом времени  $I(t)$  соответственно по процессорам, также рассмотрен другой вариант с целью получения кусочно-непрерывной выпуклой функции: обратное значение к номеру итерации  $-1/I(t)$ , которое можно рассматривать как  $1/T$ . Значение вычисляемой функции отображается цветом, например, градиентная раскраска. Изменение астрономического времени дает волну. Естественно, значение вычисляемой функции на фронте волны имеет максимальную яркость. Стабилизации взаимодействия соответствует логистическая кривая: прямая линия (в случае линейки процессоров) или плоскость. Графики волны строятся в двух вариантах: линейная аппроксимация (рис. 1) и интеграл Стратоновича (столбчатая диаграмма). Примеры представлены в практической части в виде видео.

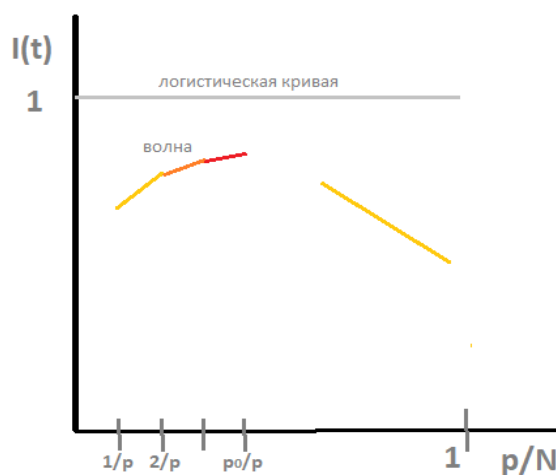


Рис. 1. Волна стабилизации взаимодействия для линейки процессоров

Основополагающей гипотезой является следующая: установившиеся триггерные автоволны могут возникнуть только в случае асинхронных итераций; в отличие от параллельных вычислений последовательные вычисления всегда синхронны, возможны только незначительные периодические колебания, взаимодействие всегда стабилизируется. В целях отладки вида отображения и более

простой асинхронной реализации рассмотрены, в первую очередь, итерации по пространственным координатам. В частности, блочное умножение матриц при количестве процессоров (блоков), стремящихся к бесконечности до этапа синхронизации. Реализация недетерминированных стационарных процессов не является стандартной схемой распараллеливания, но в принципе возможна. Приведем такой пример. Рассмотрим распараллеливание на решетке процессоров разностной схемы типа крест с использованием обещаний, рассматриваемых как условие слабой опережаемости. Предполагается, что значение вычисляемой функции явно не зависит от времени, тогда область определения функции на каждом процессоре сужается по итерациям по времени в виде пирамиды, а при получении ожидаемого сообщения область определения доопределяется. Таким образом, на каждом процессоре возможны более длительные вычисления, чем в стандартной схеме распараллеливания с синхронизацией по временному шагу, и, следовательно, средние значения времени на процессорах могут сильно отличаться. Реализации подобных схем распараллеливания будем называть сильной асинхронностью. В принципе такую сужающуюся область определения можно рассматривать как грубое множество, но основная цель – математическое моделирование; выбрав СДУ, будем придерживаться стохастической семантики [2].

### 3. Стохастическая модель Лотки-Вольтера

Известно, что автоволны, а именно спиральные волны, возникают в химической реакции Белоусова-Жиботинского, а также что подобный процесс в некоторой степени описывается моделью Лотки-Вольтера. Выходит, что, распараллеливая решение, например, системы двух линейных уравнений (модель хищник-жертва без учета нелинейной составляющей (взаимодействия)), автоматически получим численное решение стохастической модели Лотки-Вольтера [1], поскольку в параллельных вычислениях всегда имеются случайная составляющая и взаимодействие, а протяженные объемы соответствуют, например, линейке процессоров. На практике можно рассмотреть не только линейку процессоров, но и решетку процессоров (динамическая 3D-визуализация), и разные схемы взаимодействия, среди которых особый интерес представляют «недетерминированные» (без синхронизации, на бесконечном горизонте) стационарные процессы.

Стохастическая модель Лотки-Вольтера [1] имеет следующий вид:

$$\frac{d}{dt}x^{(i)}(t) = x^{(i)}(t) \left( \gamma_i + \sum_{j=1}^p \nu_{ij} x^{(j)}(t) \right) + \sigma_i x^{(i)} \xi^{(i)}(t),$$

где  $x^{(i)}$  – численность  $i$ -й популяции или в  $i$ -м объеме (в параллельных вычислениях на  $i$ -м процессоре); взаимодействие  $x^{(i)}x^{(j)}$  – передача сообщений между соответствующими процессорами; коэффициенты  $\nu_{ij}$  определяют матрицу обмена сообщениями (матрицу смежности произвольного ранга, поскольку количество процессоров  $p$  является переменным параметром). Оправданность стохастических моделей заключается в том, что, например, для химических реакций, протекающих в протяженных объемах, значительную роль играет диффузия вещества. А в последовательном решении системы двух линейных уравнений (или уравнений первого порядка (модель хищник-жертва)) возможны только периодические колебания, в параллельных вычислениях существенную роль играет явление диффузии.

### 4. Параметризованная модель белошумного случайного процесса

Система дифференциальных уравнений для параметризованной модели белошумного случайного процесса имеет вид [1]

$$\dot{x}_t = a(x_t, u(t), t) + \sum (x_t, t) \frac{1}{\mu} \xi_{\frac{1}{\mu^2}}, \quad x_0 = x(0).$$

Малый параметр  $\mu$  стремится к нулю. Стоит отметить, что случайная составляющая имеет вид  $\xi_{\frac{1}{\mu^2}}$ ,

таким образом, дисперсия установившегося возмущенного (стабилизирующегося) процесса стремится к нулю, в то время как дисперсия «наблюдаемого» процесса (без квадрата) не ограничена. Наблюдаемый процесс, а именно задача визуализации цифровой модели поверхности (ЦМП), которая

является стационарной линейной, был рассмотрен и визуализирован в работе [2]. В настоящей статье рассматривается стабилизация взаимодействия в параллельных вычислениях.

Вышеприведённая система дифференциальных уравнений была получена в контексте СДУ Р. Л. Стратоновичем. В работе [1] установившиеся возмущенные процессы сначала рассматриваются для фундаментальных последовательностей (сходимость по вероятностной мере) и дробно-рациональных функций (в качестве случайной составляющей можно рассматривать остаток любого сходящегося ряда, далее приведены типовые примеры). Влиянием малого параметра активно занимался Л. С. Понтрягин, этой тематике посвящена его первая работа с соавторами [4], в которой рассмотрены стационарные процессы с возмущениями (толчками), аналогично в среднеквадратическом смысле.

С точки зрения вариационного исчисления не имеет значения, по какому именно малому параметру дифференцировать, довольно часто в качестве малого параметра рассматривается  $1/T$ , например, в теореме о магистралах. В параллельных вычислениях это функция энтропии (малый параметр)  $p/N(p)$ , которая является гармонической (свойство среднего) и аналитической, что важно в теоретическом плане.

В случае геометрического распараллеливания вероятностное пространство определено следующим образом:  $N$  – это мощность, множество  $\Omega$ ,  $p$  – это количество непересекающихся под- $\sigma$ -алгебр  $\mathcal{F}_p$  (обозначение  $p$  является аббревиатурой от слова процессор (процесс), обычно в качестве непересекающихся под- $\sigma$ -алгебр рассматривается декартово произведение подпространств, то есть это обозначение выбрано по аналогии с  $L_p$ ), определена (редуцированная) геометрическая вероятность через отношение объемов (разной размерности).

Числовой параметр  $p$  аналогичен времени в смысле вариационного исчисления:  $\dot{x}_p = dl/dp$ , где  $l$  – длина сообщения. Важно подчеркнуть, что длина сообщения считается не в количестве байт, а в количестве буферов, так высокая латентность коротких сообщений обусловлена слишком большими помехами.

В качестве случайной составляющей можно рассматривать остаток любого сходящегося ряда. (Налицо явное отставание теории от практики, в которой преобладает применение всплесков для двумерных сигналов (изображений) и методов Монте-Карло, а не, например, унифицированного разложения Тейлора или преобразования Фурье). Если процесс по отношению к системе не является белым шумом, то рекомендуется сузить ширину пропускного канала (доверительный интервал), что можно сделать двумя способами: увеличить степень аппроксимации или увеличить количество данных. С учетом параметра  $N$  (разбиение по времени) в СДУ были ведены смешанные интегралы Стратоновича. Важно, что при доказательстве сильной сходимости численного решения СДУ интеграл, зависящий от  $N$ , должен быть мартингалом (дифференцируемость в среднеквадратическом смысле). (В том числе, можно рассматривать не стандартные кратные (по пространственным переменным) смешанные интегралы Стратоновича, а следовательно, не стандартные матрицы произвольного ранга –  $N = 2N$  (для уравнений в частных производных по каждой размерности), а не  $N = N + 1$ , что связано с практической ориентацией на  $k$ -деревья и всплески.) В работе [2] предложена формула Ито для фильтра сечения плоскостью:

$$f'(B_N) = (\vec{n}, \nabla \prod x_i)|_{x_i=N} + N.$$

Соответственно для интеграла Стратоновича  $ds = 2N$ .

**Теорема.** Пусть  $f(N, \vec{x})$  – возмущенный случайный процесс (Ито) по количеству данных. Если процесс Ито является мартингалом, то для любого фильтра  $\varphi \in \mathcal{C}^2$ , дважды непрерывно дифференцируемого,  $\varphi f(N, \vec{x})$  является мартингалом.

Доказательство этой теоремы стандартное, если фильтр сечения плоскостью рассматривать как производящий оператор [5]. На рисунке 2 представлена геометрическая интерпретация формулы Ито для фильтра сечения плоскостью (вид спереди). Малый параметр  $\Theta$  отвечает за угол наклона секущей плоскости (фильтра) относительно регулярной сетки  $0 < \Theta < \Delta z$ . Если посчитать количество данных на секущей плоскости (проекции ближайших точек), то оно будет соответствовать вышеприведённой формуле. Если рассмотреть другую окрестность, через середину которой проходит секущая плоскость, то соответственно получится интеграл Стратоновича.



Рис. 2. Геометрическая интерпретация формулы Ито для фильтра сечения плоскостью

Важно, что суперпозиция функций  $\varphi f(N, \vec{x})$ , реализованная как пересечение и проекция, при дифференцировании ведет к интегрированию по частям (свертке). Возможно, в качестве  $\varphi$  (функции многих переменных) можно рассматривать абсолютно непрерывную функцию в сочетании с условием Лифшица, тогда в контексте СДУ необходимо применение формулы Гирсанова.

Задача параллельной фильтрации данных возникает в том случае, когда данные, рассчитанные на вычислителе при фиксированном значении времени, невозможно передать в полном объеме на рабочую станцию (например, они не входят в память видеокарты) и сжатие данных не решает поставленную проблему или ведет к значительной потере точности. Таким образом, невозможно построить отображение вида  $R^3 \times \{t > 0\} \rightarrow R^3$ . Наибольший интерес представляет однопараметрическая фильтрация данных, например, сечения плоскостью (сферой, цилиндрами) или изоповерхности, сокращающая объем данных в размерность (принцип динамического программирования). В итоге надо предложить  $\varepsilon$ -проксимальную стратегию (расставить контрольные точки в зависимости от скорости анализа информации на клиенте). Задача параллельной фильтрации данных напоминает задачу транспорта мер (перемещение данных из одной области в другую с минимальной работой). На вычислителе фильтрация данных реализована как реактивное вычисление (выборка), что вселяет надежду на то, что затраты (помехи) на ее реализацию будут незначительны по отношению к основным вычислениям. Для наглядности будем рассматривать трехмерную регулярную сетку со скалярным значением функции в узлах. С целью сокращения длины сообщения передаются только значения функции, а координаты заново пересчитываются. Визуализация осуществляется сопоставимо с режимом реального времени, что ведет к ограничению на длину передаваемого сообщения. Если длина сообщения все равно большая, можно дополнительно использовать сжатие, в том числе разложение в ряд. Изменение параметра ведет к отображению другой плоскости. Вообще, не хотелось бы использовать полный перебор по параметру, в этом случае напрашивается применение методов Монте-Карло, в зависимости от задачи возможно решение некоторой оптимизационной задачи, например, отображение поверхностей экстремальных кривых.

Итак, с учетом итераций по времени надо построить отображение, абсолютно-непрерывную функцию (фильтр)  $\phi: R^3 \times \{t > 0\} \rightarrow R^2$ . Если бы эта функция имела вид  $\phi(x, t): R^3 \times \{t > 0\} \rightarrow R^1$ , то это была бы стохастическая функция Ляпунова [1], дважды непрерывно дифференцируемая по  $x$  и один раз по  $t$ , применяемая для решения задач стохастического оптимального управления. Возможно, решение поставленной задачи связано с рассмотрением задачи транспорта мер, не обязательно в контексте СДУ, другой вариант – функциональная сходимости для методов Монте-Карло.

В заключение теоретической части стоит привести простую теорему [4] в плане доказательства (дерево выбора ограничено сверху  $k$ -деревом), но важную в применении, в теории информации получившую название – оптимальное кодирование дешифруемых кодов Крафта.

Теорема 2.3. Можно указать такой способ кодирования (передачи) равномерных независимых сообщений, что

$$l_{cp} < \frac{H_{\xi}}{\ln D} + 1,$$

где  $H_{\xi}$  – средняя энтропия (количество информации).

В нашей интерпретации  $l_{cp}$  – средняя длина пути между узлами вычислителя,  $D$  – количество каналов процессора (в оригинале количество букв в алфавите). Приведем классификацию каналов по значению в знаменателе. В случае обмена сообщениями через общую шину (гауссовские каналы) можно считать, что  $D = \infty$ , следовательно, средняя длина пути между узлами вычислителя равна единице, что соответствует экспертному подходу. Причем гауссовские каналы с условием равновероятности сообщений дадут недетерминированные процессы [2].  $D = 2$  – двоичные каналы.  $D = 1$  – узкое место – вырожденный случай, необходимо рассматривать задачу стохастического оптимального управления.

Рассмотрение задачи Канторовича (транспорт мер) и принципа динамического программирования актуально не только для задачи фильтрации данных, но и для изучения автоволн на решетке процессоров. В принципе нейронные сети можно рассматривать как конвейер фильтров, где размерности подпространств соответствует количество персептронов в слое. Для обучения нейронных сетей проблематично предложить равномерное распределение по данным, поэтому актуально рассмотрение устойчивости для методов Монте-Карло. Аналогично методы эластичных сетей (карт [6]) можно рассматривать как задачу Канторовича в сочетании с принципом динамического программирования, поскольку они имеют квадратичную часть (регуляризация Тихонова), линейную часть (метод лассо) и аддитивные помехи, связанные с понижением размерности.

### Результаты экспериментов

Практическая часть находится в стадии реализации. Будем надеяться, что удастся получить визуализацию волн для блочного умножения матриц при количестве процессоров (блоков), стремящихся к бесконечности до этапа синхронизации. Поскольку рассматриваются не типовые схемы распараллеливания (в частности, применение обещаний), вероятны трудности в плане программирования.

### Обсуждение результатов

Известно, что автоволны, а именно спиральные волны, возникают в химической реакции Белоусова-Жиботинского. Подобный процесс в некоторой степени описывается моделью Лотки-Вольтера. Эта модель при рассмотрении параллельных вычислений (в частности, триггерных автоволн) является не полной, поскольку представляет интерес отображение данных на решетку процессоров, следовательно, надо рассматривать уравнение в частных производных (возможно, транспорт мер), поскольку задача имеет физический смысл, должен выполняться принцип наименьшего действия (полный дифференциал по количеству процессоров должен быть равен нулю). Возможно рассмотрение и других подходов, например уравнения неразрывности, адекватность которых неочевидна. Поэтому авторы намерены отталкиваться от распараллеливания конкретных задач, что ведет к рассмотрению функциональных уравнений. Хотелось бы начать обсуждение с практических результатов.

### Заключение

Эта работа является исследовательской, и делать какие-то выводы преждевременно. Анализ результатов показывает, что при распараллеливании намеренно допускается логическая ошибка: недетерминированные процессы сводятся к детерминированным. Именно моделирование и изучение недетерминированных процессов представляет интерес. Хотелось бы получить ответы на ряд вопросов:

1. Бывают ли стабилизирующие недетерминированные процессы?
2. Как недетерминизм повлияет на производительность?
3. Верна ли гипотеза, что триггерные автоволны могут возникнуть только в случае недетерминированных процессов?

Актуальным направлением является создание виртуального испытательного стенда (цифровых двойников). Вряд ли возможно свести, например, рой роботов к детерминированным процессам, поскольку чем больше количество роботов, тем больше время синхронизации, в результате во время синхронизации все роботы должны стоять на месте или смещаться с малым приращением.

### Список литературы

1. Кузнецов Д.Ф. Численное моделирование стохастических дифференциальных уравнений и стохастических интегралов. Санкт-Петербург: Наука, 1999. 458 с.
2. Manakov D.V., Vasev P.A. Stochastic semantics of big data (parallel computing and visualization) // Scientific visualization. 2024. Quarter 4. Vol. 16, no. 5. P. 120-150. DOI: 10.26583/sv.16.5.09
3. Стратонович Р.Л. Теория информации. М.: Сов. радио, 1975. 424 с.
4. Андронов А.А., Витт А.А., Понтрягин Л.С. О статистическом рассмотрении динамических систем // ЖЭТФ. 1933. 1.3, 3.
5. Oksendal B. Stochastic differential equations: an introduction with applications. Berlin: Springer, 2003. ISBN 3-540-04758-1.
6. Бондарев А.Е., Галактионов В.А., Шапиро Л.З. Визуальный анализ и обработка многомерных данных // GraphiCon 2017: 27-я Международная конференция по компьютерной графике и машинному зрению: труды конференции. Пермь: ПГНИУ, 2017. С. 103-107.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-524-534

## О возможности визуализации процессов подготовки катастрофических событий

Л. П. Коган<sup>1,2</sup>, А. Е. Вольвач<sup>2</sup>

<sup>1</sup>Нижегородский государственный архитектурно-строительный университет,  
Нижний Новгород, Россия

<sup>2</sup>Отдел радиоастрономии и геодинамики,  
Крымская астрофизическая обсерватория РАН, Кацивели, Крым, Россия

**Аннотация.** В статье рассматриваются изменения свойств статистики измерений физических и биологических процессов, возникающие в сложных системах перед их явлениями катастрофического типа. К изучаемым феноменам, рассматриваемым в рамках модификаций единого математического алгоритма, относятся вспышки на Солнце, сильные землетрясения и заболевания организма человека. Полученные результаты позволяют визуализировать предвестники катастрофических (или существенно негативных) событий для объектов самой разной природы.

**Ключевые слова:** прогноз катастроф, вспышки на Солнце, сильные землетрясения, опасные заболевания.

## On the possibility of visualizing the processes of preparing catastrophic events

Lev Kogan<sup>1,2</sup>, Alexandr Volvach<sup>2</sup>

<sup>1</sup>Nizhny Novgorod State University of Architecture and Civil Engineering, Nizhny Novgorod, Russia

<sup>2</sup>Radio Astronomy and Geodynamics Department of Crimean Astrophysical Observatory, Katsively, RT-22 Crimea

**Abstract.** The article examines changes in the statistical properties of physical and biological measurements that arise in complex systems prior to catastrophic events. The phenomena studied - solar flares, major earthquakes, and human diseases - are analyzed within the framework of modifications to a unified mathematical algorithm. The results obtained make it possible to visualize precursors of catastrophic (or significantly adverse) events across a wide range of systems of different nature.

**Keywords:** disaster forecast, solar flares, strong earthquakes, dangerous diseases.

### Введение

Одной из наиболее существенных проблем, стоящих перед человечеством, следует считать поиск способов прогнозирования и предупреждения разномасштабных катастроф и неблагоприятных событий, связанных с различными природными и биологическими феноменами. Ее решение рассматривается как важнейший элемент при обеспечении безопасности, защиты инфраструктуры и здоровья населения. Вспышки на Солнце, такие как солнечные штормы и корональные выбросы, являются одними из самых опасных космических явлений [1-4]. Они способны не только вызывать сбои в работе спутников, электросетей, систем связи и навигации, но и приводить к массовому ухудшению состояния здоровья. Их своевременное предсказание должно позволить свести к минимуму негативные последствия для современных технологий и инфраструктуры, равно как и обеспечить подготовленность к вероятным аварийным ситуациям. Вместе с тем на сегодня не существует вполне надежных методик прогнозирования данных природных феноменов.

Сильные землетрясения — это одни из наиболее разрушительных природных катастроф, которые могут привести к гибели очень большого числа людей, обрушению зданий и уничтожению инфраструктуры, а также вызвать возникновение цунами и оползней. При этом основной проблемой надежного прогнозирования сейсмических событий высокой магнитуды является отсутствие повторяемости тех нескольких сотен феноменов (усиление концентрации радона, возникновение линейных облачных структур, изменение поведения животных и т.д.), которые предшествуют таким природным катастрофам [5-9].

В области медицины и здравоохранения очевидны важность прогнозирования заболеваний человека, таких как обострение хронических состояний, эпидемии инфекционных болезней или развитие новых заболеваний. Использование информационных технологий позволяет выявлять риски и предсказывать предстоящее течение заболеваний. Это значительно повышает шансы на выздоровление и повышает эффективность работы системы здравоохранения [10, 11].

### Постановка задачи

В этой статье рассматривается возможность описания процессов подготовки приближающихся катастрофических событий самого разного вида. В частности, сюда следует отнести события астрофизической (сильные вспышки на Солнце и связанные с ними магнитные бури) [12–13], геофизической (землетрясения) [14–15] и медицинской [16–18] природы. При этом ставится задача создания единого подхода, в рамках которого становится возможным сформулировать общий инженерно-геометрический язык для исследования предкатастрофических изменений состояния систем самого разного типа. Соответствующий математический аппарат основан на доказанном, в частности в [12–18], феномене одинаковых законов изменения свойств статистики физических и биофизических процессов, связанных с функционированием соответствующих объектов, в период подготовки их предстоящего катастрофического преобразования.

### Теория

Далее будем считать, что любые измеряемые величины  $x(t)$ , связанные с деятельностью изучаемого объекта физической или биологической природы, могут быть представлены в виде следующей суммы:

$$x(t) = x_1(t) + x_2(t). \quad (1)$$

В формуле (1) случайная величина (СВ)  $x_1(t)$  – фоновый шум, то есть совокупность обычно наблюдаемых случайных процессов, создаваемых как исследуемой системой, так и возможным шумом самой измерительной аппаратуры. Тогда как случайное слагаемое  $x_2(t)$  полагаем вызванным явлениями подготовки предстоящей быстрой, случайной и интенсивной (то есть катастрофической) перестройки рассматриваемого объекта. По аналогии с [12–18] в роли основной гипотезы в этой статье вводится предположение о статистической независимости (или слабой зависимости) двух этих случайных слагаемых. При ее обосновании укажем, например, что применительно к землетрясениям физические поля, создаваемые явлениями разлома литосферных плит на ослабленных участках земной коры в зоне формирующегося гипоцентра, почти всегда являются независимыми от «обычных» измерений соответствующих природных феноменов.

На рисунке 1 схематически изображены плотности вероятности  $\rho_1(w)$  и  $\rho_2(w)$  для слагаемых  $x_1(t)$  и соответственно  $x_2(t)$ . Здесь вводятся обозначения:

$$\rho_{1,2}(w) = \sum_{m=0}^{N-1} P_{1,2}(m) \delta(w - m). \quad (2)$$

В выражении (2)  $\delta(Z)$  – дельта-функция Дирака, а коэффициенты  $P_{1,2}(m)$  равны вероятностям попадания значений случайных величин  $x_{1,2}(t)$  (если бы их можно было изменить по отдельности) в интервалы вида

$$\begin{aligned} x_{\min} + mh \leq x(t) < x_{\min} + (m+1)h, \\ N = \left[ \frac{x_{\max} - x_{\min}}{h} \right], \quad 0 \leq m \leq N-1. \end{aligned} \quad (3)$$

В (3) квадратные скобки означают целую часть числа, величины  $x_{\min}$  и  $x_{\max}$  есть наименьшее и соответственно наибольшее значения  $x(t)$ , а множитель  $h$  равен ширине интервала дискретности. На рисунке 1 с целью большей наглядности дискретные зависимости  $\rho_{1,2}(w)$  приведены в виде непрерывных функций. Полагаем, что в общем случае  $|x_2(t)| \ll |x_1(t)|$  (поскольку почти всегда влияние процессов подготовки катастрофы или аварии много меньше фонового шума работы соответствующей системы). Поэтому горизонтальный масштаб распределения  $\rho_2(w)$  будем считать много меньшим, чем для  $\rho_1(w)$ .

Следует подчеркнуть, что обе данные зависимости  $\rho_1(w)$  и  $\rho_2(w)$  – случайные функции, поскольку число проведенных измерений всегда является конечным. Как это изображено на рисунке 1, распределение  $\rho_1(w)$  в общем случае является «шероховатым» с учетом наличия случайных мелкомасштабных флуктуаций. Их эффективный период определяется значением ширины  $h$  интервала дискретизации, которое при необходимости может быть задано сколь угодно малым. Отметим, что аналогичные флуктуации существуют и для плотности вероятности  $\rho_2(w)$ , но при дальнейших расчетах не играют принципиальной роли в силу ее узости по сравнению с  $\rho_1(w)$ .



Для плотности вероятности  $\rho(w)$ , отвечающей измерениям суммы (1) двух независимых случайных слагаемых  $x_1$  и  $x_2$ , существует представление в виде свертки

$$\rho(w) = \int_{-\infty}^{\infty} \rho_{x_1}(w') \rho_{x_2}(w - w') dw'. \quad (4)$$

Для эффективной ширины  $S \sim 2\sigma_2$  плотности вероятности  $\rho_{x_2}(w)$ , где  $\sigma_2$  – стандарт флуктуаций случайной величины  $x_2$ , потребуем выполнения условия

$$S \gtrsim (2 \div 3)h. \quad (5)$$

Тогда, как и в [12–16], приходим к выводу, что при интегрировании вида (4) будет иметь место усреднение указанных выше мелкомасштабных флуктуаций зависимости  $\rho_1(w)$ . В результате для экспериментально полученного распределения  $\rho(w)$  ансамбля измеренных значений  $x(t)$  с большой вероятностью произойдет значительное уменьшение их эффективной амплитуды по сравнению с распределением  $\rho_1(w)$  для измерений одного фонового шума  $x_1(t)$  (если допустить, что его каким-то образом удалось измерить в отдельности от случайного слагаемого  $x_2(t)$ ). Для иллюстрации подобного эффекта на рисунке 1 также схематически изображена в форме «гладкой» кривой плотность вероятности  $\rho(w)$ , отвечающая распределению значений реально измеряемого случайного процесса  $x(t) = x_1(t) + x_2(t)$ .

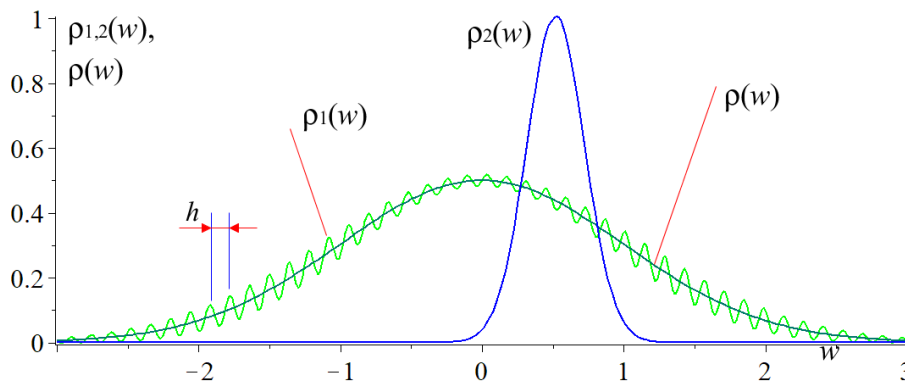


Рис. 1. Схематические изображенные плотности вероятности  $\rho_1(w)$ ,  $\rho_2(w)$  и  $\rho(w)$

Рассматриваемое изменение эффективной амплитуды мелкомасштабных флуктуаций эмпирического вероятностного распределения (равносильное переходу на рисунке 1 от «шероховатой» кривой  $\rho_1(w)$  к «гладкой» зависимости  $\rho(w)$ ), возникающее при появлении в измеряемом сигнале «дополнительного» независимого слагаемого  $x_2(t)$ , может быть математически определено с помощью ряда известных методов. В частности, удобным для использования здесь является дискретное преобразование Фурье.

Для усиления обсуждаемого эффекта в дальнейшем вместо непосредственно измеряемых значений случайной величины  $x = x(t)$  будем рассматривать свойства нелинейной функции вида

$$F[x(t)] = \sin(|Wx(t)|^\alpha), \quad (6)$$

где  $W$  и  $\alpha$  – задаваемые параметры. Статистический смысл такого перехода состоит в том, что отвечающая функции  $F[x(t)]$  плотность вероятности  $\rho_F(w)$  обладает интегрируемыми особенностями вида  $\sim [1 - w^2]^{-1/2}$ . В частности, для максимально простого случая  $W = 1$ ,  $\alpha = 1$ , при безразмерности значений  $x(t)$ , замене в (6) модуля  $|x(t)|$  на  $x(t)$  и выполнении условия  $-\pi/2 \leq x \leq \pi/2$  получаем

$$\rho_F(w) = [1 - w^2]^{-1/2} \rho_1\{\arcsin(w)\}. \quad (7)$$

Как легко понять, при  $|w| \rightarrow 1$  указанные выше мелкомасштабные шероховатости, которые будут присутствовать как для случайной функции  $\rho_1(w)$  (см. рис. 1), так и для  $\rho_1\{\arcsin(w)\}$ , значительно увеличатся по амплитуде при умножении на стремящийся к бесконечности сомножитель  $[1 - w^2]^{-1/2}$ . Поэтому при выполнении (5) с большой вероятностью произойдет усиление различий между функциями  $\rho_F(w)$  при отсутствии (см. «шероховатую» широкую кривую  $\rho_1(w)$  на рисунке 1) и наоборот, при существовании (см. «гладкую» зависимость  $\rho(w)$ ) независимого случайного слагаемого  $x_2(t)$  в правой части (1). Можно сказать, что использование нелинейного преобразования вида (6)

приводит к своего рода эффекту «увеличительного стекла», при помощи которого оказывается возможным исследовать свойства локальных участков экспериментально полученной зависимости  $\rho(w)$ . Введенные в (6) параметры  $W$  и  $\alpha$  задают протяженность интервала наиболее сильного стремления к бесконечности для интегрируемой случайной зависимости  $\rho_F(w)$ .

С учетом проведенных расчетов далее каждому отрезку реализации измеряемой случайной величины  $x(t)$ , имеющему номер  $l$ , ставится в соответствие статистический функционал вида

$$L(n) = \frac{A}{M} \sum_{l=n-M+1}^n |\mathcal{L}_l|, \quad \mathcal{L}_l = \sum_{m=0}^{N-1} (-1)^m P_{m,l}. \quad (8)$$

(Далее применительно к исследованию медицинских задач этот функционал будет несколько модифицирован.)

В (8) аргумент  $n$  отвечает моменту времени завершения отрезка реализации с этим номером, множитель  $A = 1000$  введен с целью перехода к удобной для анализа области значений данного функционала, а значение  $M$  – ширина интервала усреднения по последовательным отрезкам реализации. Подобно (3), в (8) любой множитель  $P_{m,l}$  под знаком суммы равен отвечающей данному отрезку реализации вероятности попадания значения функции  $F[x(t)]$  от измеренной случайной величины  $x = x(t)$  в интервал номер  $m$  области значений вида

$$F[x]_{\min} + mh \leq F[x(t)] < F[x]_{\min} + (m+1)h, \quad (9) \\ 0 \leq m \leq N-1, \quad N = 2[h^{-1}].$$

В (9), как и в (3), квадратные скобки в выражении  $N = 2[h^{-1}]$  для общего количества интервалов  $N$  означают целую часть числа. (В (9) принято во внимание, что разность между наибольшим ( $F[x]_{\max}$ ) и наименьшим ( $F[x]_{\min}$ ) значениями функции (6) не превосходит двух, откуда получаем соотношение для  $N$ ). Кроме того, в (8) аргумент  $p$  есть переменная Фурье. Укажем, что с учетом используемого математического аппарата предлагаемый подход можем назвать методикой спектрально-вероятностного анализа.

Отметим также, что рассматриваемое уменьшение эффективной амплитуды мелкомасштабных флуктуаций плотности вероятности означает ее приближение к неизвестному «истинному» распределению, которое соответствовало бы неограниченно большому числу измерений, проводимых за произвольно длительное время. Такое распределение означало бы переход к ситуации максимального хаоса с неограниченно высоким ростом энтропии.

Далее предлагаемый подход апробируется применительно к солнечным вспышкам высокой энергии, сильным землетрясениям и заболеваниям человеческого организма.

## Результаты экспериментов

### 1. Проведение расчетов, связанных с прогнозированием солнечных вспышек высокой энергии

На рисунке 2 приведена полученная в [12] зависимость  $L(n)$  при анализе изменений статистических свойств интенсивности рентгеновского излучения Солнца диапазона длин волн  $0.5 \div 4 \text{ \AA}$  перед вспышкой класса X2.2, которая произошла 20 апреля 2022 года (здесь и далее используется время UTC). В данном случае, как и далее на рисунке 3, в (6) принято  $W = 10^{10} \text{ (м}^2/\text{Вт)}$  и  $\alpha = 1$ . Здесь и на рисунке 3 моменты вспышек высокой энергии отмечены сплошными красными вертикальными прямыми линиями. В данном случае длительность одного отрезка реализации была принята равной  $\Delta t = 3000$  секунд, каждой из которых отвечало одно из  $N_1 = 3000$  измерений. В (8) усреднение проводилось по  $M = 10$  последовательным отрезкам реализации. Таким образом, любое значение функционала  $L(n)$  отвечает 30000 последовательным измерениям интенсивности рентгеновского излучения Солнца. При анализе данного рисунка, так же как и далее рисунка 3, используются понятия (см. [12]) «локальные тренды», «каналы», «тестирование», «скользящие границы».

(а). Назовем *локальными трендами* (ЛТ) участки зависимости  $L(n)$ , находящиеся между точками экстремумов рассматриваемой кривой и отвечающие вариациям  $\delta L$  (равным разности значений функционала  $L(n)$  в точках границ ЛТ, которые отвечают максимумам и минимумам наподобие точек  $a$  и соответственно  $b$  на рисунке 2), удовлетворяющим условию  $\delta L \leq \Delta L_{\max}$ .

Здесь  $\Delta L_{Max}$  – наибольшая вариация значений  $L(n)$  в пределах интервала времени от начала отсчета и до момента рассматриваемого события, то есть разность значений наиболее высокого максимума и самого глубокого минимума (точки  $A$  и соответственно  $B$  на рисунке 2) функционала  $L(n)$  на указанном отрезке времени. При этом вводится условие, что геометрическое отклонение  $\Delta s$  кривой  $L(n)$  от прямолинейного отрезка, соединяющего граничные точки любого локального тренда (см. рис. 2), удовлетворяет условию

$$\Delta s \leq 0.2\Delta l, \quad (10)$$

где  $\Delta l$  – длина этого отрезка. В (10) расстояния  $\Delta s$  и  $\Delta l$ , так же как и введенный далее угол  $\Delta\alpha$ , рассчитываются при нормировке горизонтальных координат на время  $n_0$  начала вспышки, а вертикальных – на наибольшее значение функционала  $L(n)$  на промежутке  $0 \leq n \leq n_0$ . Условие (10) необходимо для обеспечения квазипрямолинейности локального тренда. На рисунке 2 локальным трендом является, в частности, сегмент  $ab$  кривой  $L(n)$ ; здесь же указано расстояние  $\Delta s$ . Точки границ ЛТ будем именовать *направляющими точками*. Полагаем, что каждая такая точка по горизонтали удалена от ближайшей точки кривой  $L(n)$  не менее, чем на интервал в 10 часов.

(б). Каналом будем именовать совокупность двух прямых, каждая из которых проходит точно через две направляющие точки, и при этом угол  $\Delta\alpha$  между направлениями этих прямых удовлетворяет неравенству

$$\Delta\alpha \leq 1.4^\circ. \quad (11)$$

В (11) угол  $\Delta\alpha$  определяется при помощи скалярного произведения двух векторов, параллельных границам канала. Интервал времени между направляющими точками, через которые проведена граница канала, должен составлять не менее 10 часов. На рисунке 2 отмечены каналы: 1–2, 2–3, 4–5, 6–7, 8–9, образуемые парами наклонных пунктирных прямых отрезков, углы между которыми удовлетворяют (11).

(с). Осевой линией будем именовать прямую, проходящую через начальную и конечную точки локального тренда. Осевым сектором будем именовать пару осевых линий, удовлетворяющую условию (11). В роли момента регистрации предвестника приближающейся солнечной вспышки будем рассматривать последнюю по времени направляющую точку, через которую проходит соответствующая граница осевого сектора (точка  $d$  на рисунке 3 для осевого сектора с границами  $A_1$  и  $A_2$ ).

(д). Термин *тестирование* означает либо пересечение в некоторой точке кривой  $L(n)$  и соответствующей границы канала, либо прохождение (вне зависимости от факта пересечения или его отсутствия) границы канала на расстоянии  $\Delta L_t$  по вертикали от направляющей точки с координатами  $\{n_t, L(n_t)\}$ , удовлетворяющем условию

$$\frac{\Delta L_t}{L(n_t)} \cdot 100\% \leq \delta, \quad \delta \leq 1.2\% \quad (12)$$

В роли момента регистрации предвестника приближающегося солнечного события рассматривается либо время самого последнего тестирования границ канала зависимостью  $L(n)$  при нарушении (12) (например, точка  $c$  для канала 8–9 на рисунке 2), либо время тестирования в последней направляющей точке, через которую проведена данная граница канала и после которой границы этого канала вообще не пересекают (и не тестируют при справедливости (12)) кривую  $L(n)$ . В любом из этих двух случаев будет использоваться термин «точка последнего тестирования».

(е). Скользящая граница – прямая линия, проведенная через две направляющие точки и тестируемая кривой  $L(n)$  еще как минимум в одной направляющей точке, удовлетворяющей условию (12) (такие точки будем называть дополнительными и выделять на рисунках сплошным эллипсом). Так же как и для каналов, полагаем, что время регистрации предвестника приближающегося события высокой энергии отвечает моменту последнего тестирования (смысл этого термина эквивалентен случаю каналов) кривой  $L(n)$  со стороны скользящей границы (см, например, точку  $d$  для скользящей границы 12 на рисунке 2). Скользящие границы (на рисунке 2 это прямые 10 и 12) отмечаются сплошными наклонными прямыми линиями.





На рисунке 5 приведена зависимость  $L(n)$ , отвечающая измерениям Н-компоненты магнитного поля за 19 и 20 августа 2019 (КраО) (см. [12]). Этот период включает момент 09:34:54 20.08.2019 начала землетрясения магнитудой 4.4, которое произошло вблизи южного побережья Турции. В связи с малой магнитудой события в данном случае имеет место только одна детерминированная геометрическая структура в виде скользящей границы  $I$ , которая проведена через направляющие точки  $a$  и  $c$  при дополнительной точке  $b$ . С учетом малого (по сравнению с рисунком 4) рассматриваемого интервала времени минимальное горизонтальное расстояние от любой направляющей точки до другой точки кривой  $L(n)$  в данном случае принято равным 5 часам. Отметим, что в данном случае в (5) коэффициент  $\delta = 0.06\%$ , что свидетельствует об исключительной точности прохождения прямой  $I$  через «дополнительную» точку  $b$ . Как и ранее, момент последнего тестирования скользящей границы  $I$  (в точке  $d$ ) оказывается значительно приближен к времени начала данного землетрясения.

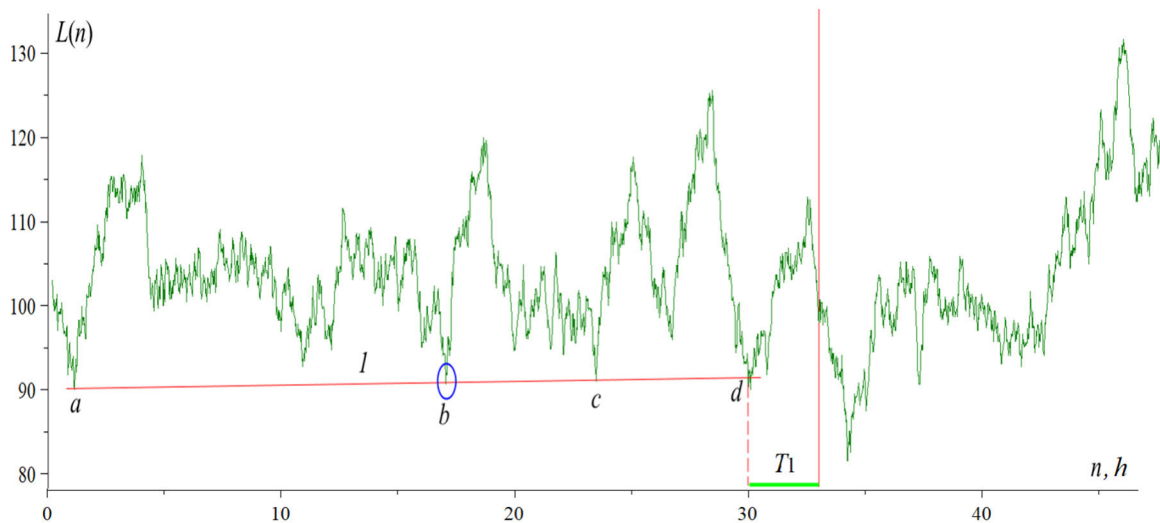


Рис. 5. Зависимость  $L(n)$ , построенная по измерениям магнитного поля за 19 и 20 августа 2019 года (КраО)

### 3. Проведение расчетов, связанных с прогнозированием изменения состояния здоровья человека

Применяемый подход, основанный на изменении уровня хаоса проводимых измерений при возникновении в системе независимого случайного процесса, оказывается вполне эффективным и при прогнозировании заболеваний человека. На рисунке 6 приведены  $N_G = 83$  зеленых кривых (рис. 6а), которые отвечают зависимостям  $L(p)$  для выздоровевших пациентов, поступивших в стационар с острой декомпенсацией сахарного диабета, а также  $N_R = 20$  красных линий (рис. 6а и 6б), соответствующих случаям неблагоприятного исхода [16].

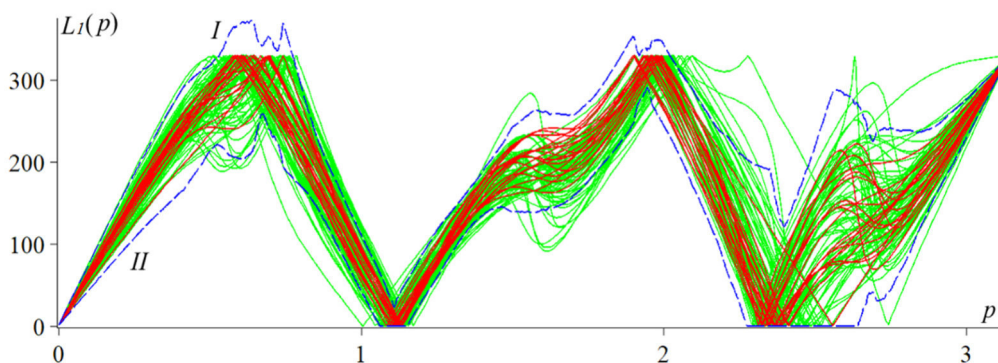


Рис. 6а. Ансамбли из  $N_G = 83$  зеленых и  $N_R = 20$  красных линий – зависимости  $L_1(p, h, W, \alpha)$  (построенные при  $h = 0.15, W = 1$  и  $\alpha = -1.495$ ) для выздоровевших пациентов и соответственно для больных с неблагоприятным исходом



На рисунке 6 каждая кривая  $L_1(p)$  отвечает статистическому функционалу вида

$$L_1(p) = L_1(p, h, W, \alpha) = \left| \text{Arg} \left( \sum_{m=0}^{N-1} P_m e^{i\alpha m} \right) \right|, \quad (13)$$

который соответствует своему больному и основывается на том же, что и (8), вероятностном эффекте. Здесь  $P_m$  есть вероятность попадания значения аналогичной (6) функции

$$F[x(j)] = \sin[(W|x(j)|)^\alpha], \quad (14)$$

от результата  $x(j)$  проведенного медицинского анализа биохимии крови пациента, полученного на момент поступления в клинику, в ячейку номер  $m$  вида (8). При этом для рисунков 6a,b полагаем  $h = 0.15, W = 1$  и  $\alpha = -1.495$ . В (14) индекс  $j$  принимает значения  $j = 1, 2, \dots, 20$ , то есть рассматривается ансамбль из  $N_a = 20$  чисел, являющихся результатами соответствующих анализов. Они включают в себя данные таких лабораторных показателей, как глюкоза плазмы при поступлении, показатели кетонурии, данные кислотно-основного состояния (рН, лактат, осмоляльность, натрий, калий), скорость клубочковой фильтрации, лейкоциты крови. В (13) символ  $\text{Arg}$  означает аргумент локального ряда Фурье  $\sum_{m=0}^{N-1} P_m e^{i\alpha m}$  на плоскости комплексной переменной. Отметим, что функционал (8) можем рассматривать как разновидность входящей в (13) суммы при условии, что в (13) Фурье-переменная  $p = \pi$ , а в (8) значение  $M = 1$ .

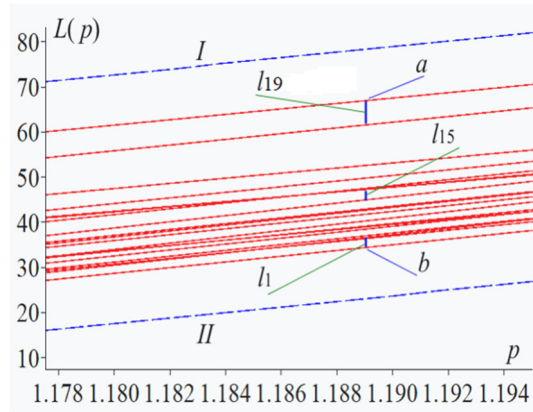


Рис. 6b. Совокупность расстояний  $l_m, m = 1, 2, \dots, 19$ , при заданном значении Фурье-переменной  $p$

На рисунках 6a,b синие пунктирные кривые I и II означают границы канала, такого, что вероятность выхода красной линии, отвечающей новому больному с неблагоприятным исходом, за эти границы при заданном значении  $p$  не превышает 5 %. Данный канал строится на основании неравенства Чебышева по ансамблю расстояний  $l_m, m = 1, 2, \dots, 19$ , изображенных на рисунке 6b.

На рисунке 6с приведена в логарифмическом масштабе совокупность значений усредненных площадей  $\mathcal{L}_{G,M}$  суммарного выхода зеленых кривых  $L_1(p)$  из пределов канала I–II (зеленые кружки; каждый из них соответствует выздоровевшему больному). При этом имеет место усреднение  $\mathcal{L}_{G,M} = \frac{1}{N} \sum_{j=1}^N G_M(h, W, \alpha_j)$  по ансамблю наборов параметров (см. (14)) вида

$$h = 0.15, W = 1, \alpha = \alpha_j = -1.479 + 0.02(j - 1), j = 1, 2, \dots, N = 15. \quad (15)$$

Здесь  $M$  – номер, сопоставляемый пациенту, а значение  $\mathcal{L}_{G,M}$  – указанная площадь выхода из границ данного канала при одном наборе параметров (15). Здесь же приведены красные кружки. Каждый из них соответствует полученной при аналогичном усреднении общей площади выхода соответствующей красной кривой из границ канала I–II, который при всех указанных значениях параметров строится без учета данной кривой.

Как следует из анализа рисунка 6с, более половины зеленых кружков находятся в верхней области, проходящей выше синего горизонтального пунктирного барьера. Здесь почти нет (кроме одного) кружков красного цвета. Это означает, что если для нового больного при поступлении в стационар результаты анализов соответствуют попаданию в данную область, то вероятность благоприятного исхода для него весьма высока. Данный факт позволяет сконцентрировать особое внимание



медицинского персонала на тех больных, которые оказываются в нижней области, где расположены 19 красных кружков. Это повышает вероятность правильного выбора пациентов с повышенной степенью риска, тем самым уменьшая для них возможность неблагоприятного исхода. Отметим, что чувствительность применяемого алгоритма существенно возрастает при увеличении числа исходных рассматриваемых результатов медицинских анализов.

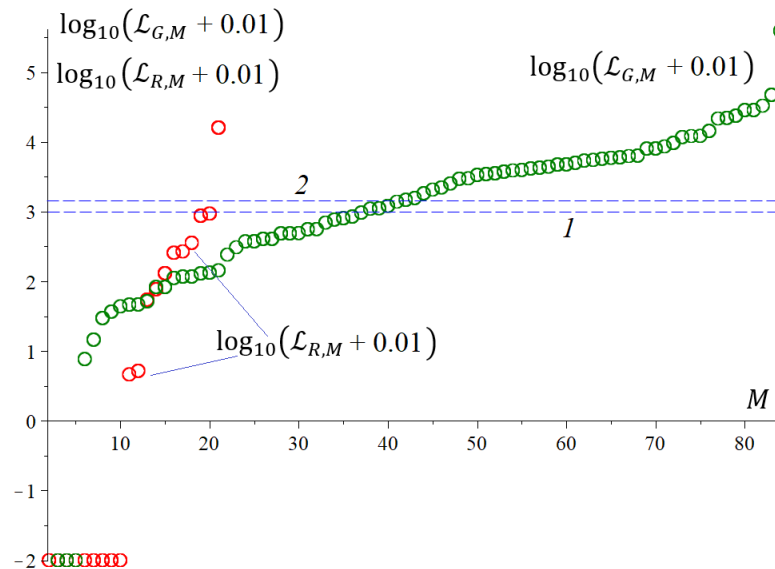


Рис. 6с. Логарифмическое представление для 83 средних значений  $\mathcal{L}_{G,M}$  и 20 подобных усреднений  $\mathcal{L}_{R,M}$  (зеленая и соответственно красная последовательности кружков)

### Обсуждение результатов

В результате применения используемой методологии во всех случаях удалось свести задачу прогнозирования состояния рассматриваемых объектов к инженерно-геометрическому анализу свойств ансамблей кривых на декартовой плоскости. Во всех случаях использование предлагаемого алгоритма было основано на феномене возникновения «естественной регуляризации» свойств соответствующих вероятностных распределений при существовании процесса подготовки катастрофического перехода любой физической или биологической системы.

Как показали проведенные расчеты, и для солнечных событий, и применительно к землетрясениям имеет место эффект возникновения геометрических структур высокой степени детерминированности. Моменты их завершения тяготеют к времени начала как солнечных вспышек, так и сейсмических событий. Возникновение большого числа точек связности на рисунках 2 и 4 позволяет сделать вывод о высокой согласованности процессов подготовки соответствующих катастрофических явлений. Концентрация рассматриваемых предвестников может считаться маркером предстоящего события. Этот эффект имеет место перед всеми исследованными солнечными вспышками X-класса и сильными землетрясениями. Однотипные и высоко детерминированные геометрические феномены, возникающие применительно к вспышкам высокой энергии и сильным землетрясениям, позволяют прийти к выводу о том, что уровень хаоса в сложных системах самой разной природы перед их катастрофическими перестройками начинает изменяться по почти детерминированным законам. С учетом взаимно подобного поведения красных кривых на рисунке 6 такой вывод оказывается справедливым и для опасных состояний организма человека. Этот факт подтверждается также и единством математического аппарата, примененного в статье для анализа катастрофических процессов совершенно разной природы.

Отметим также, что одним из главных результатов данной работы является переход к формированию языка, то есть системы взаимно согласованных понятий и определений, с применением которого становятся «читаемыми» процессы, происходящие в сложных системах в период подготовки их предстоящего катастрофического изменения.

## Закключение

В данной статье рассматривается алгоритм обоснования вывода о высокой вероятности приближения катастрофического или опасного состояния системы. В рамках модификации единого математического подхода, основанного на методике спектрально-вероятностного анализа, предлагаемая методика была апробирована как для природных катастрофических явлений (вспышки на Солнце и землетрясения), так и применительно к заболеваниям человека, несущим фактор особого риска. Во всех случаях использовался вероятностный эффект, состоящий в стремлении эмпирического вероятностного распределения значений исследуемого процесса к состоянию идеального хаоса при возникновении в системе случайного слагаемого, независимого от совокупности обычно протекающих явлений и являющегося неизбежным следствием ситуации подготовки сильного, случайного и быстрого, то есть катастрофического, преобразования свойств любой рассматриваемой системы.

## Источник финансирования

Исследование выполнено за счет гранта Российского научного фонда № 25-79-20019, <https://rscf.ru/project/25-79-20019/>.

**Научный руководитель:** Вольвач Александр Евгеньевич, Крымская астрофизическая обсерватория.

## Список литературы

1. A comparison of flare forecasting methods. I. Results from the "All-clear" workshop / G. Barnes, K. Leka, C. Schrijver, et al. // *ApJ*. 2016. Vol. 829. P. 89.
2. A Comparison of Flare Forecasting Methods. II. Benchmarks, Metrics, and Performance Results for Operational Solar Flare Forecasting Systems / K. Leka, S.-H. Park, K. Kusano, et al. // *ApJ*. 2019. Vol. 243. P. 36.
3. A physics-based method that can predict imminent large solar flares / K. Kusano, T. Iju, Y. Bamba, et al. // *Science*. 2020. Vol. 369. P. 587.
4. A Method of Forecasting Solar Activity Based on Radio Astronomical Observations / V.M. Bogod, P.M. Svidskiy, E.A. Kurochkin, et al. // *Astrophys. Bull.* 73. P. 478.
5. Uyeda S., Nagao T., Kamogawa M. Short-term earthquake prediction: current status of seismo-electromagnetics // *Tectonophysics*. 2009. Vol. 470, no. 3–4. P. 205–213. // *Geomatics, Natural Hazards and Risk*. 2010. Vol. 1, no 2. P. 115–155.
6. Добровольский И.А. Теория подготовки тектонического землетрясения. М.: ИФЗ АН СССР, 1991. 217 с.
7. Добровольский И.П. Математическая теория подготовки и прогноза тектонического землетрясения. М.: Физмат, 2010. 230 с.
8. Molchanov O. A., Hayakawa M., Rafalsky V. A. // *Physics and Chemistry of the Earth*. 2004. Vol. 29, no 4–9. P. 569–577. Lifespan variation among people with a given disease or condition / Y. Zheng, I. Permanyer, V. Canudas-Romo, et al. // *PLOS ONE*. 2023. Vol. 18, no. 9. P. e0290962.
9. Mortality among persons with obsessive-compulsive disorder in Denmark / S.M. Meier, M. Mattheisen, O. Mors, et al. // *JAMA Psychiatry*. 2016. No. 73 P. 268–274.
10. Statistical precursors of solar flares on September 6, 2017, March 23 and May 1, 2024 / A.E. Volvach, L.P. Kogan, L.N. Volvach, I.V. Yakubovskaya // *Advances in Space Research*. 2025. Vol. 76. P. 551-561.
11. On statistical precursors of solar flares / A.E. Volvach, L.P. Kogan, L.N. Volvach, I.V. Yakubovskaya // *Chinese Journal of Physics*. 2025. Vol. 97. P. 862-882.
12. On the set of deterministic phenomena preceding the earthquake June 25, 2021 with a magnitude of 5.4 near the city of Yayladere (Turkey) / A.E. Volvach, L.P. Kogan, L.N. Volvach, [et al.] // *Romanian Reports on Physics*. 2024. Vol. 76, no. 2. P. 704.
13. Changes in the properties of the statistics of physical and biophysical fields as earthquake precursor / A.E. Volvach, L.P. Kogan, V.B. Shtenberg [et al.] // *Communications in Nonlinear Science and Numerical Simulation*. 2022. Vol. 108. P. 106200.
14. Prediction of unfavorable outcome of acute decompensation of diabetes mellitus / L.P. Kogan, K.G. Korneva, A.E. Volvach, et al. // *Diabetol Metab Syndr*. 2025. Vol. 20, no. 17(1). P. 64.
15. Подтверждение клинического диагноза «шизофрения» по значениям статистических функционалов / Л. П. Коган, А. С. Гордцов, Л. Н. Касимова [и др.] // *Современные проблемы науки и образования*. 2020. № 2. С. 146.
16. Диагностика заболевания по значениям статистических функционалов, вычисленных по совокупности ИК-спектроскопических параметров крови / Л.П. Коган, Д.И. Кислицын, О.В. Красникова и др. // *Современные технологии в медицине*. 2017. № 4. С. 25-35.

## Интеграция генеративных нейросетей в математическое и трехмерное моделирование: современное состояние и перспективы развития

Н. А. Бондарева

Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В статье представлен обзор современных подходов к применению генеративных нейросетей в задачах математического и трехмерного моделирования. Рассматриваются теоретические основы генеративных нейросетей, их архитектуры и методы обучения. Анализируются существующие подходы к математическому и трехмерному моделированию, а также возможности интеграции с генеративными нейросетями. Особое внимание уделяется гибридным подходам, сочетающим преимущества генеративных нейросетей с традиционными методами и экспертными знаниями, обеспечивающими более высокую точность, надежность и контролируемость результатов. В статье обсуждаются перспективы развития и социально-экономические последствия внедрения гибридных нейросетевых технологий в инженерные и научные области.

**Ключевые слова:** генеративные нейросети, гибридный подход, математическое моделирование, трехмерное моделирование, искусственный интеллект, автоматизация, валидация, машинное обучение

## Integration of generative neural networks in mathematical and three-dimensional modeling: current state and development prospects

N. A. Bondareva

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

**Abstract.** This article provides a review of contemporary approaches to the application of generative neural networks in mathematical and three-dimensional modeling tasks. It examines the theoretical foundations of generative neural networks, their architectures, and training methodologies. Existing approaches to mathematical and three-dimensional modeling are analyzed, along with the potential for integration with generative neural networks. Particular emphasis is placed on hybrid approaches that combine the advantages of generative neural networks with traditional methods and expert knowledge, ensuring higher accuracy, reliability, and controllability of results. The article discusses the development prospects and socio-economic implications of implementing hybrid neural network technologies in engineering and scientific domains.

**Keywords:** generative neural networks, hybrid approach, mathematical modeling, three-dimensional modeling, artificial intelligence, automation, validation, machine learning

### Введение

В современную эпоху цифровой трансформации технологии искусственного интеллекта, в частности генеративные нейронные сети, демонстрируют впечатляющий потенциал в решении широкого спектра задач. Однако наряду с очевидными достижениями в области генерации контента, обработки данных и автоматизации процессов существует ряд критических ограничений, препятствующих их полноценной интеграции в производственные и научно-исследовательские процессы.

Ключевой проблемой современных нейросетевых технологий является их функционирование по принципу «черного ящика», когда результаты генерируются без прозрачного объяснения логики принятия решений. Это существенно затрудняет процессы валидации полученных результатов, выявления и корректировки ошибок, а также внесения целенаправленных изменений в финальные решения. Особенно остро данная проблема проявляется в областях, требующих высокой точности и надежности результатов, таких как математическое моделирование и трехмерное проектирование.

Анализ существующих подходов к интеграции нейросетевых технологий в производственные процессы выявляет значительный разрыв между теоретическими возможностями искусственного интеллекта и практическими требованиями промышленности. Традиционные методы внедрения часто фокусируются либо на полной автоматизации процессов с помощью AI, либо на использовании нейросетей как вспомогательного инструмента, что не позволяет полностью реализовать потенциал данных технологий.

В рамках данного исследования предлагается концепция гибридного методологического подхода, призванного преодолеть указанные ограничения. В основе подхода лежит синергия технологий обработки

естественного языка (NLP) и верифицированных программных комплексов для математического и трехмерного моделирования. Предлагаемая методология базируется на интеграции возможностей систем искусственного интеллекта в области обработки естественного языка и оперативного генерирования вариативных решений с существующими алгоритмами программных комплексов.

Особого внимания заслуживает влияние генеративных нейросетей на сферу программирования и разработки программного обеспечения. Современные языковые модели демонстрируют впечатляющие способности в генерации программного кода, автоматизации рутинных задач программирования и помощи в отладке. Однако здесь также проявляется проблема «черного ящика»: сгенерированный код требует тщательной проверки, так как нейросети могут создавать внешне корректный, но содержащий логические ошибки или уязвимости код. Это подчеркивает необходимость разработки методологических подходов к верификации и валидации генерируемых программных решений.

Внедрение нейросетевых технологий существенно трансформирует структуру рынка труда в технологическом секторе. С одной стороны, наблюдается значительное снижение порога входа в профессию: генеративные модели предоставляют начинающим специалистам мощные инструменты для обучения и решения базовых задач. Это открывает новые возможности для профессионального развития и позволяет быстрее осваивать сложные технологические области. С другой стороны, автоматизация рутинных операций приводит к сокращению потребности в низкоквалифицированных специалистах, выполняющих типовые задачи.

При этом важно отметить, что роль высококвалифицированных специалистов не только сохраняется, но и возрастает. Это обусловлено как необходимостью экспертной оценки и валидации результатов работы нейросетей, поскольку в настоящее время невозможно полностью гарантировать качество и надежность автоматически генерируемых решений, так и потребностью в разработке и оптимизации методологий применения AI-технологий.

### **Теоретические основы и современное состояние проблемы**

Генеративные нейросети – одно из наиболее динамично развивающихся направлений в ИИ. Эволюция началась с классических генеративно-состязательных сетей (GAN) в 2014 году и продолжается по сей день [1]. Современный этап характеризуется переходом к более сложным и управляемым системам.

Диффузионные модели, ставшие прорывом последних лет, предложили новый подход к генерации контента, основанный на постепенном уточнении результата через последовательное удаление шума [2]. Этот метод показал исключительную эффективность в задачах генерации изображений, трехмерных моделей и других типов данных, обеспечивая более стабильные и качественные результаты по сравнению с классическими подходами. Диффузионные модели, ставшие прорывом последних лет, предложили новый подход к генерации контента, основанный на постепенном уточнении результата через последовательное удаление шума. Параллельно с развитием диффузионных моделей произошел прорыв в области трансформер-архитектур, изначально разработанных для обработки естественного языка, но успешно адаптированных для работы с различными типами данных, включая программный код и математические формулы. Это открыло новые возможности для создания универсальных генеративных систем, способных одновременно работать с несколькими модальностями данных.

В контексте обучения современных генеративных моделей показателен пример Stable Diffusion XL, использующей многоступенчатую стратегию обучения. Модель сначала проходит предварительное обучение на массивном наборе изображений LAION-5B, после чего производится её точная настройка на специализированных датасетах для конкретных задач. Аналогичный подход применяется в CodeLlama-34b от Meta, где базовая языковая модель дообучается на специально подготовленных наборах программного кода, что позволяет достичь высокой точности в генерации специфических программных конструкций и работе с различными языками программирования.

Самоконтролируемое обучение наглядно демонстрирует свою эффективность в архитектуре GPT-4V (Visual), где модель способна извлекать семантические связи между изображениями и текстом без

явной разметки этих связей в обучающих данных. В контексте технических задач показателен пример Point-E от OpenAI, способной генерировать трёхмерные модели на основе текстовых описаний, используя промежуточное представление в виде облака точек.

Практическое применение генеративных моделей можно проиллюстрировать конкретными примерами из различных отраслей. В промышленном проектировании Autodesk использует генеративный дизайн в Fusion 360 для создания оптимизированных конструкций. Например, при проектировании кронштейна для аэрокосмической отрасли система генерирует множество вариантов, оптимизированных по весу и прочности, что позволило сократить вес деталей на 20-40 % при сохранении или улучшении механических характеристик.

В программной инженерии Amazon CodeWhisperer и GitHub Copilot демонстрируют эффективность применения генеративных моделей для автоматизации разработки. Согласно исследованиям GitHub, разработчики, использующие Copilot, завершают задачи в среднем на 55 % быстрее, при этом качество кода, измеряемое количеством успешно проходящих тестов, увеличивается на 26 % [3].

Однако внедрение подобных технологий сопряжено с конкретными техническими вызовами. Например, использование Stable Diffusion XL для промышленного дизайна требует значительных вычислительных ресурсов: минимум 16 ГБ видеопамати для базовой работы и до 24 ГБ для оптимальной производительности. При интеграции GitHub Copilot в корпоративные системы возникают вопросы безопасности и конфиденциальности кода, что требует развертывания локальных версий системы и дополнительных средств контроля.

Выбор конкретной архитектуры (см. таблицу) существенно зависит от специфики решаемых задач и доступных вычислительных ресурсов. При этом современная тенденция развития генеративных моделей направлена на создание гибридных архитектур, способных объединить преимущества различных подходов при одновременной минимизации их недостатков.

**Типы нейросетевой архитектуры**

Архитектура	Преимущества	Ограничения	Область применения
Классические GAN	<ul style="list-style-type: none"> <li>Высокая скорость генерации</li> <li>Относительная простота архитектуры</li> <li>Небольшие требования к вычислительным ресурсам</li> </ul>	<ul style="list-style-type: none"> <li>Нестабильность обучения</li> <li>Проблема mode collapse</li> <li>Сложность контроля генерации</li> </ul>	<ul style="list-style-type: none"> <li>Генерация изображений</li> <li>Аугментация данных</li> <li>Прототипирование</li> </ul>
Диффузионные модели	<ul style="list-style-type: none"> <li>Высокое качество генерации</li> <li>Стабильность результатов</li> <li>Хороший контроль над процессом</li> </ul>	<ul style="list-style-type: none"> <li>Высокие вычислительные затраты</li> <li>Медленная генерация</li> <li>Сложность архитектуры</li> </ul>	<ul style="list-style-type: none"> <li>Профессиональная генерация контента</li> <li>3D-моделирование</li> <li>Научные исследования</li> </ul>
Трансформеры	<ul style="list-style-type: none"> <li>Универсальность применения</li> <li>Хорошая масштабируемость</li> <li>Работа с разными типами данных</li> </ul>	<ul style="list-style-type: none"> <li>Большие требования к памяти</li> <li>Сложность обучения</li> <li>Высокая стоимость разработки</li> </ul>	<ul style="list-style-type: none"> <li>Генерация кода</li> <li>Мультимодальные задачи</li> <li>Комплексная автоматизация</li> </ul>
Гибридные архитектуры	<ul style="list-style-type: none"> <li>Комбинация преимуществ разных подходов</li> <li>Гибкость настройки</li> <li>Широкие возможности оптимизации</li> </ul>	<ul style="list-style-type: none"> <li>Сложность интеграции</li> <li>Повышенные требования к инфраструктуре</li> <li>Необходимость тщательной настройки</li> </ul>	<ul style="list-style-type: none"> <li>Промышленное применение</li> <li>Сложные производственные задачи</li> <li>Исследовательские проекты</li> </ul>

Особую роль в развитии генеративных технологий играет проблема валидации и верификации получаемых результатов. Если в задачах генерации изображений или текста оценка качества может производиться в том числе субъективно, то для технических задач, таких как генерация программного кода или трехмерных моделей, требуются строгие математические критерии и методики проверки. Это становится особенно актуальным при интеграции генеративных моделей в производственные процессы, где цена ошибки может быть критически высокой.

В контексте обучения современных генеративных моделей произошел значительный сдвиг парадигмы от классических методов к более комплексным подходам. Традиционные методики обучения, основанные на прямой минимизации функции потерь, уступили место многоэтапным стратегиям, включающим предварительное обучение на больших наборах данных с последующей специализированной доводкой под конкретные задачи. Особую роль в этом процессе играет концепция трансферного обучения, позволяющая эффективно адаптировать предобученные модели для решения специфических задач при существенно меньших требованиях к вычислительным ресурсам и объему обучающих данных.

Современные подходы к обучению генеративных моделей также характеризуются активным использованием техник самоконтролируемого обучения. Эти методики позволяют моделям извлекать полезные признаки и закономерности из неразмеченных данных, что особенно важно в контексте технических задач, где получение качественной разметки может быть чрезвычайно затратным или практически невозможным. При этом особое внимание уделяется проблеме регуляризации и предотвращения переобучения, что критически важно для обеспечения стабильности и надежности генерируемых результатов.

### **Современные подходы к решению задач математического моделирования**

Математическое моделирование, являясь фундаментальным инструментом научных исследований и инженерных разработок, претерпевает существенную трансформацию под влиянием ИИ. Интеграция методов машинного обучения с классическими подходами создает новую парадигму в вычислительной науке. Традиционные методы, основанные на численном решении дифференциальных уравнений, сталкиваются с ограничениями при работе со сложными нелинейными системами.

Согласно исследованию [4], опубликованному в *Nature Reviews Physics*, интеграция методов машинного обучения с классическими подходами математического моделирования создает новую парадигму в вычислительной науке. Традиционные методы математического моделирования, основанные на численном решении дифференциальных уравнений, сталкиваются с рядом существенных ограничений.

В этом контексте особый интерес представляют гибридные подходы, объединяющие классические численные методы с нейросетевыми моделями. Например, исследование [5] демонстрирует, как использование нейронных сетей в задачах гидродинамики позволяет сократить время расчетов на порядки при сохранении приемлемой точности результатов.

Математическое моделирование в 2020-х годах характеризуется активным внедрением высокопроизводительных вычислительных систем и новых методологических подходов. Ведущие научные центры, такие как национальные лаборатории США (Argonne, Lawrence Berkeley) и европейские исследовательские институты, демонстрируют устойчивый тренд к использованию гибридных вычислительных архитектур, сочетающих классические подходы с элементами искусственного интеллекта.

В промышленном секторе наблюдается активное использование коммерческих пакетов математического моделирования, среди которых лидирующие позиции занимают:

#### **1. ANSYS, предоставляющий инструменты:**

- для конечно-элементного анализа;
- вычислительной гидродинамики;
- электромагнитного моделирования;

2. COMSOL Multiphysics, который, согласно техническим отчетам компании, внедрил поддержку машинного обучения в свои решатели, что позволило существенно ускорить расчеты сложных мультифизических задач;

3. MATLAB от MathWorks, который в последних версиях значительно расширил возможности интеграции с инструментами машинного обучения.

В области научных вычислений особую роль играют суперкомпьютерные центры. Согласно данным TOP500 (рейтинг самых мощных суперкомпьютеров мира), современные системы достигают производительности в сотни петафлопс, что открывает новые возможности для решения сложных задач математического моделирования [6].

Ключевыми тенденциями развития математического моделирования, подтвержденными многочисленными публикациями в ведущих научных журналах и практическим применением, являются:

- 1) развитие методов многомасштабного моделирования, позволяющих учитывать процессы на различных пространственных и временных масштабах;
- 2) внедрение методов машинного обучения:
  - для ускорения вычислений;
  - оптимизации расчетных сеток;
  - предсказания поведения сложных систем;
- 3) создание цифровых двойников, что подтверждается успешными внедрениями в аэрокосмической промышленности (Boeing, Airbus) и энергетическом секторе (Siemens, General Electric).

### **Анализ существующих подходов к трехмерному моделированию**

Современное трехмерное моделирование характеризуется многообразием методологических подходов, каждый из которых имеет свои преимущества и области применения. В промышленном проектировании выделяются параметрическое, прямое и гибридное моделирование, а также генеративный дизайн. Облачное моделирование и интеграция с ИИ также становятся все более популярными. В различных отраслях формируются собственные подходы к 3D-моделированию, такие как BIM-моделирование в архитектуре, поверхностное моделирование в промышленном дизайне и полигональное моделирование в анимации и игровой индустрии. Современные тенденции указывают на автоматизацию процессов моделирования, интеграцию различных подходов, повышение доступности инструментов и внедрение ИИ.

В последние годы наблюдается активное внедрение технологий искусственного интеллекта в традиционные инструменты трехмерного моделирования. Autodesk, один из лидеров индустрии, интегрировал нейросетевые технологии в Fusion 360 для автоматизации процессов проектирования и оптимизации. Система использует алгоритмы машинного обучения для генеративного дизайна, что позволяет создавать оптимизированные конструкции на основе заданных параметров и ограничений. По данным компании, этот подход позволяет сократить время проектирования на 30-50 % при одновременном улучшении характеристик конечного продукта.

Siemens NX также активно развивает направление искусственного интеллекта в своих решениях. В последних версиях программного обеспечения реализованы алгоритмы машинного обучения для предсказания потенциальных проблем в конструкции, автоматической оптимизации топологии и помощи в принятии проектных решений. Особенно эффективным оказалось применение нейросетей для анализа и оптимизации сборок, где алгоритмы способны предлагать более эффективные варианты компоновки на основе накопленного опыта.

Blender, популярный инструмент для 3D-моделирования с открытым исходным кодом, интегрировал поддержку различных нейросетевых плагинов. Наиболее заметным стало внедрение инструментов для автоматической генерации текстур, оптимизации топологии моделей и создания анимации на основе машинного обучения. Сообщество разработчиков активно развивает эти технологии, создавая новые инструменты для автоматизации различных аспектов 3D-моделирования.

В контексте непосредственно нейросетевого трехмерного моделирования сформировалось несколько основных подходов. Технология Neural Radiance Fields (NeRF), представленная исследователями из UC Berkeley, произвела революцию в создании трехмерных моделей на основе фотографий. Метод позволяет создавать детальные 3D-реконструкции объектов, используя набор двумерных изображений. Основным преимуществом данного подхода является высокая точность



воспроизведения геометрии и текстур реальных объектов, однако существенным ограничением остается необходимость большого количества исходных изображений и значительные вычислительные ресурсы для обработки.

GET3D от NVIDIA представляет другой подход к нейросетевому моделированию, позволяя генерировать трехмерные модели на основе текстовых описаний или единичных изображений. Технология демонстрирует впечатляющие результаты в создании разнообразных 3D-объектов, однако точность и детализация уступают традиционным методам моделирования. Основным преимуществом является скорость создания базовых моделей и возможность быстрого прототипирования.

Point-E от OpenAI предлагает альтернативный подход, основанный на генерации облаков точек с последующей их обработкой для создания полноценных 3D-моделей. Метод отличается высокой скоростью работы и меньшими требованиями к вычислительным ресурсам по сравнению с другими нейросетевыми подходами. Однако качество получаемых моделей может быть недостаточным для промышленного применения, что ограничивает использование технологии областью быстрого прототипирования и концептуального дизайна.

Важным аспектом развития нейросетевого 3D-моделирования становится интеграция различных подходов и создание гибридных решений. Современные исследования направлены на объединение преимуществ различных методов при одновременной минимизации их недостатков. Особое внимание уделяется разработке методов валидации и верификации получаемых результатов, что критически важно для промышленного применения этих технологий.

### **Методологические основы гибридного подхода**

В современной практике наблюдается значительный потенциал для синергии нейросетевых технологий с традиционными подходами в различных областях инженерной и научной деятельности. Опыт внедрения таких решений в ведущих технологических компаниях и исследовательских центрах позволяет оценить реальные возможности и ограничения этой интеграции.

В области программирования крупномасштабное внедрение GitHub Copilot демонстрирует практическую применимость нейросетевых технологий для автоматизации разработки программного обеспечения. Согласно исследованию GitHub в 2023 году [7], использование нейросетевых ассистентов позволяет существенно ускорить процесс написания кода, особенно в задачах, связанных с созданием типовых программных конструкций и обработкой данных. При этом важно отметить, что роль программиста трансформируется от написания рутинного кода к более высокоуровневому проектированию архитектуры и валидации генерируемых решений.

В математическом моделировании наиболее перспективным направлением становится создание гибридных систем, объединяющих классические численные методы с нейросетевыми подходами. Исследования [4] демонстрируют возможность значительного ускорения расчетов при сохранении физической корректности результатов. Особенно эффективным оказывается применение таких подходов в задачах оптимизации и прогнозирования поведения сложных систем, где традиционные методы требуют существенных вычислительных ресурсов.

Трехмерное моделирование обогащается возможностями автоматической генерации и оптимизации моделей. NVIDIA с их технологией GET3D продемонстрировала возможность создания детализированных 3D-моделей на основе текстовых описаний или изображений [8]. Это открывает новые перспективы для быстрого прототипирования и концептуального дизайна. В промышленном применении особенно важной становится возможность автоматической оптимизации существующих моделей с учетом заданных параметров и ограничений.

Интеграция этих технологий создает новые возможности для междисциплинарного взаимодействия. Например, результаты математического моделирования могут автоматически преобразовываться в трехмерные модели, которые затем оптимизируются с учетом технологических ограничений. Программный код для управления такими системами может генерироваться автоматически, учитывая специфику конкретной задачи и требования к производительности.

Особого внимания заслуживает вопрос верификации результатов, получаемых с помощью нейросетевых технологий. В программировании это решается через автоматическое тестирование и

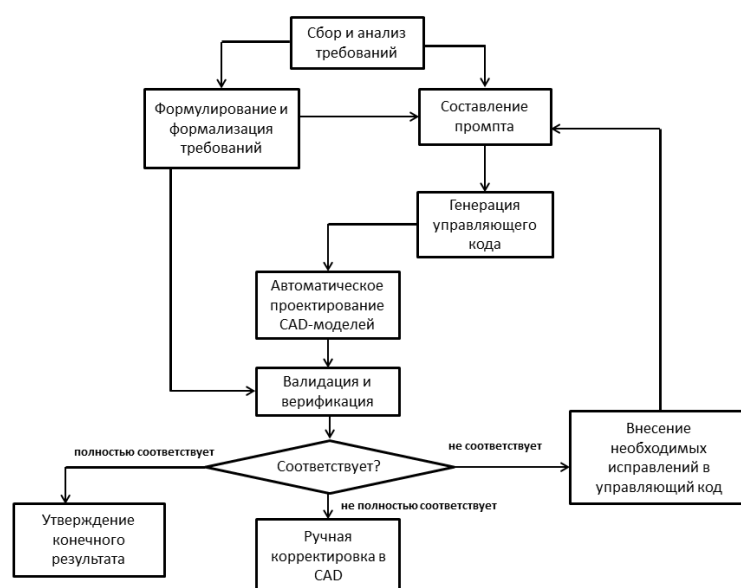
статический анализ кода. В математическом моделировании применяются методы сравнения с классическими решениями и экспериментальными данными. Для трехмерного моделирования разрабатываются специализированные методики проверки геометрической и топологической корректности генерируемых моделей.

Промышленное внедрение таких комплексных решений требует создания соответствующей инфраструктуры и методологии. Опыт компаний, успешно интегрировавших нейросетевые технологии в свои процессы, показывает необходимость поэтапного подхода с тщательной валидацией на каждом этапе. При этом ключевым фактором успеха становятся правильное определение границ применимости автоматизированных решений и сохранение контроля со стороны специалистов.

Перспективным направлением развития является создание единых платформ, интегрирующих различные аспекты проектирования и моделирования. Такие системы позволяют организовать непрерывный процесс от концептуального дизайна до готового изделия, где нейросетевые технологии играют роль интеллектуального ассистента на всех этапах работы. Это особенно важно в контексте развития цифровых двойников и умного производства.

В рамках данного исследования предлагается концепция гибридного методологического подхода [9], призванного преодолеть указанные ограничения. В основе подхода лежит синергия обработки естественного языка (NLP) и верифицированных инженерных программных комплексов. Предполагается, что сочетание этих двух подходов позволит минимизировать вероятность возникновения погрешностей и неточностей в процессе проектирования, обеспечивая при этом необходимый уровень контроля со стороны специалистов.

Предлагаемая методология базируется на интеграции возможностей систем искусственного интеллекта в области обработки естественного языка и оперативного генерирования вариативных решений с существующими алгоритмами построения CAD-моделей в отечественных системах автоматизированного проектирования, таких как КОМПАС-3D [10] и TeFlex [11].



Методология представляет собой гибридный подход к автоматизированному 3D-моделированию, сочетающий обработку естественного языка (NLP) с использованием проверенных инженерных программных комплексов (САПР), например КОМПАС-3D или TeFlex. Так называемый гибридный подход предлагает компромисс между автоматизацией и контролируемостью процесса 3D-моделирования, сочетая преимущества ИИ и проверенных инженерных инструментов. Данный подход направлен на минимизацию ошибок и повышение точности процесса моделирования по сравнению с использованием исключительно генеративных нейронных сетей. Ключевое преимущество заключается в проверке параметров скрипта, генерируемого ИИ, вместо проверки всей сгенерированной модели.

Вместо непосредственного использования нейросети для генерации 3D-модели, которое чревато скрытыми ошибками, применяется текстовый ИИ для создания управляющего скрипта на языке программирования, совместимом с выбранной САПР. Это позволяет сместить фокус контроля с проверки готовой модели на верификацию параметров, заданных в скрипте, обеспечивая более раннее выявление и исправление потенциальных ошибок. Итеративный характер процесса предполагает корректировку промпта и скрипта на основе анализа промежуточных результатов, что обеспечивает гибкость и высокую точность конечной 3D-модели.

Практическая реализация гибридных нейросетевых решений в инженерных и научных задачах требует комплексного подхода к обеспечению надежности, эффективности и контролируемости процессов. Опыт ведущих технологических компаний и исследовательских центров позволяет сформулировать ключевые требования к таким системам.

### **Перспективы развития и социально-экономические факторы**

В контексте стремительного развития ИИ и его интеграции в производственные и научные процессы особую актуальность приобретает анализ перспектив развития и социально-экономических последствий внедрения гибридных нейросетевых технологий.

Согласно прогнозам McKinsey Global Institute [12], внедрение нейросетевых технологий в инженерные и научные области приведет к существенной трансформации рынка труда в ближайшие 5-10 лет. При этом наблюдается интересный парадокс: несмотря на автоматизацию многих процессов, спрос на высококвалифицированных специалистов не только не снижается, но и возрастает. Это связано с необходимостью разработки, внедрения и контроля новых технологических решений.

В области программирования наблюдается значительное снижение порога входа в профессию благодаря инструментам вроде GitHub Copilot и подобных систем. Однако, как показывает практика ведущих технологических компаний, это не приводит к снижению требований к квалификации опытных разработчиков. Вместо этого происходит смещение фокуса их деятельности в сторону более сложных задач архитектурного проектирования, оптимизации и обеспечения качества кода.

Математическое и инженерное моделирование также претерпевает существенные изменения. Внедрение гибридных подходов позволяет решать все более сложные задачи, которые ранее были недоступны из-за вычислительных ограничений. При этом роль специалистов трансформируется от выполнения рутинных расчетов к постановке задач, выбору методологии и валидации результатов.

Особого внимания заслуживает этический аспект внедрения нейросетевых технологий. Возникают вопросы ответственности за принимаемые решения особенно в критически важных областях, таких как медицинское моделирование или проектирование ответственных инженерных конструкций. Ведущие организации, включая IEEE и ACM, активно работают над созданием этических стандартов и рекомендаций по применению AI-технологий.

Перспективы дальнейшего развития связаны с несколькими ключевыми направлениями. Во-первых, это совершенствование методов обеспечения надежности и интерпретируемости результатов работы нейросетевых компонентов. Во-вторых, развитие технологий автоматической адаптации моделей к изменяющимся условиям и требованиям. В-третьих, создание более эффективных методов интеграции экспертных знаний в процесс обучения и работы нейронных сетей.

Важным аспектом является также развитие инфраструктуры для поддержки гибридных решений. Это включает как совершенствование аппаратного обеспечения, так и создание специализированных платформ для разработки и внедрения таких систем. Ведущие технологические компании активно инвестируют в создание облачных сервисов и инструментов, упрощающих работу с гибридными системами.

### **Заключение**

Гибридный подход к применению генеративных нейросетей в математическом и трехмерном моделировании представляет собой перспективное направление, позволяющее сочетать преимущества ИИ с традиционными методами, обеспечивая более высокую точность, надежность и контролируемость результатов. Этот подход не только открывает новые возможности для

автоматизации рутинных задач и ускорения процессов проектирования, но и позволяет решать более сложные и масштабные задачи, которые ранее были недоступны из-за вычислительных ограничений или сложности ручного моделирования.

Дальнейшее развитие этого подхода требует решения ряда технологических, методологических и социально-экономических задач. Ключевыми направлениями являются разработка новых методов обеспечения надежности и интерпретируемости результатов, развитие технологий автоматической адаптации моделей к изменяющимся условиям и требованиям, а также создание более эффективных методов интеграции экспертных знаний в процесс обучения и работы нейронных сетей. Важным аспектом является также развитие инфраструктуры для поддержки гибридных решений, включая совершенствование аппаратного обеспечения и создание специализированных платформ для разработки и внедрения таких систем.

Внедрение гибридных нейросетевых технологий оказывает существенное влияние на рынок труда, требуя от специалистов новых компетенций и навыков. Образовательные учреждения должны адаптировать свои программы, чтобы подготовить специалистов, способных эффективно работать с гибридными системами и критически оценивать результаты, полученные с помощью ИИ.

В целом, гибридный подход к применению генеративных нейросетей открывает новые перспективы для развития математического и трехмерного моделирования, позволяя создавать более сложные, точные и эффективные модели, которые могут быть использованы в различных отраслях промышленности и науки. Успешное внедрение этого подхода требует комплексного подхода, включающего технологические инновации, методологические разработки и социально-экономические преобразования.

#### Список литературы

1. Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Networks. 2014. <https://doi.org/10.48550/arXiv.1406.2661>
2. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. 2020. <https://doi.org/10.48550/arXiv.2006.11239> (дата обращения 29.03.2023)
3. Rodriguez M. Research: Quantifying GitHub Copilot's impact on code quality. *GitHub Blog*. 10.10.2023. URL: <https://github.blog/news-insights/research/research-quantifying-github-copilots-impact-on-code-quality/>
4. Karniadakis G. E., Kevrekidis I. G., Lu L., Perdikaris P., Wang S., Yang L. Physics-informed machine learning. *Nature Reviews Physics*. 2021. 3(6): 422-440.
5. Recent advances on machine learning for computational fluid dynamics: A survey / H. Wang, et al. 2024 arXiv preprint arXiv:2408.12171
6. TOP500 List - June 2025. URL: [https://top500.org/lists/top500/list/2025/06/?utm\\_source=Securitylab.ru](https://top500.org/lists/top500/list/2025/06/?utm_source=Securitylab.ru)
7. Daigle K. Octoverse: The state of open source and rise of AI in 2023. *GitHub Blog*. 08.11.2023. URL: <https://github.blog/news-insights/research/the-state-of-open-source-and-ai/>
8. Isha Salian. World-Class: NVIDIA Research Builds AI Model to Populate Virtual Worlds With 3D Objects, Characters, 2022. *Blogs.nvidia*. URL: <https://blogs.nvidia.com/blog/3d-generative-ai-research-virtual-worlds/>
9. Bondareva N.A., Bondarev A.E., Andreev S.V., Ryzhova I.G. Development of a Methodology for the Application of Generative Neural Networks in Creating 3d Models. *Scientific Visualization*. 2025. 17.3: 25–34. DOI: 10.26583/sv.17.3.034.
10. KOMPAS-3 D Russian import-independent system of three-dimensional design. URL: <https://kompas.ru/> (date of access 29.04.2025).
11. T - FLEX CAD Russian engineering software for 3D design and development of design documentation. URL: <https://www.tflexcad.ru/> (date of access 04/29/2025)
12. Ellingrud K., Sanghvi S., Dandona G., Madgavkar A., Chui M., White O., Hasebe P. Generative AI and the future of work in America. 2023. URL: <https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america>

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-544-554

## Информационная плотность объектов в цифровой среде: теоретические основы и методология оценки

**Н. А. Бондарева**

Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В статье представлена теоретическая концепция оценки информационной плотности объектов в цифровой среде. Проведен анализ ограничений существующих методов количественной оценки информационного пространства, основанных преимущественно на простых метриках объема данных. Предложена концепция информационного поля объекта как совокупности всех информационных единиц, содержащих упоминания исследуемого объекта в цифровом пространстве. Рассмотрены перспективы применения методологии в информационной безопасности, анализе социальных медиа и подготовке данных для обучения нейронных сетей. Предложенный подход открывает новые возможности для комплексной оценки информационных ресурсов и может найти применение в поисковых системах, рекомендательных алгоритмах и системах анализа больших данных.

**Ключевые слова:** информационная плотность, цифровая среда, информационное пространство, анализ данных, информационная аналитика

## Information density of objects in digital environment: theoretical foundations and evaluation methodology

**N. A. Bondareva**

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

**Abstract.** The paper presents a theoretical concept for evaluating information density of objects in digital environment. An analysis of limitations of existing methods for quantitative assessment of information space, based predominantly on simple data volume metrics, has been conducted. The concept of an object's information field is proposed as an aggregate of all informational units containing mentions of the studied object in digital space.

The prospects for applying this methodology in information security, social media analysis, and data preparation for neural network training are examined. The proposed approach opens new opportunities for comprehensive evaluation of information resources and may find application in search engines, recommendation algorithms, and big data analysis systems.

**Keywords:** information density, digital environment, information space, data analysis, information analytics

### Введение

Современное цифровое пространство характеризуется экспоненциальным ростом объемов информации и качественным усложнением структуры информационных потоков. Согласно различным оценкам, ежедневно в мировом информационном пространстве генерируется свыше двух с половиной квинтиллионов байт данных [1], значительная часть которых представляет собой неструктурированную информацию в виде текстового контента, мультимедийных материалов и метайнформационных структур. В условиях такого информационного изобилия перед исследователями и практиками встает фундаментальный вопрос о разработке адекватных методов оценки информационной полноты и качественных характеристик данных, доступных относительно интересующих объектов реального мира.

Традиционные подходы к количественной оценке информационных ресурсов, опирающиеся на простые метрики объема данных, частоты упоминаний или статистических показателей встречаемости, демонстрируют существенную ограниченность при решении задач комплексной характеристики информационной ценности агрегированных данных. Существующий арсенал методов библиометрического и наукометрического анализа [2], несмотря на развитый математический аппарат для оценки научного влияния публикаций и цитирования, не предоставляет инструментария для комплексной оценки информационного окружения произвольных объектов, выходящих за рамки научно-академической сферы. Аналогично методологический арсенал анализа социальных сетей, ориентированный на оценку популярности и влиятельности узлов в сетевых структурах, не учитывает

качественные параметры информационного содержания и не решает задачу оценки уникальности информационных единиц.

Особую остроту данная проблематика приобретает в контексте интенсивного развития систем искусственного интеллекта и технологий машинного обучения. Эффективность функционирования современных нейросетевых архитектур детерминирована качественными характеристиками обучающих данных, включающими не только их количественную полноту, но и информационную достоверность, репрезентативность и отсутствие избыточности. Однако доминирующие в настоящее время подходы к формированию обучающих датасетов концентрируются преимущественно на количественных аспектах выборочных совокупностей, таких как объем выборки и сбалансированность представленности различных классов объектов, оставляя без должного внимания качественную оценку информационного насыщения данных относительно исследуемых сущностей.

Аналогичные методологические задачи возникают в смежных областях информационных технологий, включающих системы информационного поиска и извлечения знаний, рекомендательные системы, технологии автоматической обработки естественного языка и построения семантических графов знаний. Указанные системы, функциональное предназначение которых состоит в агрегации, структурировании и интеллектуальном анализе информации о реальных объектах различной природы, включающих организационные структуры, события и абстрактные концепции, испытывают потребность в разработке инструментов для адекватной оценки полноты и качественных характеристик доступной информации.

Исследовательские работы в предметной области разрешения сущностей и построения графов знаний развивают методологические подходы к связыванию и агрегации разнородной информации об объектах исследования [3, 4], однако концентрируют внимание преимущественно на технических аспектах интеграции гетерогенных данных, не предлагая решений для задачи оценки информационной плотности и качественных характеристик получаемых объектных представлений.

Научные исследования в области информационного поиска и извлечения информации выработали развитый математический аппарат для оценки релевантности документов пользовательским запросам и информационным потребностям, включающий классические метрики точности и полноты поиска, однако существующие методы не решают концептуально отличную задачу комплексной оценки информационного окружения объекта как целостного феномена цифрового пространства.

В данной работе предлагается концептуальная модель анализа информационного пространства, основанная на введении понятия плотности информационного поля объекта. Информационное поле объекта определяется как совокупность всех информационных единиц, содержащих упоминания, описания или ссылки на исследуемый объект и доступных в открытых цифровых источниках. Принципиальной характеристикой такого информационного поля является его плотность, представляющая собой интегральную меру, отражающую не только количественные параметры доступной информации, но и ее существенные качественные характеристики, включающие уникальность информационного содержания, релевантность относительно объекта исследования, авторитетность информационных источников и временную актуальность представленных данных.

Основной целью данного исследования является разработка теоретических основ концепции плотности информационного поля и формулирование методологического аппарата для ее практического применения в задачах анализа и комплексной оценки информационных ресурсов цифрового пространства. Достижение поставленной цели предполагает решение комплекса взаимосвязанных исследовательских задач, включающих формулирование строгих определений базовых понятий информационного поля объекта, плотности информационного поля и информационного кванта, разработку математической модели для расчета плотности информационного поля с учетом многомерных характеристик информационного содержания.

### **Анализ существующих подходов к оценке информационных ресурсов**

Современная научная литература содержит обширный набор исследований, посвященных различным аспектам оценки и анализа информационных ресурсов, однако существующие подходы

демонстрируют фрагментированность и ограниченную применимость к задаче комплексной оценки информационного окружения произвольных объектов реального мира.

Фундаментальные теоретические основы количественной оценки информации были заложены в классических работах Клода Шеннона по математической теории информации, где информационное содержание сообщения определяется через энтропийные характеристики источника и канала передачи. Шенноновская концепция информационной энтропии [5] предоставляет собой математический аппарат для измерения неопределенности и информационной емкости систем, однако данный подход оперирует исключительно синтаксическими характеристиками информации, не учитывая семантическое содержание сообщений и их прагматическую ценность для решения конкретных задач. Развитие данного направления в работах Андрея Колмогорова по алгоритмической теории информации [6] вводит понятие сложности строки как длины кратчайшей программы, способной ее воспроизвести, что позволяет формализовать интуитивные представления о содержательности и неслучайности информационных последовательностей.

Существенный массив современных исследований сосредоточен на проблематике оценки качества данных [7], рассматриваемой как многомерная характеристика информационных ресурсов. Концептуальные модели качества данных, развиваемые в работах исследователей данного направления, включают такие фундаментальные измерения, как точность информационного содержания, отражающая соответствие данных реальному состоянию описываемых объектов, полнота представления, характеризующая степень покрытия релевантных аспектов предметной области, согласованность данных, определяющая отсутствие внутренних противоречий в информационных структурах, актуальность сведений относительно временного контекста их использования, и релевантность информации применительно к поставленным аналитическим задачам. Указанные исследования предлагают разнообразные метрики и алгоритмические подходы к автоматической оценке качественных характеристик данных, включая статистические методы выявления аномалий и несогласованностей, техники сравнительного анализа множественных источников для верификации фактической точности, и эвристические алгоритмы оценки полноты на основе структурных особенностей предметных областей. Вместе с тем существующие исследования качества данных демонстрируют методологическую ориентацию преимущественно на структурированные информационные ресурсы, такие как реляционные базы данных и формализованные каталоги, не предлагая адекватных решений для оценки качественных характеристик неструктурированной информации, естественным образом агрегирующейся вокруг объектов в открытом цифровом пространстве.

Значительное внимание исследователей в последние десятилетия привлекают методы разрешения сущностей и построения семантических графов знаний, направленные на интеграцию и структурирование разнородной информации об объектах реального мира. Задача разрешения сущностей, состоящая в идентификации и связывании различных упоминаний одного объекта в гетерогенных информационных источниках, решается посредством комбинации методов сравнения строковых представлений, анализа контекстуального окружения упоминаний и машинного обучения на размеченных корпусах данных. Современные подходы к построению графов знаний, такие как реализованные в проектах Google Knowledge Graph, DBpedia и Wikidata [8-10], демонстрируют впечатляющие результаты в агрегации структурированной информации о миллионах сущностей различных типов, предоставляя унифицированные интерфейсы для доступа к интегрированным знаниям. Однако существующие исследования в данной области концентрируются преимущественно на технических аспектах интеграции и верификации структурированных данных, не предлагая методологических решений для оценки информационной плотности и качественных характеристик неструктурированного информационного окружения объектов, которое не может быть адекватно представлено в формате структурированных графовых отношений.

Классические исследования в области информационного поиска и извлечения информации выработали развитый математический аппарат для оценки эффективности поисковых систем и релевантности документов пользовательским запросам. Фундаментальные метрики точности и полноты поиска, а также их гармоническое среднее в виде F-меры предоставляют количественные



инструменты для оценки качества функционирования информационно-поисковых систем в контролируемых экспериментальных условиях. Развитие данного направления привело к созданию сложных моделей релевантности, учитывающих множественные факторы, включающие текстовое сходство между запросом и документом, авторитетность источников информации, временные характеристики документов, пользовательское поведение и контекстуальные особенности информационных потребностей. Современные подходы к ранжированию документов в веб-поиске, основанные на алгоритмах типа PageRank [11] и его модификациях, демонстрируют эффективность учета структурных характеристик информационного пространства, таких как паттерны гиперссылок и цитирования, для оценки авторитетности и значимости информационных ресурсов. Тем не менее существующие методы информационного поиска решают концептуально иную задачу оценки соответствия отдельных документов конкретным информационным запросам, не предлагая подходов к комплексной оценке информационного окружения объектов как целостных феноменов цифрового пространства, характеризующихся специфическими паттернами информационной агрегации и качественного разнообразия источников.

Библиометрические и наукометрические исследования развили изощренные методы оценки научного влияния и значимости публикаций, основанные на анализе паттернов цитирования и кооперативных связей между исследователями. Классические индикаторы, такие как индекс Хирша и его многочисленные модификации, предоставляют комплексные оценки продуктивности и влияния ученых, учитывающие как количественные характеристики публикационной активности, так и качественные параметры, отражаемые в частоте цитирования работ научным сообществом. Современные подходы к наукометрическому анализу включают в себя сетевые методы исследования научных коллабораций, временной анализ эволюции исследовательских направлений и междисциплинарные метрики, отражающие широту влияния научных результатов. Несмотря на методологическую развитость библиометрических подходов, их применимость ограничивается специфическим контекстом научно-академической деятельности и не распространяется на оценку информационных характеристик произвольных объектов, не связанных с формальными системами научного цитирования и экспертной оценки.

Исследования в области семантического веба и онтологического инжиниринга предлагают формализованные подходы к представлению и структурированию знаний о предметных областях, основанные на логических формализмах и стандартизированных языках описания ресурсов. Онтологические модели предоставляют выразительные средства для спецификации концептуальных структур предметных областей, включая иерархии классов объектов, свойства и отношения между сущностями, ограничения целостности и правила логического вывода [12, 13]. Технологии связанных данных, развиваемые в рамках инициативы семантического веба, демонстрируют практическую эффективность интеграции распределенных информационных ресурсов посредством стандартизированных протоколов и форматов представления данных. Методы оценки качества онтологий включают анализ логической согласованности концептуальных моделей, оценку полноты покрытия предметных областей и метрики соответствия онтологических структур реальным характеристикам описываемых феноменов. Вместе с тем онтологические подходы предполагают наличие экспертно разработанных концептуальных моделей и не решают задачу оценки качественных характеристик спонтанно формирующейся неструктурированной информации, которая не может быть адекватно описана в рамках предзаданных онтологических схем.

Анализ существующих исследовательских направлений выявляет принципиальный пробел в методологическом арсенале современных информационных технологий, связанный с отсутствием концептуальных подходов к комплексной оценке информационных полей, естественным образом формирующихся вокруг произвольных объектов в неструктурированном цифровом пространстве. Существующие методы, несмотря на их развитость в рамках специализированных предметных областей, не предоставляют адекватных инструментов для интегральной оценки качественных характеристик разнородной информации, агрегирующейся вокруг объектов различной природы в условиях информационного изобилия современного цифрового пространства. Данный методологический вакуум обуславливает необходимость разработки новых концептуальных подходов,

способных обеспечить теоретические основы для комплексной оценки информационной плотности и качественных характеристик объектно-ориентированных информационных полей.

### **Теоретические основы концепции плотности информационного поля**

Разработка концептуального аппарата для анализа информационных полей объектов требует формулирования новых теоретических конструктов, которые позволили бы преодолеть ограничения существующих подходов к оценке информационных ресурсов. Предлагаемая теоретическая модель основывается на синтезе идей классической теории информации, современных представлений о семантическом пространстве и принципов системного анализа сложных информационных структур. Центральной предпосылкой данного подхода является понимание того, что информация об объектах реального мира не существует в цифровом пространстве в виде изолированных дискретных единиц, но формирует сложные взаимосвязанные структуры, которые могут быть концептуализированы как информационные поля, обладающие специфическими топологическими и качественными характеристиками.

Концепция информационного поля объекта основывается на метафорическом переносе физических представлений о полевых структурах в область анализа информационных явлений. Подобно тому как физические поля характеризуются распределением энергии или вещества в пространстве, информационное поле объекта представляет собой распределение информационных единиц в цифровом пространстве, где каждая информационная единица обладает определенной «массой» или значимостью, а совокупность таких единиц формирует сложную топологическую структуру с различными зонами концентрации и разреженности информации. Данная концептуализация позволяет применить к анализу информационных структур математический аппарат, разработанный для изучения полевых явлений, включая понятия плотности, градиента, потока и других характеристик, описывающих пространственное распределение физических величин.

Формальное определение информационного поля объекта может быть сформулировано следующим образом: информационное поле объекта  $O$  представляет собой множество всех информационных единиц  $I = \{i_1, i_2, \dots, i_n\}$ , доступных в цифровом пространстве и содержащих прямые или косвенные упоминания, описания, ссылки или любые иные формы информационной связи с данным объектом. Математически это может быть выражено как  $IF(O) = \{i \in I \mid R(i, O) > \theta\}$ , где  $R(i, O)$  представляет функцию релевантности информационной единицы  $i$  относительно объекта  $O$ , а  $\theta$  является пороговым значением, определяющим минимальную степень релевантности для включения информационной единицы в поле объекта. Важно отметить, что границы информационного поля носят размытый характер, поскольку релевантность информации может варьироваться в широких пределах, от прямых упоминаний объекта до сложных контекстуальных ассоциаций, установление которых требует глубокого семантического анализа.

Информационное поле объекта характеризуется рядом фундаментальных свойств, определяющих его структуру и динамические характеристики. Пространственная неоднородность поля проявляется в том, что различные области цифрового пространства содержат неравномерное распределение информации об объекте, при этом некоторые источники и платформы могут аккумулировать значительные объемы релевантной информации, в то время как другие сегменты информационного пространства практически не содержат упоминаний о данном объекте. Временная динамичность представляет еще одно ключевое свойство информационных полей, поскольку они непрерывно эволюционируют под воздействием появления новых информационных источников, обновления существующих материалов, изменения популярности объекта и других факторов, влияющих на информационную активность вокруг него. Многомерность информационного поля обусловлена тем, что информация об объекте может существовать в различных форматах и на различных семантических уровнях, включая фактографические данные, аналитические материалы, мультимедийный контент, метайнформацию и контекстуальные связи.

Центральной характеристикой информационного поля является его плотность, которая представляет собой интегральную меру концентрации качественной информации об объекте в единице информационного пространства. В отличие от простых количественных метрик, таких как общий

объем данных или частота упоминаний, плотность информационного поля должна учитывать качественные параметры информационного содержания, включающие уникальность информации, ее релевантность относительно объекта исследования, авторитетность источников и временную актуальность данных. Концептуально плотность информационного поля может быть понимаема как мера информационной «массы» объекта в цифровом пространстве, отражающая не только количество доступной информации, но и ее способность формировать полное и точное представление об объекте исследования. В более общем смысле плотность информационного поля характеризует способность технологии, такой как искусственный интеллект, воссоздавать образ объекта на основе собранных данных [14], установленных паттернов и взаимосвязей.

Стоит отметить, что данная концепция представляет собой попытку отойти от простого количественного измерения информации в битах или количестве тегов, ссылок или упоминаний, так как в условиях существования современных языковых моделей особую ценность информация приобретает, когда при её интерпретации и поиске можно оперировать не просто словами, а смысловыми конструкциями и контекстами. С точки зрения количественного измерения информации, бытует традиционная установка: чем больше данных доступно, тем точнее и полнее можно воссоздать цифровой образ объекта. Однако разнородность информации, касающейся одного и того же объекта исследования в контексте разных трактовок, делает идею создать для всех нейтральную модель, ответ которой устроит каждую сторону, практически невозможной. Как говорится в отчёте NewsGuard [15-16], рост информационного охвата и подключение веб-поиска к языковым моделям и чат-ботам обернулись заметным ухудшением точности выдаваемых результатов, а во время событий, происходящих онлайн, чаты чаще усиливали ложные нарративы, подтягивая материалы из сомнительных источников и не делая разницы между авторитетными изданиями и их пропагандистскими двойниками.

Вместе с тем интегральная оценка плотности информации как истинной, так и заведомо ложной, окружающей объект, может позволить проанализировать объем источников, на которые потенциально может опираться модель, а также какой образ объекта она с большей вероятностью сформирует, какие паттерны будет использовать и какое мнение изберет наиболее приоритетным. На фоне того, что в свете развития языковых моделей в разных политических полюсах рынок, вероятно, будет подталкивать ИИ-сервисы к более ярко выраженной позиции в попытке удовлетворить ожидания своей аудитории, возникает необходимость в создании нейтральных и объективных критериев оценки объемов информации, чтобы иметь возможность интерпретировать основные направления развития цифровых образов тех или иных явлений, на которые будут ориентироваться языковые модели. В недалеком будущем развитие таких сервисов может привести к тому, что разные языковые модели, ограниченные по кардинально разнящимся принципам, будут давать диаметрально противоположные ответы на один и тот же вопрос. В свете этого собрать относительно объективную информационную картину с каждым годом будет всё труднее.

Плотность информационного поля объекта может характеризоваться количеством условных единиц (условная единица — это квант уникальной, неповторяющейся дословно информации, содержащей в себе мысль/анализ/исследование/факт/вывод, касающийся искомого объекта). Тогда условно 1000 сгенерированных мыслей, касающихся объекта, повысят плотность его информационного поля, а, к примеру, 1000 репостов (1000 упоминаний одного и того же термина без малейшего изменения формулировки) не изменят уровень плотности. В этом случае количество информации в традиционном понимании информатики, т.е. измеряемое в битах, может быть только одним из параметров, но не самым значимым.

В связи с этим можно ввести название для такой условной единицы. Инфон (*infony*, от *information* + суффикс *ony*) в первичном определении — это неповторяющиеся единицы информации, содержащие оригинальные мысли, анализ, исследования или факты об объекте.

Однако каждая условная единица (квант информации) так или иначе должна иметь свой вес или значимость. К примеру, научная статья с оригинальным исследованием объекта и случайный комментарий в социальной сети, хотя оба содержат уникальную информацию, имеют разную ценность для формирования информационного поля. Таким образом, более «тяжелые» информационные

единицы имеют более сильное влияние в информационном поле. Также важно учитывать временную составляющую. Информация имеет свойство устаревать, и её значимость может меняться со временем. Например, научная статья десятилетней давности может иметь меньший вес, чем свежее исследование, если только она не является фундаментальной для данной области.

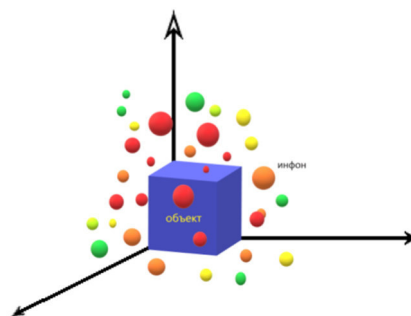
Интересным дополнением может стать концепция «информационных резонансов». Когда несколько независимых источников подтверждают одну и ту же информацию разными способами (не простое копирование), это может создавать эффект усиления, подобно тому как волны могут усиливать друг друга при резонансе. Такие резонансы могут существенно и резко повышать плотность информационного поля в определенные моменты времени. Визуализация подобных временных колебаний может существенно улучшить понимание развития того или иного явления.

Важным аспектом является также связность таких информационных квантов. Отдельные фрагменты информации, связанные между собой логическими или причинно-следственными связями, могут формировать более устойчивые и значимые структуры в информационном поле, чем разрозненные факты.

Если рассматривать инфон с точки зрения этих аспектов, то его определение уточняется: инфон — это минимальная неделимая единица уникальной информации об объекте, которая не может быть редуцирована без потери смыслового содержания. Инфон представляет собой минимальную уникальную единицу информации, содержащую законченную мысль – факт об объекте, которая не является прямым повторением существующей информации. Здесь возможно двоякое толкование в модели: можно рассматривать инфоны как единицу измерения, и тогда они должны быть одинаковыми. Из этого следует необходимость обосновать, что же является минимальной единицей уникальной информации.

Другой подход к модели допускает, что каждый инфон определяется как квант уникальной информации и может иметь свой информационный вес и релевантность (рис. 1). Инфон характеризуется атомарностью, означающей невозможность его дальнейшего деления без потери семантической и информационной целостности, уникальностью, предполагающей отсутствие точных дубликатов данной информационной единицы в других источниках, релевантностью, обеспечивающей прямую связь содержания инфона с объектом исследования, и верифицируемостью, позволяющей осуществить проверку фактической достоверности содержащейся в инфоне информации.

На рисунке приведен пример возможной визуализации такого явления, когда представленный объект окружен разнородными частицами, имеющими градиацию по цвету и размеру. Размер каждой частицы отражает размер кванта информации, цвет демонстрирует меру концептуальной связи с объектом от непосредственно относящейся к нему до имеющей весьма отдаленное отношение.



Пример визуализации плотности информационного поля объекта

Процесс выделения инфонов из общего массива информационного поля представляет собой сложную аналитическую задачу, где требуется комбинирование автоматических методов обработки естественного языка и экспертного анализа содержания информационных материалов. Алгоритмическая идентификация может основываться на методах сегментации текстового содержания с выделением семантически завершенных фрагментов, анализе новизны информации посредством сравнения с существующими базами знаний, оценке фактической уникальности через детекцию

плагиата и дублирования и определении релевантности с использованием методов семантического анализа и машинного обучения. Каждый идентифицированный инфон может быть охарактеризован набором количественных параметров, включающих меру его уникальности, степень релевантности объекту, оценку достоверности содержащейся информации и показатель информационной ценности для формирования целостного представления об объекте.

Поскольку информация в современном мире представлена в самых разных видах (визуальная, аудио, видеопоток, текст, комментарии в соцсетях, наборы численных данных), отдельной задачей становится объединение их в единую структуру. Ключ к объединению разных типов информации лежит в концепции многомерного информационного пространства. Информационное поле можно представить как многослойную структуру, где каждый тип данных формирует свой слой, но все они взаимосвязаны и влияют друг на друга. Это похоже на то, как в нейронных сетях различные модальности (текст, изображения, звук) могут быть преобразованы в единое векторное пространство.

Для унификации различных типов данных подходит концепция «информационных эмбедингов» – преобразование любого типа информации в универсальное векторное представление [17]. Современные технологии уже позволяют это делать: CLIP может находить связи между текстом и изображениями, wav2vec преобразует звук в векторы, а большие языковые модели превращают текст в многомерные представления. В этом случае важно учитывать «коэффициент информационной плотности» для разных типов данных. Например, одна секунда видео может нести больше информации об объекте, чем текстовое описание той же продолжительности, но текстовая аналитическая статья может содержать более глубокую семантическую информацию, чем простая фотография.

Также возникает вопрос «перекрестной валидации» между разными типами данных. Если информация из разных источников и форматов подтверждает друг друга, это повышает достоверность каждого отдельного информационного кванта. Например, если текстовое описание события подкрепляется видеозаписью и числовыми данными, общая достоверность информации возрастает. В контексте искусственного интеллекта такой подход открывает новые возможности для создания мультимодальных систем, способных формировать целостное представление об объектах на основе разнородных данных. Это может оказаться ещё одной системой организации данных, полезной для развития систем общего искусственного интеллекта, которые должны уметь работать с информацией целостно, как это делает человеческий мозг.

В настоящее время большие генеративные модели искусственного интеллекта склонны накапливать множество отдельных алгоритмических правил — специализированных схем для конкретных случаев, которые не объединяются в целостную систему знаний. Такие локальные закономерности часто противоречат друг другу, создавая внутренние конфликты в работе системы. Исследовательские попытки найти в структуре моделей цельные концептуальные представления [18] выявляют только разрозненные информационные фрагменты, не образующие единого связного образа. Тем не менее такие распределенные правила имеют определенную практическую ценность. Огромные параметрические возможности языковых моделей позволяют хранить подобные паттерны в больших количествах, и часто количество компенсирует отсутствие четкой структуры. Возможность создания проверяемых внутренних представлений открывает пути для борьбы с искусственными галлюцинациями, повышения надежности логических выводов и обеспечения большей прозрачности работы интеллектуальных систем.

Математическая формализация плотности информационного поля может быть представлена в виде взвешенной суммы информационных вкладов отдельных элементов поля, где каждый элемент (инфон) оценивается по множественным критериям качества. Пусть информационное поле (IF – Information Field) объекта представляет собой многомерное векторное пространство. Базовая формула могла бы выглядеть так:

$$IF = \sum (Q_i \times W_i \times T_i \times R_i) \times K_i,$$

где  $Q_i$  – некий информационный квант;  $W_i$  является весовым коэффициентом, отражающим тип и значимость информационной единицы  $i$ ;  $T(i)$  учитывает временную актуальность данных;  $R(i)$  определяет релевантность относительно объекта;  $K$  является неким нормализующим коэффициентом.

Также формула может быть дополнена такими параметрами, как  $A(i_i)$ , который отражает авторитетность источника информации, и  $U(i_i)$ , характеризующий уникальность информационного содержания.

Каждый из этих коэффициентов имеет свою методику расчета, основанную на объективных параметрах и существующих методах анализа данных. Некоторые из этих коэффициентов имеет смысл искать и заимствовать из существующих исследований по big data. Существует ряд смежных областей исследований: информационная теория поля в физике [19], семантические пространства в лингвистике [20, ссылка на Клышинского/упругие карты], цифровая экология информации и квантовая теория информации. Эти направления предоставляют полезные инструменты и методологии, которые можно адаптировать для развития теории информационных полей.

К примеру, если рассмотреть  $W_i$  (весовой коэффициент источника), то ближайшими аналогами можно определить такие сущности, как импакт-фактор научных журналов, PageRank от Google и индекс цитирования. Возможная формула расчета может быть примерно такой:

$$W_i = (A_s \times C_s \times R_s \times V_s) / N_{\max},$$

где  $A_s$  (Authority Score) является авторитетностью источника (0-1);  $C_s$  (Citation Score) отвечает за индекс цитируемости;  $R_s$  (Reliability Score) характеризует показатель надежности на основе исторических данных;  $V_s$  (Verification Score) – возможность верификации информации, а  $N_{\max}$  – некий нормализующий максимум

Также для показателя  $T_i$  можно обратиться к реально существующим примерам: научные базы данных учитывают «возраст» публикаций при расчете их значимости, а Netflix использует похожие временные коэффициенты для ранжирования контента. Метрики для оценки релевантности информации  $R_i$  используются в современных поисковых системах и системах обработки естественного языка (NLP).

Нормализующий коэффициент определяется эмпирически для конкретной предметной области и может рассчитываться как

$$K = 1 / \max(IF),$$

где  $\max(IF)$  – максимально возможное значение информационного поля в данной области.

Взаимодействие между концептами информационного поля, его плотности и составляющих инфонов формирует теоретическую модель, позволяющую осуществлять качественный анализ информационных структур в цифровом пространстве. Инфоны выступают в роли отдельных частиц информационного поля, их совокупность определяет структурные характеристики поля, а их качественные параметры детерминируют общую плотность информационного поля объекта. Данная модель предоставляет теоретические основы для разработки практических методов оценки информационных ресурсов, позволяя перейти от интуитивных представлений о «богатстве» или «бедности» информации об объекте к более строгим количественным оценкам, основанным на анализе качественных характеристик информационного содержания.

### Перспективы применения и развития теории

Предложенный подход к оценке информационной плотности в цифровой среде открывает новые возможности для практического применения и дальнейшего развития. В числе основных направлений потенциального использования методологии и перспективы её развития можно назвать такие области, как информационная безопасность, анализ социальных медиа и контента, подготовка данных для обучения нейронных сетей.

В сфере информационной безопасности методология оценки информационной плотности может найти применение в нескольких ключевых направлениях. Во-первых, анализ плотности информационного поля может позволить выявлять аномальные всплески активности, которые могут свидетельствовать о целенаправленных информационных кампаниях или атаках. Во-вторых, оценка качественных характеристик информационного пространства в перспективе поможет идентифицировать источники недостоверной информации и отслеживать распространение дезинформации. Также может оказаться актуальной задача раннего обнаружения информационных угроз путем анализа динамики изменения информационной плотности.

В контексте анализа социальных медиа предложенная методология может предоставить инструментарий для более глубокого понимания информационных процессов. Оценка информационной плотности позволяет выявлять значимые тренды и отделять их от информационного шума, что особенно важно в условиях перенасыщенности социальных сетей контентом. Анализ качественных характеристик информационного поля помогает определять реальное влияние контента и его авторов, выходя за рамки простых количественных метрик, таких как число лайков или репостов.

В свете развития данной теории можно выявить следующие направления дальнейших исследований.

1. Развитие математического аппарата представляет собой одно из ключевых направлений. Необходимы разработка более точных моделей для описания взаимодействия различных компонентов информационного поля, создание методов количественной оценки качественных характеристик информации. Особое внимание следует уделить созданию математических моделей, учитывающих временную динамику информационных процессов и нелинейный характер взаимодействия различных типов информации.

2. Экспериментальная проверка методологии требует проведения серии исследований в различных предметных областях. Необходимы подтверждение эффективности предложенных методов оценки информационной плотности на реальных данных, валидация результатов в различных контекстах применения. Важным аспектом является разработка стандартизированных методик проведения экспериментов и критериев оценки результатов.

3. Создание практических инструментов является необходимым этапом для широкого внедрения методологии. Требуется разработка программного обеспечения, способного автоматизировать процессы анализа информационной плотности, создание удобных интерфейсов для работы с данными, интеграция с существующими системами анализа информации.

Развитие предложенной методологии может оказать существенное влияние на несколько аспектов информационных технологий. В области поисковых систем появляется возможность создания более точных алгоритмов ранжирования результатов, учитывающих не только количественные, но и качественные характеристики информации. В сфере искусственного интеллекта методология может способствовать развитию более совершенных систем обработки естественного языка и анализа данных. Также подобная методология может внести вклад в развитие персонализированных рекомендательных систем, способных более точно учитывать контекст и качество информации.

## **Заключение**

Представленное исследование носит концептуальный характер и направлено на формулирование теоретических основ нового подхода к анализу информационного пространства через призму концепции плотности информационного поля. Основным результатом работы является введение системы взаимосвязанных понятий, включающих информационное поле объекта, его плотность и составляющие информационные кванты (инфоны), которые в совокупности образуют целостную концептуальную модель для качественной оценки информационных ресурсов в цифровом пространстве.

Предложенные в рамках исследования математические формализации имеют преимущественно иллюстративный характер и служат для демонстрации принципиальной возможности количественного описания качественных характеристик информационных структур. Дальнейшая разработка строгого математического аппарата потребует проведения обширных эмпирических исследований для определения конкретных параметров функций уникальности, релевантности, авторитетности и других компонентов предложенной модели плотности. Особого внимания заслуживает задача операционализации концепта инфона, требующая разработки алгоритмических процедур для автоматического выделения информационных квантов из неструктурированных текстовых массивов и их качественной оценки.

Теоретическая значимость предлагаемого подхода заключается в преодолении ограничений существующих методов оценки информационных ресурсов, которые традиционно фокусируются либо на количественных характеристиках данных, либо на узкоспециализированных аспектах



информационного качества. Практическое значение разработанной концепции определяется ее потенциальными применениями в развитии систем искусственного интеллекта нового поколения, способных формировать более надежные и верифицируемые внутренние репрезентации объектов реального мира. Переход от статистических закономерностей, извлекаемых из неконтролируемых текстовых корпусов, к систематическому анализу информационной плотности может способствовать существенному снижению частоты генерации недостоверной информации в языковых моделях и повышению прозрачности процессов принятия решений в интеллектуальных системах. Представленная концептуальная модель, таким образом, может служить отправной точкой для формирования новой исследовательской парадигмы в области анализа информационных ресурсов и их качественных характеристик.

### Список литературы

1. Austin Harris. 2.5 quintillion bytes of data are produced by people every day. 2021. URL: <https://appdeveloperomagazine.com/2.5-quintillion-bytes-of-data-are-produced-by-people-every-day/>
2. Малахов В.А. Библиометрический анализ как метод науковедческих исследований: возможности и ограничения // Науковедческие исследования. 2022. № 1. URL: <https://cyberleninka.ru/article/n/bibliometricheskii-analiz-kak-metod-naukovedcheskih-issledovaniy-vozmozhnosti-i-ogranicheniya> (дата обращения: 08.09.2025).
3. Hogan A., et al. Knowledge graphs // ACM Computing Surveys (Csur). 2021. Vol. 54, no. 4. P. 1-37.
4. Кислицына М.Ю. Analysis of the Error Structure in Identifying the Author of a Text Using the Nearest Neighbor Graphs // Научная визуализация. 2025. Т. 17.2. С. 110-122. DOI: 10.26583/sv.17.2.08
5. Shannon C.E. A Mathematical Theory of Communication (1948).
6. Kolmogorov A.N. Three approaches to the quantitative definition of information. 1965.
7. Data and information quality / C. Batini et al. Cham, Switzerland: Springer International Publishing. 2016. Vol. 63.
8. Singhal A. Introducing the Knowledge Graph: things, not strings. 2012. URL: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
9. Dbpedia: Global and Unified Access to Knowledge Graphs. URL: <https://www.dbpedia.org/>
10. Wikidata. URL: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)
11. Brin S., Page L. The anatomy of a large-scale hypertextual web search engine // Computer networks and ISDN systems. 1998. Vol. 30, no. 1-7. P. 107-117.
12. Chuprina S.I., Labutin I.A. A High-Level Adaptation Toolkit for Unified Integration of Software Systems with Neural Interfaces // Научная визуализация. 2024. Т. 16.4. С. 11-24. DOI: 10.26583/sv.16.4.02
13. Chuprina S.I. Using Data Fabric Architecture to Create Personalized Visual Analytics Systems in the Field of Digital Medicine // Научная визуализация. 2023. Т. 15.5. С. 50-63. DOI: 10.26583/sv.15.5.05
14. Bondareva N.A. The Impact of Input Data Density on the Performance of Graphic Neural Networks // Scientific Visualization. 2024. Vol. 16.5. P. 109-119. DOI: 10.26583/sv.16.5.08
15. Morrone M. Exclusive: Popular chatbots amplify misinformation. 2025. URL: [https://www.axios.com/2025/09/04/popular-chatbots-amplify-misinformation?utm\\_source=Securitylab.ru](https://www.axios.com/2025/09/04/popular-chatbots-amplify-misinformation?utm_source=Securitylab.ru)
16. AI False Information Rate Nearly Doubles in One Year. 2025. URL: <https://www.newsguardtech.com/ai-monitor/august-2025-ai-false-claim-monitor/>
17. Embedding Compression in Recommender Systems: A Survey / S. Li, H. Guo, X. Tang, R. Tang, L. Hou, R. Li, R. Zhang. 2024. arXiv:2408.02304
18. Evaluating the World Model Implicit in a Generative Model / K. Vafa, J. Y. Chen, A. Rambachan, J. Kleinberg, S. Mullainathan. 2024. <https://doi.org/10.48550/arXiv.2406.03689>
19. Torsten A. Enßlin. Information theory for fields (2018) // Annalen der Physik. 2019. Vol. 531, iss. 3. P. 1800127 DOI: 10.1002/andp.201800127
20. Визуальный анализ кластерных структур в многомерных объемах текстовой информации / А.Е. Бондарев, А.В. Бондаренко, В.А. Галактионов, Э.К. Клышинский // Научная визуализация. 2016. Т. 8, № 3. С. 1-24.

## Цифровой двойник как область научной визуализации: математическое моделирование узла ветроэнергетической установки

А. Е. Бондарев

Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

**Аннотация.** В статье представлен комплексный подход к созданию и визуализации цифрового двойника узла ветроэнергетической установки. Исследование фокусируется на интеграции математического моделирования с современными методами научной визуализации. Разработана методология, объединяющая многопараметрическое моделирование динамических процессов с их визуальным представлением в режиме реального времени. На основе дифференциальных уравнений движения и методов конечных элементов создана математическая модель, учитывающая аэродинамические и механические характеристики исследуемого узла. Представлены результаты численного моделирования с использованием современных CAE-систем, включая анализ напряженно-деформированного состояния и аэродинамических показателей. Особое внимание уделено разработке алгоритмов визуализации многомерных данных и их интерпретации.

**Ключевые слова:** цифровой двойник, научная визуализация, математическое моделирование, ветроэнергетическая установка, численные методы, CAE-системы, многопараметрическое моделирование

## Digital twin as a domain of scientific visualization: mathematical modeling of a wind turbine assembly

A. E. Bondarev

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

**Abstract.** This paper presents a comprehensive approach to the creation and visualization of a digital twin for a wind turbine assembly. The research focuses on the integration of mathematical modeling with advanced scientific visualization techniques. A methodology is developed that combines multi-parametric modeling of dynamic processes with their real-time visual representation. Based on differential equations of motion and finite element methods, a mathematical model is constructed, accounting for the aerodynamic and mechanical characteristics of the assembly under investigation. Results of numerical simulations using state-of-the-art CAE systems are presented, including analysis of stress-strain state and aerodynamic performance. Particular attention is paid to the development of algorithms for the visualization of multi-dimensional data and their interpretation.

**Keywords:** digital twin, scientific visualization, mathematical modeling, wind turbine, numerical methods, cae systems, multi-parametric modeling

### Введение

В современном мире, где цифровизация и интеллектуализация становятся ключевыми факторами конкурентоспособности в производственных отраслях, технология цифровых двойников (Digital Twin) приобретает все большее значение. Особенно актуально применение цифровых двойников на этапе разработки сложных и дорогостоящих технических систем, позволяя оптимизировать конструкцию, снизить риски и сократить сроки вывода на рынок инновационных продуктов.

Данная статья посвящена применению технологии цифровых двойников для оптимизации конструкции узла лопастей ветроэнергетической установки, находящейся на стадии разработки. В работе представлен опыт коллектива Института прикладной математики им. М.В. Келдыша РАН в создании цифрового двойника, предназначенного для проведения виртуальных испытаний и моделирования с высокой точностью.

Данное исследование, представленное в цикле работ [1-7], вносит ценный вклад в области математического моделирования, численных методов и высокопроизводительных вычислений, благодаря чему позволяет создавать цифровые двойники, адекватно отражающие физические процессы, происходящие в реальных системах. Разработанная технология позволяет проводить параметрическую оптимизацию конструкции узла лопастей, учитывая различные конструкторские ограничения и максимизируя эффективность работы ветроэнергетической установки.

Целью данной работы является демонстрация возможностей цифрового двойника, созданного коллективом ИПМ им. М.В. Келдыша РАН, для оптимизации конструкции узла лопастей

ветроэнергетической установки на этапе разработки. В работе описывается рабочий процесс и раскрывается, как применение современных методов математического моделирования и визуализации позволяет находить оптимальные геометрические параметры, обеспечивающие максимальную эффективность работы ветроэнергетической установки, а также снижать риски, связанные с внедрением новой технологии. С помощью уравнений Навье–Стокса, открытой объектно-ориентированной библиотеки OpenFOAM и современных вычислительных ресурсов, таких как гибридный вычислительный кластер K-100 в ИПМ им. М.В. Келдыша РАН, можно создать точную модель, имитирующую обтекание узла лопасти потоком воздуха. Особое внимание уделено визуализации результатов параметрического анализа, позволяющей определить оптимальную геометрию узла лопасти с учетом ограничений на массу и момент инерции.

В статье будут представлены результаты численных экспериментов, демонстрирующие эффективность разработанного подхода и перспективы его применения для проектирования других сложных технических систем.

### **Цифровой двойник: концепция и возможности**

Цифровой двойник – это виртуальная копия физического объекта или системы, которая постоянно обновляется данными, полученными от реального прототипа. Это позволяет не только отслеживать текущее состояние объекта, но и прогнозировать его поведение в различных условиях, а также оптимизировать его характеристики.

Важно отличать цифровой двойник от цифровой тени. Цифровая тень способна предсказывать поведение реального объекта только в тех условиях, в которых осуществлялся сбор данных, но не позволяет моделировать ситуации, в которых реальный объект не эксплуатировался. Цифровой двойник, напротив, позволяет моделировать широкий спектр сценариев и проводить виртуальные испытания, что особенно важно при разработке новых продуктов и оптимизации существующих.

В контексте разрабатываемых сложных и дорогостоящих установок, таких как ветроэнергетические комплексы, создание цифрового двойника становится мощным инструментом для оптимизации процесса проектирования и снижения рисков, связанных с внедрением новой технологии. На этапе разработки, когда физический прототип еще не существует или находится на ранних стадиях создания, цифровой двойник позволяет проводить виртуальные испытания и моделирование различных конструктивных решений, выявлять потенциальные проблемы и оптимизировать параметры системы до начала дорогостоящего физического производства.

Ветроэнергетические установки, даже на стадии проектирования, представляют собой сложную систему, требующую учета множества факторов, включая аэродинамику, прочность материалов, динамику нагрузок и эффективность преобразования энергии. Традиционные методы проектирования, основанные на последовательном выполнении отдельных этапов и физических испытаниях прототипов, могут быть длительными, дорогостоящими и не всегда позволяют выявить все потенциальные проблемы на ранних стадиях.

Цифровой двойник позволяет перейти к итеративному процессу проектирования, в котором виртуальные испытания и моделирование проводятся на каждом этапе разработки, позволяя быстро оценивать различные конструктивные решения и выявлять оптимальные параметры системы. Благодаря непрерывному обмену данными между цифровой моделью и результатами физических экспериментов (если таковые проводятся) цифровой двойник становится все более точным и надежным инструментом для прогнозирования поведения реальной установки. Это позволяет снизить риски, связанные с внедрением новой технологии, сократить время и затраты на разработку, а также повысить эффективность и надежность конечного продукта.

Экономическая целесообразность создания цифрового двойника на этапе разработки ветроэнергетических установок обусловлена несколькими факторами. Во-первых, снижением затрат на физические испытания и прототипированием за счет проведения виртуальных экспериментов. Во-вторых, сокращением времени разработки за счет ускорения процесса принятия решений и выявления потенциальных проблем на ранних стадиях. В-третьих, повышением эффективности и надежности конечного продукта за счет оптимизации параметров системы на основе результатов виртуальных

испытаний. В-четвертых, снижением рисков, связанных с внедрением новой технологии, за счет проведения виртуальных испытаний и моделирования различных сценариев.

Концепция цифровых двойников активно развивается в российской экономике, о чем свидетельствуют работы [8-11]. Разработана отдельная парадигма их применения в масштабах мировой экономики – Smart Digital Twin [(Simulation & Optimization)-Based Smart Big Data]-Driven Advanced (Design & Manufacturing), драйвером которой является «умный» цифровой двойник, формируемый в результате мультидисциплинарного многомасштабного численного моделирования и применения многих технологий оптимизации [8].

В контексте сложных и дорогостоящих установок, таких как ветроэнергетические комплексы, создание цифрового двойника становится не просто желательным, а практически необходимым условием для обеспечения их эффективной и безопасной эксплуатации. Ветроэнергетические установки функционируют в условиях высокой неопределенности, подвергаясь воздействию переменных ветровых нагрузок, температурных колебаний, коррозии и других факторов, которые могут привести к деградации их характеристик и возникновению аварийных ситуаций. Традиционные методы мониторинга и диагностики, основанные на периодических инспекциях и анализе исторических данных, зачастую оказываются недостаточными для своевременного выявления и предотвращения потенциальных проблем.

Создание цифрового двойника на этапе разработки ветроэнергетических установок является стратегически важным шагом, позволяющим оптимизировать процесс проектирования, снизить риски, повысить эффективность и надежность конечного продукта, а также сократить время и затраты на разработку. Эта технология открывает новые горизонты для развития ветроэнергетики и способствует созданию более эффективных и надежных энергетических систем.

### **Создание цифрового двойника узла лопастей ветроэнергетической установки**

Для эффективного создания цифрового двойника узла лопастей ветроэнергетической установки необходимо объединить усилия специалистов из разных областей: инженеров-конструкторов, технологов, материаловедов, расчетчиков и других. Это приводит к формированию нового типа инженера – системного инженера, обладающего компетенциями в различных областях знаний и способного эффективно взаимодействовать с различными инструментами моделирования и анализа.

Процесс создания цифрового двойника включает в себя несколько этапов, каждый из которых играет важную роль в обеспечении адекватности и точности виртуальной модели. Основные этапы представлены следующим списком:

1. Построение CAD-модели;
2. Построение расчетной сетки;
3. Моделирование обтекания: проводится решение задачи моделирования обтекания узла лопастей на основе уравнений Навье–Стокса с использованием специализированного программного обеспечения, такого как OpenFOAM [12-16];

4. Визуализация результатов.

Разработанная для моделирования работы и вариации формы узла лопастей модель представляет собой единую технологическую цепочку алгоритмов, включающую в себя построение CAD-модели для описания сложной геометрии узла лопастей, построение расчетной сетки на основе полученной геометрии, решение задачи моделирования обтекания на основе полной системы уравнений Навье–Стокса, а также визуализацию и анимацию результатов в моно- и стереорежимах. На основе реализованной единой технологической последовательности алгоритмов разработаны технологии вариации формы узла лопастей по ключевым параметрам для нахождения оптимальной формы узла лопастей с точки зрения силовых характеристик.

Целью данного исследования является поиск оптимальной формы узла лопастей энергетической установки с точки зрения энергонагрузки. Необходимо найти форму изделия, обеспечивающую максимальный вращающий момент, варьируя при этом три ключевых геометрических параметра узла – два угла наклона лопастей и ширину лопасти. При этом возможное увеличение массы узла и его

момента инерции не должно значительно превышать аналогичные параметры для базового варианта формы узла.

Для реализации расчетов применялась открытая объектно-ориентированная библиотека OpenFOAM, написанная на языке C++, поддерживающая механизмы массивного распараллеливания и предназначенная для численного моделирования задач механики сплошной среды [12-14]. Библиотека основана на конечно-объемных аппроксимациях, написанных в операторной форме. Элементы OpenFOAM активно используются в промышленности, в академической сфере и экспертном сообществе, в частности при численном анализе энергетических характеристик установок горизонтального типа [14-16]. Все расчеты по решению задач математического моделирования обтекания энергоустановки проводились на гибридном вычислительном кластере K-100 в ИПМ им. М.В. Келдыша РАН [17] в режиме параллельных вычислений. В качестве математической модели использовалась полная система уравнений Навье–Стокса, описывающая движение вязкого теплопроводного сжимаемого газа [18]. Для анализа результатов была построена методика анимированного визуального представления работы узла лопастей в моно- и стереорежимах [3], основанная на современных методах и концепциях визуального отображения результатов численных расчетов.

В результате ранее проведенных работ была построена компьютерная технология, позволяющая проводить математическое моделирование узла лопастей энергоагрегата с конкретно заданными геометрическими параметрами и определять силовую нагрузку, действующую на агрегат в потоке воздуха. Тем самым была создана основа для дальнейших исследований по определению оптимальной формы узла лопастей с точки зрения ветровой нагрузки с учетом технологических ограничений.

Построение работоспособной и эффективной численной технологии включает в себя несколько необходимых этапов, к которым относятся ввод геометрической информации, ее обработка, построение сеточной модели, выбор специализированных алгоритмов расчета обтекания, анализ текущих результатов, расчет функционалов (сил и моментов), визуальное представление результатов, обработка и анализ результатов многовариантных расчетов. Основой построения технологии является процесс математического моделирования узла лопастей ветроустановки с конкретно заданными геометрическими параметрами под определенной ветровой нагрузкой. Сам процесс математического моделирования предполагает создание экспериментального расчетного комплекса для расчета обтекания полной компоновки ветроустановки на основе решения уравнений Навье–Стокса и вычисление соответствующих интегральных характеристик потока.

Следующим важным этапом является создание методики проведения оптимизационных расчетов для выбора оптимальной формы ветроагрегата на основе оптимизации величин выбранных интегральных характеристик.

Процесс построения сетки начинается с описания поверхности, ограничивающей рассматриваемое трехмерное тело. Как правило, поверхность реального трехмерного объекта сложной формы, полностью или частично импортируется из CAD-пакета и в случае необходимости может быть отредактирована. К поверхностным сеткам предъявляются довольно «очевидные» требования – отсутствие самопересечений, замкнутость и некоторые другие. При работе с реальными промышленными объектами получаемые CAD-поверхности удовлетворяют названным критериям в редких случаях и для получения объемной расчетной сетки нужного качества необходима дополнительная подготовка поверхности.

Для достижения указанной цели используется ряд автоматических инструментов с широким набором средств контроля качества, такие инструменты создания и редактирования поверхностных сеток, как surface wrapping и remeshing, что позволяет на порядки сократить затрачиваемое время и практически полностью исключить подготовку сетки «вручную». Surface wrapper исправляет дефекты CAD-геометрии (закрывает отверстия, устраняет самопересечения и т. д.), что позволяет получить на выходе замкнутую поверхность с требуемой степенью детализации. Для создания начальной треугольной поверхностной сетки высокого качества применяется surface remesher, который позволяет с помощью достаточно широкого набора настроек получить поверхность, обладающую необходимыми параметрами (степень гладкости в областях с высокой кривизной, степень разрешения тонких мест,

скорость роста характерного размера поверхностных ячеек при удалении от областей с высокой детализацией, сохранение топологических особенностей, локальное измельчение сетки и т. д.).

Рисунок 1 представляет уже «сетку без ошибок» – треугольную поверхностную сетку, обладающую необходимыми параметрами (степень гладкости в областях с высокой кривизной, уровень разрешения тонких мест, скорость роста характерного размера поверхностных ячеек при удалении от областей с высокой детализацией, сохранение топологических особенностей, локальное измельчение сетки и т.д.).

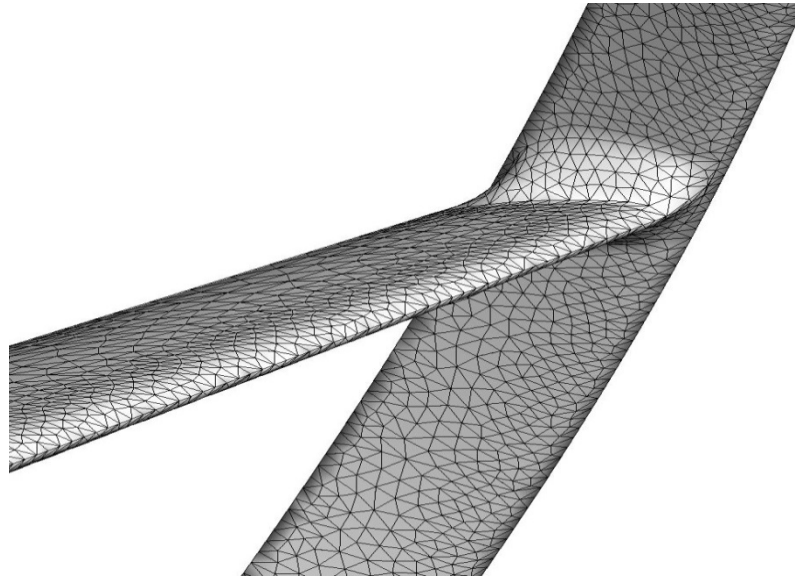


Рисунок 1. Исправленная поверхностная треугольная сетка, построенная для исходной CAD-поверхности

Результирующая поверхностная сетка является основой для построения объемной расчетной сетки. Особенности численного моделирования течений жидкостей и газов диктуют некоторые правила построения объемных сеток. При решении таких задач объемная сетка, как правило, состоит из двух основных частей: призматической вблизи обтекаемых поверхностей и произвольной многогранной на достаточном удалении от поверхностей.

В настоящее время течения газовой сплошной среды (т.е. газовой среды при предположении, что ее можно рассматривать, не учитывая индивидуальных частиц) рассчитываются на основе уравнений Навье–Стокса. Указанная система уравнений выглядит следующим образом (см., например, [18]):

$$\begin{aligned}\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{u}) &= 0, \\ \frac{\partial (\rho \vec{u})}{\partial t} + \nabla \cdot (\rho \vec{u} \otimes \vec{u}) &= -\nabla p + \nabla \cdot (\tau_m + \tau_t), \\ \frac{\partial (\rho E)}{\partial t} + \nabla \cdot (\rho \vec{u} H) &= \nabla \cdot [\vec{u} \cdot (\tau_m + \tau_t) + (\vec{q}_m + \vec{q}_t)],\end{aligned}$$

где  $\vec{u}$  – вектор скорости осредненного течения с компонентами  $(u, v, w)$ ;  $\tau_m$  и  $\tau_t$  – молекулярные и турбулентные (полученные путем осреднения различных функционалов от мелкомасштабных пульсаций) компоненты тензора вязких напряжений;  $E = e + |\vec{u}|^2/2$  – удельная полная энергия газа;  $e$  – удельная внутренняя энергия газа;  $H = E + p/\rho$  – полная энтальпия;  $\vec{q}_m$  и  $\vec{q}_t$  – молекулярные и турбулентные компоненты векторов плотности теплового потока.

Вид оставшихся турбулентных компонент уже не является универсальным, их выбор представляет собой так называемые модели турбулентности. Модели турбулентности необходимо выбирать, учитывая свойства реальных физических течений при выбранном диапазоне параметров. Заметим, что в проведенных расчетах влияние турбулентности на физические характеристики всего процесса не рассматривалось.

Для решения задачи определения оптимальной формы геометрии ветроустановки с точки зрения силовой нагрузки построен ряд моделей узла лопастей с различной геометрией исследуемого объекта.

Нами выбраны три основных параметра, варьирование которых позволяет описать большое многообразие возможных геометрических форм.

На рисунке 2 схематически показаны эти изменяемые параметры:  $\gamma_1$  – угол между направлением основной лопасти и вертикалью;  $\gamma_2$  – угловой размер основной лопасти в направлении оси вращения;  $L$  – ширина основной лопасти. Заметим, что всюду ниже мы будем приводить результаты для тройки параметров  $(\gamma'_1, \gamma_2, L)$ , где  $\gamma'_1 = 90^\circ - \gamma_1$ .

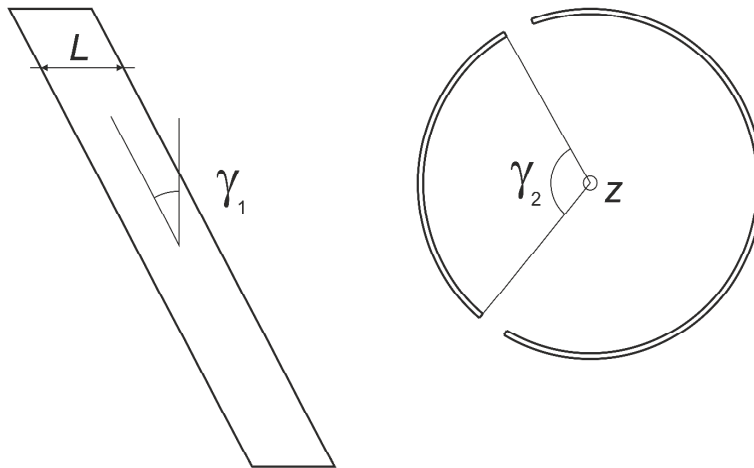


Рисунок 2. Варьируемые геометрические параметры  $(\gamma_1, \gamma_2, L)$  для определения оптимальной формы геометрии узла лопастей

Для определения основных параметров и размеров (расстояния между опорами, усредненная форма сечений лопастей и пр.) изделия использовалась геометрия ветроэнергоустановки, полученная в результате сканирования прототипного опытного образца, а построение непосредственно поверхности и ее модификации осуществлялось средствами CAD-пакета SolidWorks. Все построенные модели экспортировались в формате SLDPRТ.

Значения  $(\gamma'_1, \gamma_2, L)$  меняются в некотором диапазоне, при этом в качестве базового  $(\gamma'_1)_{\text{баз}}, (\gamma_2)_{\text{баз}}, L_{\text{баз}}$  выбран результат лазерного сканирования опытного образца:  $(\gamma'_1)_{\text{баз}} = 55$  градусов,  $(\gamma_2)_{\text{баз}} = 120$  градусов,  $L_{\text{баз}} \sim 20$  см.  $\gamma'_1$  принимает значения 45, 50, 55, 60 и 65 градусов,  $\gamma_2$  — значения 100, 110, 120 и 130 градусов,  $L$  выбирается из набора величин  $0.8 L_{\text{баз}} (L - 20\%)$ ,  $0.9 L_{\text{баз}} (L - 10\%)$ ,  $L_{\text{баз}}$ ,  $1.1 L_{\text{баз}} (L + 10\%)$ ,  $1.2 L_{\text{баз}} (L + 20\%)$ .

Дополним для наглядности схематическую картину, представленную на рисунке 2, трехмерными изображениями (см. рис. 3-4). Рисунки представляют изменение формы узла лопастей при вариации вышеописанных геометрических параметров.

Формальная общая постановка задачи оптимизации выглядит следующим образом: найти среди элементов  $x$ , образующих множество  $X$ , такой элемент  $x^*$ , на котором заданная функция  $f(x)$  достигает минимального (или максимального) значения, т.е.  $f(x^*) = \min_{x \in X} f(x)$  (или  $f(x^*) = \max_{x \in X} f(x)$  соответственно). Следовательно, для того чтобы поставить задачу оптимизации, необходимо задать допустимое множество  $X = \{\vec{x} | g_i(\vec{x}) \leq 0, i = 1, \dots, m\} \subset R^n$ , целевую функцию  $f(x): X \rightarrow R^n$  и критерий поиска – что мы ищем – (max или min).

Решить подобную задачу означает либо найти искомый экстремум  $f(x^*) = \min_{x \in X} f(x)$ , либо показать, что решения не существует. Если при задании допустимого множества  $X$  ограничения  $g_i(\vec{x})$  отсутствуют, то мы имеем дело с задачей безусловной оптимизации. Если условия  $g_i(\vec{x})$  существуют, то подобную задачу называют задачей условной оптимизации.

Предполагая возможность ряда конструкторских ограничений по сути рассматриваемой задачи, мы формально имеем дело с задачей условной многопараметрической оптимизации, которая характерна для задач конструкторской оптимизации вообще. Наш случай является многопараметрической оптимизационной задачей, так как рассматриваются вариации трех ключевых геометрических параметров. Вариации этих параметров ограничены диапазонами и, следовательно, накладывают



ограничения на варьируемые параметры. В целом задачу разведочного параметрического поиска можно сформулировать следующим образом: найти значения ключевых геометрических параметров узла лопастей  $x^* = ((\gamma_1')^*, \gamma_2^*, L^*)$ , обеспечивающие максимальное значение целевой функции  $M^* = \max M(\gamma_1', \gamma_2, L)$ . В качестве целевой функции выбирались основные аэродинамические характеристики: полная аэродинамическая сила  $\vec{F}$  и вращающий момент  $\vec{M}$ .

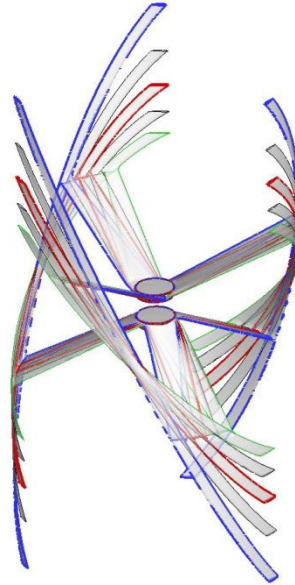


Рисунок 3. Изменение углового параметра  $\gamma_1'$ : 45 (зеленый), 50 (серый), 55 (базовый вариант, красный), 60 (серый) и 65 (синий) градусов.  $\gamma_2 = (\gamma_2)_{\text{баз}}$ ,  $L = L_{\text{баз}}$

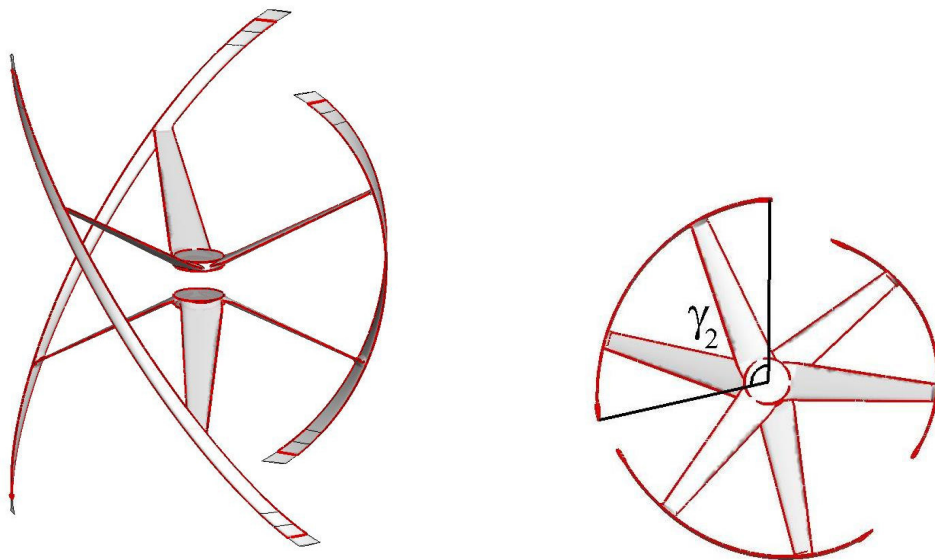


Рисунок 4. Изменение углового параметра  $\gamma_2$ : 100, 110, 120 (базовый вариант, красный) и 130 градусов. При этом параметры  $\gamma_1' = (\gamma_1')_{\text{баз}}$ ,  $L = L_{\text{баз}}$

Для решения оптимизационной задачи предполагается использовать сеточный метод, что является целесообразным для первоначальной оценки, поскольку многомерные задачи являются значительно более сложными и трудоемкими, чем одномерные. Суть предлагаемого метода поиска наименьшего значения заключается в определении значений целевой функции в дискретном наборе узлов, не выходящих за рамки допустимого множества  $X$ . Иными словами, проводится разбиение с определенным шагом диапазонов изменения для каждого из ключевых геометрических параметров.

Таким образом, область пространства, образуемая диапазонами изменения параметров, покрывается сеткой. В каждом из узлов сетки вычисляется целевая функция. Наибольшее из множества

значений целевой функции на данной сетке принимается за максимум. Ранее традиционно считалось, что для задач большей размерности данный метод практически непригоден из-за большого времени, необходимого для расчетов. Однако развитие параллельных вычислений позволяет ускорять процессы расчетов на порядки. Это обстоятельство делает самые непривлекательные и простые методы реально применимыми к практическим задачам. Более того, их простота и надежность создают им в этом случае значительные преимущества.

Проведенные массовые численные расчеты позволяют осуществить разведочный поиск и грубую оптимизационную прикидку оптимальной совокупности ключевых геометрических параметров с точки зрения максимизации вращающего момента узла лопастей. Для этого каждый из ключевых параметров  $\gamma'_1$ ,  $\gamma_2$ ,  $L$  последовательно варьировался. При этом два оставшихся параметра фиксировались.

Анализ полученных в расчетах значений объема узла лопастей и его момента инерции для различных вариантов геометрии узла показывает, что вариации угловых параметров и вариация ширины лопасти оказывают разное влияние на геометрические характеристики – объем и момент инерции узла. В частности, вариация ширины ведет к более значительному увеличению объема и момента инерции, что лишает смысла учет ограничений при совместной вариации угловых параметров и ширины лопасти. Поэтому было принято решение проводить учет конструкторских ограничений при вариации угловых параметров для каждого слоя данных, соответствующего заданной ширине лопасти, раздельно.

Ниже представлены результаты по учету конструкторских ограничений согласно ранее построенной методике [7] для вариации угловых параметров  $\gamma'_1$  и  $\gamma_2$  при заданной ширине лопасти  $L+20\%$ .

На рисунке 5 рассчитанные объемы представляются в виде трехмерной поверхности  $V_i(\gamma'_1, \gamma_2)$  в зависимости от вариаций основных углов. Плоскость, ограниченная красным, соответствует значению объема для базового варианта. Соответственно, плоскость, ограниченная синим, соответствует значению объема для базового варианта, увеличенному на 10 %. Линии пересечения обеих плоскостей с поверхностью значений объемов ограничивают вариацию формы изделия. На рисунке 6 область допустимых значений при учете ограничений по объему представлена на плоскости вариации угловых параметров и заключена в области, ограниченной толстыми красной и синей линиями.

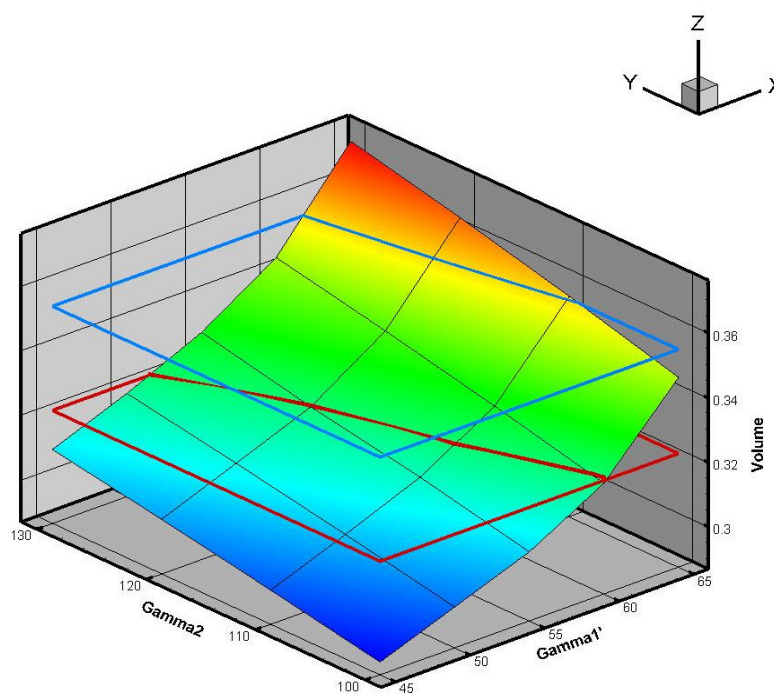


Рисунок 5. Организация учета ограничений по объему для слоя данных  $L+20\%$

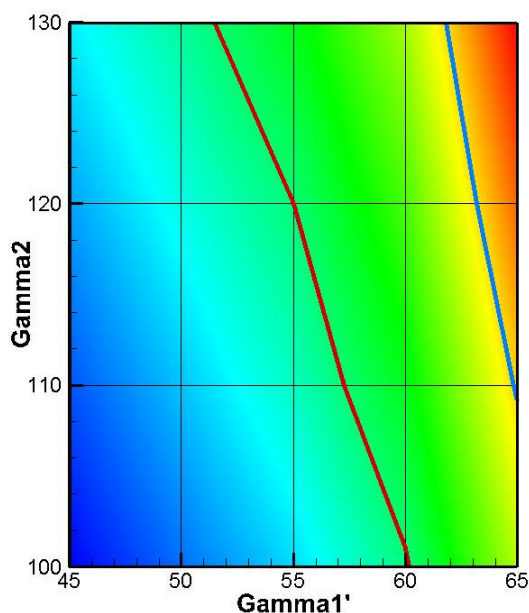


Рисунок 6. Область ограничения по объему при вариации углов для слоя данных L+20 %

Аналогичным образом определяются ограничения для момента инерции. Они показаны на рисунках 7 и 8. На рисунке 7 рассчитанные моменты инерции представляются в виде трехмерной поверхности  $M_i(\gamma_1', \gamma_2)$  в зависимости от вариаций основных углов. Плоскость, ограниченная красным, соответствует значению момента инерции для базового варианта. Соответственно, плоскость, ограниченная синим, соответствует значению момента инерции для базового варианта, увеличенному на 10 %. Линии пересечения обеих плоскостей с поверхностью значений моментов инерции ограничивают вариацию формы изделия. На рисунке 8 диапазон допустимых значений представлен на плоскости вариации угловых параметров в виде области между толстыми красной и синей линиями.

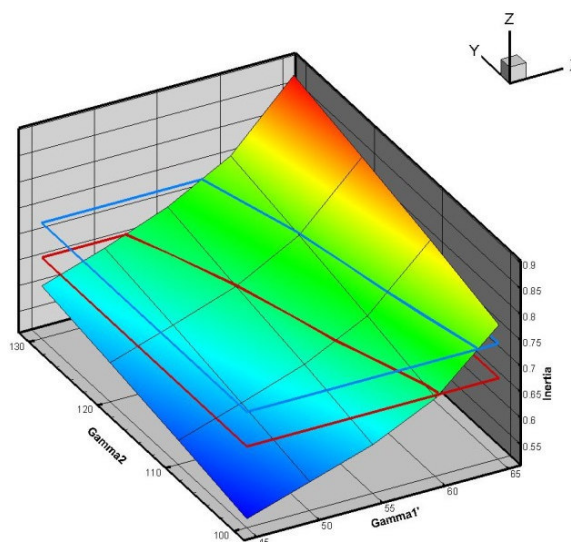


Рисунок 7. Организация учета ограничений по моменту инерции для слоя данных L+20 %

Таким образом, для заданного слоя данных, соответствующего ширине лопасти L+20 %, мы получили ограничения по объему и моменту инерции, представленные на рисунках 6 и 8. Теперь надо организовать их совместный учет. Для этой цели их следует совместить на одном изображении и выбрать диапазон вариации углов, соответствующий наиболее жесткому ограничению.

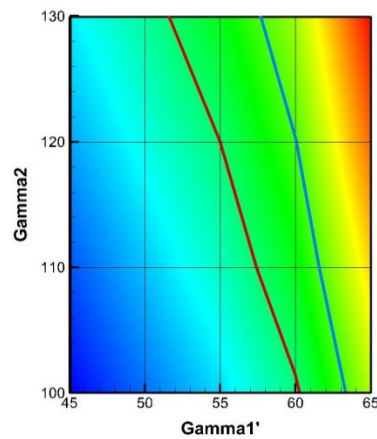


Рисунок 8. Область ограничения по моменту инерции при вариации углов для слоя данных L+20 %

Для более наглядного представления рассмотрим рисунок 9. Это представление дает нам возможность более точно определить диапазон вариации угловых параметров, обеспечивающий максимизацию вращающего момента с учетом наложенных ограничений на изменение объема и момента инерции узла лопастей.

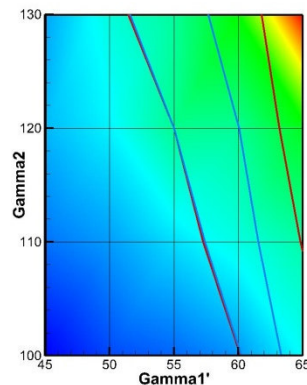


Рисунок 9. Двумерное представление распределения момента инерции с учетом ограничений при вариации угловых параметров

Рисунок 9 дает точное представление об искомом диапазоне вариации угловых параметров. Можно утверждать, что при учете выбранных ограничений по объему и моменту инерции нужный диапазон заключается в пределах вариации угла  $\gamma_1'$  от 55 до 60, а угла  $\gamma_2$  — от 120 до 125.

### **Перспективы развития: от специализированного решения к парадигме цифрового инжиниринга**

Технология создания цифровых двойников, разработанная и апробированная коллективом Института прикладной математики им. М.В. Келдыша РАН на примере узла лопастей ветроэнергетической установки, представляет собой не только эффективное решение для оптимизации конкретной конструкции, но и потенциально трансформирующую парадигму в области цифрового инжиниринга. Достигнутые результаты свидетельствуют о возможности создания масштабируемой и адаптивной платформы, способной интегрировать различные методы математического моделирования, численного анализа и визуализации данных для решения широкого круга задач в различных отраслях промышленности.

Дальнейшее развитие технологии предполагает углубленное исследование и расширение функциональных возможностей платформы в следующих направлениях:

1. Развитие методов адаптивного моделирования: необходимо разрабатывать методы адаптивного моделирования, позволяющие автоматически изменять уровень детализации модели в зависимости от

решаемой задачи и доступных вычислительных ресурсов. Это позволит оптимизировать процесс моделирования и анализа, обеспечивая требуемую точность результатов при минимальных затратах;

2. Исследование методов верификации и валидации моделей: Строгие методы верификации и валидации моделей, основанные на сравнении результатов моделирования с экспериментальными данными и результатами других независимых расчетов, позволят повысить доверие к результатам моделирования и обеспечить надежность принимаемых решений;

3. Разработка инструментов для автоматической генерации моделей: инструменты для автоматической генерации моделей на основе данных о геометрии, материалах и условиях эксплуатации системы дают потенциальные перспективы для сокращения время и затрат на создание цифровых двойников и позволяют сделать их доступными для широкого круга пользователей.

Разработанная технология может быть применена для решения широкого круга задач в различных отраслях промышленности, включая:

- проектирование и оптимизация сложных технических систем: разработка цифровых двойников для самолетов, автомобилей, энергетических установок и других сложных технических систем позволит оптимизировать их конструкцию, повысить эффективность и надежность, а также снизить затраты на разработку и эксплуатацию;
- мониторинг и диагностика технического состояния: разработка цифровых двойников для мониторинга и диагностики технического состояния оборудования позволит своевременно выявлять дефекты и предотвращать аварии, а также оптимизировать графики технического обслуживания и ремонта;
- управление сложными технологическими процессами: разработка цифровых двойников для управления сложными технологическими процессами позволит оптимизировать режимы работы оборудования, повысить эффективность производства и снизить затраты на энергию и материалы;
- разработка новых материалов и технологий: разработка цифровых двойников для моделирования свойств и поведения новых материалов и технологий позволит ускорить процесс их разработки и внедрения в производство.

Реализация этих перспективных направлений развития потребует консолидации усилий научного сообщества, промышленных предприятий и государственных органов, а также привлечения значительных инвестиций в исследования и разработки. Однако результаты этих работ позволят создать мощную платформу для цифрового инжиниринга, которая будет способствовать повышению конкурентоспособности российской промышленности и укреплению позиций России на мировом рынке высоких технологий.

## Заключение

В настоящей работе представлен опыт коллектива Института прикладной математики им. М.В. Келдыша РАН в разработке и применении технологии цифровых двойников для оптимизации конструкции узла лопастей ветроэнергетической установки. Показано, что создание цифрового двойника позволяет проводить виртуальные испытания и моделирование с высокой точностью, сокращая время и затраты на разработку и тестирование, а также находить оптимальные геометрические параметры, обеспечивающие максимальную эффективность работы установки. Разработанная технология представляет собой эффективное решение, готовое к масштабированию и применению в широком спектре инженерных задач. Она может быть использована для проектирования и оптимизации сложных технических систем в различных отраслях промышленности, а также для мониторинга и диагностики их технического состояния.

Перспективы дальнейшего развития технологии связаны с расширением функциональности платформы, разработкой специализированных библиотек моделей, интеграцией с системами автоматизированного проектирования и применением методов машинного обучения и искусственного интеллекта. Реализация этих перспективных направлений позволит создать мощную платформу для цифрового инжиниринга, которая будет способствовать повышению конкурентоспособности российской промышленности.

В заключение следует отметить, что технология цифровых двойников является одним из ключевых направлений развития современной науки и техники. Ее применение позволяет решать сложные инженерные задачи, создавать инновационные продукты и повышать эффективность производства. Дальнейшие исследования и разработки в этой области будут способствовать развитию российской экономики и повышению качества жизни населения.

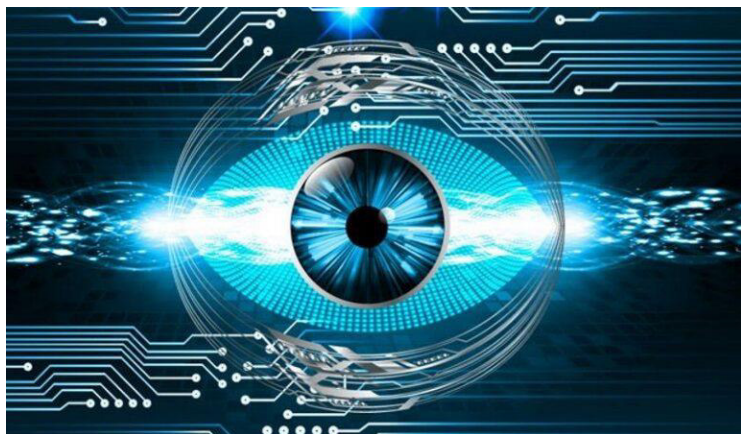
### Благодарности

Вычисления проведены с помощью гибридного суперкомпьютера K100, установленного в Суперкомпьютерном центре коллективного пользования ИПМ им. М.В. Келдыша РАН.

### Список литературы

1. Разработка и организация математического моделирования обтекания неподвижной лопатки энергетической установки / А.Е. Бондарев, В.Т. Жуков, К.В. Мануковский, Н.Д. Новикова, О.Б. Феодоритова // Препринты ИПМ им. М.В. Келдыша РАН. 2014. № 60. 19 с. URL: <http://library.keldysh.ru/preprint.asp?id=2014-60>.
2. Моделирование и визуализация работы энергетической установки сложной формы в режиме стереоанимации / С.В. Андреев, А.Е. Бондарев, А.В. Бондаренко и др. // Труды 25-й Международной конференции по компьютерной графике и зрению ГрафиКон'2015, Протвино, Россия, 22-25 сентября 2015 г. С. 183-187.
3. Моделирование и визуализация работы узла лопастей сложной формы в энергетической установке / С.В. Андреев, А.Е. Бондарев, А.В. Бондаренко и др. // Научная визуализация. 2015. Т. 7, № 4. С. 1-12. URL: <http://sv-journal.org/2015-4/01.php?lang=ru>
4. Моделирование и визуализация работы энергетической установки сложной формы / С.В. Андреев, А.Е. Бондарев, А.В. Бондаренко и др. // Математическое моделирование. 2016.
5. Разработка методики компьютерного расчета лопастей ветроэнергетической установки (ВЭУ) и создание численной модели ветроагрегата / В.А. Галактионов, А.Е. Бондарев, В.Т. Жуков и др. // Отчет ИПМ им. Келдыша РАН. 2015. № 12/15. 58 с.
6. Решение оптимизационной задачи с целью нахождения оптимальной формы конструкции изделия с точки зрения силовой нагрузки / В.А. Галактионов, А.Е. Бондарев, В.Т. Жуков и др. // Отчет ИПМ им. Келдыша РАН. 2015. № 25/15. 49 с.
7. Реализация уточненного решения задачи определения оптимальной формы узла лопастей и определение допустимого диапазона изменения геометрических параметров узла / В.А. Галактионов, А.Е. Бондарев, В.Т. Жуков и др. // Отчет ИПМ им. Келдыша РАН. 2016. № 2/16. 27 с.
8. Боровков А. И., Рябов Ю. А., Марусева В. М. Новая парадигма цифрового проектирования и моделирования глобально конкурентоспособной продукции нового поколения // Цифровое производство. Методы, экосистемы, технологии. М.: Департамент корпоративного обучения Московской школы управления Сколково, 2018. С. 5, 24-43.
9. Мировая технологическая повестка и глобальные тенденции развития промышленности в условиях цифровой экономики / А.И. Боровков, Л.А. Щербина, В.М. Марусева, Ю.А. Рябов // Инновации. 2018. № 12 (242). С. 33-42.
10. Боровков А.И., Марусева В.М., Рябов Ю.А. «Умные» цифровые двойники — основа новой парадигмы цифрового проектирования и моделирования глобально конкурентоспособной продукции нового поколения // Трамплин к успеху. 2018. № 13. С. 12–16. URL: [http://assets.fea.ru/uploads/fea/news/2018/04\\_april/12/tramplin-uspeha\\_13-16.pdf](http://assets.fea.ru/uploads/fea/news/2018/04_april/12/tramplin-uspeha_13-16.pdf)
11. Боровков А.И. Цифровые двойники: определение, подходы и методы разработки // Цифровая трансформация экономики и промышленности: сборник трудов научно-практической конференции с зарубежным участием, Санкт-Петербург, 20–22 июня 2019 года / под редакцией А.В. Бабкина. Санкт-Петербург: Санкт-Петербургский политехнический университет Петра Великого, 2019. С. 234-245. DOI: 10.18720/IEP/2019.3/25.
12. Maric T., Hopken J., Mooney K. The OpenFOAM technology primer. [www.sourceflux.de/book](http://www.sourceflux.de/book), 2014.
13. OpenFOAM. URL: <http://www.openfoam.com>
14. First Symposium on OpenFOAM® in Wind Energy. 20-21 March 2013. URL: <http://www.forwind.de/sowe/Site/Home.html>
15. Second Symposium on OpenFOAM® in Wind Energy. 19-21 May 2014. URL: <http://wind.nrel.gov/2ndSOWE/>
16. Third Symposium on OpenFOAM® in Wind Energy. 15–17 June 2015. URL: <https://www.eko.polimi.it/index.php/sowe2015/SOWE2015>
17. Гибридный вычислительный кластер K-100. URL: <http://www.kiam.ru/MVS/resourses/k100.html>
18. Ландау Л.Д., Лившиц Е.М. Гидродинамика. Москва: Наука, 1986.





Секция 3

# ***КОМПЬЮТЕРНОЕ ЗРЕНИЕ***

---

COMPUTER VISION



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия



УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-568-574

## Способ оптико-информационного обеспечения обнаружения артефактов на сложном фоне роботизированным комплексом с элементами нейронных сетей

**Р. В. Поляков**

Филиал НИУ «МЭИ» в г. Смоленске, Смоленск, Россия

**Аннотация.** Работа посвящена разработке способа обнаружения артефактов на сложном атмосферном фоне с использованием инфракрасных оптических систем и нейросетевых алгоритмов. Основное внимание уделено оптимизации параметров сегментации видеопотока для повышения точности обнаружения при сохранении приемлемой вычислительной нагрузки.

**Ключевые слова:** компьютерное зрение, инфракрасная разведка, оптико-электронная система, обнаружение и распознавание БПЛА, нейронные сети, сегментация видеопотока.

## Method of optical-information support for artifact detection on complex backgrounds using a robotic system with neural network elements

**R. V. Polyakov**

Branch of NIU "MEI" in Smolensk, Smolensk, Russia

**Abstract.** The work is devoted to the development of a method for detecting artifacts on a complex atmospheric background using infrared optical systems and neural network algorithms. The main focus is on optimizing video stream segmentation parameters to improve detection accuracy while maintaining acceptable computational load.

**Keywords:** computer vision, infrared reconnaissance, optoelectronic system, UAV detection and recognition, neural networks, video stream segmentation

### Введение

Под артефактом следует рассматривать беспилотный летательный аппарат (БПЛА). В последние годы значительное внимание уделяется проблеме их обнаружения и распознавания. Особую сложность представляют малогабаритные БПЛА, обладающие низкой эффективной площадью рассеяния и способные двигаться по сложным траекториям на малых высотах. В таких условиях традиционные радиолокационные методы демонстрируют ограниченную эффективность [1-4].

Альтернативным решением является применение инфракрасных (ИК) оптико-электронных систем. По сравнению с радиолокационными методами, ИК-детектирование обеспечивает пассивный и скрытный режим наблюдения. Кроме того, в отличие от систем видимого диапазона, ИК-системы способны функционировать круглосуточно, сохраняя работоспособность при недостаточной освещённости.

Тем не менее, эффективность существующих неавтоматизированных оптико-электронных средств ИК-диапазона, не оснащённых алгоритмами автоматического обнаружения и распознавания, остается удовлетворительной. Основные трудности связаны с выявлением малоразмерных целей на сложном атмосферном фоне (например, на фоне кучевых облаков, разрывов облачности и других неоднородностей). Дополнительным фактором, снижающим эффективность, является утомление операторов при длительном визуальном поиске.

В работе [5] предложено использование ИК-систем в сочетании с нейросетевыми алгоритмами компьютерного зрения. Предварительно обученная нейросетевая модель на основе архитектуры YOLOv11 показала высокую точность в задачах обнаружения и классификации БПЛА, однако её эффективность снижалась при работе с малоразмерными объектами на неоднородном атмосферном фоне, характеризующемся наличием облачности различных типов, разрывов и значительных градиентов водности.

Для повышения точности обнаружения было предложено сегментирование видеопотока: вместо обработки целых кадров (рис. 1) использовались их части (рис. 2), размер которых соответствовал масштабу неоднородностей фона.



Рис. 1. Исходный кадр видеопотока

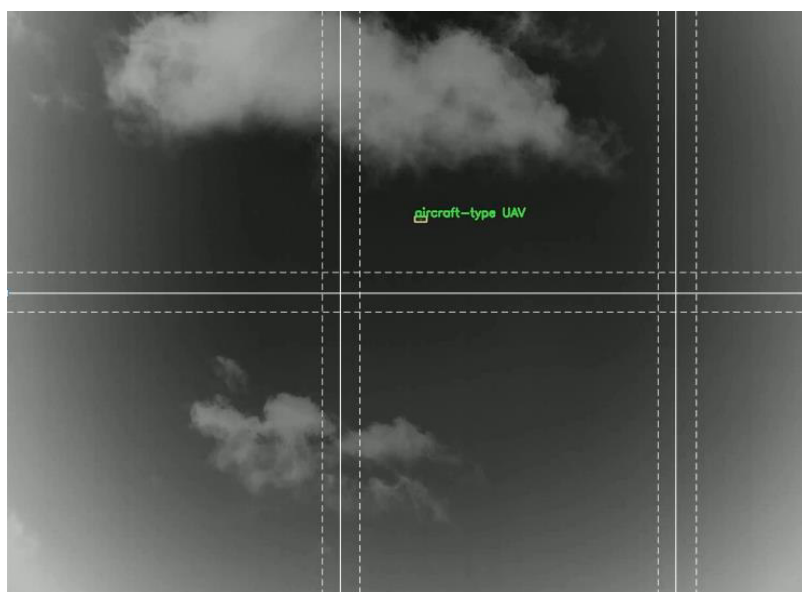


Рис. 2. Пример разделения кадра видеопотока на сегменты размером  $512 \times 512$  с перекрытием 10 % от размера сегмента

Такой подход позволил повысить точность обнаружения малоразмерных БПЛА за счёт увеличенного масштаба анализа. Однако разделение кадра на множество сегментов привело к значительному росту вычислительной нагрузки, что выразилось в увеличении времени обработки и задержках видеопотока из-за многократных прогонов модели.

### Постановка задачи

В данной работе ставятся следующие задачи:

1. Исследование влияния размера сегментов на эффективность обнаружения и распознавания БПЛА нейросетевой моделью.
2. Анализ зависимости вычислительной нагрузки системы обработки данных от размера сегментов.

## Теория

Для проведения экспериментальных исследований использовались тепловизионные видеозаписи общей продолжительностью 43 минуты, полученные с помощью специализированного аппаратного комплекса (рис. 3), включающего следующие компоненты:

- 1) В качестве оптического датчика LWIR-диапазона – тепловизор Sii AT (LWIR-диапазон), обладающий следующими характеристиками:
  - рабочий спектральный диапазон: 8...14 мкм;
  - разрешение кадра: 640×480 пикселей;
  - угол поля зрения (по горизонтали): 73,3°;
  - пространственное разрешение (iFOV): 1,999 мрад;
  - чувствительность (NETD):  $\leq 0,5$  К;
  - частота кадров: 50 Гц.
- 2) В качестве средства записи и хранения видеопотоков – видеорегистратор HIWATCH DS-H204GA, обладающий следующими характеристиками:
  - поддержка кодеков: H.265+/H.265/H.264+/H.264;
  - максимальный объём хранилища: 6 ТБ.
- 3) В качестве средства оперативного отображения информации – монитор для оперативного отображения данных.

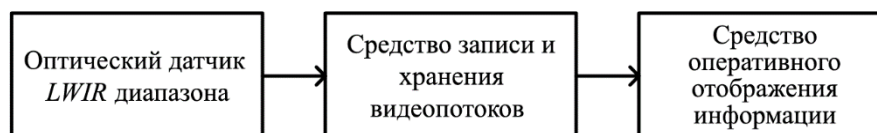


Рис. 3. Структурная схема специализированного аппаратного комплекса для регистрации видеозаписи полёта БПЛА

Видеоматериалы содержали записи полёта БПЛА (размах крыла – 5 м, длина – 2 м) на постоянной высоте 500 м с равномерной скоростью. Дистанция до точки наблюдения варьировалась от 1000 до 3700 м.

### Методика анализа

Проведён сравнительный анализ двух ключевых параметров:

1. Эффективность обнаружения нейросетевой моделью:
  - процент корректных обнаружений (без идентификации);
  - процент идентификации;
  - процент ложных срабатываний.
2. Вычислительная нагрузка:
  - время обработки видеопотока.

### Критерии сегментации

Для оптимизации обработки учитывались следующие требования:

- размеры сегментов кадра были выбраны с учётом размеров основных пространственных неоднородностей фона, это обеспечивает эффективное подавление фоновых помех;
- во избежание разделения БПЛА между сегментами и его последующей потери величина перекрытия между соседними сегментами должна превышать максимальный линейный размер проекции БПЛА на кадре, модель эффективно обнаруживает объекты без разделения на сегменты площадью  $>50$  пикселей (максимальный линейный размер около 20 пикселей) [5].

Эти требования должны реализовываться в исследовании через систему постепенно уменьшающихся размеров сегментов с соответствующим увеличением процента перекрытия для малых сегментов, что позволило бы найти баланс между точностью обнаружения и вычислительными затратами.

Исходя из этого, параметры сегментации были выбраны следующим образом:

- размер сегмента: 1280×1024 пикселей, величина перекрытия 0 % (исходный кадр);
- размер сегмента: 850×600 пикселей, величина перекрытия 5 %;

- размер сегмента: 512×512 пикселей, величина перекрытия 5 %;
- размер сегмента: 480×400 пикселей, величина перекрытия 5 %;
- размер сегмента: 400×400 пикселей, величина перекрытия 10 %;
- размер сегмента: 400×320 пикселей, величина перекрытия 10 %;
- размер сегмента: 350×350 пикселей, величина перекрытия 10 %;
- размер сегмента: 300×300 пикселей, величина перекрытия 10 %;
- размер сегмента: 340×270 пикселей, величина перекрытия 10 %;
- размер сегмента: 256×256 пикселей, величина перекрытия 10 %;
- размер сегмента: 128×128 пикселей, величина перекрытия 20 %.

### Результаты эксперимента

Анализ эффективности обнаружения и идентификации в зависимости от дальности и размера сегмента изображения представлен в таблице 1.

Таблица 1. Анализ эффективности обнаружения и идентификации БПЛА в зависимости от дальности и размера сегмента изображения

Параметры	Дальность от ОЭС до БПЛА, м										
	1000-1250	1250-1500	1500-1750	1750-2000	2000-2250	2250-2500	2500-2750	2750-3000	3000-3250	3250-3500	3500-3700
Тип атмосферного фона	ясное небо	ясное небо	ясное небо	ясное небо	ясное небо	ясное небо	мелкие АО	облако	облако	мелкие АО	ясное небо
Размер сегмента	1280 x 1024										
Кол-во обнаружений, %	100	87,5	63,5	58,5	46,5	33,5	12	0	0	1,5	0,6
Кол-во идентификаций, %	100	87,5	63,5	58,5	46,5	33,5	12	0	0	1,5	0,6
Кол-во ложных срабатываний, %	0	0	0	0	0	0	0	0	0	0	0
Размер сегмента	800 x 650										
Кол-во обнаружений, %	100	100	100	100	97,5	86,7	71	19	28,5	68	29,9
Кол-во идентификаций, %	100	100	100	100	97,5	86,7	71	19	28,5	68	29,9
Кол-во ложных срабатываний, %	0	0	0	0	0	0	0	0	0	0	0
Размер сегмента	512 x 512										
Кол-во обнаружений, %	100	100	100	100	100	100	92,5	19,5	44,5	88	56,3
Кол-во идентификаций, %	100	100	100	100	100	100	92,5	19,5	44,5	88	56,3
Кол-во ложных срабатываний, %	0	0	0	0	0	0	0	2	0	0	0,2
Размер сегмента	480 x 400										
Кол-во обнаружений, %	100	100	100	100	100	100	90	22	53	84	68,9
Кол-во идентификаций, %	100	100	100	100	100	100	90	22	53	84	68,9
Кол-во ложных срабатываний, %	0	0	0	0	0	0	0	0	0	0	0
Размер сегмента	400 x 400										
Кол-во обнаружений, %	100	100	100	100	100	100	94,5	21,5	51	91,5	76,4
Кол-во идентификаций, %	100	100	100	100	100	100	94,5	21,5	51	91,5	76,4

Окончание табл. 1

Параметры	Дальность от ОЭС до БПЛА, м										
	1000-1250	1250-1500	1500-1750	1750-2000	2000-2250	2250-2500	2500-2750	2750-3000	3000-3250	3250-3500	3500-3700
Тип атмосферного фона	ясное небо	ясное небо	ясное небо	ясное небо	ясное небо	ясное небо	мелкие АО	облако	облако	мелкие АО	ясное небо
Кол-во ложных срабатываний, %	0	0	0	0,5	0	0	0	0	0	0	0,3
Размер сегмента	400 x 320										
Кол-во обнаружений, %	100	100	100	100	100	100	93,5	20	49,5	90,5	77,6
Кол-во идентификаций, %	100	100	100	100	100	100	93,5	20	49,5	90,5	77,6
Кол-во ложных срабатываний, %	0	0	0,5	0	0	0,5	0	1	2,5	0	0,2
Размер сегмента	350 x 350										
Кол-во обнаружений, %	100	100	100	100	100	97,3	90,5	11,5	40,5	91,5	82,2
Кол-во идентификаций, %	100	100	100	100	100	97,3	90,5	11,5	40,5	91,5	82,2
Кол-во ложных срабатываний, %	0	0	0	0	0	0	0,5	0	0	0	0
Размер сегмента	300 x 300										
Кол-во обнаружений, %	100	100	100	100	100	94,7	80,5	2	27,5	84,5	72,9
Кол-во идентификаций, %	100	100	100	100	100	94,7	80	2	27,5	84,5	72,9
Кол-во ложных срабатываний, %	0	0	0	0,5	0	0	0	0	0	0	0
Размер сегмента	340 x 270										
Кол-во обнаружений, %	100	100	100	100	100	93	81	14,5	30	87,5	75,1
Кол-во идентификаций, %	100	100	100	100	100	93	81	14,5	30	87,5	75
Кол-во ложных срабатываний, %	0	0	0	0,5	0	0	0	0	0	0,5	0
Размер сегмента	256 x 256										
Кол-во обнаружений, %	100	100	100	100	96,5	84	69	0	13	84,5	39,5
Кол-во идентификаций, %	100	100	100	100	96	83	69	0	13	84,5	39
Кол-во ложных срабатываний, %	3	5,5	4,5	2	7,5	11	2,5	2,5	0	0	0,4
Размер сегмента	128 x 128										
Кол-во обнаружений, %	97	89,5	76	56	44,5	27,5	22,5	0	0	8	4,35
Кол-во идентификаций, %	96,5	89,5	76	56	43	27,5	22,5	0	0	8	4,26
Кол-во ложных срабатываний, %	12	15,5	14	4	12,5	10	16,5	1	13,5	5	5,8

Экспериментальные данные показывают устойчивое 100 % обнаружение на ближних дистанциях (1000...2000 м) для большинства параметров сегментации. Однако с увеличением расстояния до 2000...3500 м наблюдается заметное снижение эффективности, достигающее минимальных значений (0...28,5 %) в условиях облачности на дистанциях 2750...3000 м. На предельных расстояниях

3500...3700 м удовлетворительные результаты (76,4...94,5 %) демонстрируют только размеры сегментов 350×350...400×400 пикселей.

Наилучший компромисс между точностью обнаружения и уровнем ложных срабатываний обеспечивают сегменты 400×400 и 400×320 пикселей. В то время как крупные форматы (1280×1024) теряют эффективность на дистанциях свыше 2000 м, а чрезмерно мелкие (128×128) приводят к неприемлемому росту ложных срабатываний (до 16,5 %) и падению обнаружений ниже 50 %.

Наибольшие сложности для системы создает облачный фон, особенно в диапазоне 2750...3000 м, где эффективность обнаружения падает до нуля. Менее выраженное, но значимое влияние оказывают мелкие атмосферные объекты на дистанциях 3250...3500 м.

Зависимость роста вычислительных затрат от параметров сегментации видеопотока представлена в таблице 2.

Таблица 2. Зависимость роста вычислительных затрат от параметров сегментации видеопотока

Размер сегмента, в пикселях	Размер кадра, в сегментах	Размер перекрытия между соседними сегментами, %	Количество сегментов в одном кадре	Усреднённая вычисленная кратность увеличения вычислительных затрат
1280 x 1024	1 x 1	0	1	1
800 x 650	1.6 x 1.6	5	4	3.19
512 x 512	2.5 x 2	5	9	6.23
480 x 400	2.7 x 2.6	5	12	6.44
400 x 400	3.2 x 2.6	10	12	8.78
400 x 320	3.2 x 3.2	10	16	9.85
350 x 350	3.7 x 2.9	10	16	11.8
300 x 300	4.3 x 3.4	10	20	14.49
340 x 270	3.8 x 3.8	10	25	14.89
256 x 256	5 x 4	10	30	20.3
128 x 128	10 x 8	20	130	90.58

Проведённый анализ данных показывает существенную зависимость между параметрами сегментации и ростом вычислительных затрат. Наибольший практический интерес представляет нелинейный характер этой зависимости – при уменьшении размера сегмента с 1280×1024 до 128×128 пикселей нагрузка возрастает с 1 до 90.58 условных единиц.

Ключевым фактором, определяющим вычислительную сложность, выступает количество сегментов в кадре. Так, обработка 130 фрагментов размером 128×128 требует в 90 раз больше ресурсов по сравнению с анализом целого кадра. При этом даже незначительное уменьшение размера сегмента (например, с 512×512 до 400×400 пикселей) приводит к заметному росту нагрузки – с 6.23 до 8.78 единиц.

### Обсуждение результатов

Особого внимания заслуживает компромисс между точностью обнаружения и производительностью системы. Использование сегментов размером 400×400...512×512 пикселей обеспечивает разумный баланс между вычислительной нагрузкой (6-9 единиц) и эффективностью обнаружения, подтверждённой предыдущими исследованиями.

Полученные результаты подчёркивают необходимость тщательного выбора параметров сегментации с учётом как требований к точности обнаружения, так и доступных вычислительных ресурсов. Особенно это актуально для систем реального времени, где критически важна производительность.

## Выводы

Проведенное исследование подтвердило эффективность сегментации видеопотока для обнаружения малоразмерных БПЛА в ИК-диапазоне на сложном атмосферном фоне. Экспериментальные данные показывают устойчивое обнаружение на ближних дистанциях (1000...2500 м) для большинства параметров сегментации. На дистанциях свыше 2500 м наилучшие результаты достигнуты при использовании сегментов  $350 \times 350 \dots 400 \times 400$  пикселей, обеспечивающих наиболее высокую точность при умеренной вычислительной нагрузке.

Эксперименты выявили ключевые факторы, влияющие на эффективность системы: атмосферные условия, дальность до цели и размеры сегментов. Установлено, что уменьшение сегментов ниже  $300 \times 300$  пикселей приводит к недопустимому росту вычислительных затрат при одновременном снижении качества обнаружения и возрастанию ложных обнаружений.

Полученные результаты могут быть рекомендованы для практического применения в системах мониторинга воздушного пространства.

**Научный руководитель:** д. т. н, проф. Якименко Игорь Владимирович, филиал НИУ «МЭИ» в г. Смоленске, Смоленск, Россия

## Список литературы

1. Ростопчин В. В. Ударные беспилотные летательные аппараты и противовоздушная оборона – проблемы и перспективы противостояния // Беспилотная авиация. 2019. URL: [https://www.researchgate.net/publication/331772628\\_Udarnye\\_bespilotnye\\_letatelnye\\_apparaty\\_i\\_protivovozdusnaa\\_oborona\\_-\\_problemy\\_i\\_perspektivy\\_protivostoania](https://www.researchgate.net/publication/331772628_Udarnye_bespilotnye_letatelnye_apparaty_i_protivovozdusnaa_oborona_-_problemy_i_perspektivy_protivostoania) (дата обращения 20.05.2019).
2. Самойлов П. В., Иванов К. А. Угрозы применения малоразмерных БПЛА и определение наиболее эффективного способа борьбы с ними // Молодой ученый. 2017. № 45. С. 59-65. URL: <https://moluch.ru/archive/179/46398/> (дата обращения: 21.09.2019).
3. Митрофанов Д. Г., Шишков С. В. Инновационный подход к вопросу обнаружения малогабаритных беспилотных летательных аппаратов // Известия ЮФУ. Технические науки. 2018. № 1 (195). С. 28-40.
4. Еремин Г. В., Гаврилов А. Д., Назарчук И. И. Малоразмерные беспилотники – новая проблема для ПВО // Отвага. 29.01.2015. № 6 (14). URL: <http://otvaga2004.ru/armiya-i-vpk/armiya-i-vpk-vzglyad/malorazmernye-bespilotniki> (дата доступа 11.12.2019).
5. Оптико информационное обеспечение обнаружения артефактов на сложном фоне роботизированным комплексом с элементами нейронных сетей / И.В. Якименко, Ю.И. Якименко, Р.В. Поляков, Н.Н. Лавринков // МСАРД. 2025. С. 38-39.



## Методика полуавтоматической разметки данных: снижение трудоемкости формирования масок семантической сегментации

В. В. Лаптев, О. М. Гергет

Томский политехнический университет, Томск, Россия

**Аннотация.** В данной статье рассматривается методика полуавтоматической генерации сегментационных масок, основанная на использовании карт активации, извлекаемых из классификационных моделей глубокого обучения. Проведен анализ современных архитектур сверточных нейронных сетей (ResNet50, EfficientNetB5, Xception и ConvNeXtBase) и методов визуализации значимых признаков (GradCAM, HiResCAM, LayerCAM и др.) для выделения морфологических признаков на изображениях оптической когерентной томографии. Для повышения точности предложен этап постобработки с использованием модели FBA-Net, эффективно уточняющей границы объектов. Эксперименты показали, что предложенный подход позволяет достичь метрики IoU до 71,3 % на объектах с четкими структурными признаками, что существенно снижает трудоемкость разметки. Метод особенно актуален для медицинской визуализации, где объем ручной аннотации ограничен.

**Ключевые слова:** семантическая сегментация, маски изображений, карта внимания, карта признаков, сверточные нейронные сети, CAM, FBA-Net.

## Semi-Automatic generation of segmentation masks using deep learning activation maps

V. V. Laptev, O. M. Gerget

Tomsk Polytechnic University, Tomsk, Russia

**Abstract.** This article discusses a methodology for semi-automatic generation of segmentation masks based on the use of activation maps extracted from deep learning classification models. We conduct an analysis of modern convolutional neural network architectures (ResNet50, EfficientNetB5, Xception, and ConvNeXtBase) and salient feature visualization methods (GradCAM, HiResCAM, LayerCAM and others) for highlighting morphological features in optical coherence tomography images. To improve accuracy, a post-processing stage using the FBA-Net model is proposed, effectively refining object boundaries. Experiments demonstrated that the proposed approach achieves an Intersection over Union (IoU) metric of up to 71,3 % on objects with clear structural features, significantly reducing annotation workload. The method is particularly relevant for medical imaging, where manual annotation availability is limited.

**Keywords:** semantic segmentation, image masks, activation map, feature map, convolutional neural networks, CAM, FBA-Net.

### Введение

Сегментация изображений представляет собой одну из фундаментальных задач компьютерного зрения, направленную на разделение изображения на семантически значимые области. В медицинской визуализации эта задача приобретает критическую важность, так как точное выделение патологий, органов или анатомических структур напрямую влияет на качество диагностики и планирования лечения. Современные методы сегментации, основанные на глубоком обучении, демонстрируют высокую эффективность, однако их внедрение сталкивается с рядом ограничений, связанных с необходимостью масштабных аннотированных данных. Подготовка таких данных требует значительных временных и экспертных ресурсов, особенно в условиях пиксельной точности, необходимой для медицинских исследований. Например, разметка одного изображения с поражениями COVID-19 занимает до 20 минут, а для наборов из тысяч изображений трудозатраты достигают сотен часов [1]. Для сравнения: подготовка аналогичного объема данных для классификации требует на порядок меньше времени, что создает существенный дисбаланс в развитии методов анализа изображений. Эта проблема усугубляется в условиях ограниченности экспертных ресурсов и возрастающих требований к качеству медицинской диагностики, где ошибки сегментации напрямую влияют на достоверность клинических выводов.

Традиционные подходы к сегментации можно условно разделить на методы, основанные на анализе гистограмм, кластеризации, пороговой обработке и нейронных сетях. Последние, благодаря способности к обобщению и адаптации, становятся доминирующими, однако их эффективность напрямую зависит от качества и объема обучающих данных. В этом контексте актуальной остается

проблема разработки методов, минимизирующих зависимость от ручной разметки. Перспективным направлением является использование парадигм слабо контролируемого обучения, таких как обучение с частичными аннотациями или автоматическая генерация масок на основе промежуточных результатов классификационных моделей.

Актуальность работы определяется необходимостью минимизировать зависимость современных алгоритмов сегментации от ручной разметки без ущерба для их точности. Большинство существующих подходов, таких как U-Net [2] или DeepLab [3], демонстрируют высокую эффективность только при наличии полноценно размеченных данных. Однако в реальных сценариях, особенно в медицине, получение таких данных часто затруднено из-за этических, временных или финансовых ограничений. В качестве альтернативы предлагается использовать классификационные модели, обученные на слабо аннотированных данных (например, бинарных метках наличия/отсутствия объекта), для генерации первичных сегментационных масок. Ключевая гипотеза заключается в том, что карты активации признаков, извлекаемые из промежуточных слоев таких моделей, содержат достаточно информации для выделения областей интереса, а их последующее уточнение с помощью специализированных архитектур позволит достичь качества, сопоставимого с полностью контролируруемыми методами.

### Постановка задачи

В данной работе предложен инновационный подход, сочетающий преимущества классификационных моделей и методов постобработки для формирования сегментационных масок. Основная гипотеза исследования заключается в том, что карты активации признаков, извлекаемые из глубоких нейронных сетей, могут служить основой для генерации масок «низкого качества», которые затем могут быть уточнены с помощью специализированных архитектур. Для этого проведен сравнительный анализ методов активационного картирования (Class Activation Mapping, CAM), включая GradCAM [4], HiResCAM [5], GradCAM++ [6] и др., направленных на визуализацию значимых областей изображения. Эти методы, изначально разработанные для интерпретации решений нейросетей, адаптированы для формирования первичных масок, что позволяет избежать трудоемкого этапа ручной разметки. Ключевым этапом предложенного метода является использование трехканальной карты (Trimap), структурирующей изображение на три семантические зоны:

- детерминированный передний план;
- фон;
- неопределенную область.

Trimap формализует пространственные отношения между объектом и окружением, что особенно важно для медицинских данных с низкой контрастностью и зашумленностью. Для финального уточнения масок применена модель FBA-Net (Feature Boundary-Aware Network) [7], сочетающая механизмы внимания к границам и семантическим признакам. Архитектура FBA-Net, интегрирующая подходы U-Net и attention-механизмы, демонстрирует высокую эффективность в задачах, где критична точность определения контуров, таких как выделение опухолей или органов в условиях артефактов изображения.

Основная задача исследования формулируется следующим образом: разработка метода автоматической генерации сегментационных масок на основе карт активации классификационных моделей с их последующей оптимизацией с использованием FBA-Net.

### Теория

Исходные данные получены во время диагностических процедур в НИИ комплексных проблем сердечно-сосудистых заболеваний (Кемерово) и Тюменском кардиологическом научном центре (Тюмень). Общий объем данных составил 103 исследования оптической когерентной томографии (ОКТ): 77 исследований пациентов мужского пола и 26 исследований пациентов женского пола. Каждое ОКТ-исследование представляет собой трехканальный видеофайл с разрешением кадра 704x704 пикселя. Глубина обзора каждого ОКТ-исследования варьируется от 215 до 270 кадров.

Полученные 103 ОКТ-исследования, представленные 25698 RGB-срезами, имеют четыре морфологических признака атеросклеротической бляшки: сосудистый просвет (Lumen, LM),

фиброзная капсула (Fibrous cap, FC), липидное ядро (Lipid core, LC) и vasa vasorum (VV) [8]. Представленные характеристики имеют решающее значение для понимания состава бляшки и ее клинических последствий, в частности для оценки риска сердечно-сосудистых событий.

Все выявленные морфологические характеристики атеросклеротических бляшек играют важную роль в понимании структуры и прогрессирования атеросклеротического заболевания. Каждая характеристика дает критическое представление о различных аспектах состава бляшки, влияя на клиническую оценку стабильности бляшки и риск сердечно-сосудистых событий. Сосудистый просвет отражает степень стеноза, влияя на кровоток, в то время как фиброзная покрывка, липидное ядро и vasa vasorum являются важными маркерами для оценки уязвимости бляшки.

Предлагаемая методика представляет собой полуавтоматический подход к генерации сегментационных масок на основе карт активации, извлекаемых из классификационных моделей глубокого обучения, с последующей постобработкой. Методика состоит из четырех основных этапов, показанных на рисунке 1.

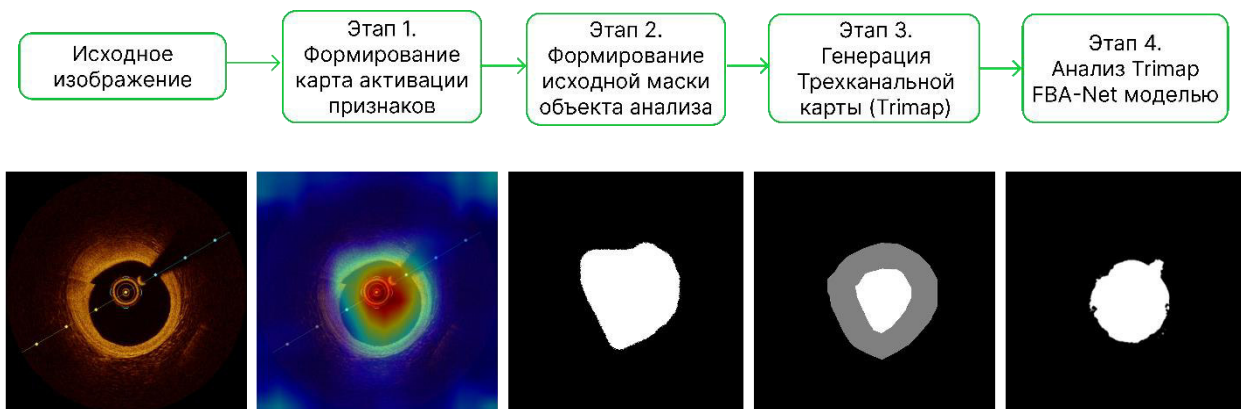


Рис. 1. Методика полуавтоматической генерации сегментационных масок

*Карта признаков.* Задача классификации – это получение категориального ответа на основе набора признаков. Концептуальная схема работы модели классификации изображения на примере сверточной нейронной сети приведена на рисунке 2.

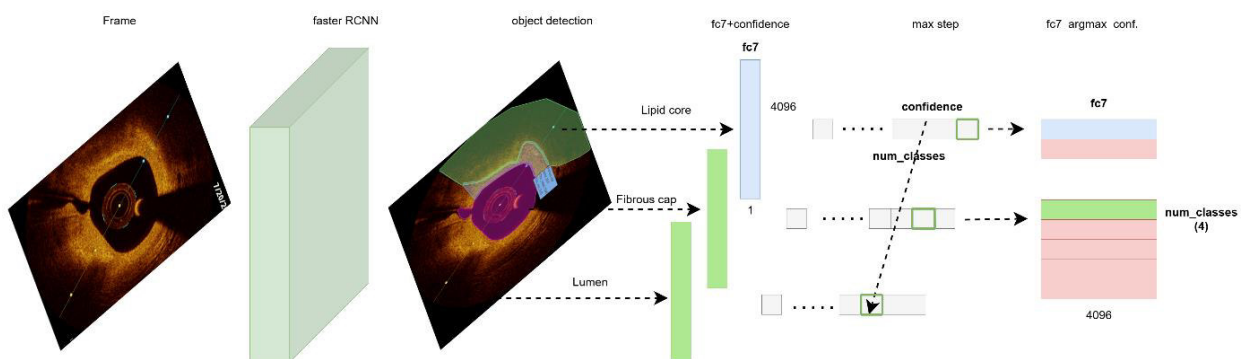


Рис. 2. Концептуальная схема работы нейросетевой модели классификации (на примере модели Faster RCNN)

Для получения категориального ответа на выходе входное изображение проходит большой путь «сверток» с целью обобщения носимой информации. На каждом этапе свертки неинформативные характеристики изображения отбрасываются, а ключевые характеристики изображения сохраняются. Непосредственное определение класса изображения происходит, как правило, на основе не более 1024 признаков.

В данной работе раскрывается концепция анализа тех самых ключевых признаков, отобранных моделью искусственной нейронной сети, на основе которых был произведен финальный ответ. Для генерации первичных сегментационных масок на основе классификационных моделей были

проанализированы и адаптированы методы активационного картирования (Class Activation Mapping, CAM), позволяющие визуализировать области изображения, наиболее значимые для предсказания модели. В исследовании рассмотрены 9 различных методов: GradCAM [4], HiResCAM [5], GradCAMElementWise [4], GradCAM++ [6], XGradCAM [9], EigenCAM [10], EigenGradCAM [11], ScoreCAM [12] и LayerCAM [13]. Перечисленные методы, основанные на анализе последнего сверточного слоя были выбраны из-за их способности локализовать морфологические признаки на ОКТ-изображениях, несмотря на низкую контрастность и наличие спекл-шума.

Полученные результаты (рис. 3) представляют собой вариант некой «тепловой карты следования за объектом», а также имеют схожее описание с сегментационной маской изображения. Данный результат позволяет выдвинуть гипотезу, что вектор признаков нейросетевой модели классификации можно интерпретировать как сегментационную маску изображения с обозначением позиции класса анализа. Полученная описанным способом маска изображения не имеет четких границ, а лишь примерно определяет местоположение анализируемого класса. С целью уточнения границ объекта в исследовании предложен подход, основанный на анализе Trimap карты FBA-Net моделью.

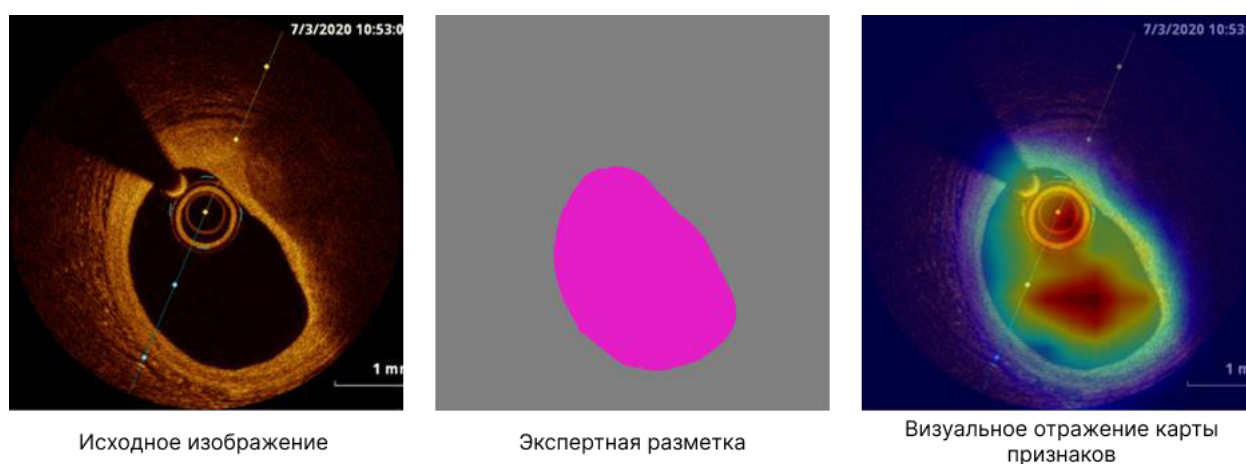


Рис. 3. Визуализация карты признаков (на примере анализа класса Lumen моделями Xception и LayerCAM)

Функционально Trimap выступает в качестве априорной маски, задающей контекст для алгоритмов матирования, которые вычисляют альфа-канал — полупрозрачную маску, количественно определяющую степень принадлежности каждого пикселя к объекту. Смешанная область, являясь буферной зоной, фокусирует вычислительные ресурсы на анализе переходных участков, где традиционные методы сегментации демонстрируют низкую точность из-за неоднозначности цветовых или текстурных характеристик. Ключевая роль Trimap в компьютерном зрении обусловлена ее способностью снижать сложность задач разделения объектов и фона, особенно в условиях ограниченной контрастности, наличия шумов или динамического освещения. В современных подходах, таких как AlphaGAN [14] или методах, основанных на глубоком обучении, Trimap используется как входной сигнал для обучения нейросетей предсказанию альфа-каналов. Данный подход позволяет достичь высокой детализации при выделении объектов с нежесткими границами, таких как волосы, дым или полупрозрачные материалы, что критически важно в приложениях видеомонтажа, дополненной реальности и медицинской визуализации. Таким образом, трехканальная карта объекта служит не только инструментом разметки, но и математическим аппаратом, формализующим пространственные отношения между объектом и фоном, что обеспечивает воспроизводимость и интерпретируемость результатов в задачах анализа изображений.

Процесс составления Trimap включает несколько этапов:

1. Инициализация: первоначально пользователю или алгоритму необходимо задать начальные области для фона и объекта. Это можно сделать вручную, применяя инструменты выделения в графических редакторах, или автоматически, используя предварительные алгоритмы сегментации.

2. Определение границ: границы между объектом и фоном часто определяются с помощью алгоритмов, таких как градиентный анализ или методы машинного обучения. Пиксели, которые находятся рядом с границами, помещаются в неопределённую область.

3. Создание маски Trimap: на основе выделенных областей создаётся маска, где каждому пикселю присваивается одно из трёх значений: 0 (фон), 128 (неопределённая область), 255 (объект).

Для решения поставленной задачи Trimap будет использоваться в качестве валидатора. Полученные на предыдущем этапе маски изображений, сформированные из карты активации классификационной модели, подвергаются следующим преобразованиям:

1. Изображение приводится к размеру 1024x1024px.
2. Последовательное применение морфологических операций открытие и закрытие для устранения случайных шумов и незначительных пропусков.
3. Создается копия маски изображения.
4. Один из вариантов подвергается морфологической операции – дилатация. Данная маска в будущем будет представлять вариант Unknown.
5. Второй вариант подвергается морфологической операции – эрозия. Данная маска в будущем будет представлять вариант Foreground.
6. Заключительным этапом является наложение полученных масок с использованием представленной выше цветовой схемы.

Полученное Trimap изображение анализируется FBA-Net моделью [7]. FBA-Net (Feature Boundary-Aware Network) — это глубокая нейронная сеть, разработанная для задач сегментации медицинских изображений с акцентом на точное определение границ объектов. Основная цель архитектуры — улучшить качество выделения сложных анатомических структур (опухолей, органов) за счёт комбинации механизмов внимания к границам и семантическим признакам. Модель была предложена как решение проблем, связанных с неоднородностью интенсивности пикселей, шумами и размытыми границами в медицинских данных.

Интеграция FBA-Net обеспечивает существенное повышение точности автоматически формируемых сегментационных масок объектов. Способность модели выделять границы объектов в условиях зашумленности и низкой контрастности напрямую коррелирует с целями исследования.

### Результаты эксперимента

Для решения поставленной задачи – определения присутствия классов Lumen, Fibrous cap, Lipid core и Vasa vasorum на кадре видеопоследовательности – было обучено 4 CNN модели (ResNet50, EfficientNetB5, Xception, ConvNeXtBase). Выбранные модели являются современными архитектурами глубокого обучения, доказавшими свою эффективность в задачах анализа медицинских изображений [15, 16]. Их ключевые преимущества включают способность работать с низкоконтрастными данными, малыми объектами и неоднородными текстурами, характерными для медицинских снимков. Данные модели покрывают широкий спектр медицинских задач: от анализа микроскопических объектов до обработки 3D-данных [17]. Для достижения качественного результата за кратчайшие сроки мы использовали технологию трансферного обучения [18]. С помощью моделей, обученных на 1,4 млн изображений ImageNet и хранящих веса для определения 17000 классов, в кратчайшие сроки были подготовлены модели классификации.

Все модели обучались в равных условиях с заданным набором гиперпараметров, представленным в таблице 1. Обучение проводилось в течение 75 эпох с использованием предварительно размеченных 25698 изображений. Результаты, полученные в ходе тестирования моделей, приведены в таблице 2.

Таблица 1. Гиперпараметры, использованные для обучения моделей классификации

Гиперпараметр	Значение
Batch size	32
Input size	512x512
Optimizer	Adam
Learning rate	$12^{-4}$

Таблица 2. Результаты работы моделей классификации на тестовом множестве

Model name	Accuracy, %	Precision, %	Recall, %
ResNet-50	74,3	70,5	70,3
EfficientNetB5	78,2	68,4	67,2
Xception	97,8	91,7	91,7
ConvNeXtBase	69,9	65,4	65,7

Согласно полученным результатам, можно сделать вывод, что лидером среди моделей классификации атеросклеротических бляшек стала модель Xception с точностью классификации по метрике Accuracy 97,8 % на тестовом множестве. Время обучения модели составило 5,5 ч, а время обработки одного изображения – 0,0013 с. Полученная эффективность модели позволяет использовать ее в дальнейших экспериментах создания сегментационной маски изображения.

Для оценки интерпретируемости моделей классификации и их ориентации на релевантные морфологические признаки бляшек были применены различные методы CAM. Полученные сравнительные карты активации демонстрируют вариативность работы алгоритма CAM для признаков разного размера и сложности. Стоит отметить, что метрическая оценка не проводилась, так как полученные карты активации не позволили сформировать сегментационные маски с четкими очертаниями. В связи с этим было принято решение провести качественную оценку результатов экспертами, фокусируясь на способности методов локализовать морфологические признаки. Ниже приведены результаты по каждой из характеристик бляшек, полученные в ходе анализа.

- Lumen:

- LayerCAM, HiResCAM, GradCAMElementWise и EigenGradCAM постоянно демонстрировали превосходную производительность в определении границ просвета с высокой точностью. Эти алгоритмы эффективно подчеркивали контуры и минимизировали ложные активации за пределами просвета, что свидетельствует о надежной локализации признаков;

- GradCAM, GradCAM++ и XGradCAM продемонстрировали умеренную производительность. Хотя они улавливали общую форму просвета, случайные активации на углах изображения или фоновые артефакты снижали их надежность;

- EigenCAM показал низкие результаты, часто ошибочно классифицируя окружающие области как часть просвета, что снижало интерпретируемость.

- Fibrous cap:

- LayerCAM, HiResCAM, GradCAMElementWise и EigenGradCAM отлично выделяли тонкий слой фиброзной капсулы, что является критической характеристикой для оценки уязвимости бляшки. Эти методы точно определяли границы фиброзной капсулы даже в сложных случаях с диффузной текстурой;

- GradCAM, GradCAM++ и XGradCAM показали средние результаты, при этом результаты иногда включали ложные активации на краях изображения;

- EigenCAM показал низкие результаты, часто ошибочно активируя области липидного ядра, а не на самой фиброзной структуре, что подрывает его клиническую значимость.

- Lipid core:

- Сегментация липидного ядра оказалась сложной для всех протестированных алгоритмов. Слабые или неправильные активации наблюдались во всех методах CAM, включая GradCAM, HiResCAM, GradCAMElementWise, GradCAM++, XGradCAM, EigenCAM, EigenGradCAM и ScoreCAM. Данное ограничение, вероятно, связано со сложной морфологией липидного ядра, нечеткими границами и встраиванием в окружающие ткани.

- Vasa Vasorum:

- Большинство алгоритмов, за исключением EigenCAM, показали адекватные результаты при идентификации класса vasa vasorum. Исследуемые методы правильно выделили мелкие, менее отчетливые сосудистые структуры с незначительными отклонениями. EigenCAM демонстрировал значительные ошибки в классификации, часто активируясь в области просвета, где структуры vasa vasorum отсутствовали, что снижает его надежность при выявлении мелких анатомических особенностей.



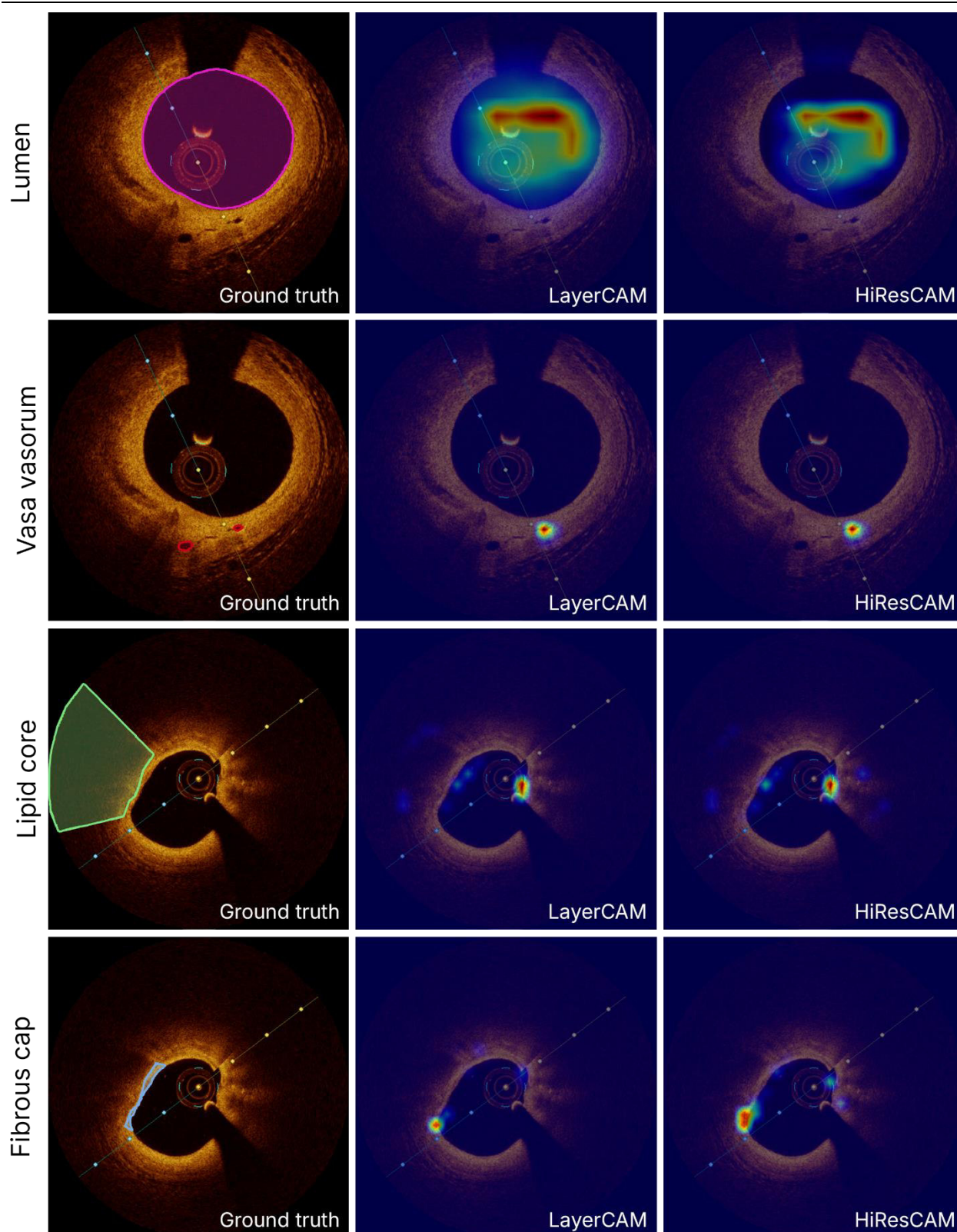


Рис. 4. Результат анализа карты признаков классификационной модели методами LayerCAM и HiResCAM

Полученные карты активации подтверждают ориентацию моделей на анатомически значимые области и обеспечивают механизм для выявления потенциальных областей неправильной классификации или двусмысленности. Примечательно, что LayerCAM и HiResCAM продемонстрировали стабильную производительность в создании точных визуализаций по всем признакам бляшек, превосходя другие методы благодаря своей способности агрегировать многомасштабные активации и локализовать тонкие структуры (например, тонкие фиброзные



капсулы) с минимальным количеством ложных срабатываний. Их градиентно-взвешенные механизмы внимания дополнительно подавляют артефакты изображения, присущие ОКТ, такие как спекл-шум, сохраняя при этом важнейшие морфологические детали. Применение этих методов позволяет в кратчайшие сроки получить сегментационные маски анатомических объектов удовлетворительной точности, что дает возможность составить большой обучающий набор данных и повысить уверенность в клинической применимости будущей модели. Визуализация результата для методов LayerCAM и HiResCAM приведена на рисунке 4.

Заключительным этапом формирования сегментационной маски изображения является уточнение границ анатомической структуры. Представленный в работе алгоритм формирования трехканальной маски отлично подходит для решения поставленной задачи. Визуализация этапа создания трехканальной маски и результата применения FBA-Net модели по результатам работы метода LayerCAM представлена на рисунке 5.

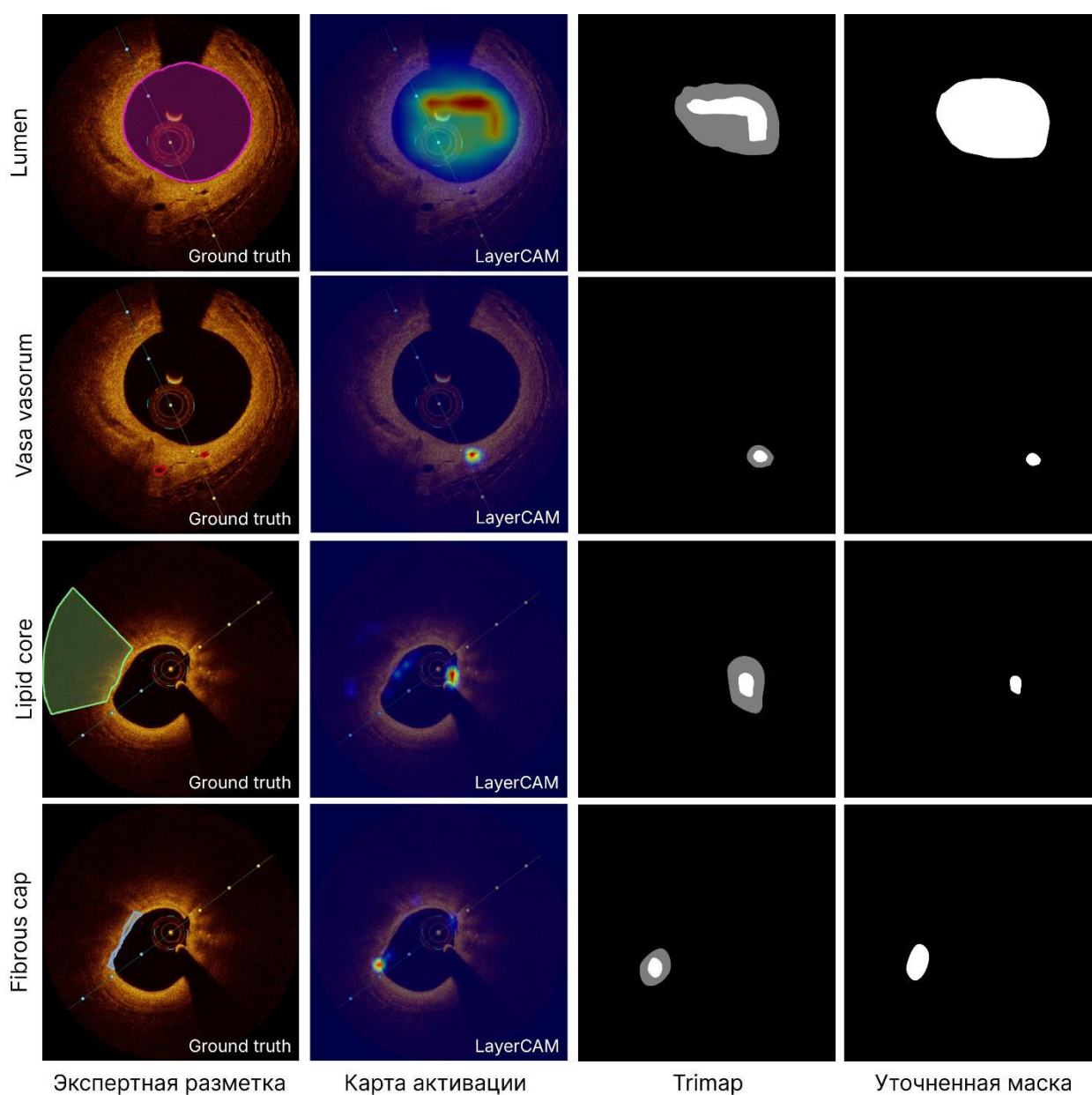


Рис. 5. Визуализация этапов работы представленного метода

Из результатов, представленных на рисунке 5, видно, что представленная методика формирования сегментационных масок изображений подходит для тех объектов исследования, которые имеют выраженные очертания, такие как Lumen и Vasa vasorum. Для объектов, что не обладают данным

признаком, такие как Fibrous cap и Lipid core, представленный метод формирования сегментационной маски не подходит. Основной проблемой для представленных классов выступает отсутствие явных границ объекта, в связи с чем FBA-Net модель не может провести детальное выделение области интереса. Также стоит отметить, что именно для классов Fibrous cap и Lipid core создать исходную маску объекта смогло наименьшее число рассматриваемых алгоритмов синтеза карт внимания. Несмотря на полученные результаты, представленные в таблице 3, данная методика является эффективным инструментом формирования обучающего множества.

Таблица 3. Показатели эффективности метода формирования сегментационных масок изображений на примере классов Lumen и Vasa vasorum в расчете на одно изображение

Algorithm	IoU, %	Speed, s
Ручная разметка	100,0	176,30
Автоматическое создание маски объекта	71,3	0,14
Автоматическое создание маски объекта + экспертная доработка	96,7	54,50

Согласно результатам, представленным в таблице 3, предложенная в работе методика формирования сегментационных масок изображений оказалась в 3 раза эффективнее ручного аннотирования. Также стоит отметить, что для формирования качественного обучающего множества, как правило, требуется не менее трех экспертов, что обеспечивает объективность мнения. Важным моментом является ограниченность времени таких экспертов и загруженность прочей работой. Так, на разметку используемого в работе набора данных ушло 4 месяца. На обучение классификационной модели и формирования сегментационных масок для классов Lumen и Vasa vasorum с учетом всех проводимых экспериментов было затрачено менее 1 месяца, что еще раз подчеркивает эффективность данного метода.

### Обсуждение результатов

Экспериментальная часть работы подтверждает, что интеграция автоматически генерируемых масок и их последующее уточнение с помощью FBA-Net позволяет достичь метрики Intersection over Union (IoU) на уровне 87,3 % для классов объектов с выраженными структурными признаками. Это свидетельствует о потенциале комбинирования методов интерпретируемости глубоких нейронных сетей и продвинутых архитектур сегментации для снижения зависимости от ручной разметки. Полученные результаты открывают новые возможности для применения подхода в областях, где экспертные аннотации ограничены, включая онкологию, кардиологию и анализ гистологических срезов.

Работа вносит вклад в развитие методов слабо контролируемой сегментации, предлагая практический метод для минимизации трудозатрат на подготовку данных без ущерба для точности. Дальнейшие исследования могут быть направлены на адаптацию метода для работы с мультимодальными медицинскими данными и оптимизацию вычислительной эффективности архитектуры FBA-Net.

Научная новизна подхода заключается в комбинации интерпретируемости классификационных моделей и продвинутых методов постобработки. В отличие от традиционных методов слабого обучения, где маски генерируются эвристически, предложенный метод опирается на данные, извлеченные непосредственно из внутренних представлений нейросети, что повышает их семантическую согласованность. Использование FBA-Net, в свою очередь, позволяет компенсировать недостатки первичных масок за счет фокуса на граничных признаках и механизмах внимания.

### Заключение

В рамках исследования разработан метод полуавтоматической генерации масок семантической сегментации, сочетающий интерпретируемость классификационных моделей и постобработку с использованием FBA-Net. Данный подход позволяет формировать обучающие данные с приемлемым качеством при значительном снижении временных затрат на ручную разметку более чем в 3 раза. Это

особенно критично в области медицинской визуализации, где качество аннотаций напрямую влияет на клинические выводы.

Проведенные эксперименты продемонстрировали высокую точность на объектах с выраженными морфологическими границами (Lumen, Vasa vasorum), достигнув метрики IoU до 71,3 % в автоматическом режиме и до 96,7 % с экспертной доработкой. В то же время для объектов с нечеткими границами (Fibrous cap, Lipid core) методика демонстрирует качество ниже, чем традиционные подходы, что открывает перспективы для его дальнейшей оптимизации.

Предложенная методология обладает высоким потенциалом для применения в условиях ограниченных ресурсов аннотирования и может быть адаптирована к другим медицинским задачам и мультимодальным данным. В будущем планируется расширение метода для работы с 3D-данными и снижение вычислительной нагрузки архитектуры FBA-Net, что сделает его еще более применимым в клинической практике.

### Источник финансирования

Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации, проект «Наука FSWW-2023-0007».

### Список литературы

1. Временные методические рекомендации «Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID-19)». Версия 19 (27.05.2025). URL: [https://www.consultant.ru/document/cons\\_doc\\_LAW\\_347896](https://www.consultant.ru/document/cons_doc_LAW_347896).
2. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, Springer, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4\_28.
3. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Trans Pattern Anal Mach Intell*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.
4. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Oct. 2016, doi: 10.1007/s11263-019-01228-7.
5. R. L. Draelos, and L. Carin, “Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks,” *ArXiv*, Nov. 2020, Accessed: May 04, 2024. Available: <http://arxiv.org/abs/2011.08891>
6. A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Mar. 2018, pp. 839–847, doi: 10.1109/WACV.2018.00097.
7. Y. Chung, C. Lim, C. Huang, N. Marrouche, and J. Hamm, “FBA-Net: Foreground and Background Aware Contrastive Learning for Semi-Supervised Atrium Segmentation,” Jun. 2023, Accessed: Apr. 13, 2025. Available: <http://arxiv.org/abs/2306.15189>
8. S. S. Virani, *et al.* “Heart Disease and Stroke Statistics - 2021 Update: A Report From the American Heart Association,” *Circulation*, vol. 143, no. 8, pp. E254–E743, Feb. 2021, doi: 10.1161/CIR.0000000000000950.
9. R. Fu, Q. Hu, X. Dong, Y. Guo, Y. Gao, and B. Li, “Axiom-based Grad-CAM: Towards Accurate Visualization and Explanation of CNNs,” *31st British Machine Vision Conference, BMVC 2020*, Aug. 2020, Accessed: May 04, 2024. Available: <http://arxiv.org/abs/2008.02312>
10. M. B. Muhammad and M. Yeasin, “Eigen-CAM: Class Activation Map using Principal Components,” *Proceedings of the International Joint Conference on Neural Networks*, Aug. 2020, doi: 10.1109/IJCNN48605.2020.9206626.
11. A. Chaddad, Y. Hu, Y. Wu, B. Wen, and R. Kateb, “Generalizable and Explainable Deep Learning for Medical Image Computing: An Overview,” *Curr Opin Biomed Eng*, vol. 33, Mar. 2025, doi: 10.1016/j.cobme.2024.100567.
12. H. Wang, *et al.* “Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, pp. 111–119, Oct. 2019, doi: 10.1109/CVPRW50498.2020.00020.
13. J. Zheng, *et al.* “A Novel Defense Against Poisoning Attacks on Federated Learning: LayerCAM Augmented with Autoencoder,” Jun. 2024, Accessed: Apr. 13, 2025. [Online]. Available: <https://arxiv.org/abs/2406.02605v1>
14. S. Lutz, K. Amplianitis, and A. Smolic, “AlphaGAN: Generative adversarial networks for natural image matting,” *British Machine Vision Conference 2018, BMVC 2018*, Jul. 2018, Accessed: Apr. 13, 2025. Available: <https://arxiv.org/abs/1807.10088v1>

- 
15. C. Chen, *et al.* “Improving protein-protein interactions prediction accuracy using XGBoost feature selection and stacked ensemble classifier,” *Comput Biol Med*, vol. 123, Aug. 2020, doi: 10.1016/j.combiomed.2020.103899.
  16. D. Wang, H. He, and C. Wei, “Cellular and potential molecular mechanisms underlying transovarial transmission of the obligate symbiont *Sulcia* in cicadas,” *Environ Microbiol*, vol. 25, no. 4, pp. 836–852, Apr. 2023, doi: 10.1111/1462-2920.16310.
  17. B.-E. Munkhsaikhan, “SOEs in China and AI Development: Legal Constraints for Foreign Investments towards China&#8217;s Digital Transformation,” *Open J Soc Sci*, vol. 12, no. 12, pp. 146–157, 2024, doi: 10.4236/JSS.2024.1212010.
  18. H.-Y. Zhou, A. Oliver, J. Wu, and Y. Zheng, “When Semi-Supervised Learning Meets Transfer Learning: Training Strategies, Models and Datasets,” *ArXiv*, Dec. 2018, Accessed: Mar. 13, 2024. Available: <http://arxiv.org/abs/1812.05313>

## Разработка метода повышения плотности и качества лидарного сканирования на основе зеркального усиления механических LiDAR-систем

М. Р. Вишневский, И. Ю. Фёдоров  
НИТУ «МИСИС», Москва, Россия

## Development of a method to enhance LiDAR scanning density and quality through mirrors in mechanical systems

M. R. Vishnevskiy, I. Yu. Fyedorov  
NUST MISIS, Moscow, Russia

**Аннотация.** В работе рассмотрен инновационный подход к повышению плотности и качества облаков точек, создаваемых механическим LiDAR, путем введения настраиваемого пассивного оптического элемента — зеркала. Анализируется влияние положения зеркала на точность и детализацию сканируемых объектов. Экспериментально продемонстрировано, что применение зеркального усиления позволяет существенно увеличить плотность точек и обеспечить более равномерное покрытие пространства, сохраняя простоту конструкции самого датчика сканирования. Приведены результаты практических исследований и моделирования, демонстрирующих высокую эффективность предложенного метода.

**Ключевые слова:** LiDAR, облака точек, зеркальное сканирование, плотность точек.

### Введение

Современные робототехнические системы, входящие в состав сложных аппаратных и программных комплексов на базе ИИ [1], наряду с технологиями глубокого обучения, оснащаются многочисленными настраиваемыми датчиками, обладающими возможностью адаптации к изменяющимся условиям физической среды [2]. Это предполагает наличие все возрастающих требований к производительности, миниатюризации и энергоэффективности LiDAR-сенсоров, определяющих их ключевую роль в формировании облаков точек в 3D-пространстве наблюдаемой сцены и создании областей с повышенной плотностью сканирования [3] для повышения ситуационной осведомленности, необходимой для безопасной навигации и предотвращения возможных столкновений. Существующие решения, направленные на повышение плотности облаков точек, включают микро-электромеханические системы (MEMS) [4, 5] и оптические фазированные решётки (OPA-LiDAR) [6, 7]. Они обладают рядом достоинств, среди которых высокая точность измерений, надежность, возможность электронного управления лучом и отсутствие механических движущихся элементов [8, 9]. В частности, OPA-LiDAR предусматривает интеграцию лазерного источника, схем распределения мощности, фазовращателей, амплитудных модуляторов, массива излучателей и волноводных соединений, занимающих значительную площадь чипа. Для обеспечения низкого уровня перекрестных помех элементы размещают на заданном минимальном расстоянии друг от друга. Таким образом, достижение высокой плотности излучения требует нахождения тонкого баланса между сложностью схемы, быстродействием, масштабируемостью и уровнем потерь энергии в многоканальной распределённой сети питания [10]. Несмотря на перспективность и потенциал OPA-LiDAR стать высокоэффективной и доступной технологией, на сегодняшний день её сложность и высокая стоимость ограничивают массовое распространение в мобильных устройствах и бюджетном сегменте рынка [11, 12].

Механические LiDAR-устройства остаются наиболее распространенными благодаря приемлемому соотношению стоимости и производительности, однако они сталкиваются с существенными недостатками по пространственному разрешению и качеству покрытия сцены [13]. Ограниченность вызвана периодичностью сканирующих лучей и появлением визуальных артефактов при высоких скоростях перемещения и сложных условиях окружающей среды [14, 15].

Традиционным способом преодоления указанных ограничений является установка нескольких LiDAR в тех местах сканирующего модуля, где информация об окружающей среде может быть

получена в максимально возможном объеме. Углы и положение каждого сенсора влияют на зоны покрытия и мертвые зоны. Увеличение числа сканирующих датчиков ведет к росту плотности данных и увеличению потока информации, но также усложняет конструкцию, увеличивая затраты и необходимость решения сложной оптимизационной задачи оптимального размещения в рамках конструктивных ограничений установки [16].

Перспективной альтернативой увеличения числа сенсоров и применения дорогостоящих ОРА является использование пассивных оптических компонентов, таких как зеркало [17, 18, 19], которые позволяют расширить область сканирования и создать дополнительный виртуальный канал без значительного роста стоимости системы. Подобные схемы уже применялись для повышения детализации и улучшения покрытия сложных сцен [20, 14], особенно в задачах, где часть объектов экранирована или затенена.

Несмотря на накопленный опыт, в большинстве работ основной акцент делается на аппаратную реализацию или локальные геометрические настройки, тогда как формальный количественный анализ прироста плотности и полноты покрытия сцены остаётся недостаточно проработанным.

В данном исследовании представлен простой и воспроизводимый метод повышения плотности LiDAR-сканирования с помощью пассивного зеркала, не требующий сложной калибровки или оптимизации положения. На примере как симуляционного, так и реального эксперимента демонстрируется, что подобное решение позволяет существенно увеличить информативность и равномерность облаков точек при минимальных затратах на доработку аппаратной части.

### Постановка задачи

В последние годы задача повышения плотности и качества облаков точек, получаемых с помощью лидарных систем, становится всё более актуальной для широкого спектра приложений — от обнаружения и классификации объектов [21, 22] и автономной навигации [23, 24] до создания высокодетализированных трёхмерных моделей и инспекции инфраструктурных объектов [25, 26, 27, 28]. Существующие подходы к увеличению плотности точек включают активные методы на основе микроэлектромеханических систем (MEMS) [8, 4, 5] и оптических фазированных решёток (ОРА) [6], а также пассивные схемы с использованием зеркал и призм [19, 18, 17]. При всех достоинствах, активные MEMS и ОРА решения требуют сложной аппаратной реализации и отличаются высокой стоимостью и чувствительностью к внешним воздействиям [29, 6, 14, 9, 11], что ограничивает их применение в мобильных и бюджетных системах [30, 13]. Пассивные методы, напротив, обладают преимуществами по простоте интеграции и стоимости, сохраняя при этом возможность значительного улучшения характеристик сканирования.

Целью настоящего исследования является разработка и экспериментальная верификация простого и воспроизводимого метода повышения плотности и качества облаков точек для механических LiDAR-сенсоров за счёт использования пассивной зеркальной оптики. Такой подход должен обеспечить увеличение пространственной информативности и детализации облаков точек при минимальном усложнении аппаратной конструкции, что особенно важно для задач, требующих высокой точности и однородности восприятия окружающей среды (например, автономная навигация роботов, точное картирование и реконструкция сложных объектов).

Для достижения поставленной цели в работе были решены следующие задачи:

- проведён анализ современных подходов к увеличению плотности и качества LiDAR-сканирования, включая методы с использованием зеркальной оптики [14, 19, 18];
- предложена и реализована методика размещения зеркала относительно сенсора, позволяющая существенно расширить охват сцены при минимальных потерях мощности сигнала. Данный подход базируется на инженерных решениях, аналогичных тем, что рассматривались в ряде публикаций [31, 32], где анализируется влияние параметров зеркала на качество регистрации;
- экспериментально подтверждена применимость предложенного метода для повышения пространственной информативности облаков точек на примере механического LiDAR Velodyne VLP-16 в симуляционной среде Gazebo;
- оценена эффективность метода на основе формальных количественных метрик: плотности точек и заполненности воксельной сетки [33, 15].

## Методы и алгоритмы

В данной работе реализован альтернативный подход к повышению плотности и качества сканирования: для создания дополнительного виртуального канала используется статичное зеркало, отражающее часть лазерных импульсов без аппаратных модификаций базового LiDAR [17, 18]. Такой метод позволяет сохранить ключевые преимущества механических лидаров — надёжность, доступную стоимость и достаточную точность [15, 13, 34], а также преодолеть ограничения регулярного сканирующего паттерна и проблемы с «мертвыми зонами» при быстром движении или наличии мелких объектов [35, 24]. Принцип работы предлагаемого метода схематично представлен на рисунке 1.

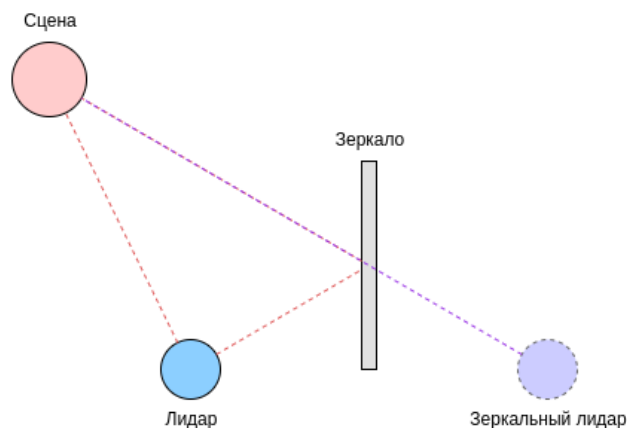


Рис. 1. Схема предлагаемой методики: LiDAR Velodyne VLP-16 (слева) испускает лазерные импульсы, часть из которых отражается от статичного зеркала, образуя дополнительный виртуальный канал сканирования

**Аппаратная конфигурация.** В качестве базового сенсора выбран LiDAR Velodyne VLP-16, широко применяемый благодаря хорошему соотношению цены и качества, а также проверенной надёжности [13, 30]. Зеркало размещалось таким образом, чтобы обеспечить расширение вертикального угла обзора. Угол наклона определялся с учётом минимизации потерь мощности лазерного сигнала и исключения зон с тенями [31, 17, 18].

**Процесс сбора данных.** Для экспериментов использовалась симуляционная среда Gazebo с фреймворком ROS, что позволяло удобно моделировать различные сцены и эффективно работать с облаками точек [36, 37]. Данные от LiDAR собирались в формате PointCloud2, а разделение на прямые и зеркально отражённые каналы осуществлялось программно.

**Алгоритмы обработки и калибровки.** Для регистрации прямых и отражённых облаков точек применялись методы библиотеки Point Cloud Library (PCL), в частности алгоритм ICP для регистрации и RANSAC для сегментации плоскостей [38, 26].

Отметим, что вопросы детальной калибровки отражающей поверхности и высокоточного совмещения облаков точек в рамках данной работы сознательно не рассматривались: основной целью было экспериментально подтвердить сам принцип повышения плотности и пространственного охвата облака точек за счёт зеркального канала. Дальнейшая проработка вопросов калибровки и оптимизации параметров планируется в будущих исследованиях.

Таким образом, предложенная методика позволяет повысить плотность и пространственную информативность облаков точек без аппаратных усложнений сенсора, что делает её привлекательной для широкого класса прикладных задач.

## Результаты экспериментов

**Постановка эксперимента.** В экспериментах сравнивались облака точек, полученные как в симуляционной среде Gazebo, так и при реальном сканировании сцены с использованием LiDAR с пассивным зеркалом.

Для объективной количественной оценки информативности данных применялась **вокселизация**: пространство сцены разбивалось на регулярную трёхмерную сетку одинаковых кубических элементов (вокселей) с размером от 0,1 до 1,0 м. Воксель считался занятым, если в его пределах находится хотя бы одна точка облака. Далее рассчитывались две метрики:



- **плотность облака точек** — среднее количество точек на занятый воксель, характеризующее детализацию и разрешение сканирования;
- **заполненность воксельной сетки** — доля занятых вокселей от общего их числа, характеризующая полноту и равномерность покрытия сцены.

Такой анализ позволяет формально сравнивать различные конфигурации сенсора, выявлять «мертвые зоны» и количественно оценивать прирост информативности при использовании зеркального канала.

### Эксперимент в симуляторе

Для наглядности на рисунке 2а и 2г приведена визуализация исходной симуляционной сцены в Gazebo. Далее на рисунках 2б, 2в и 2д, 2е показаны сравнения облаков точек по разным проекциям для стандартной и зеркальной конфигураций.

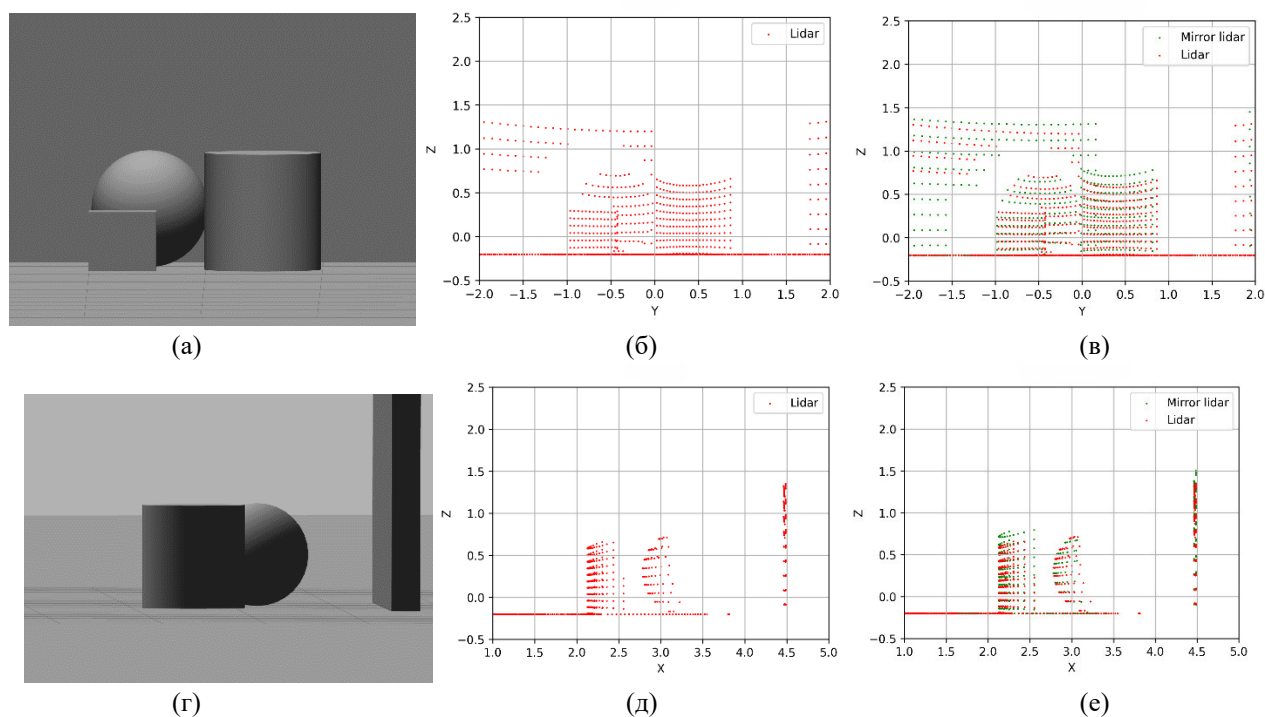


Рис. 2. Визуализация облаков точек в симуляторе (проекция X (а-в) и Y (г-е): слева — вид с виртуальной камеры, в центре — облако точек с оригинального лидара, справа — с зеркалом

Наибольший прирост заполненности сетки (табл. 2) отмечается при размерах вокселя от 0,1 до 0,25 м, где увеличение числа занятых вокселей составляет от 20 до 50 %. Например, при размере 0,1 м количество занятых вокселей увеличилось с 86 до 129, а относительный прирост составил 50 %. Для более грубой дискретизации (0,5 и 1,0 м) разница между конфигурациями практически исчезает.

Как видно из таблицы 1, зеркальный канал позволяет существенно увеличить как полноту покрытия, так и плотность точек при малых и средних размерах вокселя.

Таблица 1. Симуляция: сравнение количества занятых вокселей и плотности точек (L – один лидар; LM – лидар + зеркало)

Воксель, м	Занято вокселей (L)	Занято вокселей (LM)	Прирост, %	Точек на воксель (L)	Точек на воксель (LM)
0,10	86	129	50,0	3,16	3,18
0,20	46	61	32,6	5,91	6,72
0,25	40	48	20,0	6,80	8,54
0,50	16	16	0,0	17,0	25,63
1,00	4	4	0,0	68,0	102,5

### Эксперимент с реальным лидаром

Аналогичный эксперимент был проведён на реальной установке (рис. 4). На рисунках 3а, б показаны облака точек в проекции по оси X, а на рисунках 3в, г — в проекции по оси Y; в обоих случаях слева приведены данные без зеркала, справа — с зеркалом. Метрики рассчитывались по тем же правилам, а ключевые результаты приведены в таблице 2.

Для размера вокселя 0,25 м число занятых вокселей возросло с 37 до 52 (+40,5 %), а для 0,10 м — с 179 до 230 (+28,5 %). Плотность точек на воксель также существенно увеличилась для всех разрешений.

Таблица 2. Реальный эксперимент: сравнение количества занятых вокселей и плотности точек (L – один лидар; LM – лидар + зеркало)

Воксель, м	Занято вокселей (L)	Занято вокселей (LM)	Прирост, %	Точек на воксель (L)	Точек на воксель (LM)
0,10	179	230	28,5	10,62	15,09
0,20	44	58	31,8	43,20	59,84
0,25	37	52	40,5	51,38	66,75
0,50	11	11	0,0	172,82	315,55
1,00	4	4	0,0	475,25	867,75

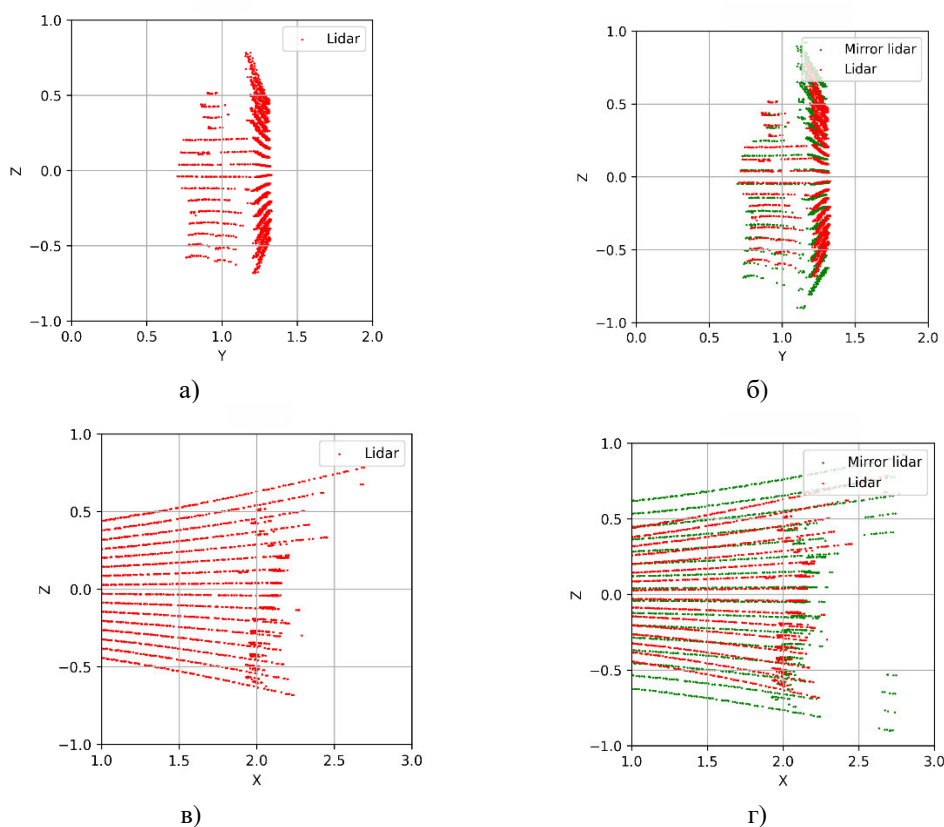


Рис. 3. Вид облаков точек с реального эксперимента (проекция X (а, б) и Y (в, г): слева — без зеркала, справа — с зеркалом)

**Результаты оценки.** Результаты как симуляционного, так и реального эксперимента подтверждают, что введение зеркала существенно увеличивает плотность и полноту облаков точек, особенно на мелких и средних размерах вокселей (от 0,1 до 0,25 м). Так, максимальный прирост заполненности сетки достигал от 28,5 до 50 % в зависимости от сцены и выбранного разрешения, что наиболее важно для задач высокой детализации — например, инспекции сложных объектов или построения 3Э-моделей с минимальным числом «слепых зон».

Наибольшее относительное увеличение плотности точек отмечалось при грубой дискретизации (1,0 м), где средняя плотность возросла с 68 до 102,5 точек на воксель в симуляции и с 475 до 868 точек на воксель

в реальном эксперименте. Однако именно при мелких размерах вокселя (0,10-0,25 м) зеркало обеспечивает заполнение ранее недоступных областей и существенный прирост информативности облака.

Повышение плотности и заполненности облака точек непосредственно влияет на качество последующей реконструкции поверхности, сегментации и извлечения признаков, а также на устойчивость алгоритмов при наличии затенённых или сложно структурированных объектов.

### Обсуждение результатов

Полученные в ходе экспериментов результаты подтверждают первоначальную гипотезу о том, что дополнительный зеркальный канал является эффективным методом повышения качества сканирования без необходимости усложнять конструкцию базового механического лидара [17, 18]. Значительный прирост плотности точек и заполненности воксельной сетки свидетельствует о существенном улучшении детализации и равномерности покрытия сцены, что подтверждается также ранее опубликованными результатами, полученными при использовании многоракурсных или пассивно-модифицированных систем [19, 22]. Особенно наглядно эти преимущества проявляются при высоких разрешениях, что делает предложенное решение перспективным для применений, требующих высокой точности и пространственной полноты, таких как инспекция инфраструктуры [28], создание цифровых двойников объектов [25, 26], а также картографирование и ориентация в условиях без доступа к спутниковой системе навигации [39].

Особого внимания заслуживает различие в характере изменения метрики заполненности при варьировании размера вокселя: в симуляции прирост уменьшается по мере укрупнения, тогда как в реальном эксперименте наблюдается обратная зависимость. В первом случае дополнительные точки зеркального канала заполняют главным образом мелкие детали сцены и при увеличении размера вокселя их вклад сглаживается. Во втором случае на данные существенно влияют шумы и небольшие погрешности регистрации: при мелкой дискретизации часть ячеек остаётся пустой, тогда как при укрупнении вокселя шум усредняется и эффект от зеркального канала проявляется более отчётливо.



Рис. 4. Фотографии реальной экспериментальной установки с зеркалом и LiDAR

Предложенный метод прост в реализации и не требует конструктивных изменений сенсора, что облегчает его интеграцию в существующие аппаратные платформы. Это особенно актуально в условиях ограниченных бюджетов и энергоэффективных сценариев, где использование более дорогих и сложных технологий, таких как MEMS и оптические фазированные решётки (OPA), может быть затруднено или нецелесообразно [6, 11, 4].

Таким образом, проведённые эксперименты демонстрируют не только техническую реализуемость, но и высокую практическую ценность зеркального канала сканирования для мобильных

робототехнических систем, задач инспекции, ЭБ-моделирования и автономной навигации, где требуется высокая детализация, равномерность и надёжность пространственного восприятия [40, 41, 42].

### Заключение

В рамках настоящего исследования разработан и экспериментально подтверждён эффективный метод повышения плотности и качества LiDAR-сканирования за счёт использования пассивной зеркальной оптики. Метод продемонстрировал высокую эффективность, выразившуюся в значительном увеличении плотности точек и полноты покрытия сцены без аппаратного усложнения сенсора [35]. Одним из потенциальных ограничений предложенного метода является искажение геометрии из-за неточной калибровки зеркала, что может повлиять на точность конечных измерений. Это ограничение должно быть изучено в дальнейших работах.

Предложенное решение обеспечивает доступное и практичное улучшение характеристик механических LiDAR-систем, открывая широкие перспективы применения в различных областях робототехники, автономной навигации и ЭБ-моделирования [23, 24, 12]. Дальнейшие исследования могут быть направлены на оптимизацию геометрии зеркал и разработку адаптивных систем для повышения качества сканирования в различных условиях эксплуатации [43, 44].

### Благодарность

Исследование выполнено при поддержке кафедры инженерной кибернетики НИТУ «МИСИС».

### Список литературы

1. V. Anand, B. Lohani, G. Pandey, R. Mishra, Advancing lidar intensity simulation through learning with novel physics-based modalities, *IEEE Transactions on Intelligent Transportation Systems* (2025).
2. V. Bharati, Lidar+ camera sensor data fusion on mobiles with ai-based virtual sensors to provide situational awareness for the visually impaired, in: *2021 IEEE Sensors Applications Symposium (SAS)*, IEEE, 2021, pp. 1-6.
3. A. Shomer, S. Avidan, Prior based sampling for adaptive lidar, *arXiv preprint arXiv:2304.07099* (2023).
4. D. Wang, C. Watkins, H. Xie, Mems mirrors for lidar: A review, *Micromachines* 11 (2020) 456. doi:10.3390/mi11050456.
5. D. Yang, Y. Liu, Q. Chen, M. Chen, S. Zhan, N.-k. Cheung, H.-Y. Chan, Z. Wang, W. J. Li, Development of the high angular resolution 360° lidar based on scanning mems mirror, *Scientific Reports* 13 (2023). doi:10.1038/s41598-022-26394-6.
6. W. Xu, Q. Yuan, Y. Yang, L. Lu, J. Chen, L. Zhou, Progress and prospects for lidar-oriented optical phased arrays based on photonic integrated circuits, *npj Nanophotonics* 2 (2025). URL: <https://www.nature.com/articles/s44310-025-00059-4>. doi:10.1038/s44310-025-00059-4.
7. Y. Fu, B. Chen, W. Yue, M. Tao, H. Zhao, Y. Li, X. Li, H. Qu, X. Li, X. Hu, et al., Target-adaptive optical phased array lidar, *Photonics Research* 12 (2024) 904-911.
8. A. Kasturi, V. Milanovic, D. Lovell, F. Hu, D. Ho, Y. Su, L. Ristic, Comparison of mems mirror lidar architectures, in: *MOEMS and Miniaturized Systems XIX*, volume 11345 of *Proc. SPIE*, 2020, p. 113450G. doi:10.1117/12.2556248.
9. C. Altuntas, Review of scanning and pixel array-based lidar point-cloud measurement techniques to capture 3d shape or motion, *Applied Sciences* 13 (2023) 6488. doi:10.3390/app13116488.
10. W. Xu, Q. Yuan, Y. Yang, L. Lu, J. Chen, L. Zhou, Progress and prospects for lidar-oriented optical phased arrays based on photonic integrated circuits, *npj Nanophotonics* 2 (2025) 14.
11. J. Zhou, A review of lidar sensor technologies for perception in automated driving, *Academic Journal of Science and Technology* 3 (2022) 255-260.
12. S. Royo, M. Ballesta-Garcia, An overview of lidar imaging systems for autonomous vehicles, *Applied Sciences* 9 (2019) 4093. doi:10.3390/app9194093.
13. J. Schulte-Tigges, M. Forster, G. Nikolovski, M. Reke, A. Ferrein, D. Kaszner, D. Matheis, T. Walter, Benchmarking of various lidar sensors for use in self-driving vehicles in real- world environments, *Sensors* 22 (2022) 7146. URL: <https://www.mdpi.com/1424-8220/22/19/7146>. doi:10.3390/s22197146.
14. T. Raj, F. H. Hashim, A. B. Huddin, M. F. Ibrahim, A. Hussain, A survey on lidar scanning mechanisms, *Electronics* 9 (2020) 741. doi:10.3390/electronics9050741.
15. M. Gunn, Lidar sensors simplified: Part 1 - lidar system point density, returns and beam divergence, <https://aevox.com/lidar-sensors-simplified-part-1-lidar-system-point-density-returns-and-beam-divergence/>, 2022.
16. T.-H. Kim, T.-H. Park, Placement optimization of multiple lidar sensors for autonomous vehicles, *IEEE Transactions on Intelligent Transportation Systems* 21 (2019) 2139-2145.

17. K. Matsubara, K. Nagatani, Y. Hirata, Improvement in measurement area of 3d lidar for a mobile robot using a mirror mounted on a manipulator, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, pp. 9523-9530. URL: <https://doi.org/10.1109/IROS45743.2020.9340822>. doi:10.1109/IROS45743.2020.9340822.
18. J. E. Kibii, A. Dreher, P. L. Wormser, H. Gimpel, Design and calibration of plane mirror setups for mobile robots with a 2d-lidar, *Sensors* 22 (2022) 7830. URL: <https://www.mdpi.com/1424-8220/22/20/7830>. doi:10.3390/s22207830.
19. M. Pogacnik, M. Munih, Towards a multi-perspective time of flight laser ranging device based on mirrors and prisms, *Applied Sciences* 12 (2022) 7121. URL: <https://www.mdpi.com/2076-3417/12/14/7121>. doi:10.3390/app12147121.
20. J. Morales, V. Plaza-Leiva, A. Mandow, J. A. Gomez-Ruiz, J. Seron, A. Garria-Cerezo, Analysis of 3d scan measurement distribution with application to a multi-beam lidar on a rotating platform, *Sensors* 18 (2018) 395. doi:10.3390/s18020395.
21. Y. Zhou, et al., End-to-end multi-view fusion for 3d object detection in lidar point clouds, in: Proceedings of the 3rd Conference on Robot Learning (CoRL 2019), volume 100 of *Proceedings of Machine Learning Research*, 2019, pp. 923-932. URL: <https://proceedings.mlr.press/v100/zhou20a.html>.
22. K. Schrab, F. Hilgerdenaar, R. Protzmann, I. Radusch, Improving object detection through multi-perspective lidar fusion, in: Proceedings of the 11th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS), 2025, pp. 448-457. doi:10.5220/0013283600003941.
23. C. Huang, LiDAR-based 3D Perception from Multi-frame Point Clouds for Autonomous Driving, Ph.d. thesis, University of Waterloo, Waterloo, Ontario, Canada, 2025.
24. D. Lee, M. Jung, W. Yang, A. Kim, Lidar odometry survey: Recent advancements and remaining challenges, *Intelligent Service Robotics* 17 (2024) 95-118. doi:10.1007/S11370-024-00515-8.
25. U. R. Kumar, I. Coudron, S. Puttemans, P. Vandewalle, Quality assessment for 3d reconstruction of building interiors, *Journal of Imaging Science and Technology* 62 (2018) 010402-1-010402-10. URL: <https://library.imaging.org/ei/articles/62/1/010402>. doi:10.2352/J. ImagingSci.Technol.2018.62.1.010402.
26. S. O. Elberink, G. Vosselman, Quality analysis on 3d building models reconstructed from airborne laser scanning data, *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (2012) 157-165. doi:10.1016/j.isprsjprs.2011.10.005.
27. B. Mitka, P. Klapa, P. Gawronek, Laboratory tests of metrological characteristics of a non-repetitive low-cost mobile handheld laser scanner, *Sensors* 24 (2024) 6010. URL: <https://www.mdpi.com/1424-8220/24/18/6010>. doi:10.3390/s24186010.
28. B. Ramlal, D. Archibald, R. Al-Tahir, M. Sutherland, D. Davis, Assessment of multi-perspective and multi-sensor data for urban strata mapping, *Research Square* (2023). doi: 10.21203/rs.3.rs-2599958/v1.
29. D. Yang, W. J. Li, M. Chen, X. Yu, Q. Chen, G. Zhang, Enhancing point cloud density using mems mirror-based focusing for lidar systems, *Research Square* (2024). doi:10.21203/rs.3.rs-4652241/v1.
30. C. Kelly, B. Wilkinson, A. Abd-Elrahman, O. Cordero, H. A. Lassiter, Accuracy assessment of low-cost lidar scanners: An analysis of the velodyne hdl-32e and livox mid-40's temporal stability, *Remote Sensing* 14 (2022) 4220. doi: 10.3390/rs14174220.
31. A. Shaker, W. Y. Yan, N. El-Ashmawy, The effects of laser reflection angle on radiometric correction of the airborne lidar intensity data, in: *ISPRS Archives*, volume XXXVIII-5/W12, 2011, pp. 213-218.
32. J. Goettsche, S. Alexopoulos, A. Dummmler, S. K. Maddineni, Multi-mirror array calculations with optical error, in: *SolarPACES 2020 International Conference on Concentrating Solar Power and Chemical Energy Systems*, 2020. URL: <https://www.solarpaces-conference.org/>, paper ID 02005.
33. TopoDOT, Establishing requirements, extracting metrics and evaluating quality of lidar data, 2015. URL: <https://blog.topodot.com/establishing-requirements-extracting-metrics-and-evaluating-quality-of-lidar-data/>, techNote #1021.
34. Y. Li, L. Ma, C. Wang, J. Li, M. Chapman, An overview of lidar requirements and techniques for autonomous driving, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, pp. 1-7. URL: <https://ieeexplore.ieee.org/document/9435580>. doi:10.1109/ICRA48506.2021.9435580.
35. P. Trybala, J. Szrek, B. Dfbogorski, B. Ziftek, J. Blachowski, J. Wodecki, R. Zimroz, Analysis of lidar actuator system influence on the quality of dense 3d point cloud obtained with slam, *Sensors* 23 (2023) 721. URL: <https://www.mdpi.com/1424-8220/23/2/721>. doi:10.3390/s23020721.
36. V. Kroumov, F. Smarandache, Simulation environment for mobile robots testing using ros and gazebo, in: 2016 20th International Conference on System Theory, Control and Computing (ICSTCC), 2016, pp. 96-101. doi:10.1109/ICSTCC.2016.7790647.
37. A. Lazaro Enguita, UGV Navigation in ROS using LIDAR 3D, Master's thesis, Universidad de Alcala, 2016.
38. M. Wang, Y.-H. Tseng, Automatic segmentation of lidar data into coplanar point clusters using an octree-based split-and-merge algorithm, *Photogrammetric Engineering & Remote Sensing* 76 (2010) 407-420.
39. H. Sier, et al., A benchmark for multi-modal lidar slam with ground truth in gnss-denied environments, *Remote Sensing* 15 (2023) 3314. doi:10.3390/rs15133314.

- 
40. S. Manivasagam, S. Wang, K. Wong, W. Zeng, M. Sazanovich, S. Tan, B. Yang, W.-C. Ma, R. Urtasun, Lidarsim: Realistic lidar simulation by leveraging the real world, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11167-11176. URL: [https://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Manivasagam\\_LiDARsim\\_Realistic\\_LiDAR\\_Simulation\\_by\\_Leveraging\\_the\\_Real\\_World\\_CVPR\\_2020\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2020/html/Manivasagam_LiDARsim_Realistic_LiDAR_Simulation_by_Leveraging_the_Real_World_CVPR_2020_paper.html).
  41. Y.-C. Lin, J. Shao, S.-Y. Shin, Z. Saka, M. Joseph, R. Manish, S. Fei, A. Habib, Comparative analysis of multi-platform, multi-resolution, multi-temporal lidar data for forest inventory, *Remote Sensing* 14 (2022) 649. doi:10.3390/rs14030649.
  42. L. Ma, Y. Li, J. Li, C. Wang, R. Wang, M. A. Chapman, Mobile laser scanned point-clouds for road object detection and extraction: A review, *Remote Sensing* 10 (2018) 1531. URL: <https://www.mdpi.com/2072-4292/10/10/1531>. doi:10.3390/rs10101531.
  43. H. You, F. Xu, Y. Ye, P. Xia, J. Du, Adaptive lidar scanning based on rgb information, *Automation in Construction* (2024). doi: 10.1016/j.autcon.2024.105265. URL: <https://www.sciencedirect.com/science/article/pii/S0926580524000736>.
  44. F. Noichl, D. D. Lichti, A. Borrmann, Automating adaptive scan planning for static laser scanning in complex 3d environments, *Automation in Construction* 165 (2024) 105511. doi:10.1016/j.autcon.2024.105511.

## Сегментация изображений геологических аншлифов с использованием дообученных визуально-языковых моделей

С. Д. Загайнов<sup>1</sup>, Д. М. Коршунов<sup>2</sup>, А. В. Хвостиков<sup>1</sup>

<sup>1</sup> Московский государственный университет имени М.В. Ломоносова, Москва, Россия

<sup>2</sup> Геологический институт РАН, Москва, Россия

**Аннотация.** В работе предложен метод семантической сегментации минералов на изображениях геологических аншлифов, основанный на текстовых описаниях. Стандартные подходы к сегментации с открытым словарём (Open-Vocabulary Segmentation), использующие модели типа CLIP, показывают низкую эффективность на узкоспециализированных данных из-за того, что обучающие выборки общего назначения, на которых они обучались, не содержат достаточного количества примеров из целевой предметной области. Для решения этой проблемы предложен двухэтапный подход: сначала модель CLIP дообучается на целевом домене с использованием набора пар «изображение-текст», собранных с ресурса Mindat.org. Затем дообученная модель интегрируется в качестве энкодера признаков в фреймворк Trident, не требующий дополнительного обучения. Эксперименты показывают, что предложенный подход значительно улучшает качество сегментации по сравнению с использованием базовой модели CLIP. Метрика F1 для классификации минералов на тестовых изображениях выросла с 0.16 до 0.43, а индекс Жаккара – с 0.10 до 0.29.

**Ключевые слова:** семантическая сегментация, сегментация с открытым словарём, геологические аншлифы, CLIP, глубокое обучение, компьютерное зрение.

## Segmentation of geological polished section images using fine-tuned vision-language models

S. D. Zagaynov<sup>1</sup>, D. M. Korshunov<sup>2</sup>, A. V. Khvostikov<sup>1</sup>

<sup>1</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> Geological Institute, Russian Academy of Sciences, Moscow, Russia

**Abstract.** This paper proposes a method for semantic segmentation of minerals in geological thin section images based on text descriptions. Standard Open-Vocabulary Segmentation approaches using CLIP-like models show low performance on highly specialized data due to the lack of relevant knowledge in general-purpose training datasets. To address this issue, a two-stage approach is proposed: first, the CLIP model is fine-tuned on the target domain using a custom dataset of image-text pairs collected from Mindat.org. Second, the fine-tuned model is integrated as a feature encoder into the state-of-the-art Training-Free framework, Trident. Experiments demonstrate that the proposed approach significantly improves segmentation quality compared to using the baseline CLIP model. The F1-score for mineral classification on test images increased from 0.16 to 0.43, and the Jaccard index increased from 0.10 to 0.29.

**Keywords:** semantic segmentation, open-vocabulary segmentation, geological thin sections, CLIP, deep learning, computer vision.

### Введение

Семантическая сегментация – задача разделения изображения на семантически значимые области – является фундаментальной проблемой в компьютерном зрении с широким спектром применений. Цель состоит в том, чтобы для каждого пикселя изображения определить, к какому классу объектов он принадлежит. Однако традиционные подходы к семантической сегментации сталкиваются с существенными ограничениями. Во-первых, они оперируют в рамках закрытого набора предопределенных классов, что делает невозможным сегментацию новых, ранее не встречавшихся категорий. Во-вторых, создание точных попиксельных масок для обучения таких моделей представляет собой чрезвычайно трудоемкий и дорогостоящий процесс.

Эта проблема особенно актуальна для специализированных областей, таких как анализ геологических аншлифов, где важно иметь попиксельную карту для изображений аншлифов, в которой каждый цвет соответствует определенному минералу (рис. 1). Однако создание таких эталонных масок,



как на примере из набора LumenStone<sup>1</sup>, требует глубоких экспертных знаний в области минералогии и является главным препятствием для использования стандартных методов обучения с учителем.

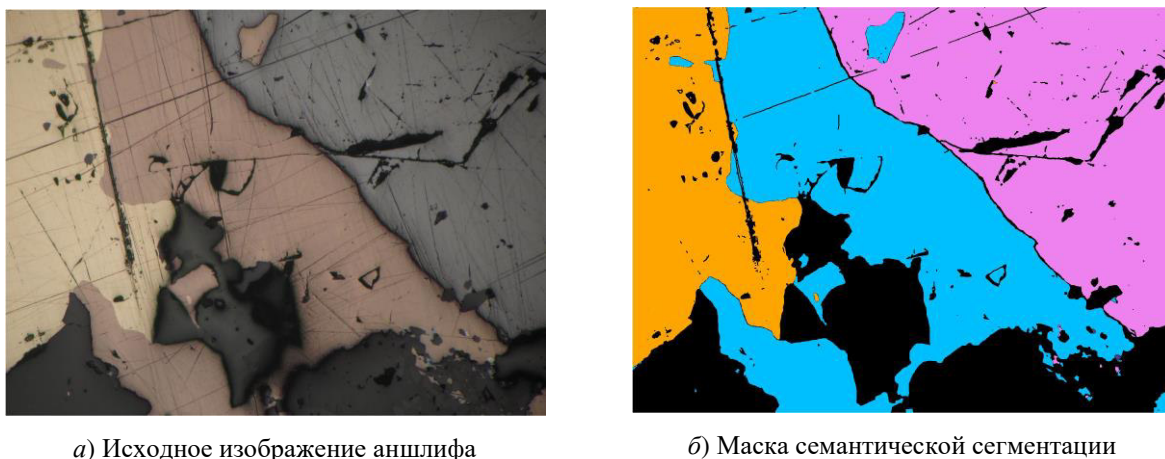


Рис. 1. Пример желаемого результата семантической сегментации для геологического изображения. Каждому цвету на маске (б) соответствует определенный минерал, присутствующий на исходном изображении (а). Данные из набора LumenStone S1

В настоящее время уже предложено несколько сегментационных моделей, построенных на классических методах компьютерного зрения (ResUNet [1, 2]), однако такие модели способны лишь определять классы минералов, а задачи структурно-текстурного анализа остаются за пределами их возможностей. Хотя именно на выделении характерных «текстурных рисунков» строятся генетические интерпретации процессов рудообразования.

Для преодоления этих ограничений активно развивается направление семантической сегментации с открытым словарём (Open-Vocabulary Semantic Segmentation, OVS). Цель OVS – создание моделей, способных сегментировать на изображении объекты любых произвольных классов, заданных текстовым описанием, а не только тех, которые были заранее определены в обучающем наборе данных. Это достигается за счёт использования мощных предварительно обученных визуально-языковых моделей (Vision Language Models, VLMs), таких как CLIP [3], которые научились сопоставлять визуальные концепции с текстом на огромных массивах данных из интернета.

Несмотря на успехи, прямая адаптация моделей типа CLIP для задач сегментации сопряжена с трудностями. Их архитектура нацелена на классификацию всего изображения и порождает пространственно-инвариантные признаки, что мешает точной локализации объектов. Современные подходы, не требующие дополнительного обучения (Training-Free), пытаются обойти это ограничение. Одним из передовых решений является фреймворк Trident [4], который эффективно комбинирует возможности CLIP (для семантического понимания), DINO [5] (для получения пространственно-локализованных признаков) и Segment Anything Model (SAM) [6] (для генерации высококачественных масок). Такие методы позволяют достичь высокой гибкости и снизить затраты на подготовку данных.

Однако общим недостатком всех методов, основанных на CLIP, является их зависимость от знаний, заложенных в модель при предварительном обучении. Если целевые классы объектов очень специфичны и редко встречаются на фотографиях из интернета (например, геологические минералы, видимые под микроскопом), то качество сегментации будет низким. Модель может неверно интерпретировать текстовые запросы или не находить соответствующие визуальные признаки.

Данная работа посвящена решению проблемы сегментации минералов на изображениях геологических аншлифов. Актуальность этой задачи подтверждается активными исследованиями в данной области, в том числе разработкой специализированных нейросетевых методов, использующих оптические свойства минералов при съемке в разных режимах поляризации для повышения точности [7]. Однако фундаментальным ограничением таких подходов является их зависимость от обучения с

<sup>1</sup> <https://imaging.cs.msu.ru/en/research/geology/lumenstone>

учителем, которое требует наличия больших наборов данных с точной попиксельной разметкой. Создание подобных эталонных наборов, как, например, LumenStone [1], само по себе является отдельной трудоёмкой научной задачей. Эта зависимость от дорогостоящей ручной разметки ограничивает масштабируемость и адаптацию существующих методов к широкому спектру разнообразных и редких минералов, что и обуславливает актуальность разработки принципиально иных подходов, исследуемых в данной работе.

Основной вклад авторов данной статьи заключается в разработке и оценке метода, который адаптирует передовой OVS-подход к узкоспециализированной геологической предметной области. В данной работе предлагается двухэтапный подход.

1. Дообучение (fine-tuning) модели CLIP на целевом домене с использованием специализированного набора данных пар «изображение-текст», собранного из геологических источников.

2. Интеграция дообученной модели в качестве энкодера признаков в фреймворк Trident для выполнения семантической сегментации минералов.

### Обзор существующих методов

#### 1. Семантическая сегментация с открытым словарем без дополнительного обучения

Подходы, не требующие дополнительного обучения (Training-Free), представляют особый интерес, поскольку они позволяют избежать затрат на аннотирование данных и сохраняют обобщающие способности исходных моделей.

Trident [4] является одной из передовых работ в этой области. Фреймворк направлен на решение проблем пространственной инвариантности и ограниченного разрешения CLIP при работе с изображениями высокого разрешения. Trident предлагает парадигму «Splice-then-Segment»: сначала извлекаются признаки из частей изображения с помощью CLIP и DINO, а затем эти признаки объединяются в единую карту. Ключевым шагом является использование энкодера SAM для создания матрицы корреляции, которая применяется к этой карте, обеспечивая глобальную агрегацию информации. Это позволяет успешно сегментировать объекты даже на изображениях высокого разрешения. Кроме того, Trident предлагает стратегию уточнения грубых результатов сегментации, преобразуя их в промпты (точки, рамки, маски) для декодера SAM.

Другой Training-Free подход представлен в работе TextRegion [8]. Вместо работы с признаками на уровне отдельных пикселей или патчей, этот метод сначала использует мощную модель сегментации (например, SAM2 [9]) для деления изображения на множество регионов-кандидатов. Затем для каждого региона признаки из визуально-языковой модели (например, CLIP) агрегируются (путем взвешенного суммирования) в единый «региональный токен». Этот токен напрямую сравнивается с эмбедингом текстового запроса, превращая задачу сегментации в задачу классификации регионов. Однако применимость этого подхода к нашей задаче анализа геологических аншлифов ограничена, что проявляется в двух основных аспектах:

- зависимость от качества начальной сегментации. Эффективность метода полностью определяется тем, насколько хорошо SAM2 может выделить все потенциально значимые минералы. На сложных текстурах аншлифов, где границы между минералами могут быть нечеткими или сложными, SAM2 может допускать ошибки, которые невозможно будет исправить на последующих этапах;

- проблема отсутствия фона. Подход TextRegion использует специальные техники, чтобы отфильтровать признаки, относящиеся ко всему изображению в целом (глобальные признаки), и сосредоточиться на локальных. На изображениях аншлифов часто нет фона в привычном понимании: вся область снимка заполнена разными минералами. В таких условиях, во-первых, фильтрация глобальных признаков может работать некорректно. Во-вторых, усреднение признаков внутри маски, которая по ошибке захватила несколько разных минералов, приведет к потере уникальных текстурных характеристик каждого из них. В результате модель не сможет правильно классифицировать такой регион.

## 2. Слабоконтролируемая семантическая сегментация с открытым словарем

Методы этого типа стремятся снизить зависимость от полной попиксельной разметки, используя более доступные формы контроля, такие как пары изображение-текст или неразмеченные данные.

OVSegmentor [10] предлагает решение, обучаемое исключительно на парах «изображение-текст». В его основе лежит трансформерная архитектура, которая с помощью механизма slot-attention [11] группирует визуальные патчи в групповые токены, представляющие семантические области, и сопоставляет их с эмбедами текста. Для обучения используются прокси-задачи, такие как «завершение замаскированных сущностей», что заставляет модель устанавливать точное соответствие между визуальными группами и текстовыми понятиями, и «согласованность масок между изображениями»: модель учится генерировать схожие сегментационные маски для одной и той же сущности на разных изображениях.

CLIP-DINOiser [12] улучшает локализационные способности признаков из MaskCLIP [13], интегрируя в них информацию от модели DINO. Для этого обучаются легковесные свёрточные слои, которые учатся предсказывать «карты объектности» и улучшать пространственные характеристики признаков CLIP, используя DINO и метод FOUND [14] в качестве «учителей» на неразмеченных данных.

TCL [15] решает проблему несоответствия между обучением на уровне всего изображения и сегментацией на уровне регионов. Ключевым элементом TCL является обучаемый декодер, выполняющий функцию семантической локализации (text grounding). Он генерирует маску, указывающую на области изображения, которые семантически связаны с данным текстом. Визуальные эмбединги из этих областей напрямую сопоставляются с текстом с помощью контрастивной функции потерь. Однако данный метод плохо подходит для задачи сегментации минералов, так как его функция потерь предполагает, что на изображении присутствует только один целевой объект, и наказывает за наличие других объектов, которые могут быть целевыми в других примерах батча (пакета).

Общий вывод из анализа существующих решений заключается в том, что, несмотря на прогресс, их эффективность в узкоспециализированных областях остаётся ограниченной из-за отсутствия этих знаний у базовой модели CLIP. Наша работа нацелена на устранение именно этого недостатка.

## Методология

Предлагаемый в данной работе метод состоит из трёх основных этапов: сбор и подготовка данных для геологического домена, дообучение модели CLIP на этих данных и интеграция адаптированной модели в фреймворк Trident для выполнения сегментации.

### 1. Сбор и подготовка набора данных

Для адаптации модели к предметной области был сформирован специализированный набор данных. Источником послужила онлайн-база по минералогии Mindat.org<sup>1</sup>. Основное внимание было уделено изображениям полированных срезов (аншлифов) и тонких срезов (шлифов) из категорий «Polished Section (Polarized Light)» и «Polished Section (Slice or Surface)», каждый такой снимок сопровождается списком присутствующих минералов и описанием.

Процесс сбора и обработки данных включал следующие шаги.

1. Сбор «сырых» данных: с помощью автоматизированного парсинга были извлечены изображения, их оригинальные текстовые описания и списки присутствующих минералов.

2. Извлечение названий минералов: с помощью регулярных выражений из текстовой информации были извлечены наименования минералов.

3. Фильтрация классов: был проведён частотный анализ, по результатам которого отобраны топ-20 наиболее часто встречающихся минералов (Arsenopyrite, Bismuth, Bornite, Calcite, Chalcedony, Chalcocite, Chalcopyrite, Covellite, Dolomite, Galena, Goethite, Hematite, Magnetite, Nickeline, Pyrite, Pyrrhotite, Quartz, Rammelsbergite, Sphalerite, Tetrahedrite Subgroup). Это было сделано для обеспечения достаточного количества примеров для каждого класса и упрощения задачи на начальном этапе исследования.

---

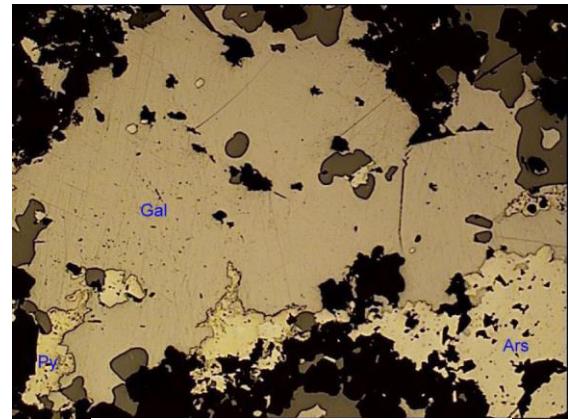
<sup>1</sup> <https://www.mindat.org/>

4. Формирование текстовых описаний: для каждого изображения, содержащего отображенные минералы, было сформировано стандартизированное описание: "Minerals on this photo: <минерал\_1, минерал\_2, ...>. Description: <оригинальное текстовое описание>". Такой формат предоставляет модели как явную информацию о целевых классах, так и дополнительный геологический контекст.

В результате был сформирован набор данных, состоящий из 9884 пар «изображение-текст», который далее использовался для дообучения модели CLIP. Примеры таких пар показаны на рисунке 2.



**а)** Minerals on this photo: Galena. Description: Felted acicular zinkenite crystals (greenish- to blueish- to brownish-grey anisotropy with some specs of red internal reflections) in association with galena (grey, bottom). Vertically reflected plane polarized light digital image in air, with crossed polarizers, width 0.5 mm



**б)** Minerals on this photo: Galena, Pyrite, Arsenopyrite. Description: Polished section with galena (Gal) and some pyrite (Py) and arsenopyrite (Ars)

Рис. 2. Примеры пар «изображение-текст» из собранного набора данных (текстовые описания изображений на английском языке)

## 2. Архитектура и дообучение модели CLIP

Модель CLIP [3] состоит из двух энкодеров: визуального (Image Encoder) и текстового (Text Encoder), которые преобразуют изображения и текст в эмбединги в общем векторном пространстве. В данной работе в качестве визуального энкодера использовалась архитектура Vision Transformer (ViT-V/16) [16], а в качестве текстового – трансформер на основе архитектуры GPT-2 [17].

Обучение модели основано на контрастивном подходе, целью которого является максимизация сходства между эмбедингами соответствующих пар «изображение-текст» и минимизация для несоответствующих. Для пакета из  $N$  пар вычисляется матрица косинусных сходств  $S$  размером  $N \times N$ . Затем вычисляется симметричная контрастивная функция потерь:

$$L_{\text{img}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(S_{ii}/\tau)}{\sum_{j=1}^N \exp(S_{ij}/\tau)},$$

$$L_{\text{text}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(S_{ii}/\tau)}{\sum_{j=1}^N \exp(S_{ji}/\tau)},$$

$$L_{\text{CLIP}} = \frac{L_{\text{img}} + L_{\text{text}}}{2},$$

где  $\tau$  – обучаемый параметр.

Для эффективной адаптации модели с меньшими вычислительными затратами и для снижения риска катастрофического забывания была применена техника параметро-эффективной настройки (Parameter-Efficient Fine-Tuning, PEFT), а именно метод LoRA (Low-Rank Adaptation) [18]. Ключевая идея LoRA заключается в модификации выходного вектора  $h$  каждого адаптируемого слоя путем добавления к нему низкоранговой поправки. Вместо прямого дообучения исходной матрицы весов  $W_0$  её оставляют замороженной, а выход слоя вычисляется по формуле

$$h = W_0 x + \frac{\alpha}{r} B A x.$$



В данной формуле  $x$  – это входной вектор, а  $W_0x$  – результат работы исходного, предварительно обученного слоя. Адаптация достигается за счёт второго члена, где обучаемыми являются только две новые, низкоранговые матрицы  $A \in \mathbb{R}^{r \times k}$  и  $B \in \mathbb{R}^{d \times r}$ . Ключевой гиперпараметр, ранг адаптации  $r$  (где  $r \ll C \min(d, k)$ ), определяет внутреннюю размерность этих матриц, то есть контролирует количество обучаемых параметров и сложность аппроксимации. В работе было выбрано стандартное значение  $r = 32$ . Второй гиперпараметр,  $\alpha$ , является коэффициентом масштабирования, который регулирует итоговую величину вклада от адаптера, позволяя управлять балансом между исходными знаниями модели и новыми. Использовалось стандартное значение  $\alpha = 64$ .

LoRA-адаптеры с указанными гиперпараметрами были добавлены к линейным слоям в блоках внимания и полносвязных слоях (MLP) обоих энкодеров: визуального и текстового. Дообучение проводилось 5 эпох с размером пакета 32, начальной скоростью обучения  $1e-6$ , оптимизатором AdamW [19] и планировщиком OneCycleLR [20].

### 3. Интеграция в фреймворк Trident

Дообученная модель CLIP была интегрирована в фреймворк Trident, заменив собой базовую модель. Trident реализует парадигму Splice-then-Segment, общая архитектура которой показана на рисунке 3. Метод состоит из четырёх основных этапов, которые были адаптированы для рассматриваемой задачи.

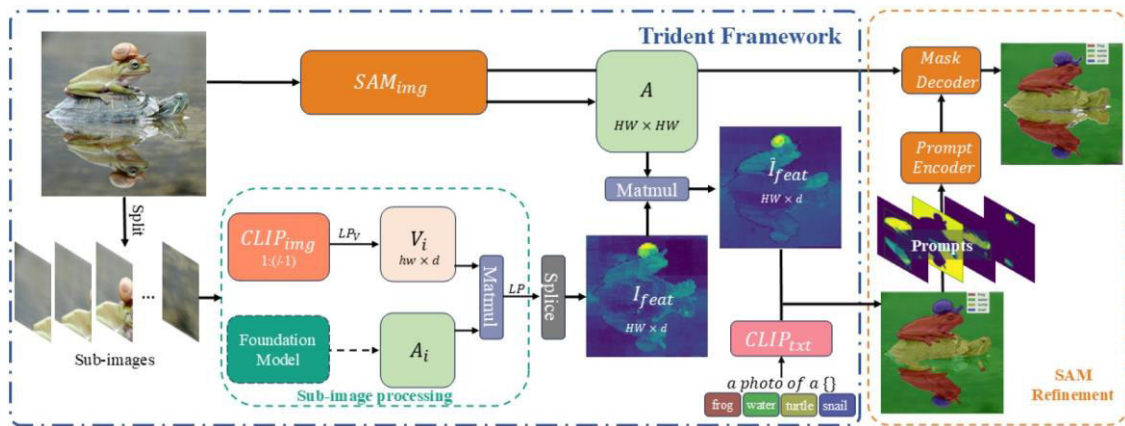


Рис. 3. Общая архитектура фреймворка Trident. Дообученная модель CLIP используется на этапе извлечения семантических признаков, которые затем обрабатываются и уточняются с помощью DINO и SAM (иллюстрация из статьи [4])

1. **Обработка субизображений.** Исходное изображение высокого разрешения  $I^{src}$  делится на множество  $n$  перекрывающихся субизображений (патчей) с помощью скользящего окна. Размер патчей соответствует входному разрешению энкодера CLIP ( $224 \times 224$  пикселя), а перекрытие обеспечивает сохранение контекста между соседними областями. Для каждого патча извлекаются два типа признаков: семантические признаки  $V_i$  с помощью энкодера дообученной модели CLIP и пространственно-локализованные признаки  $A_i$ ; от модели DINO. Эти признаки объединяются, чтобы дополнить высокоуровневые семантические признаки пространственной детализацией, и затем соединяются в одну общую карту признаков  $I^{feat}$ .

2. **Глобальная агрегация.** Энкодер SAM обрабатывает полное исходное изображение  $I^{src}$  для генерации корреляционной матрицы  $A$ , которая улавливает семантические отношения между пикселями. Эта матрица  $A$  затем умножается на общую карту признаков  $I^{feat}$ , агрегируя информацию в глобальном масштабе:  $\tilde{I}^{feat} = A \cdot I^{feat}$ , где  $\cdot$  – матричное умножение. Это позволяет эффективно расширить рецептивное поле и учесть контекст всего изображения.

3. **Сегментация.** Агрегированные признаки  $\tilde{I}^{feat}$  сравниваются с текстовыми эмбедами  $T_k^{emd}$  для каждого целевого класса  $k$  (названия минерала), которые также генерируются текстовым энкодером (дообученной моделью CLIP). Начальная карта сегментации  $S$  получается путём присвоения каждому пикселю метки класса с максимальным косинусным сходством:

$$S(x, y) = \underset{k}{\operatorname{argmax}} \cos(\tilde{I}^{feat}(x, y), T_k^{emd}).$$

4. **Уточнение с помощью SAM.** Полученная карта сегментации  $S$  может быть грубой и неточной по краям. Для её уточнения Trident использует декодер и промпт-энкодер модели SAM. Поскольку SAM наиболее эффективен при работе с точечными и рамочными промптами, Trident преобразует начальные маски для каждого класса в комбинацию из трёх типов промптов.

1) **Точечные промпты:** точки внутри найденной области, где уверенность модели в предсказании класса максимальна. Это помогает SAM сфокусироваться на наиболее репрезентативных частях объекта.

2) **Рамочные промпты:** минимальный ограничивающий прямоугольник (bounding box), который полностью охватывает найденную область (задаёт общие границы объекта).

3) **Масочные промпты:** грубая бинарная маска

$$B_k(x, y) = \begin{cases} 1, & \text{если } S(x, y) = k \\ 0, & \text{иначе} \end{cases},$$

которая указывает на примерное расположение и форму объекта.

Одновременная подача этих промптов в SAM позволяет сгенерировать финальные, значительно более точные и качественные, маски для каждого объекта.

### Эксперименты и результаты

#### 1. Детали реализации

Все эксперименты по дообучению модели проводились на одном графическом ускорителе NVIDIA A6000. Процесс дообучения на собранном наборе данных длился 10 эпох и занял около 3 часов. Оценка производительности и визуализация результатов выполнялись на той же аппаратной конфигурации.

#### 2. Настройка экспериментов

Эксперименты проводились для сравнения двух конфигураций.

1) **Базовый Trident:** использует оригинальную модель CLIP (ViT-B/16), обученную на LAION-400M.

2) **Предложенный метод (Trident + Fine-tuned CLIP):** использует модель CLIP, дообученную на наборе данных, полученном на основе Mindat.org.

Оценка проводилась на отложенной тестовой выборке из собранного набора данных. Для каждого изображения в качестве текстовых запросов подавались названия минералов, которые, согласно разметке, присутствуют на нём. Таким образом, задача семантической сегментации минералов при таком упрощении сводится к задаче классификации.

#### 3. Метрики оценки

Поскольку получение точных попиксельных масок для тестового набора затруднительно, то задача оценивалась как мультиклассовая классификация на уровне всего изображения. Для каждого изображения получался набор предсказанных классов (минералов, для которых были сгенерированы непустые маски) и сравнивался его с истинным набором классов. Использовались следующие метрики:

1) **Jaccard (индекс Жаккара):**  $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$ , где  $A$  – набор предсказанных,  $B$  – набор истинных классов;

2) **Recall (полнота):**  $R = \frac{TP}{TP + FN}$ , доля найденных истинных классов;

3) **Precision (точность):**  $P = \frac{TP}{TP + FP}$ , доля верных среди предсказанных классов;

4) **F1-score:**  $F1 = 2 \cdot \frac{P \cdot R}{P + R}$ , гармоническое среднее точности и полноты.

#### 4. Количественные результаты

Результаты количественной оценки эффективности предложенного метода на отложенной тестовой выборке из собранного нами набора данных с Mindat.org представлены в таблицах 1 и 2.

Таблица 1. Метрики классификации для базового подхода Trident

Название метрики	Значение
Jaccard	0.1006
Recall	0.3106
Precision	0.1278
F1-score	0.1645

Таблица 2. Метрики классификации для предложенного метода

Название метрики	Значение
Jaccard	0.2929
Recall	0.5795
Precision	0.4405
F1-score	0.4326

Таблицы показывают, что дообучение модели CLIP на целевом домене привело к значительному улучшению всех метрик. F1-score увеличился (с 0.1645 до 0.4326), что указывает на существенное повышение способности модели правильно идентифицировать минералы. Увеличение Recall говорит о том, что модель стала реже пропускать присутствующие минералы, а рост Precision – о снижении количества ложных срабатываний.

#### 5. Визуальные результаты

Визуализация результатов (рис. 4-7) подтверждает полученные количественные оценки. На них продемонстрировано, как дообучение улучшает качество сегментации.

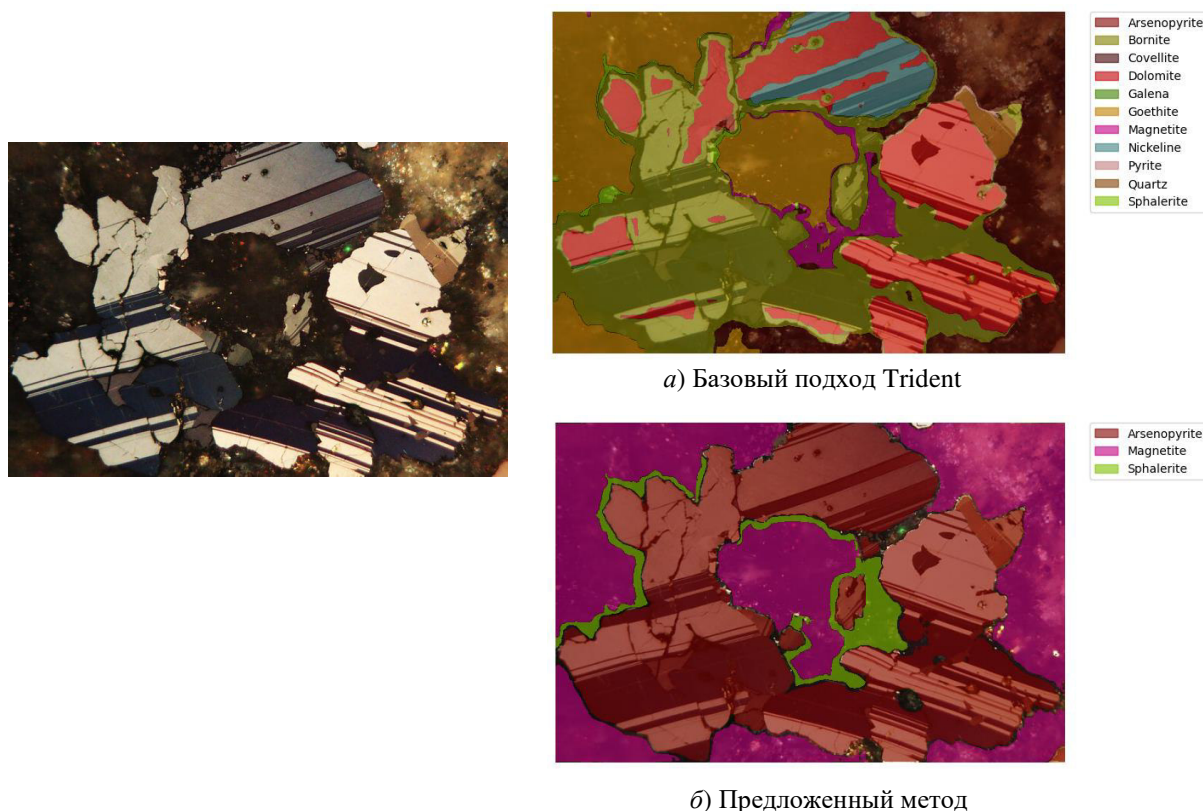
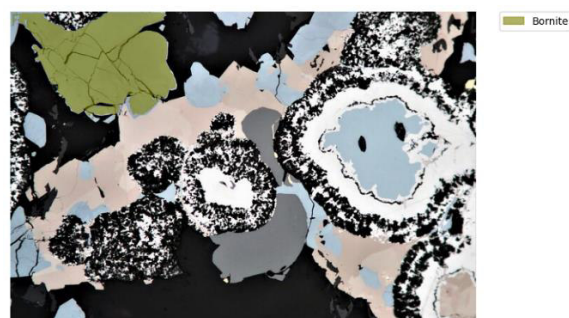
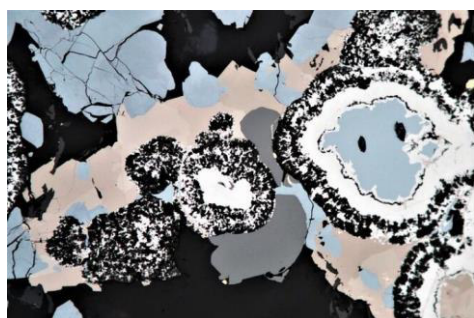
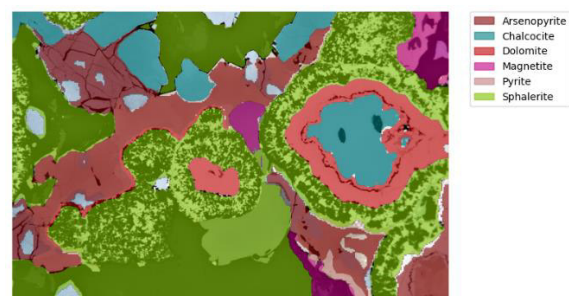


Рис. 4. Минералы, присутствующие на изображении: Sylvanite, Pyrite, Galena, Quartz. Слева – исходное изображение. Справа – изображения с наложенной маской сегментации



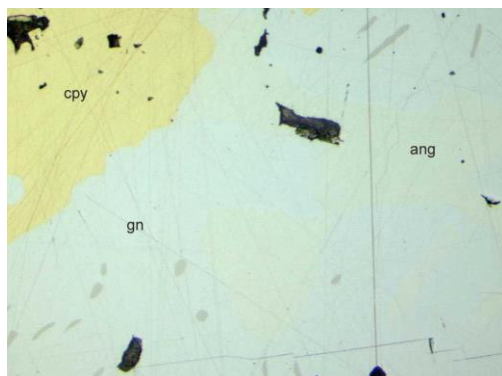


а) Базовый подход Trident



б) Предложенный метод

Рис. 5. Минералы, присутствующие на изображении: Chalcopyrite, Sphalerite. Слева – исходное изображение. Справа – изображения с наложенной маской сегментации



а) Базовый подход Trident



б) Предложенный метод

Рис. 6. Минералы, присутствующие на изображении: Chalcopyrite, Galena. Слева – исходное изображение. Справа – изображения с наложенной маской сегментации

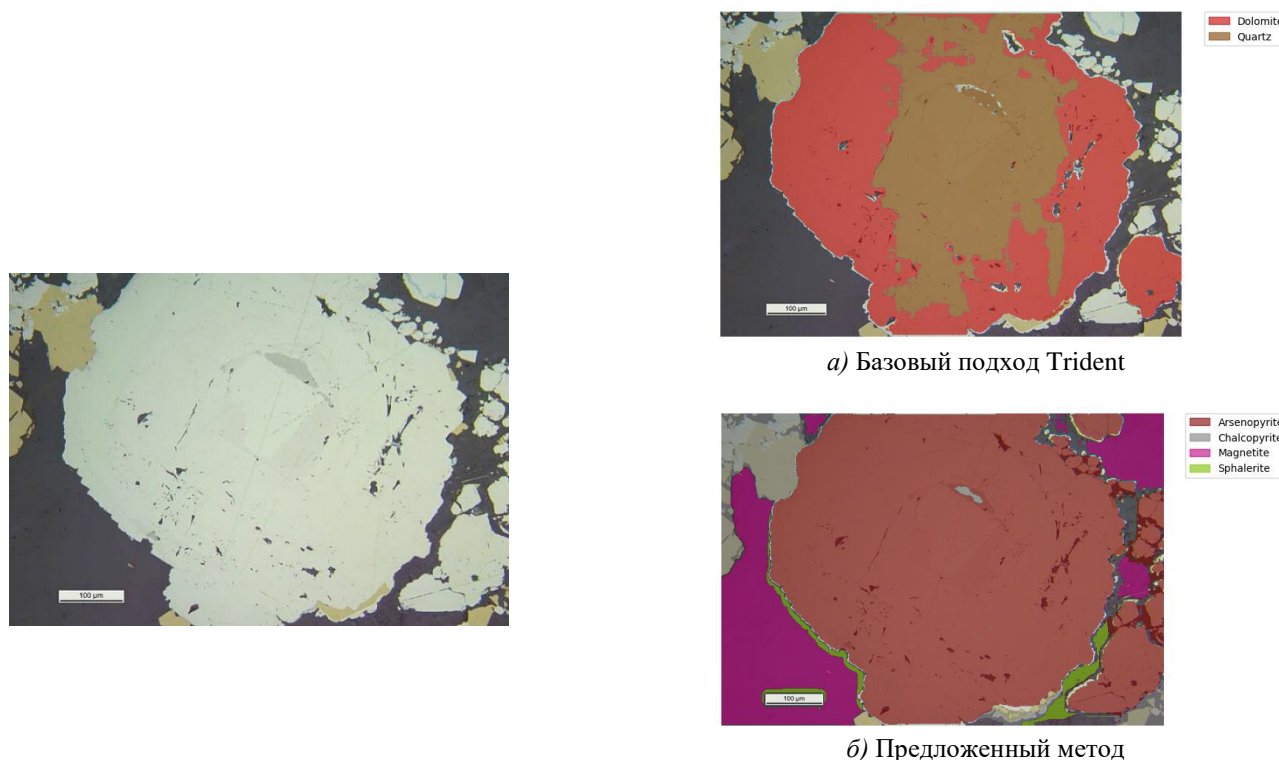


Рис. 7. Минералы, присутствующие на изображении: Pyrite, Chalcopyrite. Слева – исходное изображение. Справа – изображения с наложенной маской сегментации

Анализ визуальных результатов (см. рис. 4-7) позволяет качественно оценить преимущества предложенного подхода. На примерах видно, что дообученная модель демонстрирует более глубокое понимание геологического контекста, текстур и структур.

На рисунке 4 представлено изображение со сложной структурой, базовый метод (a) генерирует сильно фрагментированную маску с большим количеством ложных классов. В то же время предложенный метод (б) лучше справляется с задачей, формируя семантически верные маски, но также ошибается в определении классов.

На рисунке 5 базовый подход Trident (a) не может распознать сложную структуру, в то время как предложенный метод (б) сегментирует различные минералы, что говорит о его способности определять не только цвет, но и геологические паттерны.

Особенно ярко преимущество предложенного метода (a) видно на рисунке 6. Базовый метод допускает ошибки в классификации и генерирует шумные маски, тогда как предложенный метод (б) правильно определяет основные минералы, присутствующие на изображении.

На рисунке 7 базовый метод (a) ошибочно разделяет крупный фрагмент минерала на несколько разных классов, нарушая его целостность. Предложенный метод (б), напротив, корректно идентифицирует весь фрагмент как единый объект.

Таким образом, визуальные результаты показывают, что дообучение CLIP позволяет модели не просто лучше классифицировать минералы, но и генерировать более точные и геологически осмысленные маски сегментации, сохраняя целостность объектов и распознавая сложные структуры.

### Заключение

В данной работе был предложен и исследован метод семантической сегментации для узкоспециализированной области – анализа изображений геологических аншлифов. Ключевой идеей метода является адаптация мощной визуально-языковой модели CLIP к целевому домену путём дообучения на специально собранном наборе данных, с последующей её интеграцией в передовой OVS-фреймворк Trident.

Проведённые эксперименты убедительно доказывают применимость предложенного метода визуально-языковой модели сегментации для целей анализа изображений аншлифов. Дообучение CLIP на доменных данных значительно повышает качество сегментации по сравнению с использованием базовой, универсальной модели. Это подтверждается как существенным ростом количественных метрик (F1-score вырос с 0.16 до 0.43), так и улучшением визуальных результатов сегментации. Предложенный подход позволяет эффективно адаптировать существующие OVS-решения для задач в специализированных областях, где сбор больших наборов данных с попиксельной разметкой затруднён или невозможен.

При использовании данной сегментационной модели совместно с уже разработанными традиционными методами компьютерного зрения (например, ResUNet[1]), обученными на больших наборах данных аннотированных изображений аншлифов, в перспективе получится создать полноценный механизм автоматического описания руд под микроскопом, который будет способен не только определять минералы и статистически обрабатывать их распределения (содержания и группы по размерам), но и проводить полноценный структурно-текстурный анализ.

Предлагаемый метод был программно реализован на языке Python с использованием библиотек PyTorch, NumPy, Pillow, Skimage, а также открытых реализаций OpenCLIP и Trident.

### Дальнейшее развитие

Несмотря на положительные результаты, работа на текущий момент имеет некоторые ограничения. Во-первых, набор данных ограничен 20 наиболее частыми минералами. Во-вторых, оценка проводилась на уровне классификации изображений, а не попиксельного совпадения масок, из-за отсутствия точной разметки.

В качестве дальнейшего развития работы можно выделить следующие направления:

- расширение набора данных для включения более редких минералов;
- создание небольшого валидационного набора с точной попиксельной разметкой для оценки качества сегментационных масок с помощью метрики mIoU;
- исследование влияния различных архитектур VLM на итоговое качество сегментации минералов.

### Источник финансирования

Исследование выполнено за счёт гранта Российского научного фонда (проект № 24-21-00061).

### Список литературы

1. Automatic identification of minerals in images of polished sections / AV Khvostikov, DM Korshunov, AS Krylov, MA Boguslavskiy // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2021. Vol. 44. Pp. 113-118.
2. An improved mineral image recognition method based on deep learning / Huaming Tang, Hongming Wang, Ling Wang, et al. // *Jom*. 2023. Vol. 75, no. 7. Pp. 2590-2602.
3. Learning transferable visual models from natural language supervision / Alec Radford, Jong Wook Kim, Chris Hallacy, et al. // *International conference on machine learning / PmLR*. 2021. Pp. 8748-8763.
4. Shi Yuheng, Dong Minjing, Xu Chang. Harnessing Vision Foundation Models for High-Performance, Training-Free Open Vocabulary Segmentation // *arXiv preprint arXiv:2411.09219*. 2024.
5. Emerging properties in self-supervised vision transformers / Mathilde Caron, Hugo Touvron, Ishan Misra, et al. // *Proceedings of the IEEE/CVF international conference on computer vision*. 2021. Pp. 9650-9660.
6. Segment anything / Alexander Kirillov, Eric Mintun, Nikhila Ravi, et al. // *Proceedings of the IEEE/CVF international conference on computer vision*. 2023. Pp. 4015-4026.
7. Registration and segmentation of PPL and XPL images of geological polished sections containing anisotropic minerals / DI Razzhivina, DM Korshunov, MA Boguslavsky, et al. // *Computational Mathematics and Modeling*. 2023. Vol. 34, no. 1. Pp. 16-26.
8. TextRegion: Text-Aligned Region Tokens from Frozen Image-Text Models / Yao Xiao, Qiqian Fu, Heyi Tao, et al. // *arXiv preprint arXiv:2505.23769*. 2025.
9. Sam 2: Segment anything in images and videos / Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, et al. // *arXiv preprint arXiv:2408.00714*. 2024.

10. Learning open-vocabulary semantic segmentation models from natural language supervision / Jilan Xu, Junlin Hou, Yuejie Zhang, et al. // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023. Pp. 2935-2944.
11. *Object-Centric Learning with Slot Attention* / Locatello Francesco, Weissenborn Dirk, Unterthiner Thomas, et al. 2020.
12. Clip-diy: Clip dense inference yields open-vocabulary semantic segmentation for-free / Monika Wysoczanska, Michael Ramamonjisoa, Tomasz Trzcinski, Oriane Simeoni // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024. Pp. 1403-1413.
13. Learning to prompt for vision-language models / Kaiyang Zhou, Jingkang Yang, Chen Change Loy, Ziwei Liu // *International Journal of Computer Vision*. 2022. Vol. 130, no. 9. Pp. 2337-2348.
14. Unsupervised object localization: Observing the background to discover objects / Oriane Simeoni, Chloe Sekkat, Gilles Puy, et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. Pp. 3176-3186.
15. *Cha Junbum, Mun Jonghwan, Roh Byungseok*. Learning to generate text-grounded mask for open-world semantic segmentation from only image-text pairs // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. Pp. 11165-11174.
16. An image is worth 16x16 words: Transformers for image recognition at scale / Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, et al. // *arXiv preprint arXiv:2010.11929*. 2020.
17. Language models are unsupervised multitask learners / Alec Radford, Jeffrey Wu, Rewon Child, et al. // *OpenAI blog*. 2019. Vol. 1, no. 8. P. 9.
18. Lora: Low-rank adaptation of large language models / Edward J Hu, Yelong Shen, Phillip Wallis, et al. // *ICLR*. 2022. Vol. 1, no. 2. P. 3.
19. Loshchilov Ilya, Hutter Frank. *Decoupled weight decay regularization* // arXiv preprint arXiv:1711.05101. 2017.
20. *Smith Leslie N, Topin Nicholay*. Super-convergence: Very fast training of neural networks using large learning rates // Artificial intelligence and machine learning for multi-domain operations applications / SPIE. Vol. 11006. 2019. Pp. 369-386.

УДК 004.93'11 : 004.93'12 : 004.932.2 : 004.932.72'1

DOI: 10.25686/978-5-8158-2474-4-2025-607-618

## Модернизация системы управления для обнаружения края металлической полосы, свёрнутой в рулон, на агрегате продольной и поперечной резки листа с применением компьютерного зрения

Е. О. Зубова<sup>1,2</sup>, М. Н. Шамшин<sup>1</sup>, А. В. Рыбаков<sup>1,2</sup>

<sup>1</sup>АО «ОМК», Москва, Россия

<sup>2</sup>ВФ НИТУ «МИСИС», Выкса, Россия

**Аннотация.** В статье проведён анализ возможности применения компьютерного зрения для задачи обнаружения края металлической полосы, свёрнутой в рулон, на агрегате продольной и поперечной резки листа (АПР) при задаче этого края в агрегат. Для определения положения края полосы применяется цифровая видеокамера. В статье рассмотрены два алгоритма для обнаружения края полосы: алгоритм с применением Canny Edge Detector с последующим преобразованием Хафа и применением детектора углов Ши-Томаси и алгоритм с применением нейронной сети. Для первого алгоритма был предварительно проведён сравнительный анализ различных алгоритмов компьютерного зрения, критериями которого являлись точность обнаружения объекта и время обработки алгоритмом одного кадра. Для алгоритма с применением нейронной сети были предварительно выполнены обучение двух версий YOLO, YOLOv5 и YOLOv8 и их сравнительный анализ на основании полученных метрик. Проведён сравнительный анализ алгоритма с применением детекторов и алгоритма с применением нейронной сети для определения наиболее подходящего для решения поставленной задачи. Разработано программное обеспечение с интерфейсом на основе выбранного алгоритма, проведена интеграция разработанного программно-аппаратного комплекса с системой АСУ ТП АПР и осуществлен ряд экспериментов для определения качества работы системы.

**Ключевые слова:** компьютерное зрение, нейронные сети, цифровая обработка изображений, программно-аппаратный комплекс, система управления, агрегат продольной резки листа.

## Modernization of the control system for detecting the edge of a rolled metal strip on a sheet longitudinal and transverse cutting unit using computer vision

E. O. Zubova<sup>1,2</sup>, M. N. Shamshin<sup>1</sup>, A. V. Rybakov<sup>1,2</sup>

<sup>1</sup>OMK JSC, Moscow, Russia

<sup>2</sup>VF NUST MISIS, Vyksa, Russia

**Abstract.** The article analyzes the possibility of using computer vision for the task of detecting the edge of a rolled metal strip on a sheet longitudinal and transverse cutting unit (APR) when inserting this edge into the unit. A digital video camera is used to determine the position of the edge of the strip. The article discusses two algorithms for detecting the edge of the strip: an algorithm using the Canny Edge Detector followed by the Hough transform and the use of the Shi-Tomasi angle detector and an algorithm using a neural network. For the first algorithm, a preliminary comparative analysis of various computer vision algorithms was carried out, the criteria of which were the accuracy of object detection and the processing time of a single frame by the algorithm. For the algorithm using a neural network, two versions of YOLO, YOLOv5 and YOLOv8, were preliminarily trained, and their comparative analysis was based on the obtained metrics. A comparative analysis of the algorithm using detectors and the algorithm using a neural network is carried out to determine the most suitable one for solving the problem. Software with an interface based on the selected algorithm was developed, the integration of the developed software and hardware complex with the automated process control system of the APR was carried out, and a number of experiments were conducted to determine the quality of the system.

**Keywords:** computer vision, neural networks, digital image processing, software and hardware complex, control system, sheet slitting unit.

### Введение

В настоящее время компьютерное зрение (Computer Vision, CV) всё чаще применяется в промышленности, в том числе в металлургии [1]. Например, благодаря обработке кадра путём моделирования и вычитания фона (модель фона строится на основе усреднения кадров) можно автоматически обнаруживать поверхностные дефекты сварных соединений труб из нержавеющей стали [2]. Технологии CV используются для обнаружения сварного шва при лазерной сварке, для чего применяется нейронная сеть и классификатор для устранения аномальных элементов [3]. Компьютерное зрение применяется для обработки изображений спечённых поверхностей на основе алгоритма семантической сегментации и схемы полууправляемого глубокого обучения для

классификации, основанной на псевдометках [4]. При помощи Computer Vision осуществляют обнаружение поверхностных дефектов горячекатаной стали в роботизированных системах, основанное на применении механизмов внимания и расширенной свёртке [5]. Кроме того, разрабатываются системы обнаружения дефектов на малых зубчатых передачах, изготовленных методом порошковой металлургии, с применением свёрточных нейронных сетей [6]. Авторы [7] предлагают для решения проблем, связанных с низкой эффективностью и сильной субъективностью применяемых методов обнаружения дефектов сварных швов с помощью радиографической визуализации, использовать метод обнаружения дефектов сварных швов на основе многоканальной свёрточной нейронной сети. В [8] авторы решают проблему обработки рентгеновских снимков для обнаружения дефектов сварных швов, связанную с низкой контрастностью и большим разнообразием форм и размеров дефектов, при помощи модуля повышения контрастности кривой значений серого (GCE) и модели, специально разработанной для обнаружения дефектов на сварных швах (WD-YOLO). Авторы [9] предлагают повысить эффективность обнаружения поверхностных дефектов на металлических деталях при помощи метода, основанного на нейронной сети IE-YOLOv5. В работе [10] авторы предлагают новую сеть для обнаружения дефектов на металлических поверхностях под названием YOLOv-5s-FRN, которая решает проблему неэффективного обнаружения, характерную для традиционной системы ручного контроля. Другие исследователи [11] предлагают усовершенствованный метод обнаружения дефектов на металлических поверхностях на основе Cascade R-CNN. В работе [12] авторами предложено решение проблемы обнаружения дефектов на металлических поверхностях, связанной с нехваткой размеченных данных для новых целевых дефектов, при помощи модели, разработанной на основе контрастного обучения с использованием конвейера аугментации и облегчённого свёрточного кодировщика. В [13] авторы предлагают feature-aware network (FaNet) для классификации дефектов по нескольким изображениям, которая может эффективно различать новые классы при небольшом количестве размеченных образцов. Авторами [14] представлена новая модель для обнаружения дефектов на поверхности металла, которая сочетает в себе преимущества модуля CSPlayer и механизма глобального улучшения внимания на основе модели YOLOv5s. В статье [15] приведен алгоритм обнаружения микродефектов на поверхности металла YOLO-MD. Для обнаружения дефектов на поверхности металла авторами [16] предложена модель MeDERT, повышающая эффективность и точность процесса обнаружения. В работе [17] авторами проведён сравнительный анализ нескольких предварительно обученных нейронных сетей для задачи обнаружения и классификации трещин.

### Постановка задачи

В данной работе рассматривается одна из задач металлургического производства, в которой применяется компьютерное зрение – это автоматизация процесса позиционирования края полосы, смотанной в рулон, на агрегате продольной и поперечной резки листа. Агрегат продольной и поперечной резки листа применяется на литейно-прокатном комплексе (ЛПК) для раскроя штрипса на мерную длину и/или ширину и включает в себя операцию размотки рулона на разматывателе [18]. В ходе этой операции осуществляется поворот рулона в положение, удобное для разрезания упаковочной ленты на рулоне, отгибки края полосы выдвижным откидным столом (заправочным столом) и задачи этого края в линию агрегата [19].

Была поставлена следующая задача: разработать программно-аппаратный комплекс (ПАК) для определения края полосы (далее конец рулона), смотанной в рулон, на АПР [19].

Автоматизацию процесса определения положения конца рулона можно осуществить как с использованием датчиков, так и с применением системы компьютерного зрения. В качестве альтернативных вариантов были проанализированы индуктивный, ёмкостный, ультразвуковой и лазерный датчики перемещения. В результате анализа выявлено, что применение датчиков для обнаружения конца рулона неэффективно, так как сложно подобрать необходимый тип датчика, который работал бы без влияния помех, не имел жёстких технических требований к установке и мог определить конец рулона, имеющего относительно маленькую толщину полосы. Если же использовать комплекс датчиков, то такая система будет сложной и дорогой в установке и эксплуатации. Поэтому



для обнаружения конца рулона было принято решение использовать систему компьютерного зрения, поскольку она удовлетворяет всем описанным выше требованиям.

В процессе работы были опробованы алгоритмы обнаружения конца рулона как с применением детекторов, так и с использованием нейронной сети.

## Теория

### 1. Анализ контура управления вращением рулона на барабанах размотывателя

В рамках работы более подробно была рассмотрена часть автоматизированной системы управления, связанная с процессом вращения рулона на барабанах размотывателя при помощи асинхронного электродвигателя при позиционировании конца рулона для его задачи в линию АПР, для чего была составлена схема контура управления, представленная на рисунке 1.

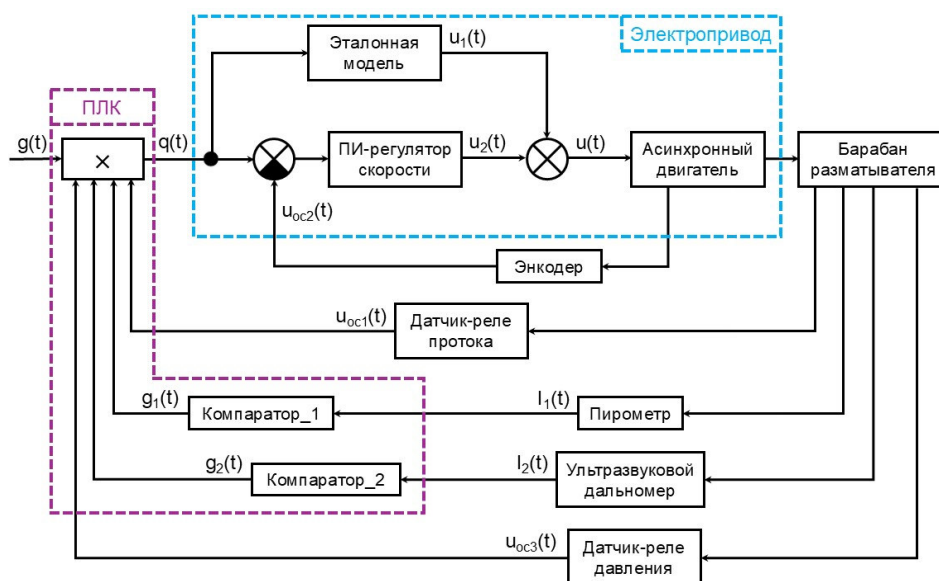


Рис. 1. Схема контура управления вращением рулона на барабанах размотывателя при позиционировании конца рулона для его задачи в линию АПР

Рассматриваемый контур управления состоит из программируемого логического контроллера (ПЛК) (далее контроллер), электропривода, барабана размотывателя, являющегося исполнительным органом рабочей машины, и датчиков (энкодер, пирометр, ультразвуковой дальномер и датчик-реле давления).

На схеме контроллер представлен в виде элемента умножения и двух компараторов, которые реализованы в контроллере в виде кода.

На вход контроллера поступают:

- входное воздействие  $g(t)$ , представляющее собой нажатие оператором кнопки начала вращения барабана размотывателя;
- сигналы обратной связи в виде величин напряжения  $u_{oc1}(t)$  и  $u_{oc3}(t)$  от датчика-реле протока жидкой смазки редуктора размотывателя и от датчика-реле давления (обратная связь от барабана размотывателя) соответственно;
- сигналы обратной связи в виде величин тока  $I_1(t)$  и  $I_2(t)$  от пирометра (обратная связь от рулона, расположенного на барабанах размотывателя) и ультразвукового дальномера (обратная связь от рулона, расположенного на барабанах размотывателя) соответственно (в контроллере эти сигналы обрабатываются при помощи элементов сравнения (компараторов), на выходе которых формируются сигналы  $g_1(t)$  и  $g_2(t)$ ).

В результате обработки всех входных сигналов на выходе контроллера формируется заданное значение скорости  $q(t)$ .



Электропривод на схеме представлен блоком, включающим в себя пропорционально-интегральный регулятор (ПИ-регулятор) скорости, эталонную модель, асинхронный двигатель, энкодер и сумматоры. Принцип действия электропривода следующий: на вход электропривода поступает сигнал заданного значения скорости  $q(t)$ , который параллельно подаётся в эталонную модель (предназначена для подавления переходных процессов и устранения отклонения между заданным и фактическим значением скорости, обусловленного запаздыванием, при постоянном значении скорости), формируется выходной сигнал  $u_1(t)$  в виде  $q(t)$  и в ПИ-регулятор скорости в виде разности  $q(t)$  и фактического значения скорости  $u_{oc2}(t)$  (поступает по обратной связи от энкодера) (формируется выходной сигнал  $u_2(t)$ ); далее полученные выходные сигналы  $u_1(t)$  и  $u_2(t)$  суммируются, образуя заданное значение момента вращения, которое поступает на асинхронный электродвигатель, обеспечивающий вращение барабана размотчика.

## 2. Алгоритм обнаружения конца рулона при помощи детекторов

В настоящее время существует большое количество алгоритмов компьютерного зрения и цифровой обработки изображений (ЦОИ). Для определения алгоритма, наиболее подходящего для решения поставленной задачи, был проведён сравнительный анализ различных алгоритмов компьютерного зрения и ЦОИ при помощи разработанной ранее мини-программы, в которой представлены различные алгоритмы компьютерного зрения и ЦОИ с возможностью интерактивного изменения параметров методов, применяемых в этих алгоритмах, а также с выводом времени работы каждого из алгоритмов.

При анализе бралось исходное изображение рулона с ровным концом. Далее к этому изображению применялись разные алгоритмы компьютерного зрения и ЦОИ с подбором подходящих параметров методов, которые применяются в этих алгоритмах [19].

Определение точности работы каждого из алгоритмов проводилось путём их применения к ряду изображений и дальнейшего расчёта по формуле для определения ассигуры на основании полученных данных.

При применении к изображению ряда алгоритмов, в том числе Canny Edge Detector с последующим применением Line Segment Detector (LSD) (рис. 2), билатерального фильтра и некоторых других, кроме самого конца полосы обнаруживались также и другие объекты, которые сложно отфильтровать, либо к ним нужно применять дополнительную фильтрацию, что может сказаться на быстродействии программы [19].

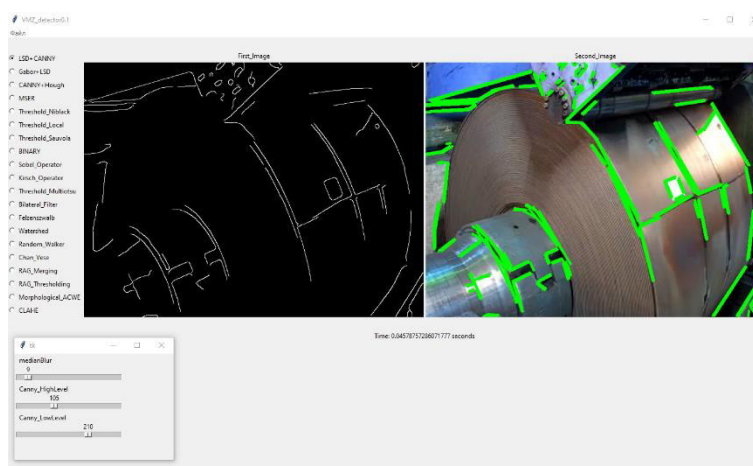


Рис. 2. Применение Canny Edge Detector с последующим использованием Line Segment Detector (LSD)

Некоторые алгоритмы, такие как Region adjacency graph (RAG) Thresholding (пороговый фильтр по значению графов смежности регионов) (рис. 3), Region adjacency graph (RAG) Merging (объединение графов смежности регионов) и некоторые другие, несмотря на то, что хорошо выделяли конец полосы, длительны по времени (обработка одного изображения занимает 1,69 с (см. табл. 1)) и к ним нужно применять дополнительную фильтрацию, что ещё сильнее усугубило бы ситуацию, связанную с быстродействием программы [19].

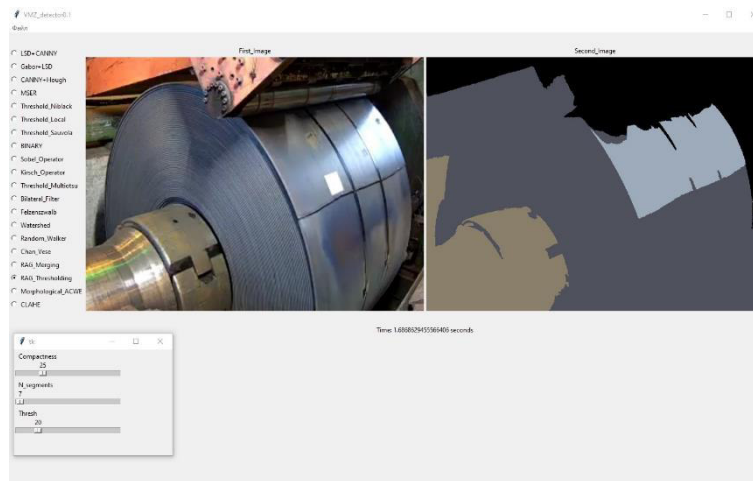


Рис. 3. Применение Region adjacency graph (RAG) Thresholding

Таблица 1. Сравнение результатов обработки изображений рулона алгоритмами компьютерного зрения и ЦОИ

Параметр	LSD + Canny	Canny + Hough	Bilateral filter	RAG Merging	RAG Thresholding	Threshold Niblack	Threshold Binary
Точность	0,65	0,7	0,6	0,6	0,65	0,68	0,65
Время обработки, с	0,046	0,031	0,579	1,772	1,687	0,05	0,018

Единственным алгоритмом, в котором не было ложных срабатываний и, соответственно, к которому не нужно было в дальнейшем применять дополнительную фильтрацию, а также который имел допустимое время по быстродействию (0,03 с), являлся алгоритм Canny Edge Detector с последующим применением преобразования Хафа (рис. 4) [19].

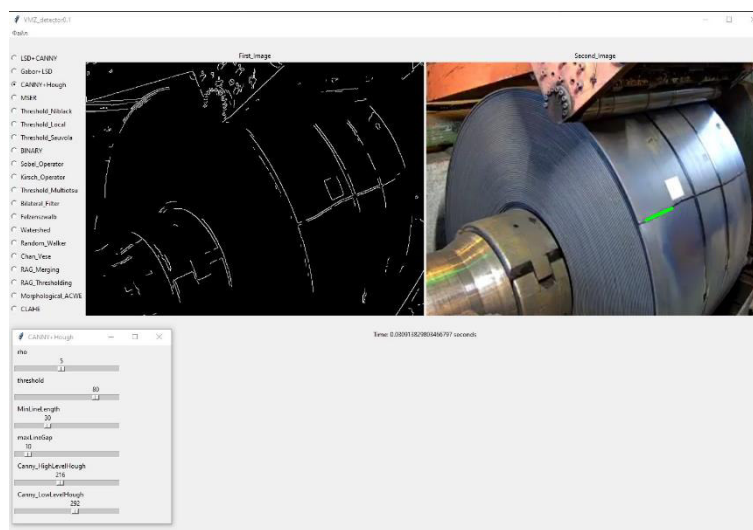


Рис. 4. Применение Canny Edge Detector с последующим использованием преобразования Хафа

Поэтому было принято предварительное решение использовать этот алгоритм для поиска прямолинейных концов полосы [19].

Оператор Canny Edge Detector использует алгоритм, который состоит из четырёх основных этапов: сглаживание входного изображения и удаление шума, вычисление градиентов, «не максимальное подавление» и порог гистерезиса [20, 21].

Преобразование Хафа (Hough) – это преобразование, используемое для обнаружения прямых линий, в основе которого лежит утверждение, что любая точка бинарного изображения может быть частью некоторого набора возможных линий в пространстве Хафа [22].

Однако конец полосы может иметь криволинейную форму, поэтому к выбранному алгоритму ещё был добавлен детектор углов Ши-Томаси. Данный детектор является модификацией детектора углов Харриса. Углом на изображении являются те области, где происходят значительные изменения яркости по всем направлениям. Именно эта идея лежит в основе детектора Харриса и соответственно детектора Ши-Томаси. Отличие детектора Ши-Томаси от детектора Харриса заключается в том, что в детекторе углов Ши-Томаси оценка того, содержит окно угол или нет, осуществляется по формуле (1)

$$R = \min(\lambda_1, \lambda_2), \quad (1)$$

где  $\lambda_1$  и  $\lambda_2$  – собственные значения матрицы;  $R$  – мера отклика.

Если она больше порогового значения, то эта область изображения считается углом [23].

### 3. Алгоритм обнаружения конца рулона при помощи нейронной сети

Далее был опробован алгоритм с применением нейронной сети YOLO. Было проведено обучение двух версий нейронной сети YOLO: YOLOv5 и YOLOv8. Данные версии были выбраны, так как на момент создания ПО версии модели выше YOLOv8 не были выпущены, а модели ниже YOLOv5 сильно уступают YOLOv5 и YOLOv8 по точности и скорости. YOLOv6 и YOLOv7 не были выбраны, так как они уступают YOLOv5 по скорости и производительности.

Для обучения моделей было собрано 659 различных изображений с камеры, расположенной на реально действующем агрегате. Разметка изображений проводилась в веб-платформе CVAT.

Результаты обучения двух версий YOLO представлены ниже (рис. 5-7).

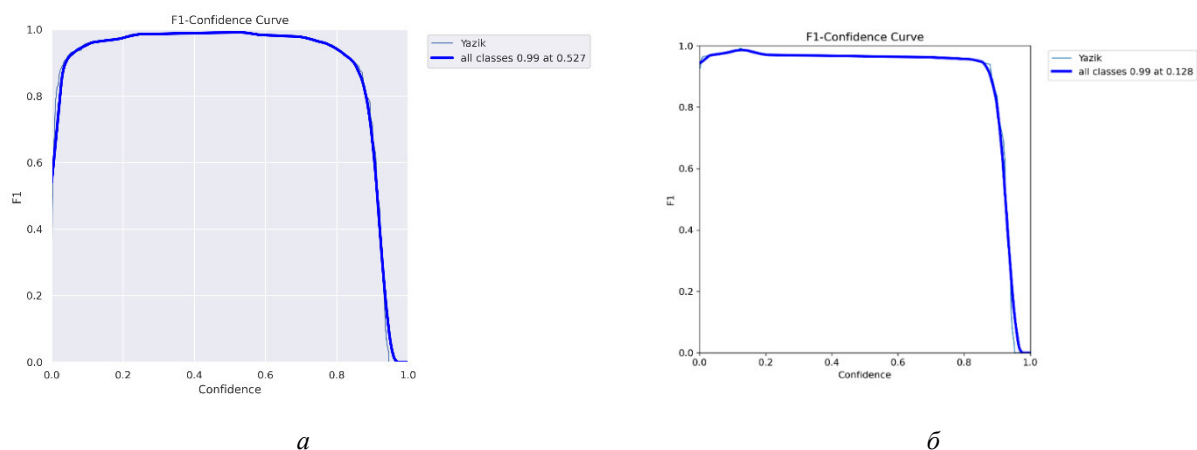


Рис. 5. Кривые F1-меры для YOLOv5 (a) и YOLOv8 (б)

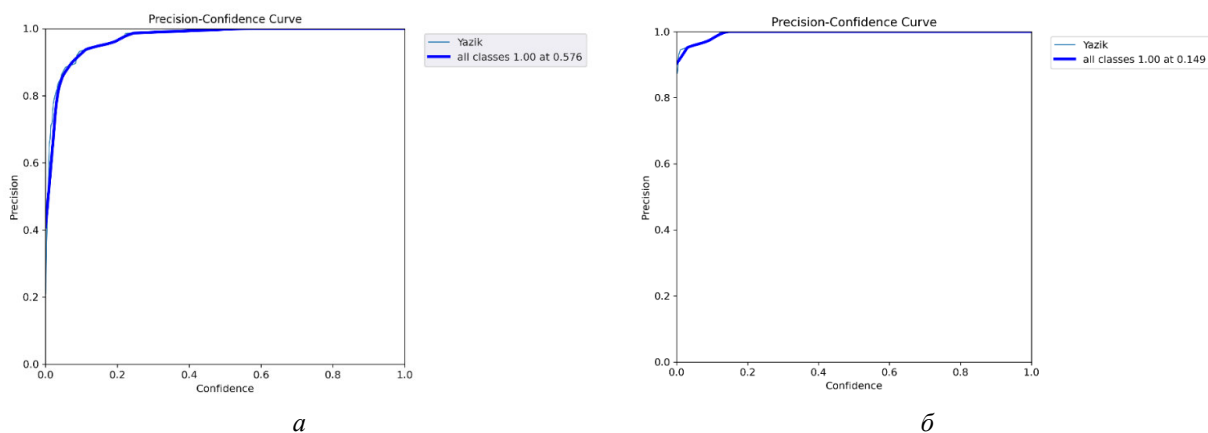


Рис. 6. Кривые precision для YOLOv5 (a) и YOLOv8 (б)

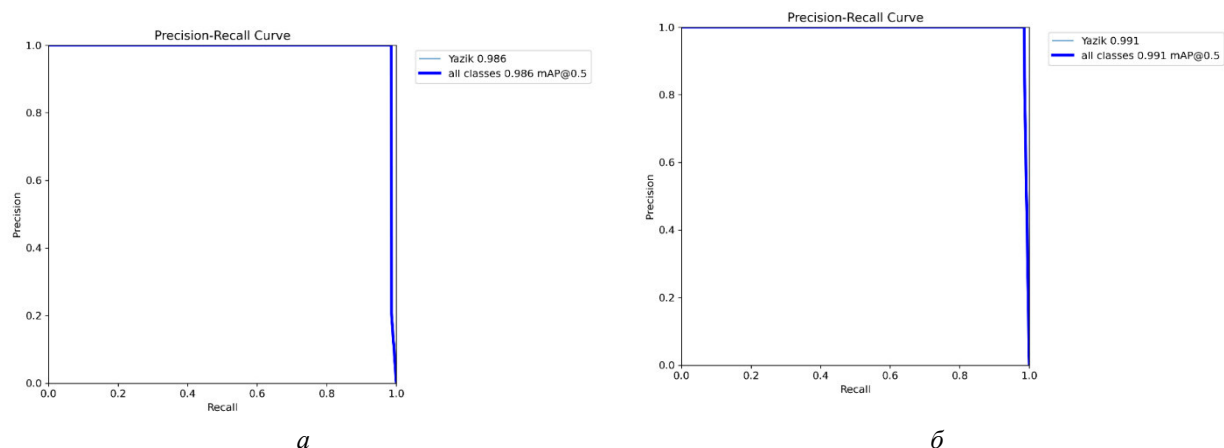


Рис. 7. Кривые recall для YOLOv5 (а) и YOLOv8 (б)

По результатам обучения были сформулированы следующие выводы:

- обе версии показали высокую способность обнаруживать все экземпляры класса и низкое количество ложных срабатываний. Кривые F1 имели значение 0,99, что свидетельствовало о сбалансированности точности и полноты в обеих версиях;
- однако YOLOv8 отличалась более высокой точностью при меньшем пороге достоверности и большим быстродействием.

#### 4. Сравнительный анализ алгоритма с применением детекторов и алгоритма с применением нейронной сети

Если сравнивать алгоритм на основе Canny Edge Detector и преобразования Хафа и алгоритм на основе нейронной сети, то алгоритм с применением детекторов работал быстрее (0,055 с), чем алгоритм с применением нейронной сети (0,3 с), однако он выдавал большое количество ложных срабатываний (точность работы составила 65 %), которые никак не удавалось устранить. В этом плане алгоритм с нейросетью отработал точнее (90 %), поэтому для дальнейшей разработки было принято решение использовать именно его. Недостаток, связанный со временем работы, было решено устранить путём использования более мощного компьютера с видеокартой, а также благодаря оптимизации алгоритма. Повысить точность обнаружения в случае нейронной сети можно при помощи дообучения её на дополненном наборе данных.

На основе выбранного алгоритма обнаружения конца рулона было разработано программное обеспечение с интерфейсом.

### Результаты экспериментов

Разработка ПО включала в себя проектирование архитектуры и написание кода программы.

#### 1. Проектирование архитектуры ПО

Изначально программа по поиску конца рулона была реализована на основе монолитной архитектуры, однако при её тестировании выявлены значительные отставания по времени, поэтому было принято решение перестроить программу в соответствии с микросервисной архитектурой, где некоторые модули работают в отдельных потоках. В результате была спроектирована диаграмма потоков данных (рис. 8).

В данной программе есть один управляющий модуль, который запускается через модуль интерфейса (именно этот модуль запускается через main при запуске программы) и управляет запуском других модулей (захват кадра, обработка кадра, сохранение кадра, взаимодействие с контроллером), каждый из которых работает в отдельном потоке. Взаимодействие между потоками осуществляется при помощи потокобезопасных очередей. Поток захвата кадров непосредственно взаимодействует с камерой через протокол RTSP, после чего передаёт полученный кадр в поток основной обработки. В потоке основной обработки последовательно выполняются следующие процессы, реализованные в виде отдельных классов: предобработка кадра, обработка кадра нейронной сетью, вычисление угла

поворота в случае обнаружения конца рулона. Данные, полученные из потока основной обработки, поступают на интерфейс, в поток сохранения кадра и в поток взаимодействия с контроллером. Поток сохранения кадра включает в себя отрисовку на изображении необходимых элементов и сохранение этих изображений в случае необходимости. Поток взаимодействия с контроллером производит непосредственный приём и передачу данных в контроллер через протокол S7.

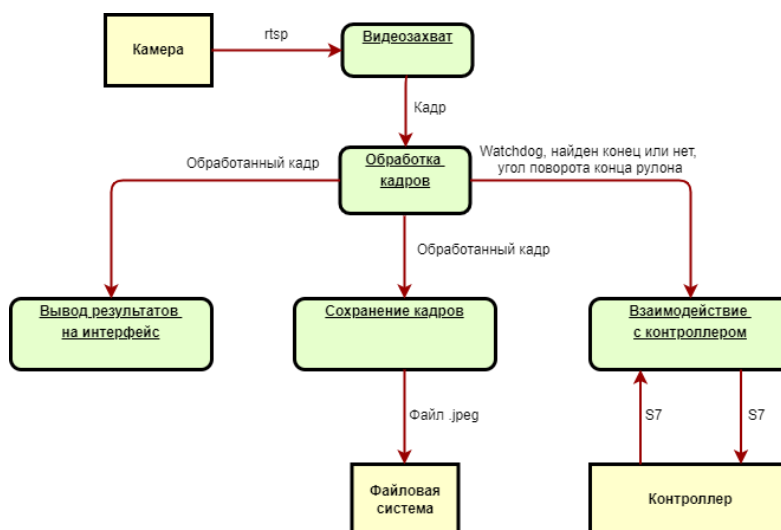


Рис. 8. Схема взаимодействия программных модулей ПАК

## 2. Интерфейс

Для программы был разработан пользовательский интерфейс, однако в процессе выполнения работы выявлено отсутствие необходимости в активном взаимодействии пользователя с интерфейсом, так как программа должна работать постоянно в автоматическом режиме. В результате чего было принято решение использовать в качестве интерфейса окно, на котором отображаются видеопоток с камеры и необходимые фигуры (рис. 9).

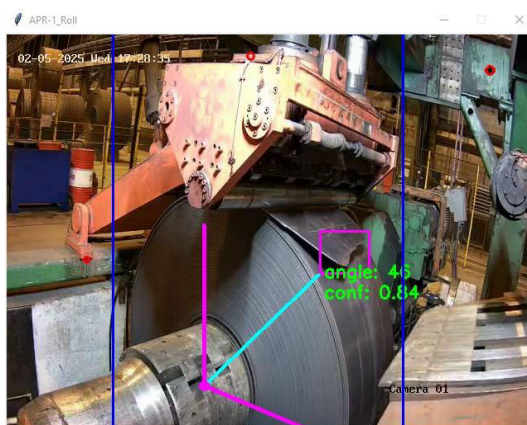


Рис. 9. Интерфейс программы поиска конца рулона

На этом интерфейсе отображаются:

- три маленьких красных окружности, по которым в случае необходимости можно отрегулировать положение камеры;
- вертикальные синие линии, показывающие область, в которой система производит поиск конца рулона;
- точка начала координат и координатные оси, в соответствии с которыми происходит расчёт угла поворота конца рулона;
- другие фигуры и надписи, которые появляются в случае обнаружения конца рулона (ограничивающая рамка, значение угла поворота и т. д.).



### 3. Модернизация автоматизированной системы управления узла разматывателя АПР

После разработки ПАК была проведена его интеграция с контроллером системы АСУ ТП АПР.

Схема контура управления вращением рулона на барабане разматывателя после модернизации путём интеграции системы позиционирования конца рулона с применением компьютерного зрения представлена на рисунке 10.

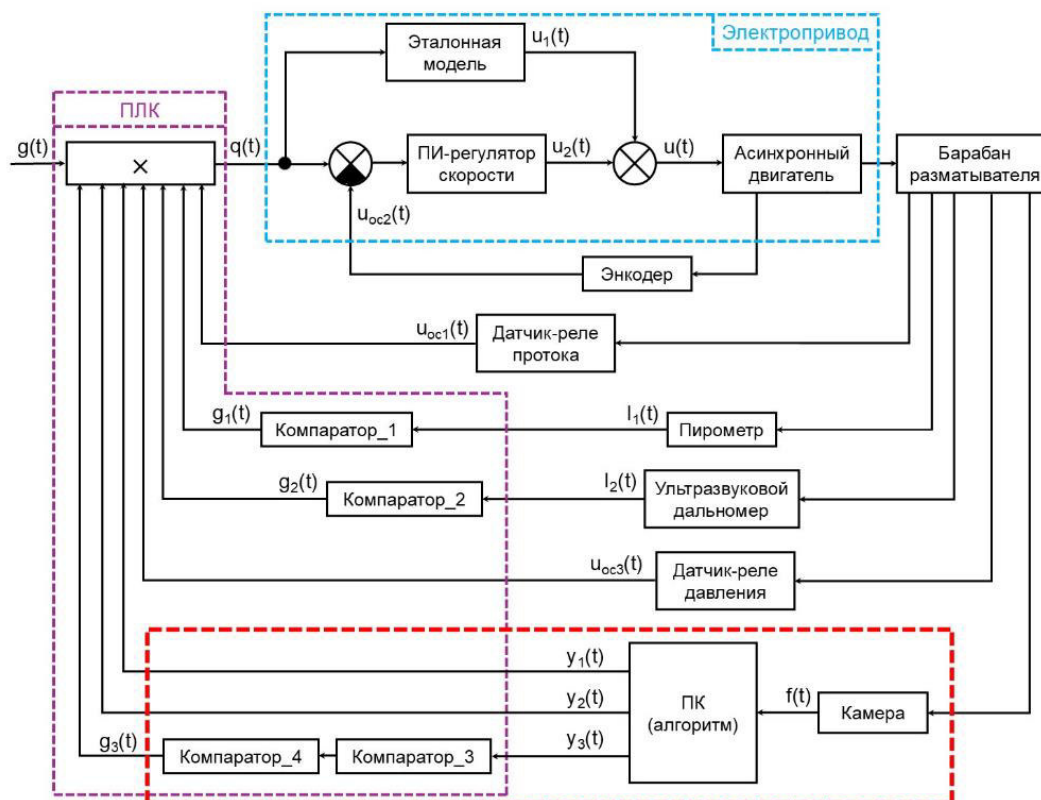


Рис. 10. Схема контура управления после модернизации

В данную схему были добавлены следующие элементы:

- камера, которая получает изображение рулона (кадр,  $f(t)$ ) и передаёт его на ПК с алгоритмом;
- ПК с алгоритмом после получения изображения производит его обработку, в результате которой на выходе формируются три сигнала: watchdog ( $y_1(t)$ ), найден/не найден конец рулона ( $y_2(t)$ ) и угол поворота найденного конца рулона ( $y_3(t)$ ). Далее эти сигналы поступают на вход контроллера;
- угол поворота поступает на компараторы, реализованные в виде программного кода на ПЛК. На выходе из компараторов формируется сигнал  $g_3(t)$ .

### 4. Эксперименты

Был проведён ряд экспериментов, а именно проверка правильности передачи сигналов в контроллер, определение времени запаздывания поступления сигналов в контроллер, проверка работы системы после интеграции (проверка совместной работы кода ПАК и кода ПЛК), а также проверка точности вычисления угла.

Для обработки результатов этих экспериментов применялась программа (средство диагностики) IBA Analyzer (рис. 11). Благодаря этой программе происходила проверка того, правильно ли приходят сигналы от ПАК, например, watchdog должен изменяться раз в секунду (не реже и не чаще). Также в ней определялось время запаздывания поступления сигналов, например, при проверке того, произведётся ли остановка двигателя при достижении определённого значения угла. Кроме того, с помощью IBA Analyzer можно было определить, что могло стать причиной того, что система, например, не нашла конец рулона: сам ПАК дал сбой, или помешал заправочный стол.

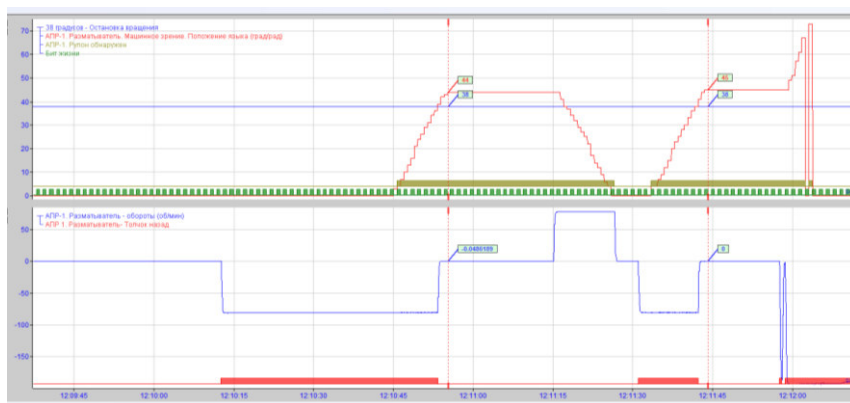


Рис. 11. Результат эксперимента по проверке корректности работы системы управления после модернизации

Кроме IBA Analyzer также применялось нанесение разметки на сам рулон. Проводилось это для экспериментов по определению точности вычисления угла поворота конца рулона. На рулон наносились линии, соответствующие концу рулона, а также линии, находящиеся в позиции 45 и 90 градусов от него, после чего осуществлялась обработка изображения этого рулона ПАК. Далее происходило вычисление отклонения значения угла, посчитанного системой от фактического угла поворота, а также корректировка работы программы в связи с полученными отклонениями (рис. 12).

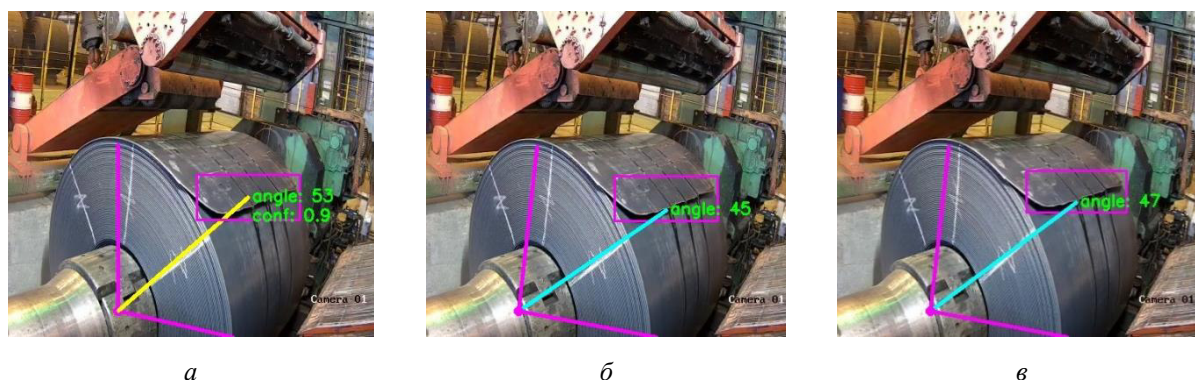


Рис. 12. Результаты эксперимента по определению точности вычисления угла поворота конца рулона и корректировки кода программы на основе этих результатов:

а) результат эксперимента (отклонение в 8 градусов); б) результат корректировки кода программы по данным эксперимента; в) результат проверки программы после корректировки

### Обсуждение результатов

В результате проведённой работы был разработан программно-аппаратный комплекс для определения положения конца рулона на АПР.

Для определения качества работы системы проведены эксперименты в реальных условиях на 100 изображениях рулонов, по результатам которых была составлена матрица ошибок (табл. 2).

Таблица 2. Матрица ошибок

		Предсказанное состояние	
		Positive	Negative
Реальное состояние	Positive	True positive (TP) = 70	False negative (FN) = 1
	Negative	False positive (FP) = 2	True negative (TN) = 27



При помощи этой матрицы для задачи детекции объектов на изображении рассчитываются следующие метрики качества работы системы:

- *precision* – доля истинных положительных результатов среди всех положительных обнаружений. Оценивается способность системы избегать ложных срабатываний. Вычисляется по формуле (2).

$$Precision = \frac{TP}{TP + FP}; \quad (2)$$

- *recall* – доля истинных положительных результатов среди всех реальных. Оценивается способность системы обнаруживать все экземпляры объекта. Вычисляется по формуле (3).

$$Recall = \frac{TP}{TP + FN}. \quad (3)$$

По результатам экспериментов точность обнаружения конца рулона (*precision*) составила 97 %, а полнота составила 99 %. Причиной низкого значения точности по сравнению со значением полноты является наличие ложных срабатываний, которые появляются из-за нестандартных рулонов (например, с большим количеством упаковочной ленты).

Точность вычисления угла поворота конца рулона для больших диаметров (1,8...2,1 м) составила в среднем 3°...4°, для средних (1,5...1,8 м) – 5°...7° и для малых (менее 1,5 м) – 8°...10°. Такой разброс по точности, связан с тем, что система координат, относительно которой происходит расчёт угла поворота является фиксированной, в то время как рулоны имеют разную ширину и диаметр, и соответственно, в этом случае нужна разная система координат. Поэтому далее будет проведена корректировка алгоритма для того, чтобы система координат подстраивалась под текущую ширину и диаметр рулона.

### Заключение

В ходе работы проведён анализ возможности применения компьютерного зрения для задачи обнаружения конца рулона металлической полосы на агрегате продольной и поперечной резки листа при задаче этого конца в агрегат. Проанализированы два алгоритма обнаружения конца рулона (с применением детекторов и с применением нейронной сети). Проведён сравнительный анализ различных алгоритмов компьютерного зрения с детекторами. Проведён сравнительный анализ YOLOv5 и YOLOv8 для обнаружения конца полосы. Проведена оценка качества работы каждого из алгоритмов при решении поставленной задачи. Алгоритм с применением детекторов работал быстрее (0,055 с), чем алгоритм с применением нейронной сети (0,3 с.), однако он выдавал большое количество ложных срабатываний (точность работы составила 65 %), которые никак не получалось устранить. В этом плане алгоритм с нейросетью отработал точнее (90 %), поэтому для дальнейшей разработки было принято решение использовать его. Недостаток, связанный со временем работы, было решено устранить путём использования более мощного компьютера с видеокартой, а также путём оптимизации алгоритма. Повысить точность обнаружения в случае нейронной сети можно при помощи дообучения её на дополненном наборе данных. На основе выбранного алгоритма было разработано программное обеспечение с интерфейсом, проведены интеграция разработанного программно-аппаратного комплекса с системой АСУ ТП АПР и ряд экспериментов для определения качества работы системы.

### Благодарности

Авторы благодарят ведущего инженера управления по автоматизации ООО «Эколант» Дмитрия Евгеньевича Антипова, ведущего инженера участка по автоматизации ЛПК АО «ВМЗ» Алексея Юрьевича Абрамова и начальника участка по автоматизации ЛПК АО «ВМЗ» Андрея Евгеньевича Филатова.

### Список литературы

1. Digitalization as the Most Important Tool for the Improvement of Metallurgical Technologies / A. V. Muntin, M. N. Shamshin, A. G. Ziniagin [et al.] // *Metallurgist*. 2023. Vol. 66, no. 9-10. P. 1051-1067.
2. Емельянова М.Г., Смаилова С.С., Бакланова О.Е. Обнаружение поверхностных дефектов сварных соединений при визуальном контроле методами машинного зрения // *Компьютерная оптика*. 2023. Т. 47, № 1. С. 112-117. DOI: 10.18287/2412-6179-CO-1137.
3. Rabia Ali, Muhammad Sarmad, Jawad Tayyub, Alexander Vogel. 2023. Accurate detection of weld seams for laser welding in real-world manufacturing. *AI Mag*. 44, 4 (Winter 2023), pp. 431–441. <https://doi.org/10.1002/aaai.12134>.

4. Yang, Y.; Chen, T.; Zhao, L. From Segmentation to Classification: A Deep Learning Scheme for Sintered Surface Images Processing. *Processes* 2024, 12, 53. <https://doi.org/10.3390/pr12010053>.
5. Yu, Y.; Chan, S.; Tang, T.; Zhou, X.; Yao, Y.; Zhang, H. Surface Defect Detection of Hot Rolled Steel Based on Attention Mechanism and Dilated Convolution for Industrial Robots. *Electronics* 2023, 12, 1856, pp. 1856–1868. <https://doi.org/10.3390/electronics12081856>.
6. Liu, C.; Su, S.; Chen, Z. Application of Convolutional Neural Networks in Defect Detection System for Powder Metallurgy Small Gears. *Journal of the Japan Society of Powder and Powder Metallurgy* 2025, 72, pp. 1255–1258. <https://doi.org/10.2497/jjspm.16C-T11-12>.
7. Geng, H.; Li, Z.; Zhou, Y. Research on Weld Defect Detection and Evaluation Technology based on Deep Learning 2023, 1(1):1. DOI:10.56028/fesr.1.1.1.2023.
8. Pan, K.; Hu, H.; Gu, P. WD-YOLO: A More Accurate YOLO for Defect Detection in Weld X-ray Images. *Sensors* 2023, 23, 8677. DOI:10.3390/s23218677.
9. Zhao, Y.; Wang, H.; Xie, X.; Xie, Y.; Yang, C. An Enhanced YOLOv5-Based Algorithm for Metal Surface Defect Detection. *Appl. Sci.* 2023, 13, 11473. DOI:10.3390/app132011473.
10. Vengaloor, R. Muralidhar, R. Deep Learning Based Feature Discriminability Boosted Concurrent Metal Surface Defect Detection System Using YOLOv-5s-FRN. *The International Arab Journal of Information Technology* 2024, 21 (1). pp. 94–106. DOI:10.34028/iajit/21/1/9.
11. Wang, Y.; Wang, X.; Hao, R.; Lu, B. Metal Surface Defect Detection Method Based on Improved Cascade R-CNN. *Journal of Computing and Information Science in Engineering* 2024, 24(4):041002. DOI:10.1115/1.4063860.
12. Mahe Zabin, Anika Nahian Binte Kabir, Muhammad Khubayeb Kabir, Ho-Jin Choi, Jia Uddin. Contrastive self-supervised representation learning framework for metal surface defect detection. *Journal of Big Data* (2023), 10 (1), p. 145. DOI:10.1186/s40537-023-00827-z.
13. Zhao, W.; Song, K.; Wang, Y.; Liang, S. FaNet: Feature-aware Network for Few Shot Classification of Strip Steel Surface Defects. *Measurement* 2023, 208(12). p.112446. DOI:10.1016/j.measurement.2023.112446.
14. Zhou, C.; Lu, Z.; Lv, Z.; Meng, M. Metal surface defect detection based on improved YOLOv5. *Scientific Reports* 2023, 13(1), p. 20803. DOI:10.1038/s41598-023-47716-2.
15. Li Jiao, Muhammad Irsyad Abdullah. Neutrosophic Enhancement of YOLO-MD Algorithm for Automated Metal Surface Micro Defect Detection. *International Journal of Neutrosophic Science* 2024, 23(2), p. 308–316. DOI:10.54216/IJNS.230225.
16. Wang, C.; Xie, H. MeDERT: A metal surface defect detection model. *IEEE Access* 2023, PP(99):1–1. DOI:10.1109/ACCESS.2023.3262264.
17. Qayyum, W.; Ehtisham, R.; Bahrami, A.; Camp, C.; Mir, J.; Ahmad, A. Assessment of Convolutional Neural Network Pre-Trained Models for Detection and Orientation of Cracks. *Materials* 2023, 16, pp. 826. DOI:10.3390/ma16020826.
18. Muntin A. V. Advanced Technology of Combined Thin Slab Continuous Casting and Steel Strip Hot Rolling // *Metallurgist*. 2019. Vol. 62, no. 9–10. P. 900–910.
19. Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2024» / отв. ред. И.А. Алешковский, А.В. Андриянов, Е.А. Антипов, Е.И. Зимакова. М.: МОО СИПНН Н.Д. Кондратьева, 2024. ISBN 978-5-901-64042-5.
20. Zubova E., Pakhalev A., Shamshin M., Mortin K., Gladishev M. Application of Methods for Detecting the Boundaries of an Object in an Image to Search for a Pilot Line // 2023 7th International Conference on Information, Control, and Communication Technologies (ICCT). Astrakhan, Russian Federation, 2023. P. 1–4. doi: 10.1109/ICCT58878.2023.10347100.
21. Canny J. A computational approach to edge detection // *Pattern Analysis and Machine Intelligence*, IEEE Transactions on. 1986. No. 6. P. 679–698.
22. Matas J. Kittler C., J. Robust detection of lines using the progressive probabilistic hough transform // *GalambosComputer Vision and Image Understanding*. 2000. No. 1. P. 119–137.
23. Shi J., Tomasi C. Good features to track // *Computer Vision and Pattern Recognition*. 1994. P. 593–600.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-619-624

## Разработка модульной системы технического зрения для соревновательного колесного робота

Д. В. Морохин, А. В. Иванов

Поволжский государственный технологический университет, Йошкар-Ола, Россия

**Аннотация.** В данной статье представлена разработка модульной системы технического зрения для колесного соревновательного робота, ориентированного на образовательные задачи. Основная цель исследования — создание доступной, но функциональной платформы, сочетающей невысокую стоимость с возможностями автономной навигации и обработки визуальной информации. Проведены комплексные испытания разработанной системы и сравнение ее с коммерческими аналогами. Результаты подтвердили конкурентоспособность решения при значительно меньшей стоимости. Особое внимание уделено модульности архитектуры, позволяющей легко модернизировать систему, включая интеграцию нейросетевых алгоритмов. Практическая значимость работы подтверждена успешным использованием платформы в образовательных соревнованиях.

**Ключевые слова:** техническое зрение, ROS 2, OpenCV, автономная навигация, машинное обучение, компьютерное зрение, одноплатный компьютер.

## Modular vision system for a competitive wheeled robot development

D. V. Morokhin, A. V. Ivanov

Volga State University of Technology, Yoshkar-Ola, Russia

**Abstract.** This paper presents the development of a modular autonomous control and computer vision system for wheeled competition robots. The research aims to create an affordable yet functional platform combining autonomous navigation capabilities with real-time visual data processing for educational applications. The system architecture integrates an Orange Pi Zero 2 single-board computer and ESP-32 microcontroller, providing sufficient computational power for processing data from a Gxvision-SVGA camera and VL53L0X laser distance sensors. The software implementation utilizes Python and C++ with OpenCV libraries and ROS 2 framework for module interoperability. Comprehensive testing demonstrated line recognition accuracy up to 96 % under various lighting conditions. Comparative analysis with commercial solutions (DJI RoboMaster, VEX Robotics) confirmed the system's competitiveness at significantly lower cost (<\$300). The modular design facilitates future upgrades, including neural network algorithm integration. Practical significance was validated through successful deployment in educational robotics competitions. The solution offers an optimal balance of performance and affordability for STEM education applications.

**Keywords:** wheeled robot, computer vision, ROS 2, OpenCV, autonavigation, machine learning, microcomputer.

### Введение

Актуальность данной работы обоснована растущим интересом к робототехнике как в научном, так и в образовательном контексте. Разработка алгоритмов управления для соревновательных роботов способствует развитию навыков программирования, проектирования и инженерного мышления у студентов и молодых специалистов [1]. Это, в свою очередь, стимулирует интерес к STEM-дисциплинам (наука, технологии, инженерия и математика) и формирует основу для будущих исследований и разработок в области робототехники.

Современная образовательная робототехника требует доступных платформ с функционалом автономной навигации и технического зрения. Анализ коммерческих решений выявил их ключевые ограничения: высокую стоимость (400-1000 тыс. руб.), сложность калибровки и недостаточную гибкость для массового образовательного применения [5, 9-11]. В работах [2, 4] предложены бюджетные альтернативы, однако их производительность недостаточна для обработки видеопотока в реальном времени.

Исследование, проведенное в данной работе, направлено на необходимость создания недорогой, модульной системы с открытой архитектурой, совместимой с Robot Operating System 2 и современными библиотеками компьютерного зрения [8].

## Постановка задачи

Целью работы является разработка модульной системы для колесного робота, включающей:

1) модуль автономной навигации на базе стека Orange Pi Zero и ESP-32 с датчиками расстояния и освещения;

2) систему технического зрения на основе микрокомпьютера Orange Pi Zero 2;

3) ROS2 – совместимый программный стек.

Задачи:

- Разработка структуры аппаратного и программного обеспечения колесной платформы.
- Разработка механизма взаимодействия между головным вычислителем и микроконтроллером нижнего уровня через micro-ROS.
- Оптимизация алгоритмов обработки изображений для слабых вычислительных ресурсов.
- Сравнение эффективности с коммерческими аналогами.

## Проектирование системы

### 1. Аппаратная архитектура

Система реализована по модульному принципу (рис. 1). Каждый отдельный блок связан с блоком управления, реализованным на микрокомпьютере Orange Pi Zero 2 [7]. Он выполняет функцию управления остальными блоками.

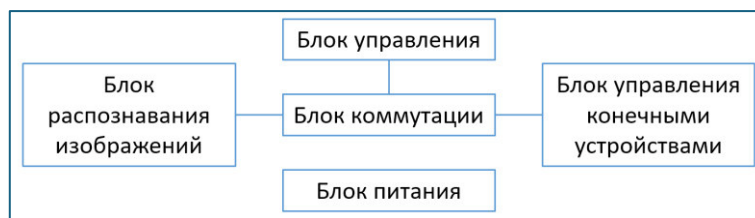


Рис. 1. Структурная схема робота

Данный микрокомпьютер сочетает в себе хорошую производительность и энергоэффективность, имея 4-ядерный процессор Cortex-A53 1.5 ГГц и графический процессор Mali G31 MP2 с поддержкой OpenGL. Наличие Ethernet, Wi-Fi, Bluetooth, USB-интерфейсов позволяет реализовать блок передачи данных без дополнительных внешних модулей. Поддерживает ОС Linux с большим количеством библиотек для компьютерного зрения. Установленный графический процессор и 1ГБ DDR3 оперативной памяти ограничивает возможности по сравнению с LPDDR4 в других современных решениях. Однако по соотношению цена-качество для задач обработки изображения и выполнения задач с моделями нейросети для распознавания объектов данная модель является оптимальным выбором.

Orange Pi Zero 2 поддерживает подключение USB-камер через стандартные UVC-драйверы, что упрощает интеграцию по сравнению с CSI-интерфейсами. Наличие 26-контактного GPIO обеспечивает прямую интеграцию с датчиками робота.

В числе рассматриваемых аналогов были такие модели как Raspberry Pi, OpenMV-H7, NVIDIA Jetson Nano, Repka Pi, ESP32-CAM. В сравнении с Orange Pi Zero 2 они имеют либо высокую производительность, но высокую стоимость, либо достаточно слабые характеристики для возможности использования обученных моделей нейросетей и распознавания. Одним из главных преимуществ Orange Pi Zero 2 является его гибкость и настраиваемость. Его можно легко адаптировать для работы с различными компонентами системы технического зрения, такими как камеры, сенсоры, устройства вывода и другие.

Блок распознавания является ключевым компонентом системы, отвечая за полный цикл обработки каждого изображения видеопотока. Модуль камеры Gxivision передает изображение в программное обеспечение блока управления для распознавания. Главным преимуществом данной камеры является поддержка стандарта UVC (USB Video Class), которая обеспечивает совместимость по принципу «подключи и работай» с Orange Pi без необходимости установки дополнительных драйверов.

Блок питания обеспечивает стабильное напряжение 5В всем элементам системы. В его роли выступает аккумулятор с понижающим напряжение модулем.

Для организации беспроводной коммуникации используется блок передачи данных, оснащенный Wi-Fi модулем. Этот компонент выполняет двусторонний обмен информацией между управляющими устройствами, обеспечивая эффективную передачу данных. Благодаря встроенным Wi-Fi модулям у Esp-32 и Orange Pi достаточно просто организовать их взаимодействие через фреймворк ROS 2. Команды с одноплатного компьютера будут передаваться на контроллер управления роботом, а данные с датчиков – в обратном направлении.

Для сбора данных с датчиков (лидар VL53L0X, лазерные дальномеры VL53L0X-V2, ИК-бампер FLASH-I2C) и управления моторами (ZK-5AD) используется микроконтроллер ESP-32.

## 2. Алгоритмы компьютерного зрения

Для тестирования системы была реализована задача следования колесного робота по контрастной линии. При её решении были использованы и исследованы следующие алгоритмы:

### 2.1. Пороговая бинаризация. Пример кода:

```
gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
_, binary = cv2.threshold(gray, 100, 255, cv2.THRESH_BINARY_INV)
```

### 2.2. Контурный анализ с вычислением моментов. Пример кода:

```
M = cv2.moments(largest_contour)
cx = int(M["m10"]/M["m00"])
```

### а. ПИД-регулятор для движения по линии. Пример кода:

$$u(t) = 0.8 * e(t) + 0.05 \int e(t) dt + 0.2 * de(t)/dt$$

## 3. Программная архитектура

Состав и взаимодействие между программными модулями показаны на рисунке 2.

Основными программными функциями одноплатного компьютера являются:

- распознавание объектов на изображении;
- взаимодействие с основной системой управления колесным роботом;
- возможность транслировать обработанное изображение с камеры.

Микроконтроллер выполняет свои задачи:

- получение данных от датчиков по протоколу I2C;
- управление коллекторными двигателями.

Также важной общей задачей является обмен данными и командами.

Программное обеспечение системы управления реализовано на Python для одноплатного компьютера. Основная задача программы — обнаружение и отслеживание тёмных объектов на светлом фоне (линия движения) в режиме реального времени с последующей передачей команд микроконтроллеру для управления колесным роботом. Тестирование направлено на проверку корректности работы алгоритма обработки видеопотока, выделения объектов и интеграции модуля с роботом.

Для обеспечения модульности не только в аппаратном, но и программном аспекте было принято решение об использовании Robot Operating System 2 – фреймворка для создания робототехнических систем [3]. Его основные особенности: работа в реальном времени, поддержка IoT-устройств, средства безопасности и интеграция с промышленными стандартами. Это не операционная система в классическом понимании, а набор инструментов, библиотек и соглашений, упрощающих разработку сложных робототехнических приложений.

ROS 2 помогает разделить функционал на независимые узлы (nodes), которые обмениваются данными через топики (topics) и сервисы (services). Полноценная версия ROS в проекте выполняется на микрокомпьютере OrangePi. Но для управления роботом требуется включить в обмен микроконтроллер ESP-32. На нём для функционирования полноценной ROS недостаточно ресурсов. Однако существует реализация Micro-ROS – адаптация ROS 2 для микроконтроллеров, позволяющая им участвовать в ROS-сети как полноценные узлы. Одноплатный компьютер выполняет обработку изображения с камеры (в данном случае определяет линии следования) и публикует команды управления в топики ROS 2. Micro-ROS Agent — промежуточное ПО, запущенное на Orange Pi, которое

преобразует сообщения ROS 2 в формат, понятный micro-ROS на ESP-32, и наоборот. ESP-32 подписывается на топик, получает команды и преобразует их в PWM-сигналы для управления моторами.

На рисунке 2 также показаны некоторые из «топиков» разработанной системы. Ниже опишем ещё один из пакетов системы, который реализует алгоритм движения по линии (ROS 2 пакета Line follower).

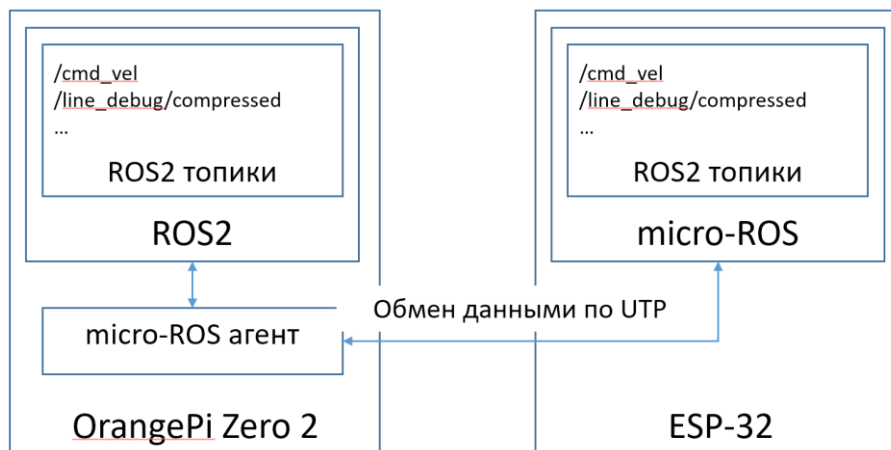


Рис. 2. Структура программного обеспечения системы

ROS 2 пакет состоит из двух python файлов cameraPublisher.py и subscriberImage.py.

В cameraPublisher.py происходит получение и обработка изображения с веб-камеры, а затем её отправка на отладочный топик linedebug/compressed, к которому подписывается subscriberImage.py.

Далее происходит расчет управления роботом, а именно определение того, в какой части кадра по горизонтали находится объект в виде линии.

Затем в работу включается программа micro ros на ESP32, она подписывается на twistmsg и получает значения переменных для управления моторами linear.x и angular.z:

- twist.linear.x – линейная скорость вдоль оси X (движется вперед/назад);
- twist.angular.z – угловая скорость вокруг оси Z (вращение по часовой стрелке/против часовой стрелки).

Значения параметров обычно указываются в метрах в секунду (м/с) для linear.x и радианах в секунду (рад/с) для angular.z. Для дифференциальных роботов эти два параметра являются основными для управления движением.

### Результаты экспериментов

В результате работы реализована колесная платформа с возможностью расширения (рис. 3). С ней был проведен ряд экспериментов.

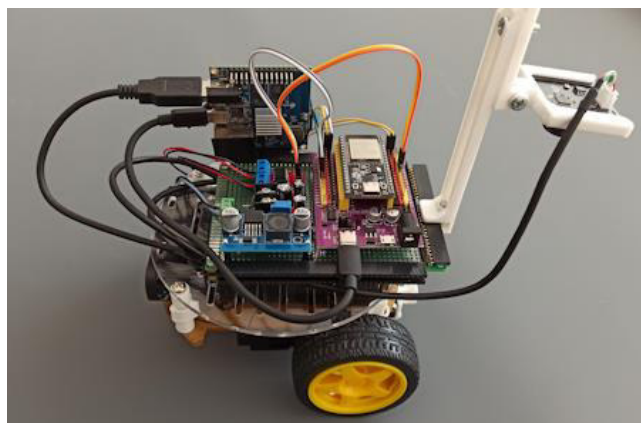


Рис. 3. Прототип колёсной платформы

Основной задачей экспериментов была оценка количественных и качественных характеристик работы системы и сравнение их с аналогами.

Первый эксперимент – оценка точности распознавания и задержки получения данных от алгоритма распознавания в зависимости от освещения (табл. 1). Здесь результаты ожидаемые: камера лучше и быстрее работает при хорошем освещении.

Таблица 1. Тесты технического зрения

Условия	Точность	Задержка
>500 люкс	96 %	110 мс
300-500 люкс	94 %	120 мс
<300 люкс	85 %	150 мс

Также производилось сравнение с аналогами. В их качестве использовались DJI RoboMaster EP от компании DJI, VEX Robotics, система технического зрения Robotiq Wrist Camera для роботов Universal Robots. Разработанная система значительно меньше (в 5-10 раз) стоит, имеет более низкое (в 3-4 раза) энергопотребление, однако показывает более низкую (в 3-4 раза) скорость обработки.

Кроме того, разработанная платформа прошла испытания на трассе. Здесь сравнение с аналогами не проводилось в связи с разными видами приводов, датчиков и контроллеров, что в большинстве случаев говорит о разных областях использования. Кроме того, результаты сильно зависят от качества алгоритмов распознавания и управления, а также их программной реализации.

Разработанная система показала следующие усредненные результаты: время движения по прямой линии (10 м) – 32 с, точность движения по линии сложной траектории – 91 %, среднее время прохождения лабиринта с максимальной длиной маршрута 10 м – 2 мин. 15 с.

Одним из важнейших результатов стало использование разработанной платформы при проведении Всероссийской студенческой олимпиады «Программирование контроллеров» в 2025 году [6]. Было собрано 15 комплектов платформы. Олимпиада продолжалась три дня, во время которых оборудование использовалось в достаточно жестких условиях. При этом не было зарегистрировано ни одного серьезного отказа оборудования.

### Обсуждение результатов

1. В качестве базовых ограничений разработанной системы можно отметить, что черно-белая камера не поддерживает цветные маркеры, а также то, что достаточно большая задержка при распознавании линии ограничивает максимальную скорость движения.

2. Важнейшим преимуществом разработанной системы является более низкая стоимость по сравнению с аналогами. Заложенная модульность позволяет заменять компоненты и использовать новые программные обработчики основных событий.

3. В качестве перспективных путей развития проекта можно предложить интеграцию нейросети для детекции объектов и использование скоростной цветной камеры. Рассмотрим подробнее реализацию нейросети.

Orange Pi Zero 2 позволяет реализовать распознавание объектов, используя различные методы машинного обучения, например, сверточные нейронные сети (Convolutional Neural Networks, CNN). Они используются для классификации изображений и детектирования объектов на изображении. Для реализации этого подхода на Orange Pi Zero 2 используется установленный фреймворк TensorFlow Lite, который предоставляет оптимизированные модели глубокого обучения, специально созданные для работы на встраиваемых устройствах.

Для реализации технического зрения с распознаванием на колесном роботе можно сформировать свою модель машинного обучения или применять уже существующие и проверенные модели.

Наиболее популярные модели и их применимость к платформе.

- EfficientDet Lite (например, `efficientdet_lite0.tflite`):
  - легковесная модель, оптимизированная для работы на устройствах с ограниченными ресурсами;
  - поддерживается TensorFlow Lite и хорошо интегрируется с OpenCV;



- позволяет распознавать несколько классов объектов с приемлемой точностью.
- MobileNet SSD (Single Shot Detector):
- быстрая и компактная модель для распознавания объектов;
- часто применяется на одноплатных компьютерах и мобильных устройствах;
- может работать в режиме реального времени при разрешении 300×300 или 640×480;
- поддерживается TensorFlow Lite.
- YOLOv5n (Nano) и Tiny YOLO:
- упрощённые версии YOLO, оптимизированные для маломощных устройств;
- позволяют достичь порядка 10-20 FPS;
- требуют оптимизации и квантования моделей для повышения производительности.

На данный момент наибольшие перспективы ожидаются от применения YOLOv5n.

## Выводы

Разработана система технического зрения для применения в колесных робототехнических платформах, которая может быть использована в образовательных целях и для проведения соревнований. Точность распознавания линии составила примерно 96 %. Важной характеристикой подобных систем является стоимость. Удалось достичь значения этого показателя около 25 тыс. руб.

На основании проведенных испытаний доказана эффективность связки микрокомпьютера Orange Pi Zero 2 и микроконтроллера ESP-32 для решения образовательных задач. В области программной части системы предложенный стек ROS 2 и micro-ROS обеспечил стабильное взаимодействие модулей.

## Список литературы

1. Джозеф Лентин. Изучение робототехники с использованием Python. СПб.: Просвещение, 2019. 250 с.
2. Малогабаритный колесный автономный робот (МКАР). URL: <https://spbpu.com/proekty/malogabaritnyj-kolesnyj-avtonomnyj-robot-mkar>
3. Основные понятия о ROS и ее компонентах. URL: <https://cyberleninka.ru/article/n/osnovnye-ponyatiya-o-robot-operating-system-i-ee-komponentah/viewer>
4. Робототехническая платформа Omegabot: Education. URL: <https://robotbaza.ru/product/robototekhnicheskaya-platforma-omegabot-education>
5. Техническое зрение FANUC iRVision. URL: <http://www.robotosvarka.ru/tehnicheskoe-zrenie-fanuc-irvision>
6. VII ВСЕРОССИЙСКАЯ СТУДЕНЧЕСКАЯ ОЛИМПИАДА «Программирование контроллеров». URL: <https://robots.vlgatech.net/>
7. Orange Pi Zero – подключение и настройка. URL: <https://istarik.ru/blog/orangepi/82.html>.
8. OpenCV. URL: <https://ru.wikipedia.org/wiki/OpenCV>
9. DJI RoboMaster EP. URL: <https://dji-rus.ru/robomaster-ep/robomaster-ep/>
10. VEX Robotics. URL: <https://educube.ru/products/robototekhnicheskij-komplekt-na-baze-vex-iq-startovyy-s-tekhnicheskim-zreniem/>
11. Wrist Camera - Robotiq. URL: <https://robotiq.com/products/wrist-camera>

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-625-630

## Алгоритм оценки точности совмещения изображений после компенсации проективного преобразования

Н. А. Ходаков, Е. С. Жгутова, П. В. Бабаян

Рязанский государственный радиотехнический университет имени В.Ф. Уткина, Рязань, Россия

**Аннотация.** Рассматривается проблема оценки ошибок совмещения изображений. Приводится подход для оценки параметров проективного преобразования. Предложены математическая модель и алгоритм расчёта корреляционной матрицы параметров, построенной на якобиане нелинейной модели. Разработано программное обеспечение (с использованием пакета MATLAB), позволяющее производить анализ для соответствия точка-точка. Полученные формулы позволяют строить карты дисперсии  $Dx$ ,  $Dy$  и ковариации  $Kxy$  для каждого пикселя. Научная новизна заключается в том, что впервые аналитически исследовано влияние расположения опорных точек на точность оценки параметров проективных преобразований.

**Ключевые слова:** совмещение изображений, опорные точки, геометрические преобразования, стабилизация изображений, проективное преобразование, корреляционная матрица.

## Algorithm for assessing the accuracy of image matching after projective transformation compensation

N. A. Khodakow, E. S. Zhgutova, P. V. Babayan

Ryazan State Radio Engineering University named after V. F. Utkin, Ryazan, Russia

**Abstract.** The issue of estimating image alignment errors has been addressed. An approach to estimating the parameters of a projective transformation has been proposed. A mathematical model and algorithm for calculating a correlation matrix of parameters based on a Jacobian nonlinear model have been presented. Software has been developed using the MATLAB package that allows for point-to-point correspondence analysis. The obtained formulas make it possible to construct maps of variance  $Dx$ ,  $Dy$ , and covariance  $Kxy$  for each pixel. The scientific novelty of this work lies in the analytical investigation of the influence of reference point location on the accuracy of parameter estimation in projective transformations for the first time.

**Keywords:** image alignment, reference points, geometric transformations, image stabilization, projective transformation, correlation matrix.

### Введение

Современные алгоритмы оценки параметров геометрических преобразований демонстрируют неплохие результаты, однако в силу различных ограничений, которые накладываются областью применения в каждом конкретном случае, ошибки оценивания параметров всё равно присутствуют [1, 2]. Во многих исследованиях влияние этих ошибок полагается незначительным либо (также в качестве упрощения) считается, что статистические характеристики ошибок постоянны для всего кадра или не зависят от расположения опорных участков. Степень корректности этих предположений зависит от множества факторов.

Во многих задачах компьютерного зрения возникает необходимость совмещения изображений, полученных с различных ракурсов, то есть претерпевающих геометрические искажения. Для того чтобы скомпенсировать такие искажения, требуется оценить их параметры [3]. Анализ литературных источников выявляет три ключевых подхода к оценке этих параметров:

- глобальные методы, требующие полного попиксельного сравнения изображений;
- локальные методы, основанные на сопоставлении структурных элементов;
- нейросетевые подходы, использующие свёрточные архитектуры.

Сравнение этих подходов позволяет сделать вывод о том, что на сегодняшний день не существует методов, которые давали бы существенный выигрыш в качестве работы, скорости, аппаратной реализации и др. Однако целесообразным оказывается оценка параметров преобразований на основе выбора опорных элементов/точек.

Несмотря на разнообразие существующих методов оценки параметров преобразования, возникает необходимость более тщательной проработки вопросов, связанных с чувствительностью таких

алгоритмов к структуре опорных элементов и характеру преобразования. Это становится особенно актуальным при переходе от простых случаев — таких как параллельный сдвиг или поворот — к более сложным моделям отображения, включающим перспективные искажения.

### Постановка задачи

Для улучшения различных характеристик работы многих алгоритмов, использующих совмещение изображений, может быть применена оценка статистических характеристик точности совмещения. В наших работах [2, 3] были представлены алгоритмы и программы, позволяющие получить такие оценки для каждого пикселя изображения. Однако полученные результаты могут быть применены лишь в ограниченном числе случаев. Алгоритм и программное обеспечение, представленные в работе [2] могут компенсировать различные виды геометрических преобразований, которые включают:

- аффинное;
- подобия;
- евклидово;
- билинейное.

Однако все они являются линейными, что позволяет применить для их анализа общий подход. Среди них нет проективного преобразования, которое является одним из самых востребованных в системах компьютерного зрения, так как позволяет наиболее точно описывать искажения, которые претерпевает изображение при смене положения камеры.

Основной проблемой для описания статистических характеристик точности совмещения изображений при компенсации проективного преобразования является нелинейность самого преобразования. В данной работе предлагается алгоритм, решающий эту задачу и позволяющий получить оценки точности.

Пусть известны координаты опорных точек на исходном  $(x_i, y_i)$  и на текущем  $(x'_i, y'_i)$  изображениях. Между изображениями существует проективное преобразование, то есть опорные точки связаны уравнениями

$$\begin{aligned} x'_i &= \frac{h_1 x_i + h_2 y_i + h_3}{h_7 x_i + h_8 y_i + 1}, \\ y'_i &= \frac{h_4 x_i + h_5 y_i + h_6}{h_7 x_i + h_8 y_i + 1}, \end{aligned}$$

где  $h_i, i = 1 \dots 8$  — параметры проективного преобразования.

Требуется:

1. Найти оценку вектора параметров  $\mathbf{h} = [h_1 \dots h_8]^T$ .
2. Получить корреляционную матрицу  $\mathbf{K}_h$  ( $8 \times 8$ ).
3. Построить по ней карты дисперсии и ковариации для каждого пикселя.

### Теория

#### 1. Математическая модель и якобиан

Для того чтобы описать, какое преобразование было применено к каждой из  $N$  опорных точек, зададим  $N$  вектор-функций вида

$$f_i(\mathbf{h}) = \begin{pmatrix} \frac{h_1 x_i + h_2 y_i + h_3}{h_7 x_i + h_8 y_i + 1} \\ \frac{h_4 x_i + h_5 y_i + h_6}{h_7 x_i + h_8 y_i + 1} \end{pmatrix}, \begin{pmatrix} x'_i \\ y'_i \end{pmatrix} = f_i(\mathbf{h}) + \boldsymbol{\varepsilon}_i, \quad (1)$$

где  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2 I_{2N})$  — некоррелированный шум с единичной дисперсией (далее  $\sigma^2 = 1$ ),  $I_{2N}$  — единичная матрица размерности  $2N \times 2N$ .

Для описания чувствительности этих функций к ошибкам определения параметров преобразования используется якобиан относительно  $\mathbf{h}$ :

$$J_{xy}(\mathbf{h}) = \frac{1}{D} \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -\frac{xx'}{D} & -\frac{yx'}{D} \\ 0 & 0 & 0 & x & y & 1 & -\frac{xy'}{D} & -\frac{yy'}{D} \end{pmatrix}, \quad D = h_7x + h_8y + 1.$$

Такое описание для всех пар соответствующих друг другу точек может быть представлено в виде глобального якобиана, который имеет вид

$$\mathbf{J} = [\mathbf{J}_1^T; \dots; \mathbf{J}_N^T],$$

а его размерность равна  $2N \times 8$ .

## 2. Ковариационная матрица параметров

Пусть  $\hat{\mathbf{h}}$  — оценка максимального правдоподобия, то есть

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \sum_{i=1}^N \left\| \begin{pmatrix} x'_i \\ y'_i \end{pmatrix} - f_i(\mathbf{h}) \right\|^2.$$

Она может использоваться для компенсации перспективных искажений на изображении. Для получения ковариационной матрицы преобразование (1) необходимо линеаризовать. Произведём это посредством разложения в ряд Тейлора. Разложение будет выполняться в окрестности точки  $\hat{\mathbf{h}}$  до первого порядка:

$$f_i(\mathbf{h}) \approx f_i(\hat{\mathbf{h}}) + \mathbf{J}_i(\hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}}).$$

Для удобства дальнейшего анализа соберём вектор из тождеств для всех опорных точек:

$$\mathbf{z} = \mathbf{f}(\hat{\mathbf{h}}) + \mathbf{J}(\hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}}) + \boldsymbol{\varepsilon}.$$

Простыми преобразованиями можно привести модель к виду

$$\mathbf{y} = \mathbf{z} - \mathbf{f}(\hat{\mathbf{h}}) = \mathbf{J}(\hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}}) + \boldsymbol{\varepsilon},$$

$$\mathbf{y} = \mathbf{J}(\hat{\mathbf{h}})\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\delta} = \mathbf{h} - \hat{\mathbf{h}},$$

для которого удобно получить оценку  $\boldsymbol{\delta}$  (обозначим её как  $\hat{\boldsymbol{\delta}}$ ) методом наименьших квадратов (сократим запись, полагая  $\mathbf{J} = \mathbf{J}(\hat{\mathbf{h}})$ ):

$$\hat{\boldsymbol{\delta}} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \mathbf{y}.$$

Пользуясь теоремой Гаусса-Маркова для линейной регрессии

$$\text{Cov}(\hat{\boldsymbol{\delta}}) = \sigma^2 (\mathbf{J}^T \mathbf{J})^{-1}$$

и принимая  $\sigma^2 = 1$ , получим ковариационную матрицу системы

$$\mathbf{K}_h = (\mathbf{J}^T \mathbf{J})^{-1}.$$

## 3. Визуализация ошибки совмещения

Зная, как оценивать  $\mathbf{K}_h$ , можно получить значения дисперсий  $D_x$ ,  $D_y$  и их ковариации  $K_{xy}$  для любого пикселя изображения (и даже за его пределами, так как в оценке не учитываются размеры изображения). Для получения карт точности совмещения изображений используется якобиан  $J_{xy}(2 \times 8)$ .

$$\mathbf{K}_z(x, y) = J_{xy} \mathbf{K}_h J_{xy}^T = \begin{pmatrix} D_x(x, y) & K_{xy}(x, y) \\ K_{xy}(x, y) & D_y(x, y) \end{pmatrix}.$$

Остаётся только выполнить расчёт для каждого пикселя, который входит в область интереса. В качестве оптимизации можно проводить расчёт не для каждой точки изображения, а в узлах квадратной сетки с тем большим шагом, чем сильнее требования к быстродействию и чем слабее требования к точности.

### Программное обеспечение для оценивания точности совмещения изображений

Авторы данной статьи продолжают исследование, начатое в [5-7]. Используется модифицированное программное обеспечение, представленное в [5], доработанное для работы с проективным преобразованием.

Разработанная программа позволяет оценивать параметры проективного преобразования и производить совмещение изображений по опорным точкам, которые в экспериментах расставляются вручную. Также программа позволяет получать карты точности совмещения, построенные на основе математической модели и алгоритма, изложенных выше. Есть возможность визуализировать их как в виде контурных диаграмм, так и в виде трёхмерных графиков.

### Результаты экспериментов

Экспериментальное исследование проводилось с целью сравнения точности оценки параметров проективного преобразования при различном расположении опорных точек на изображениях. Основываясь на результатах предшествующих исследований [4-7], было выдвинуто предположение о том, что точность будет выше, если точки равномерно распределены по изображению, и ниже, если сконцентрированы в какой-либо локальной области (в экспериментах точки ставились в пределах левой верхней четверти изображения).

Выбранное для опыта изображение имело размер  $809 \times 543$  пикселя. На рисунке 1, *а* представлен кадр, который считается исходным. На рисунке 1, *б* — искажённое проективным преобразованием исходное изображение, которое считается текущим.



Рис. 1. Сюжет для экспериментов: *а* — исходный кадр; *б* — преобразованный

Эксперимент проводился для каждого варианта выбора области расположения опорных точек. Рассчитывались коэффициенты преобразования для переноса исходного изображения в систему координат текущего и корреляционная матрица ошибок ( $\sigma_z^2$  принималась равной 1), по которой строились карты точности совмещения изображений.

Корреляционная матрица  $K_z$  для каждого эксперимента представлена ниже.

При равномерном расположении точек:

0	0	-0.0044	0	0	-0.0048	0	0
0	0	-0.0071	0	0	-0.0033	0	0
-0.0044	-0.0071	1.9575	0	-0.0078	0.8614	0	0
0	0	0	0	0	-0.0015	0	0
0	0	-0.0078	0	0.0001	-0.0080	0	0
-0.0048	-0.0033	0.8614	-0.0015	-0.0080	1.7371	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Точки расположены только в левом верхнем углу изображения:

0.0012	-0.0001	-0.0231	0.0004	0.0004	-0.0405	0	0
-0.0001	0.0001	-0.0078	-0.0001	0.0002	-0.0126	0	0
-0.0231	-0.0078	2.2350	-0.0041	-0.0158	1.3107	0	0
0.0004	-0.0001	-0.0041	0.0002	0.0001	-0.0138	0	0
0.0004	0.0002	-0.0158	0.0001	0.0007	-0.0546	0	0
-0.0405	-0.0126	1.3107	-0.0138	-0.0546	5.7684	-0.0001	-0.0001
0	0	0	0	0	-0.0001	0	0
0	0	0	0	0	-0.0001	0	0

Значения элементов матриц демонстрируют, что во втором случае ошибки выражены значительно сильнее.

На рисунке 2 представлены карты точности совмещения, представляющие значения дисперсии по координате  $Y$  для каждой точки изображения:  $a$  — для первого эксперимента;  $b$  — для второго. По шкале справа от каждого графика можно определить предельные значения дисперсии и оценить, насколько точно произведено совмещение. Из рисунков видно, что в первом эксперименте разброс значений дисперсии меньше на несколько порядков.

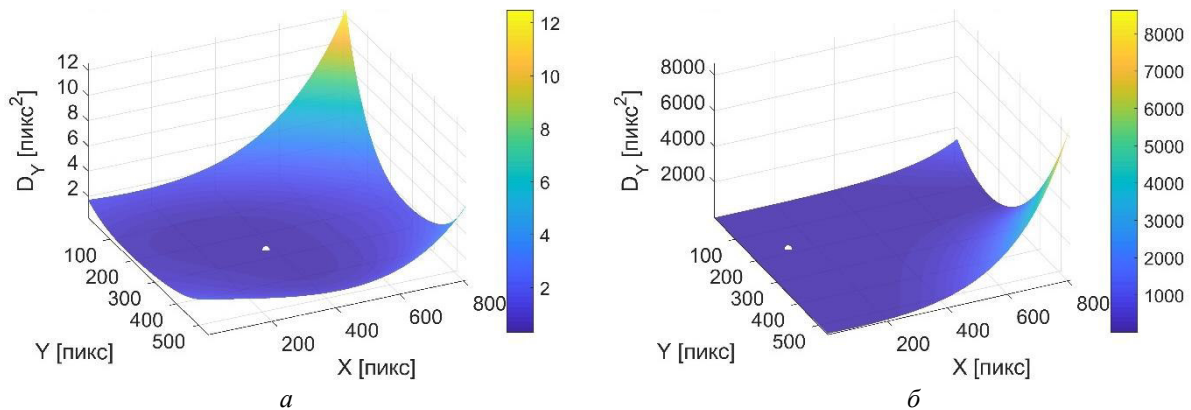


Рис. 2. Значение дисперсии по оси  $Y$  в каждой точке изображения:  
 $a$  — для первого эксперимента;  $b$  — для второго эксперимента

Проведённые эксперименты показывают зависимость точности оценки параметров изображений от характера области расположения опорных точек, что подтверждает выдвинутую гипотезу. Для лучшего совмещения изображений необходимо добиваться расположения опорных точек равномерно по всей площади кадра, отдавая предпочтение тем позициям, в которых точки находятся как можно дальше друг от друга.

Для того чтобы однозначно оценить параметры проективного преобразования, требуются, как минимум, четыре опорные точки. Для избежания получения тривиальных результатов в проведённых экспериментах их устанавливалось по пять для каждого варианта размещения.

### Выводы

В наших предшествующих работах [4-7] было показано, что точность оценивания параметров линейных геометрических преобразований локальными методами в первую очередь зависит от расположения опорных точек. Проведённые эксперименты показывают, что то же верно и для проективного преобразования, которое является нелинейным. Лучших результатов удаётся добиться, располагая опорные элементы таким образом, чтобы они отстояли друг от друга на как можно большее расстояние и покрывали как можно большую площадь.

В работах [4-7] указывается, что точность оценки параметров во многих случаях может быть повышена посредством увеличения количества опорных точек. Однако добиться таким способом точного совмещения изображений будет невозможно, если все опорные точки сконцентрированы в небольшой области изображения.

Полученные результаты могут быть использованы для улучшения работы алгоритмов, основанных на совмещении изображений или требующих совмещения для получения более корректных результатов (например, для обнаружения изменений сцен и выделения движущихся объектов). Информация о величине ожидаемой ошибки может использоваться при интерпретации результатов, полученных такими алгоритмами, отсеивая ложные срабатывания.

### Список литературы

1. Object-Based Change Detection of Very High-Resolution Images by Fusing Pixel-Based Change Detection Results Using Weighted Dempster–Shafer Theory / Y. Han, A. Javed, S. Jung, S. Liu // *Remote Sens*, 2020. 12. 983. <https://doi.org/10.3390/rs12060983>
2. Babayan P.V., Kozhina E.S. A Study of the Automatic Object Extraction Algorithm on the Image Sequence Under Scaling Transformations // 2023 25-th International Conference on Digital Signal Processing and its Applications (DSPA). IEEE, 2023. P. 1-4.
3. Обработка изображений в системах обнаружения и сопровождения объектов. Классические методы / П.В. Бабаян, С.А. Смирнов, В.С. Муравьев, В.В. Стротов. М.: Горячая линия – Телеком, 2024. 168 с.
4. Жгутова Е.С., Бабаян П.В. Алгоритмы и программное обеспечение для оценки точности совмещения изображений // *GraphiCon-2024: материалы 34-й Междунар. конф. по компьютерной графике и машинному зрению (Россия, Омск, 17–19 сент. 2024 г.)* / Ом. гос. техн. ун-т ; редкол.: Е. В. Любчинов (отв. ред.) [и др.]. Омск: Изд-во ОмГТУ, 2024. С. 484-490. DOI: 10.25206/978-5-8149-3873-2-2024-484-490
5. Бабаян П.В., Кожина (Жгутова) Е.С. Влияние точности оценивания координат опорных участков на точность совмещения изображений // *GraphiCon-2023: труды 33-й Междунар. конф. по компьютерной графике и машинному зрению (Москва, 19–21 сент. 2023 г.)*. М.: Институт прикладной математики им. М.В. Келдыша РАН, 2023. С. 474-481. DOI: 10.20948/graphicon-2023-474-481
6. Бабаян П.В., Кожина (Жгутова) Е.С. Влияние точности оценивания координат опорных участков на точность совмещения изображений в системах дистанционного зондирования Земли // 8-я международная научно-техническая конференция «В.Ф. Уткин – 100 лет со дня рождения. Космонавтика. Радиоэлектроника. Геоинформатика»: материалы докладов / Рязан. гос. радиотехн. университет им. В.Ф. Уткина. Рязань, 2023. С. 360-363. ISBN 978-5-7722-0388-0
7. Babayan P.V., Kozhina (Zhgutova) E.S. Estimation of the Accuracy of Determining the Parameters of Affine Transformations // 2024 26th International Conference on Digital Signal Processing and its Applications (DSPA). IEEE, 2024. P. 1-5. DOI: 10.1109/DSPA60853.2024.10510022



## Пороговая сегментация изображений на основе генетического алгоритма

Я. А. Буянкина, П. В. Бабаян

Рязанский государственный радиотехнический университет имени В.Ф. Уткина, Рязань, Россия

**Аннотация.** В работе рассматривается актуальная проблема сегментации изображений, полученных в условиях наблюдения, характеризующихся мультимодальностью гистограммы объекта. Для решения этой проблемы используется генетический алгоритм. В рамках исследования представлена реализация алгоритма, а также проведены эксперименты с различными изображениями, что позволяет оценить его эффективность. Результаты экспериментов включают в себя расчёт характеристик, которые демонстрируют улучшение качества сегментации по сравнению с традиционными методами, такими как метод Отцу и энтропийные методы Капура и Тсаллиса.

**Ключевые слова:** пороговая сегментация изображения, генетический алгоритм, функция распределения яркости, ROC-характеристика

## Threshold image segmentation based on genetic algorithm

Y. A. Buyankina, P. V. Babayan

Ryazan State Radio Engineering University named after V.F. Utkin, Ryazan, Russia

**Abstract.** The paper considers the current problem of segmentation of images obtained under observation conditions characterized by multimodality of the object histogram. To solve this problem, a genetic algorithm is used. The study presents the implementation of the algorithm, as well as experiments with various images, which allows us to evaluate its effectiveness. The results of the experiments include the calculation of characteristics that demonstrate an improvement in the quality of segmentation compared to traditional methods such as the Otsu method and the entropy methods of Kapur and Tsallis.

**Keywords:** threshold image segmentation, genetic algorithm, brightness distribution function, ROC characteristic

### Введение

В области обработки изображений метод сегментации применяется для выделения интересных объектов из цифровых кадров.

Одним из простейших подходов является пороговая сегментация, при которой гистограмма яркости изображения делится на две (или более) части с использованием единого глобального порога. Затем изображение сканируется попиксельно, и каждый пиксель классифицируется как принадлежащий объекту или фону в зависимости от того, превышает ли его яркость пороговое значение. Эффективность этого метода определяется степенью, до которой гистограмма яркости поддается разделению на отдельные компоненты [1].

Проблема пороговой сегментации в современной обработке изображений заключается в том, что этот метод может быть недостаточно эффективным при работе с изображениями, имеющими сложные или неоднородные гистограммы яркости. Например, в случае мультимодальных или слабо различимых распределений яркости метод может привести к некорректному разделению объектов и фона, ошибкам в классификации пикселей и потере деталей изображения.

Кроме того, выбор единого глобального порога может не учитывать локальные особенности изображения, такие как изменения освещённости или текстуры. Это может привести к тому, что важные детали будут потеряны или, наоборот, нежелательные элементы будут выделены.

Также стоит учесть, что некоторые изображения могут содержать шумы или артефакты, которые могут повлиять на точность определения порога и, соответственно, на качество сегментации.

Для улучшения результатов сегментации в таких случаях часто требуется применение более сложных и вычислительно затратных методов, таких как адаптивная пороговая обработка или использование алгоритмов машинного обучения.

В данной работе исследование направлено на пороговую сегментацию особого класса изображений, содержащих тёмный фон и контрастный объект интереса, включающий несколько зон яркости. К таким изображениям относятся, например, изображения летательных аппаратов, полученные на фоне

ночного неба, изображения предметов, наблюдаемых в 3D-сканерах, или изображения нагретых предметов, формируемые камерами ИК-диапазона. При наблюдении таких объектов в неравномерном или малом освещении их изображения содержат зоны разных уровней яркости, зеркальные или блестящие поверхности, блики. Применение традиционных методов сегментации к таким изображениям обычно не обеспечивает приемлемого качества. В настоящей работе для сегментации изображений, относящихся к рассматриваемому классу, предлагается использовать генетический алгоритм.

### Классические методы пороговой сегментации и их недостатки

Метод Отцу — это алгоритм вычисления порога бинаризации для полутонового изображения, используемый в области компьютерного распознавания образов и обработки изображений для получения бинарных изображений. С помощью данного метода вычисляется порог  $t$ , минимизирующий среднюю ошибку сегментации, т.е. среднюю ошибку от принятия решения о принадлежности пикселей изображения объекту или фону. Значения яркостей пикселей можно рассматривать как случайные величины, а их гистограмму — как оценку плотности распределения вероятностей [2]. Метод Отцу ищет порог, уменьшающий внутриклассовую дисперсию, которая определяется как взвешенная сумма дисперсий двух классов:

$$\sigma_{\omega}^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t), \quad (1)$$

где веса  $\omega_1, \omega_2$  — это вероятности принадлежности точки классам 1 и 2, разделённым порогом  $t$ ;  $\sigma_1^2, \sigma_2^2$  — дисперсии этих классов. Отцу показал, что минимизация внутриклассовой дисперсии равносильна максимизации межклассовой дисперсии.

Метод Капура относится к энтропийным методам бинаризации изображений [3]. Энтропия Капура — это метрика для многоуровневой сегментации изображений, которая разделяет изображение на разные классы, а величина энтропии определяет, является ли класс однородным. Предполагается, что  $T = [t_1, t_2, \dots, t_{k-1}]$  — представляет комбинацию пороговых значений, которые разделяют изображение на различные классы. Тогда энтропия Капура будет равна

$$f_{\text{kapur}}(t) = \sum_{j=1}^k H_j^C. \quad (2)$$

Как правило, каждое значение энтропии вычисляется независимо на основе конкретного значения  $T$ . Для многоуровневой пороговой обработки это можно выразить как

$$H_k^C = \sum_{j=t_k+1}^L \frac{Ph_j^C}{\omega_{k-1}^C} \ln \left( \frac{Ph_j^C}{\omega_{k-1}^C} \right), \quad (3)$$

где  $Ph_j^C$  — распределение вероятностей уровней интенсивности;  $C$  — единица (1) для черно-белых изображений;  $\omega_{k-1}^C$  вероятность принадлежности точки к классу  $k$ .

Недавние разработки статистической механики, основанные на концепции неэкстенсивной энтропии, также называемой энтропией Тсаллиса, усилили интерес к исследованию возможного распространения энтропии Шеннона на алгоритмы обработки изображений [4]. Энтропия Тсаллиса основана на сегментации изображений с помощью принципа максимальной энтропии. Функция Тсаллиса изначально была выведена из теории Шеннона и может быть определена как

$$S_q = \frac{1 - \sum_{j=1}^t (p_j)^q}{q-1}, \quad (4)$$

где  $t$  — пороговое значение;  $q$  — индекс энтропии [5]. Метод был создан для обработки медицинских изображений (первоначально — маммография) [6].

Каждый из этих методов имеет свои особенности и применяется в зависимости от задачи. Однако все они обладают определёнными недостатками и ограничениями, которые необходимо учитывать при выборе метода для конкретной задачи обработки изображений.

1. Чувствительность к равномерности распределения интенсивности пикселей и к шуму. Классические методы могут быть менее эффективны на изображениях с неоднородным

распределением яркости или сложными текстурами, где оптимальный порог может зависеть от локальных условий; шум на изображении способен привести к искажению результатов сегментации.

2. Возможность неправильного выбора порога при наличии нескольких пиков в гистограмме в случае, когда гистограмма интенсивности пикселей имеет несколько выраженных пиков.

3. Методы могут быть неэффективны для изображений с низкой контрастностью, сложными текстурами, неоднородным фоном, перекрывающимися объектами и т.д.

4. Часто перед применением метода пороговой сегментации может потребоваться предварительная обработка изображения, например, фильтрация шума или нормализация яркости и контрастности. Это может усложнить процесс сегментации и потребовать дополнительных вычислительных ресурсов.

Данные методы пороговой сегментации демонстрируют недостаточную эффективность при обработке изображений с тёмным фоном и контрастными объектами интереса, исходя из вышеописанных пунктов. Для решения подобных задач требуется применение альтернативных методов сегментации, более точно учитывающих специфику изображений с указанными характеристиками. В данной работе был применён генетический алгоритм в силу его особенностей:

1. Гибкость генетических алгоритмов позволяет настраивать параметры в зависимости от конкретных условий и требований к сегментации.

2. Способность находить глобальные оптимумы благодаря своему механизму популяции. Локальные оптимумы (такие как в методе Отцу) могут застревать.

3. Генетические алгоритмы, как правило, более устойчивы к шуму в изображениях, так как они используют популяцию решений, что позволяет фильтровать некачественные решения и выбирать лучшие.

4. Генетические алгоритмы могут обрабатывать задачи многоклассовой сегментации, что сложно реализовать с помощью таких методов, как метод Отцу, которые часто предназначены для бинарной сегментации.

### Генетический алгоритм

Генетический алгоритм (GA) – это естественная метаэвристика, которая имитирует гены. Генетический алгоритм используется для поиска оптимальных решений, когда не существует детерминированного метода или если детерминированный метод сложен в вычислительном отношении. GA – это алгоритм, основанный на популяции (т.е. он генерирует несколько решений на каждой итерации). Количество решений на итерации называется размером популяции. Каждое решение представлено в виде хромосомы, и каждая хромосома состоит из генов [7]. Блок-схема алгоритма представлена на рисунке 1.

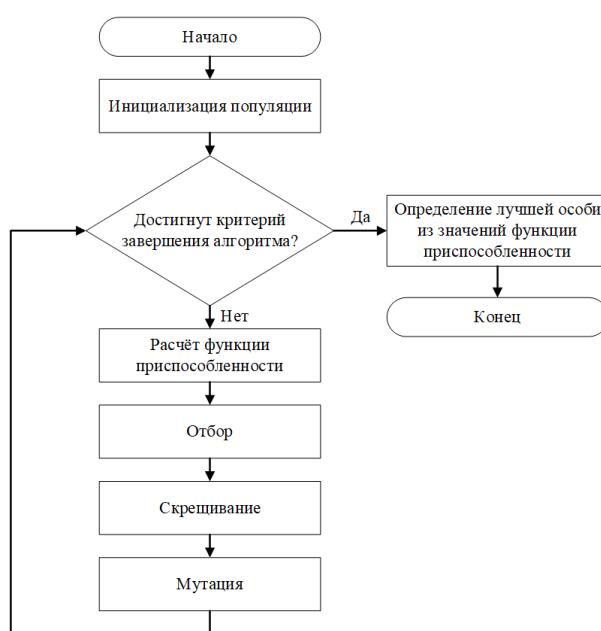


Рис. 1. Блок-схема общего генетического алгоритма

Генетический алгоритм начинается с инициализации популяции, состоящей из  $N$  случайных решений, каждое из которых представляет собой набор хромосом. Хромосомы, в свою очередь, состоят из генов, кодирующих параметры решаемой задачи. После инициализации выполняется оценка приспособленности каждого решения, что позволяет определить их качество. На основе этих оценок отбираются лучшие решения для дальнейшего использования в процессе воспроизводства.

Процесс воспроизводства состоит из четырех этапов: отбор, скрещивание, мутация и принятие решения.

На этапе отбора происходит селекция наиболее приспособленных особей из текущей популяции для формирования новой генерации. Менее приспособленные особи также могут быть отобраны, что способствует генетическому разнообразию. Существует множество механизмов реализации этапа отбора. Одним из распространенных методов является использование ролевого колеса. В этом методе колесо делится на  $N$  секций, каждая из которых соответствует отдельной хромосоме. Площадь каждой секции определяется пропорционально её приспособленности. Затем случайным образом генерируется число в диапазоне от 0 до 360, которое соответствует позиции на колесе рулетки. Это обеспечивает более высокую вероятность отбора для хромосом с более высокой приспособленностью, но также оставляет небольшой шанс для менее приспособленных особей [8].

На практике процесс отбора осуществляется следующим образом: сначала оценивается приспособленность каждой хромосомы, затем значения приспособленности нормализуются таким образом, чтобы их сумма равнялась 1. Это преобразование позволяет интерпретировать приспособленность как вероятность. Далее формируется кумулятивное распределение вероятностей, на основе которого генерируется случайное число в диапазоне от 0 до 1. Хромосома, соответствующая интервалу, в который попадает это случайное число, отбирается для формирования новой генерации. Этот процесс повторяется дважды для выбора двух хромосом, что позволяет создать новую пару хромосом для продолжения эволюционного процесса.

После выбора хромосом выполняется операция скрещивания путем выбора случайной точки в хромосомах и обмена генами после этой точки. Процесс скрещивания может застревать в локальных оптимумах. Для преодоления этой проблемы необходим «разрыв связей», который достигается с помощью операции мутации. В рамках этой операции ген выбирается случайным образом, и его значение изменяется. Гены часто представляются в виде битов, где каждый ген соответствует одному биту. В этом случае мутация заключается в небольшом случайном изменении положения гена в хромосоме.

После операций скрещивания и мутации генерируются две новые хромосомы. На заключительном этапе эти хромосомы включаются в новую популяцию. Новые хромосомы обычно принимаются в популяцию, если они превосходят по качеству своих родителей.

Завершающий этап генетических алгоритмов заключается в остановке итерационного процесса при достижении определенного критерия. Наиболее распространенным критерием остановки является количество итераций. Когда заданное число итераций выполнено, генетический алгоритм завершает свою работу.

Основная задача многопороговой сегментации найти оптимальные пороги, которые максимально точно разделят изображение на нужные сегменты. Поиск таких порогов — это задача оптимизации, которая может быть сложной из-за большого пространства возможных решений, особенно при большом количестве порогов. Для этого и применяется генетический алгоритм, который рассматривает набор порогов как индивидuum (хромосому) и с помощью операций селекции, скрещивания и мутации ищет наиболее подходящие комбинации порогов, а функция приспособленности является критерием, по которому оценивается качество каждого набора порогов.

### **Экспериментальные исследования алгоритмов. Вычисление ROC-характеристики**

Экспериментальные исследования проводились с целью оценивания качества выделения объекта интереса на изображениях. Использовался набор изображений различного характера и формата. Пример одного из изображений представлен на рисунке 2.

Для определения правильности выделения был применён метод ROC-графика (Receiver Operating Characteristic – Рабочая характеристика приёмника). Методология ROC основана на теории статистических решений и была разработана в контексте электронного обнаружения сигналов и проблем с радарами в начале 1950-х годов. Кривые истинного и ложного выделения могут быть использованы для выбора рабочей точки, которая позволила бы обеспечить оптимальный баланс между положительным и ложноположительным результатами [9].

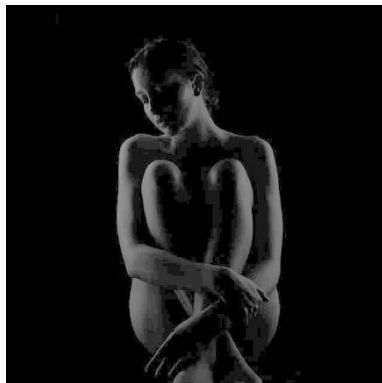


Рис. 2. Оригинальное изображение  
(контрастность -300, яркость +100)

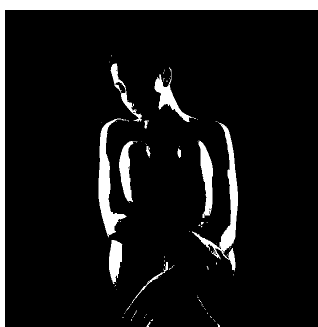


Рис. 3. Сегментированное изображение, метод Отцу (порог 7)

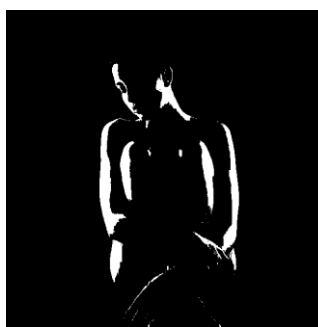


Рис. 4. Сегментированное изображение, метод Капура (порог 8)

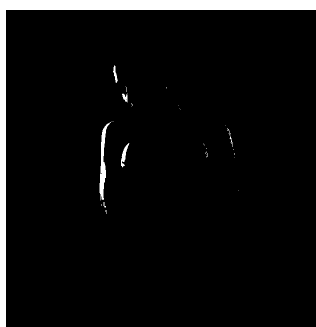


Рис. 5. Сегментированное изображение, метод Тсаллиса (порог 22)

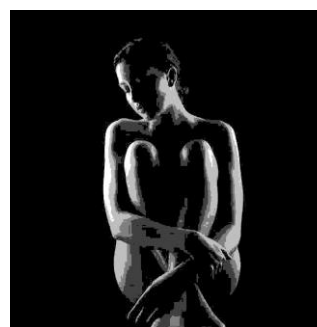


Рис. 6. Сегментированное изображение, генетический алгоритм (пороги 2 4 9 16)

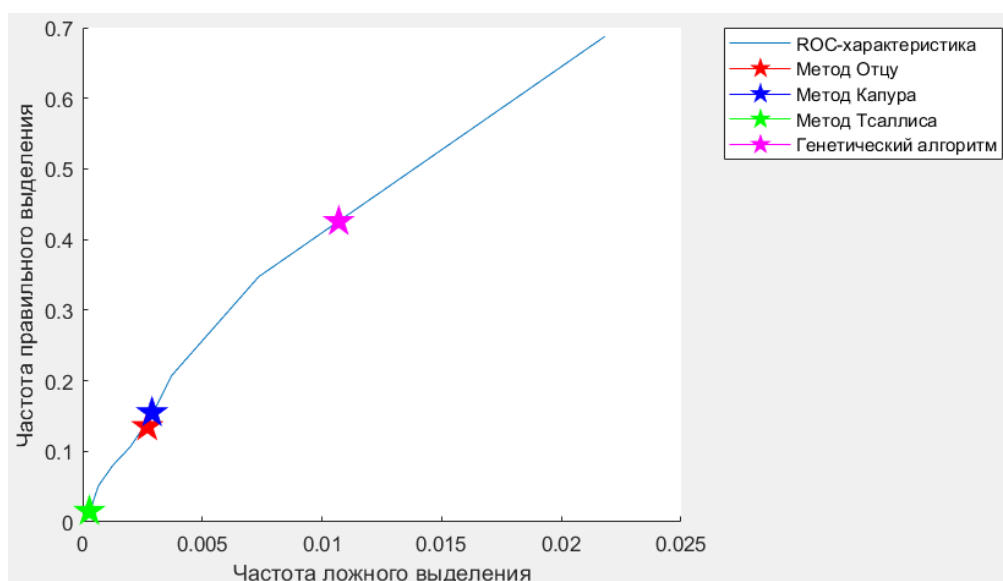


Рис. 7. ROC-характеристика с отмеченными маркерами разных методов

Результаты были сопоставлены двумя методами: визуальным и количественным. Для каждого сюжета (рис. 2) вручную создавались маски объектов, определялся порог сегментации по четырём методам (рис. 3-6), после чего определялись частоты правильного (TP – true Positive) и ложного выделения (FP – false Positive). График, отображающий зависимость TP от FP (рис. 7), помогает оценить качество выделения объектов: чем выше находится точка – тем выше значение правильного выделения, чем левее – тем ниже значение ложного выделения.

В ходе испытаний использовался набор малококонтрастных изображений, на которых нечётко выделен объект интереса (стенд). Результаты исследования, расчёт TP и FP представлены в таблице 1.

Таблица 1. Экспериментальное исследование алгоритмов сегментации

Изображение	Частоты выделения	Метод пороговой сегментации			
		Отцу	Капур	Тсаллис	Генетический алгоритм
1	TP	0,8973	0,1026	0,8703	0,8419
	FP	0,0528	0,0068	0,0499	0,056
2	TP	0,0066	0,0099	0,0116	0,7183
	FP	0,00027	0,000503	0,000701	0,0589
3	TP	0,9194	0,0344	0,8091	0,7013
	FP	0,0244	0,00019	0,0157	0,0267
4	TP	0,9164	0,2193	1	0,746
	FP	0,0357	0,0055	0,2673	0,0409
5	TP	0,5439	0,6459	1	0,694
	FP	0,0234	0,0292	0,245	0,0583
6	TP	0,0166	0,021	0,1717	0,7006
	FP	4,502e-05	6,548e-05	0,0043	0,0478
7	TP	0,045	0,1043	0,2046	0,8767
	FP	0,00026	0,00049	0,00069	0,1722

Ниже приведены графики частот правильного и ложного выделения. Можно заметить, что линия графика, соответствующая результатам генетического алгоритма, стремится к линии. Таким образом GA ищет оптимальные решения задач пороговой сегментации при разных параметрах контрастности и яркости по сравнению с классическими методами.

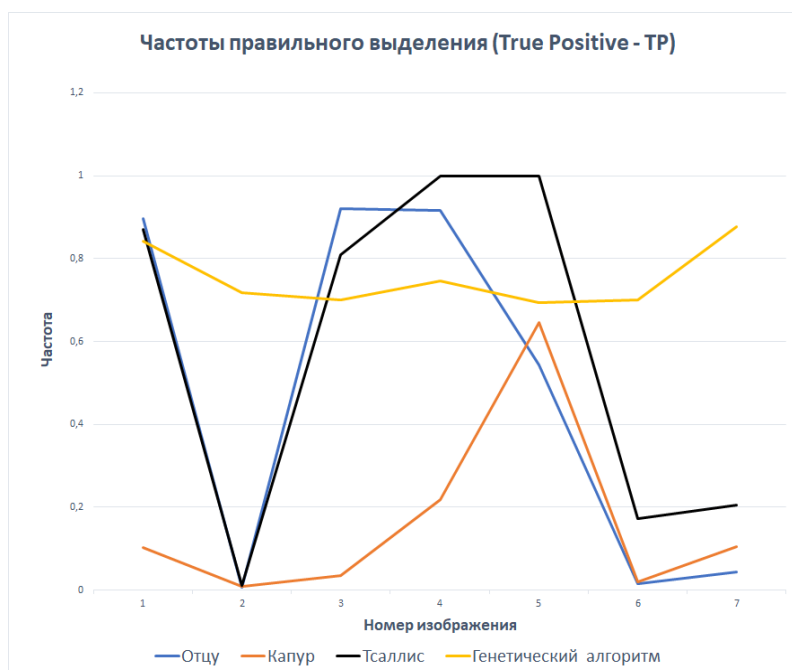


Рис. 8. График частот правильного выделения алгоритмов сегментации

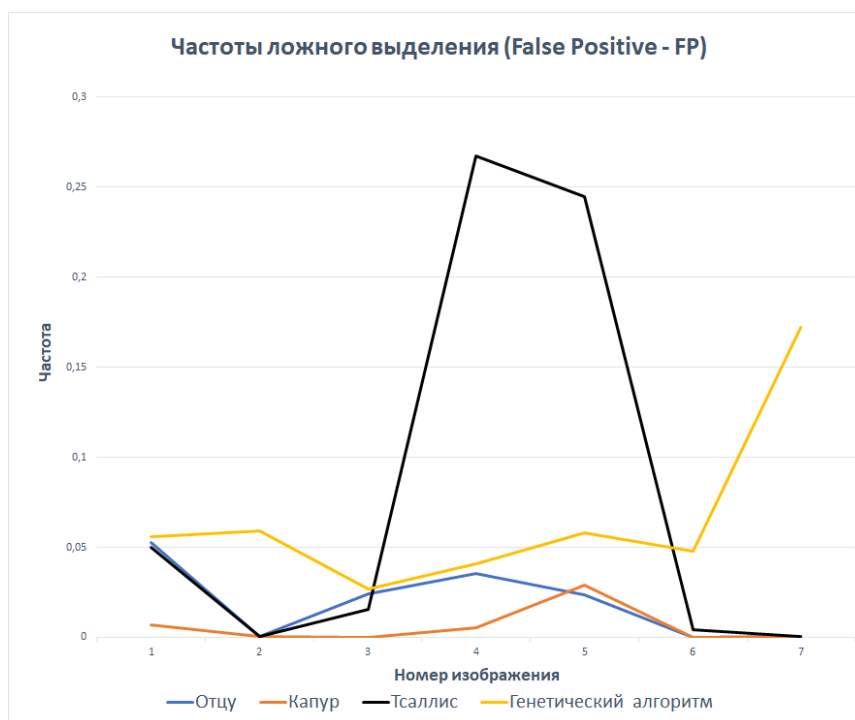


Рис. 9. График частот ложного выделения алгоритмов сегментации

Результаты данной выборки были усреднены и показаны на диаграмме на рисунке 10. Можно заметить, что самые высокие показания правильного выделения наблюдаются у генетического алгоритма.

Так как изображения включают шум и иные области с тем же уровнем яркости, что и основной объект интереса, то и показания FP будут увеличены. Для точности экспериментов необходима дополнительная обработка результатов сегментации на наличие включений и областей шума.

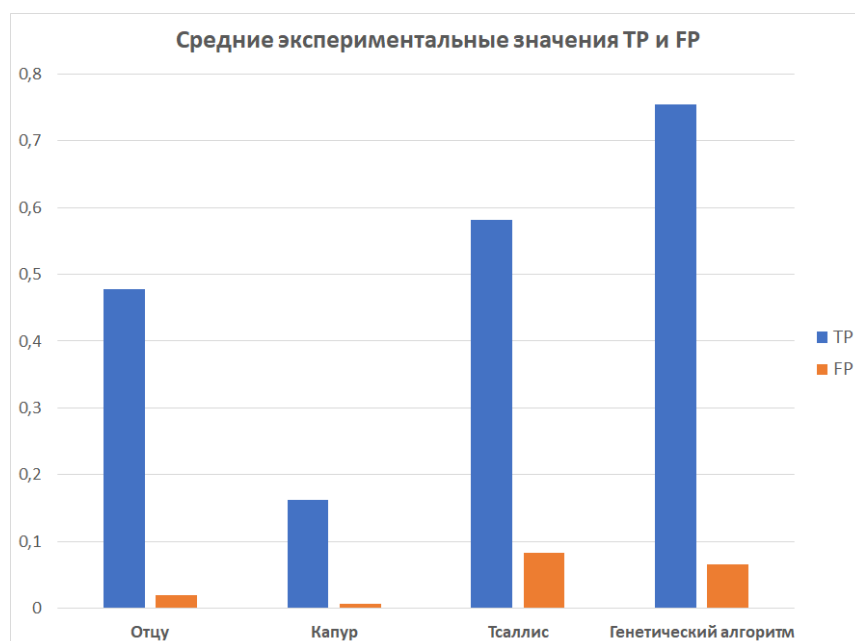


Рис. 10. Диаграмма усреднённых экспериментальных значений TP и FP

Оценка эффективности алгоритма требует комплексного подхода, учитывающего не только истинно положительные и ложноположительные результаты. Важно понимать, что простое сравнение по значениям может привести к некорректным выводам о качестве сегментации изображений.



Поскольку невозможно произвольно регулировать соотношение TP и FP, то необходимо использовать метрики, учитывающие ещё показатели FN (False Negatives — количество неправильно классифицированных отрицательных объектов) и TN (True Negatives — количество правильно классифицированных отрицательных объектов).

Accuracy (точность, правильность) — это метрика, которая показывает долю правильных классификаций от общего числа случаев [10]. Рассчитывается как отношение количества верно классифицированных объектов (как положительных, так и отрицательных) к общему количеству объектов по формуле

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision (точность) — это метрика, которая показывает, какая доля положительных предсказаний алгоритма действительно верна. Рассчитывается как отношение TP к сумме TP и FP:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

Recall (полнота) — это метрика, которая показывает, какая доля действительно положительных объектов была корректно идентифицирована алгоритмом. Рассчитывается как отношение TP к сумме TP и FN:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

F1-метрика — гармоническое среднее между точностью (precision) и полнотой (recall). Это сбалансированная метрика для оценки моделей классификации, особенно в условиях несбалансированных данных, где один класс встречается гораздо чаще, чем другой. Высокое значение F1-score (близкое к 1) указывает на сбалансированную производительность: модель может одновременно достигать высокой точности и полноты.

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

Результаты данной выборки были усреднены (табл. 2) и показаны на диаграмме (рис. 11). Можно заметить, что самые высокие показания наблюдаются у генетического алгоритма.

Таблица 2. Усреднённые значения метрик

Метрика	Метод Отцу	Метод Капура	Метод Тсаллиса	Генетический алгоритм
Accuracy	0,831771429	0,828714286	0,798742857	0,927642857
Precision	0,739757143	0,766885714	0,584042857	0,662385714
Recall	0,477885714	0,162485714	0,581042857	0,688928571
F1-score	0,429963302	0,208722784	0,411680505	0,596029788

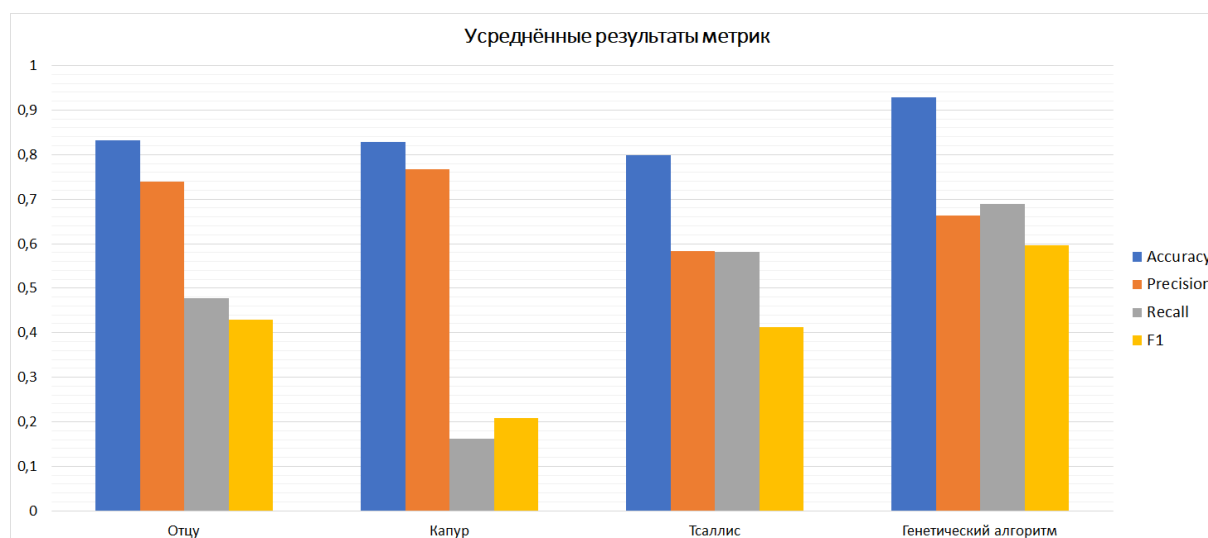


Рис. 11. Диаграмма усреднённых экспериментальных значений метрик

На основе полученных данных можно сделать вывод, что генетический алгоритм обеспечивает повышение качества сегментации по метрикам ассигасу на 11,53 % и по F1 на 38,62 % по сравнению с методом Отцу. Значения метрик precision и recall практически равны, из чего можно судить о том, что алгоритм демонстрирует сбалансированную работу между точностью и полнотой: хорошо избегает ложных срабатываний и эффективно обнаруживает большинство истинных положительных объектов.

Также нельзя не сказать о вычислительной сложности алгоритмов — ключевом аспекте при проектировании и создании программных алгоритмов, особенно при работе с большим объемом данных или выполнении вычислительно сложных операций. Понимание, как оценивать сложность алгоритмов, помогает принимать обоснованные решения в выборе алгоритмов и структур данных, а также оптимизировать производительность своих приложений. В таблице 3 представлены усреднённые затраты времени.

Таблица 3. Анализ вычислительной сложности алгоритмов

Размер изображения (px)	Затраченное время на обработку изображений, с			
	Метод Отцу	Метод Капура	Метод Тсаллиса	Генетический алгоритм
250×250	0,018451	0,022705	3,870947	0,069
640×640	0,019144	0,026152	31,719965	0,081
1024×1024	0,02907	0,034318	78,20478	0,097

На основе полученных данных можно сделать вывод, что генетический алгоритм демонстрирует значительно меньшую вычислительную сложность по сравнению с методом Тсаллиса, особенно при увеличении размера изображения. В данном случае алгоритм Тсаллиса настроен на изображения различного характера, если задавать точно предел изменяемого индекса энтропии  $q$  (см. формулу (4), то можно сократить время обработки в несколько раз. По сравнению с методами Отцу и Капура, генетический алгоритм имеет меньшую скорость обработки изображений, хотя разница не существенна.

Таким образом, можно сделать вывод, что генетический алгоритм пороговой сегментации изображений можно применять в обработке видеопоследовательностей в режиме реального времени, если камера будет снимать не более 10-12 кадров в секунду (зависит от разрешения камеры).

### Заключение

В ходе проведенного исследования был реализован алгоритм сегментации, проведены сравнительные исследования с классическими методами пороговой сегментации. Построены ROC-характеристики и графики частот правильного и ложного выделения. Рассчитаны некоторые метрики, характеризующие качество сегментации.

Применение сегментации с использованием генетического алгоритма (GA) продемонстрировало значительное улучшение качества выделения контрастных объектов интереса на рассматриваемой выборке изображений. Генетический алгоритм позволяет эффективно идентифицировать и разделять различные компоненты изображения, включая шум, объекты интереса и яркие блики. Это обеспечивает возможность целенаправленного выбора нужной части изображения, содержащей требуемый объект интереса.

В дальнейшем планируется расширение возможностей алгоритмов для улучшения качества сегментации; увеличение выборки, что позволит рассмотреть другие случаи дефектов качества изображений, включение постобработки результатов для повышения параметров правильного выделения.

### Список литературы

1. Буянкина Я.А., Бабаян П.В. Пороговые методы выделения объекта интереса на изображении на основе методов Отцу, Капура и Тсаллиса // Новые информационные технологии в научных исследованиях: материалы XXIX Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. Рязань: ИП Коняхин А.В., 2024. 260 с. URL: <https://rsreu.ru/faculties/fvt/kafedri/saprvs/konferentsiya-nit/16380-item-16380>

2. Nobuyuki Otsu A. Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, vol. 1(9), pp. 62-66. DOI: 10.1109/TSMC.1979.4310076. Available at: <https://ieeexplore.ieee.org/document/4310076/authors#authors>
3. Amelia Carolina Sparavigna. On the Role of Tsallis Entropy in Image Processing. *International Scientific Research Journal*, 2015, vol. 1(6), pp. 16-24. DOI: 10.18483/IRJSci.79. Available at: <https://hal.science/hal-01352704v1>
4. P.D. Sathya, R. Kayalvizhi PSO-Based Tsallis Thresholding Selection Procedure for Image Segmentation. *International Journal of Computer Applications* (0975 – 8887), August 2010, vol. 5, no. 4, pp. 39-46. Available at: <https://www.ijcaonline.org/archives/volume5/number4/903-1279/>
5. K. Suresh Manic, R. Krishna Priya, V. Rajinikanth. Image Multithresholding based on Kapur/Tsallis Entropy and Firefly Algorithm. *Indian Journal of Science and Technology*, 2016, vol. 9(12), pp. 1-6. DOI: 10.17485/ijst/2016/v9i12/89949. Available at: <https://indjst.org/articles/image-multithresholding-based-on-kapurtsallis-entropy-and-firefly-algorithm>
6. Prem Kumar Kalra, Nirmal Kumar An automatic method to enhance microcalcifications using normalized Tsallis entropy. *Signal Processing*, March 2010, vol. 90, pp. 952-958. DOI: 10.1016/j.sigpro.2009.09.012. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0165168409004034?via%3Dihub>
7. Omar Banimelhem, Yahya Ahmed Yahya. Multi-Thresholding Image Segmentation Using Genetic Algorithm. *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, 2011. Available at: <https://www.semanticscholar.org/paper/Multi-Thresholding-Image-Segmentation-Using-Genetic-Banimelhem-Yahya/589b57ba67ae9b87c256ec45ad36b36d89f110d5>
8. Yu W., Huang M., Zhu D., Li X. A Method of Image Segmentation Based on Improved Adaptive Genetic Algorithm. *Foundations of Intelligent Systems. Advances in Intelligent and Soft Computing*, vol 122. Springer, Berlin, Heidelberg, 2011. Available at: [https://doi.org/10.1007/978-3-642-25664-6\\_60](https://doi.org/10.1007/978-3-642-25664-6_60)
9. Zweig M. H., Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, 1993, vol. 39, no. 4, pp. 561-577. DOI: 10.1093/clinchem/39.4.561. Available at: <https://doi.org/10.1093/clinchem/39.4.561>
10. Кугаевских А.В., Муромцев Д.И., Кирсанова О.В. Классические методы машинного обучения. СПб.: Университет ИТМО, 2022. 53 с. URL: <https://books.ifmo.ru/file/pdf/3075.pdf>

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-641-648

## Опыт автоматического распознавания объектов на видеоконтенте, распространяемых в социальных сетях, с применением YOLOv12

Е. А. Белов, Е. Б. Замятина

Национальный исследовательский университет "Высшая школа экономики", Пермь, Россия

**Аннотация.** В статье рассмотрены возможности применения модели YOLOv12 для автоматического распознавания объектов на видеоконтенте, размещённом в социальных сетях. Особое внимание уделено распознаванию нестандартных объектов, таких как оружие, пламя, маскировка. Описаны этапы формирования обучающей выборки, дообучения модели и проведения тестирования на реальных видеозаписях из VK, Telegram и YouTube. Представлены количественные результаты по точности (mAP), полноте, скорости обработки и устойчивости модели к низкому качеству видео. Обоснована применимость подхода для задач визуального поиска, автоматической модерации и мультимедийной аналитики.

**Ключевые слова:** YOLO, компьютерное зрение, видеоаналитика, социальные сети, распознавание объектов.

## Methods for automatic recognition of objects in video content from social networks using YOLOv12

E. A. Belov, E. B. Zamyatina

National Research University Higher School of Economics, Perm, Russia

**Abstract.** This article explores the applicability of the YOLOv12 model for automatic object recognition in video content published on social media platforms. Particular focus is placed on the detection of non-standard object classes such as weapons, fire, and concealment. The paper describes the stages of training dataset preparation, model fine-tuning, and evaluation on real videos from VK, Telegram, and YouTube. Quantitative results are provided in terms of detection accuracy (mAP), recall, processing speed, and robustness under degraded video quality. The proposed approach is shown to be applicable to visual search, automated moderation, and multimedia analytics tasks.

**Keywords:** YOLO, computer vision, video analytics, social media, object detection.

### Введение

За последние годы задачи автоматического анализа видеоконтента приобрели особую актуальность в связи с экспоненциальным ростом объёмов пользовательских видео, публикуемых в социальных сетях [1, 2]. Согласно отчётам Cisco и Statista, доля видеотрафика в мировом интернете к 2025 году превысит 85 %. Современные исследования в области компьютерного зрения и мультимедийной аналитики сосредоточены на разработке эффективных методов извлечения смысловой информации из видеопотока.

Несмотря на активное развитие методов объектной детекции, большинство академических исследований фокусируется на задачах обработки формализованных датасетов — COCO, Pascal VOC, Open Images и др. Работы, представленные на таких конференциях, как CVPR, ECCV, NeurIPS, ICIP, а также в специализированных журналах (IEEE TPAMI, MDPI Sensors, Computer Vision and Image Understanding), редко затрагивают задачи детекции в условиях реальных пользовательских видеороликов, размещённых в интернете [3-5]. Таким образом, существует исследовательский пробел в области применения детекторов объектов на неструктурированных, стихийно загруженных видеопотоках из соцмедиа.

Широкое распространение получили сверточные нейросетевые архитектуры, ориентированные на детекцию и классификацию объектов в изображениях и видео. Среди них особое внимание уделяется семейству моделей YOLO (You Only Look Once), демонстрирующих высокую скорость и достаточную точность при решении задач одновременного распознавания и локализации объектов [6].

Проблема эффективного поиска и систематизации видеоматериалов исключительно по текстовым признакам (аннотациям, хештегам) уже продемонстрировала свою ограниченность. Пользовательские описания подвержены субъективности, не отражают визуального содержания и не позволяют автоматически выявлять критически важные объекты (например, огонь, оружие, символику, опасные действия) [7]. В то же время существующие сервисы компьютерного зрения, такие как AWS

ReKognition или Microsoft Video Indexer, зачастую не поддерживают интеграцию с российскими социальными платформами, не обладают достаточной гибкостью для обработки специфических классов объектов и требуют высокого порога технической подготовки.

Кроме того, данное направление представляет собой мультиотраслевую исследовательскую и прикладную задачу. Оно находит применение:

- в **безопасности**: автоматическое выявление потенциальных угроз, признаков чрезвычайных ситуаций;
- в **модерации**: идентификация нежелательного контента на платформах;
- в **маркетинге и бренд-аналитике**: отслеживание присутствия логотипов или рекламных вставок;
- в **поведенческом и социологическом анализе**: изучение визуальных трендов, событий, активности пользователей.

Отдельного внимания заслуживает формирующееся направление — **видеоаналитика в социальных сетях**, в рамках которого осуществляется автоматический анализ визуального контента в реальном времени или ретроспективно, с опорой на API-платформы и адаптированные модели глубокого обучения. Эта область объединяет достижения в сфере компьютерного зрения, машинного обучения, архитектуры программных систем и работы с потоковыми данными.

Таким образом, возникает потребность в создании и исследовании методов, адаптированных к условиям пользовательского видеоконтента, способных эффективно выделять целевые объекты в потоке с учётом его нестабильности, разнообразия форматов и отсутствия предварительной структуры.

### Постановка задачи

Задача автоматического распознавания объектов в пользовательском видеоконтенте из социальных сетей представляет собой частный случай задачи объектной детекции в условиях нестабильного качества данных, нестандартизированных форматов и высокой визуальной вариативности. В рамках настоящей работы рассматривается видеопоток  $V$ , представленный в виде упорядоченной последовательности кадров  $\{F_1, F_2, \dots, F_n\}$ , на каждом из которых необходимо выполнить локализацию и классификацию целевых объектов.

Целевыми объектами в данном исследовании считаются элементы из множества классов  $C = \{c_1, c_2, \dots, c_k\}$ , включающих:

- опасные предметы: оружие, пламя, дым;
- признаки потенциально нарушающего контента: маскировка лица, нестандартные поведенческие сцены;
- стандартные контрольные классы: человек, транспорт, животное.

На каждом кадре  $F_i$  необходимо найти множество предсказаний:

$$O_i = (x_j, y_j, w_j, h_j, c_j, s_j)_{j=1}^{m_i},$$

где  $(x_j, y_j, w_j, h_j)$  — координаты ограничивающей рамки (bounding box);  $c_j \in C$  — предсказанный класс,  $s_j \in [0, 1]$  — вероятность (confidence score).

Постановка задачи формализуется следующим образом:

- Требуется обучить и адаптировать модель  $M$ , способную на каждом кадре  $F_i$  детектировать объекты из  $C$  с высокой точностью  $\text{Precision} \geq 0.85$  и полнотой  $\text{Recall} \geq 0.75$ , при этом обеспечивая привязку результата к временной метке  $t_i$ , соответствующей кадру.
- Дополнительно необходимо обеспечить устойчивость модели к изменениям в качестве видео (разрешение, шум, фильтры) и обеспечить возможность масштабирования модели на более широкий спектр видеосцен.

Для решения поставленной задачи предлагается использовать модель YOLOv12 как базовую архитектуру объектного детектора, применимую к последовательности кадров, извлечённых из пользовательских видеороликов [8].

Целью данной работы является исследование эффективности модели YOLOv12 для задач автоматического распознавания объектов в видеоконтенте из социальных сетей с акцентом на нестандартные классы и условия съёмки.

## Теория

Объектная детекция в изображениях и видеопотоках является одной из ключевых задач в области компьютерного зрения. Её цель — локализовать и классифицировать все значимые объекты на изображении, что используется в таких приложениях, как автономное вождение, видеонаблюдение, медицина, промышленность и цифровые медиа. В последние годы основной прорыв в этой области обеспечили глубокие нейронные сети, в частности сверточные архитектуры, способные извлекать и обобщать сложные визуальные признаки.

Развитие методов детекции прошло несколько этапов:

**Классические методы** (до 2012 г.). На раннем этапе применялись алгоритмы ручного извлечения признаков [9, 10]:

- **SIFT (Scale-Invariant Feature Transform)**,
- **HOG (Histogram of Oriented Gradients)**,
- классификаторы на основе **Support Vector Machine (SVM)** и **Adaboost**.

Эти методы обладали ограниченной устойчивостью к поворотам, масштабам и шумам, а также не могли масштабироваться к большим наборам данных.

Переход к глубокому обучению. Прорыв был достигнут после победы AlexNet на ImageNet в 2012 году, что открыло путь к использованию сверточных нейронных сетей (CNN) для извлечения признаков.

**Эволюция нейросетевых детекторов:**

- **R-CNN (2014)** — первый успешный подход с использованием CNN: предложил регионы интереса (ROI) и классифицировал каждый из них [11];
- **Fast R-CNN** ускорил процесс путём объединения извлечения признаков и классификации [11];
- **Faster R-CNN** ввёл Region Proposal Network (RPN), сделав весь процесс end-to-end [11];
- **YOLO (You Only Look Once)** кардинально изменил подход, выполняя детекцию и классификацию за один проход по изображению. Это обеспечило высокую скорость и применимость в реальном времени;
- **SSD (Single Shot MultiBox Detector)** предложил альтернативную однопроводную архитектуру с несколькими уровнями разрешения [12];
- **DETR (DEtection TRansformer)** — трансформерный подход, обеспечивающий глобальное внимание, но требующий больше вычислительных ресурсов [13].

Современные подходы к задачам детекции объектов в изображениях и видеопотоке базируются на использовании глубоких сверточных нейронных сетей, способных автоматически извлекать пространственные и семантические признаки из визуальных данных. Одним из наиболее эффективных и широко применяемых семейств моделей в этой области является YOLO (You Only Look Once), ключевая особенность которого — выполнение локализации и классификации объектов за один проход по изображению. Это отличает YOLO от двухэтапных архитектур, таких как R-CNN и его производные, где сначала генерируются области интереса (Region Proposals), а затем для каждой области выполняется классификация.

YOLO воспринимает изображение как единую сущность, разделяет его на регулярную сетку и предсказывает для каждой ячейки ограничивающие рамки (bounding boxes), классы объектов и коэффициенты уверенности. Это обеспечивает высокую скорость обработки и делает модель подходящей для применения в реальном времени, включая видеопотоки и мобильные вычислительные платформы. Такие качества особенно важны в условиях пользовательского контента из социальных сетей, где задержка обработки и вычислительная нагрузка являются критичными параметрами.

Последняя на момент исследования версия YOLOv12 представляет собой дальнейшее развитие архитектуры, объединяющее в себе идеи трансформеров и сверточных слоёв. В качестве механизма извлечения признаков (бэбона) используется модифицированный Swin Transformer, позволяющий эффективно агрегировать контекстные зависимости между объектами в различных частях изображения. Это особенно важно при наличии сложных визуальных сцен, характерных для пользовательских видеороликов. Кроме того, YOLOv12 реализует многомасштабную обработку (multi-scale prediction), позволяющую точно детектировать как крупные, так и мелкие объекты, а также

оптимизацию вычислений за счёт использования смешанной точности (mixed precision), что ускоряет обучение и инференс на современных GPU [14, 15].

Обработка видеоконтента с использованием моделей детекции требует предварительного разбиения видеопотока на отдельные кадры. Обычно выбирается определённый временной интервал (например, 1 кадр в секунду), что позволяет снизить объём данных и обеспечить репрезентативность выборки. Каждому кадру присваивается временная метка, и на его основе выполняется объектная детекция. Это позволяет зафиксировать не только факт появления целевого объекта, но и его точное положение во времени, что критично для задач последующего анализа, таких как построение временной шкалы событий, автоматическая модерация, мониторинг или видеописк.

Важно отметить, что модели общего назначения (обученные, например, на наборах данных COCO или Open Images) не демонстрируют достаточной точности при распознавании специфических объектов, таких как оружие, огонь, дым или маскировка лица. Кроме того, в условиях низкого разрешения, шумов, нестабильной экспозиции и применённых фильтров такие модели склонны ошибаться или не распознавать объекты вовсе. Для повышения эффективности детекции в пользовательских видеороликах требуется дополнительная адаптация модели путём дообучения (fine-tuning) на специализированных датасетах, собранных с учётом целевых классов и визуальных искажений, характерных для контента из соцсетей.

Оценка качества модели детекции объектов производится с использованием набора стандартных метрик. Основной является mAP (mean Average Precision), рассчитываемая как усреднённая точность при разных порогах пересечения ограничивающих рамок (IoU). Дополнительно используются такие показатели, как точность (Precision), полнота (Recall), а также скорость обработки (Frames Per Second, FPS), позволяющая оценить пригодность модели для работы в реальном времени.

В контексте настоящего исследования YOLOv12 выбрана как базовая модель на основании баланса между качеством предсказаний и вычислительной эффективностью. Её архитектура позволяет точно и быстро обрабатывать видео с нестабильными условиями съёмки, характерными для пользовательского контента, а возможность кастомизации и дообучения делает её гибким инструментом для задач, требующих распознавания редких или чувствительных классов объектов.

Таким образом, теоретическая основа проведённого исследования опирается на применение современной глубокой нейросетевой архитектуры YOLOv12, адаптированной к задачам потоковой видеоаналитики в условиях нестандартизированных и искажённых данных, с акцентом на расширяемость, интерпретируемость и производительность.

### Результаты экспериментов

Экспериментальное исследование проводилось с целью количественной оценки эффективности модели YOLOv12 при распознавании объектов на видеоконтенте из открытых цифровых платформ. В качестве исходных данных использовались видеоролики, опубликованные пользователями в социальных сетях VK, Telegram и YouTube Shorts. Основной акцент был сделан на ролики с нестандартными визуальными условиями: вертикальной ориентацией, низким разрешением (480p и ниже), шумами, фильтрами, нестабильным освещением и быстрой сменой сцены.

Для дообучения модели была вручную собрана и размечена выборка, включающая 8 классов объектов:

- человек,
- транспорт,
- животное,
- огнестрельное оружие,
- пламя,
- маскировка лица,
- сцены агрессии,
- элементы военной атрибутики.

Всего было размечено **3 127 кадров**, извлечённых из **320 видеороликов**, с общим числом **5 842 объектов**. Разметка производилась в формате YOLOv5/Ultralytics с последующей конвертацией под



формат входных данных YOLOv12. Выборка была разбита на обучающую и тестовую части в пропорции 80/20. Дополнительно использовалась встроенная система аугментаций (random brightness, mosaic, blur), направленная на повышение устойчивости к шумам и искажениям.

Дообучение модели производилось с использованием библиотеки **Ultralytics YOLOv12** в среде **Python 3.10**, с применением **PyTorch 2.1.0**. Эксперименты выполнялись на GPU **NVIDIA RTX 3090** с 24 GB VRAM. Параметры обучения:

- число эпох: 100;
- размер батча: 16;
- размер входного изображения: 640×640;
- оптимизатор: AdamW;
- стратегия early stopping по метрике mAP.

Оценка результатов проводилась на тестовой подвыборке с использованием стандартных метрик (таблица).

Результаты оценки модели

Метрика	Значение
mAP@0.5	0.871
mAP@0.5:0.95	0.663
Precision	0.894
Recall	0.752
FPS (на RTX 3090)	59.7

Модель уверенно распознаёт стандартные классы (человек, транспорт), а также демонстрирует приемлемую точность по редким и нестандартным классам (оружие, пламя, маскировка). Визуальный анализ результатов показал, что большинство ошибок возникает в условиях чрезмерной зашумлённости изображения, сильных искажений из-за фильтров или наличия пересекающихся объектов на кадре.

На рисунках 1 и 2 представлены примеры работы модели в процессе обучения на кадрах из реальных видеороликов Telegram и VK с реальными и предсказанными метками соответственно. Объекты успешно локализованы даже при наличии артефактов и нестандартных углов обзора. Каждый детектированный объект сопровождается указанием класса и confidence score, а также привязкой к временной отметке кадра.



Рисунок 1. Аннотированные модели на валидационных изображениях

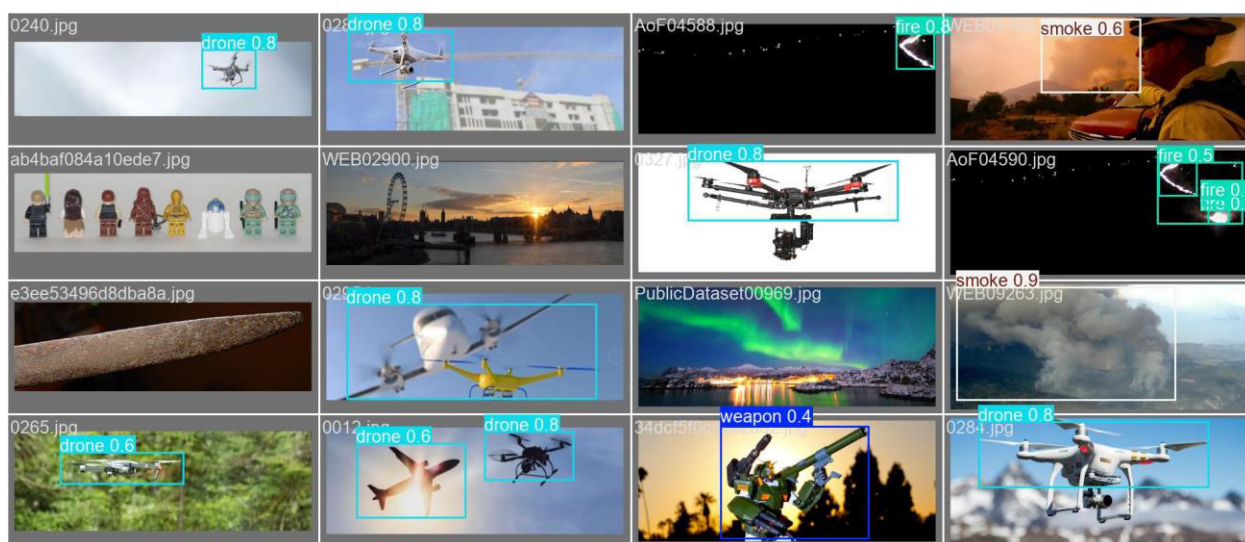


Рисунок 2. Предсказания модели на валидационных изображениях

### Обсуждение результатов

Полученные экспериментальные результаты подтверждают гипотезу о применимости модели YOLOv12 для задач объектной детекции в условиях пользовательского видеоконтента из социальных сетей. Достигнутые значения метрик ( $mAP@0.5 = 0.871$ ,  $Precision = 0.894$ ,  $FPS \approx 60$ ) свидетельствуют о высокой эффективности подхода при наличии предварительного дообучения на предметно-ориентированной выборке.

Важно отметить, что модель продемонстрировала устойчивость к ряду типичных искажений, характерных для видео в открытых цифровых медиа: вертикальное соотношение сторон, сжатие, шумы, нестабильная экспозиция, агрессивные цветовые фильтры. В частности, при тестировании на данных из Telegram с сильными визуальными артефактами точность обнаружения нестандартных классов (оружие, пламя) снижалась в среднем на 10–15 %, однако оставалась на уровне, превышающем порог допустимого для автоматической фильтрации контента.

Сравнение с базовыми предобученными версиями YOLOv8 и YOLOv5 показало, что без дополнительной адаптации их производительность по нестандартным классам была неудовлетворительной ( $mAP < 0.3$ ). Это подчёркивает значимость кастомной дообученной модели, а также важность предварительной подготовки обучающей выборки с учётом специфики задач.

Модель также показала высокую скорость обработки (почти 60 FPS на RTX 3090), что делает её применимой не только для офлайн-аналитики, но и для внедрения в системы потоковой видеообработки, где критична задержка в распознавании.

Ограничения выявлены при следующих сценариях:

- сильная частичная окклюзия объектов (например, лицо скрыто рукой или предметами);
- резкая смена сцены в пределах одного видеокadra;
- наложение текста, эмодзи или интерфейсных элементов поверх изображения;
- очень малые объекты (меньше 3–5 % площади кадра) при низком разрешении.

Также можно отметить, что увеличение количества редких классов требует значительно большего объёма размеченных данных, поскольку модель склонна к переобучению или игнорированию плохо представленных категорий.

Тем не менее в большинстве реалистичных пользовательских сценариев модель успешно справляется с задачей и обеспечивает приемлемое качество предсказаний даже при ограниченном объёме дообучения. Это делает её применимой в таких прикладных областях, как:

- автоматическая модерация видеоконтента в социальных медиа;
- визуальный мониторинг рисков и инцидентов;
- видеописк по содержанию на платформе;
- аналитика поведения и контекста (например, агрессия, опасные ситуации).

Таким образом, экспериментально подтверждена применимость архитектуры YOLOv12 в задачах, выходящих за рамки классических датасетов и охватывающих сценарии, связанные с анализом пользовательского видеоконтента из открытых сетей.

### Заключение

В данной работе рассмотрена задача автоматического распознавания объектов на пользовательских видеороликах из социальных сетей с использованием модели YOLOv12. Основное внимание уделялось выявлению нестандартных и слабо представленных классов объектов (оружие, пламя, маскировка), а также оценке устойчивости модели к условиям низкого качества и нестабильного видеопотока.

На основе сформированной размеченной выборки и проведённого дообучения модели получены следующие ключевые результаты:

- достигнута высокая точность детекции объектов на реальных видеоданных из VK, Telegram и YouTube Shorts ( $mAP@0.5 = 0.871$ ;  $Precision = 0.894$ );
- обеспечена скорость обработки, достаточная для применения в системах реального времени (более 59 кадров в секунду на GPU уровня RTX 3090);
- подтверждена способность модели выявлять нестандартные визуальные признаки даже в условиях шума, фильтров и нестабильной съёмки;
- продемонстрирована применимость подхода для задач автоматической модерации, визуального мониторинга и аналитики пользовательского контента.

Проведённый анализ показал, что даже ограниченная по объёму кастомная выборка, включающая целевые классы и искажения, позволяет существенно повысить эффективность нейросетевого детектора по сравнению с универсальными предобученными моделями. Вместе с тем сохранение стабильного качества распознавания при расширении набора редких классов требует системного подхода к формированию датасета и регулярной валидации модели на новых видеосценах.

Перспективы дальнейшей работы включают:

- автоматизацию процесса извлечения и разметки кадров с привлечением активного обучения;
- интеграцию модели в потоковые конвейеры обработки видео;
- расширение исследуемых классов объектов и переход к мультимодальному анализу (учёт аудио и текста поверх видео);
- адаптацию и внедрение в инфраструктуру отечественных платформ для анализа пользовательского медиаконтента.

Таким образом, проведённое исследование подтверждает целесообразность и эффективность применения современных нейросетевых архитектур, таких как YOLOv12, для решения прикладных задач анализа видеоданных в условиях реальных пользовательских сценариев.

### Список литературы

1. Wang S., Ji Q. Video Affective Content Analysis: A Survey of State-of-the-Art Methods // IEEE Transactions on Affective Computing. 2015. Vol. 6. P. 410–430.
2. Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study / H. Egger, G. Dawson, J. Hashemi, et al. // NPJ Digital Medicine. 2018. Vol. 1.
3. A real-time object detection algorithm for video / S. Lu, B. Wang, H. Wang, L. Chen, X. Zhang // Computers and Electrical Engineering. 2019. Vol. 77. P. 398–408.
4. New Generation Deep Learning for Video Object Detection: A Survey / L. Jiao, R. Zhang, F. Liu, et al. // IEEE Transactions on Neural Networks and Learning Systems. 2021. Vol. 33. P. 3195–3215.
5. Video anomaly detection based on spatio-temporal relationships among objects / Y. Wang, T. Liu, J. Zhou, J. Guan // Neurocomputing. 2023. Vol. 532. P. 141–151.
6. Diwan T., Anirudh G., Tembhurne J. Object detection using YOLO: challenges, architectural successors, datasets and applications // Multimedia Tools and Applications. 2022. Vol. 82. P. 9243–9275.
7. A comprehensive review of the video-to-text problem / J. Perez-Martin, B. Bustos, S. Guimarães, et al. // Artificial Intelligence Review. 2021. Vol. 55. P. 4165–4239.

8. Terven J., Córdova-Esparza D., Romero-González J. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS // *Machine Learning and Knowledge Extraction*. 2023. Vol. 5. P. 1680–1716.
9. Efficient object detection and classification on low power embedded systems / S. Jagannathan, K. Desappan, P. Swami, et al. // *IEEE Int. Conf. on Consumer Electronics (ICCE)*. 2017. P. 233–234.
10. Multi-vehicle detection algorithm through combining Harr and HOG features / Y. Wei, Q. Tian, J. Guo, W. Huang, J. Cao // *Mathematics and Computers in Simulation*. 2018. Vol. 155. P. 130–145.
11. Cai Z., Vasconcelos N. Cascade R-CNN: High Quality Object Detection and Instance Segmentation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019. Vol. 43. P. 1483–1498.
12. SSD: Single Shot MultiBox Detector / W. Liu, D. Anguelov, D. Erhan, et al. // *Computer Vision – ECCV 2016*. 2016. P. 21–37.
13. End-to-End Object Detection with Transformers / N. Carion, F. Massa, G. Synnaeve, et al. // *European Conf. on Computer Vision (ECCV)*. 2020. P. 213–229.
14. Zhao L., Zhu M. MS-YOLOv7: YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography // *Drones*. 2023. Vol. 7. Art. 188.
15. Swin-Transformer-Enabled YOLOv5 with Attention Mechanism for Small Object Detection on Satellite Images / H. Gong, T. Mu, Q. Li, et al. // *Remote Sensing*. 2022. Vol. 14. Art. 2861.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-649-658

## Метод быстрого распознавания неверного поворота фотографий детей для сайта «усыновите.рф»

Д. С. Потапов

Федеральный институт цифровой трансформации в сфере образования, Москва, Россия  
Институт развития профессионального образования, Москва, Россия

**Аннотация.** Распространённой проблемой при загрузке пользовательских фотографий на сайт является их неправильная ориентация, чаще всего поворот на 90 градусов. Причиной обычно является отсутствие сенсора ориентации камеры или его неправильное срабатывание, игнорирование метаданных о повороте программным обеспечением или ошибочные действия пользователя. В работе представлен способ значительного ускорения проверки на поворот за счёт простейших преобразований изображения: обрезки и уменьшения изображения. Для решения проблемы ложного срабатывания на лежащих детях-инвалидах разработана процедура обучения свёрточной нейронной сети (СНС) для автоматического отсеивания изображений данного класса. Подход показал высокую точность решения поставленной задачи на реальных данных и был успешно интегрирован в программное обеспечение для обновления данных на сайте.

**Ключевые слова:** распознавание неверной ориентации фотографий, компьютерное зрение, нейронные сети, распознавание изображений, дисбаланс классов, ROC-кривая

## Fast method for wrong rotation recognition in orphan children photos for the website <https://усыновите.рф>

D. S. Potapov

Federal institute for digital transformation in education, Moscow, Russia  
Institute for development of professional education, Moscow, Russia

**Abstract.** When a user uploads a photo to the web, a common problem is the wrong orientation of the file, usually a 90-degree rotation. This happens due to the absence of the automatic orientation sensor of the camera or its mistakes, due to ignoring of the orientation metadata by the processing software or because of the user faulty action. In this work we present a technique for a significant speed-up of the rotation check due to the use of simple image transformations: cropping and downsampling. We have developed a procedure for training of a convolutional neural network (CNN) for automatic elimination of false alarms coming from images of bedridden handicapped children. The proposed approach solves the formulated problem with high accuracy and has been successfully intergrated in the software for updating the website content.

**Keywords:** recognition of wrongly oriented photos, computer vision, neural networks, image classification, imbalanced classes, ROC-curve

### Введение

При поддержке веб-ресурсов с большим объёмом данных часто встаёт проблема автоматического нахождения ошибок в данных, поскольку ручная проверка невозможна. Проблема особенно актуальна в связи с тем, что даже 0,1 % ошибок уже могут быть замечены 50-500 пользователями в день. Хотя и часть проблем может быть решена на этапе ввода данных, это не всегда возможно, поскольку интерфейс ввода может быть несовершенен либо из-за того, что ошибки семантические и не могут быть легко обнаружены или требуют анализа всех данных в целом. Для сайта [усыновите.рф](https://усыновите.рф) данные собираются с органов опеки всех регионов России, частично проверяются региональными и федеральными операторами, а мы проводим дополнительные проверки на этапе загрузки на сайт. В качестве одной из проверок в данной работе предлагается распознавание неверно повернутых фотографий.

**Задача определения ориентации фотографий** давно известна в области компьютерного зрения [1-4]. Она подразумевает нахождение наиболее правильной ориентации из четырёх вариантов: 0°, +90°, -90°, 180°. Решение этой задачи предполагает использование признаков изображения, по которым можно определить корректную ориентацию фотографии. Если на изображении точно есть человек, то задачу можно решать, основываясь на обнаружении лица, поскольку в большинстве случаев лицо находится в естественной ориентации (лоб сверху, подбородок снизу).

В 2009 году был опубликован патент [5] на определение ориентации фотографии на основе детектора лиц того времени. Метод предполагает нахождение лиц на четырёх разных ориентациях фотографии и выбор ориентации с наибольшим количеством лиц или наибольшей уверенностью. В работе [4] сравнивается определение ориентации изображений лиц на основе детектора лиц и использование признаков из предобученной нейронной сети. Мы показываем, что такой метод работает медленнее, чем методы на основе детекторов лиц, поэтому мы применяем его только на этапе отсеивания ложных срабатываний.

**Методы обнаружения лиц** развивались благодаря их важным применениям. Во-первых, они используются в фотосъёмке для фокусировки на лицах и срабатывании затвора по улыбке. Во-вторых, большую популярность приобрела задача распознавания лиц. В связи с особенностями строения тела человека обычно лицо отклоняется не более чем на  $45^\circ$  в плоскости фотографии от вертикального положения. Традиционно детекторы обнаруживали лица, близкие к вертикальному положению в плоскости изображения, что навело нас на мысль, что эту особенность можно использовать для определения ориентации фотографии.

Задача обнаружения лиц приобрела популярность в связи с увеличением количества цифровых фотографий в начале 2000-х годов [6, 7]. В изначальной постановке подразумевалось обнаружение только околорасположенных лиц без значительного поворота в плоскости изображения. В широко известном методе Viola Jones [6] задача решается с использованием скользящего окна, быстро вычисляемых признаков Хаара, алгоритма машинного обучения AdaBoost и каскада бинарных классификаторов. Метод был реализован во многих библиотеках и устройствах из-за его высокой скорости работы (например, в OpenCV [8]).

Начиная с ключевой статьи, опубликованной в 2012 году [9], в которой была существенно повышена точность в задаче распознавания изображений, в мире сильно возросла популярность искусственных нейронных сетей. Постепенно они были адаптированы для решения задачи обнаружения лиц. Метод MTCNN [10] решает сразу несколько задач: обнаружение лица, предсказание прямоугольника лица и нахождение ключевых точек. За счёт дополнительных данных при обучении метод превосходит предыдущие алгоритмы по точности. Метод также отличается высокой скоростью, благодаря чему был реализован в нескольких библиотеках [11, 12]. В статье [13], вышедшей в 2020 году, описан подход RetinaFace для обнаружения лиц, в котором в дополнение к ключевым точкам, также моделируется 3D-форма лица. В работе [14] описан быстрый детектор лиц YuNet, который достигает наилучшего компромисса между точностью и скоростью. Метод включает в себя высокопроизводительную основу и упрощённый способ смешивания признаков с разных масштабов. Метод был интегрирован в библиотеку OpenCV [8].

Мы применили обнаружение лиц для распознавания неправильно повернутых фотографий на сайте [usyynovite.rf](http://usyynovite.rf), и сразу выяснилась одна особенность данных: в банке данных детей-сирот присутствует значительное число фотографий лежащих детей-инвалидов, у которых голова ориентирована на кадре горизонтально, поэтому возникают ложные срабатывания. Мы решили их отсеивать с помощью методов распознавания изображений.

**Задача распознавания изображений** активно развивалась в последние 30-40 лет. В её традиционной формулировке необходимо классифицировать неизвестное модели изображение на несколько категорий (классов). Изначально рассматривались относительно простые задачи, например распознавание рукописных цифр от 0 до 9 [15]. Впоследствии задачи стали усложняться и размер изображений увеличивался. В работе [16] применяется метод скользящего окна для классификации подизображения на два класса (пешеход и фон) с использованием гистограмм ориентированных градиентов. В 2000-е годы стал популярным подход Bag of Visual Words [17], в котором локальные признаки изображения (например, SIFT [18]) агрегировались в дескриптор всего изображения без учёта положения признаков. В дальнейшем появились более качественные методы агрегации [19] и учёт приблизительного положения признаков на изображении [20, 21]. Сильный толчок к развитию методов распознавания изображений дало соревнование Pascal Visual Object Challenge [22], проходившее в 2008-2012 годах.



Основополагающая работа [9] практически совершила революцию в мире распознавания изображений. Этому способствовало появление больших коллекций изображений [23] и возможности значительного ускорения вычислений на графических картах (GPU). С этого момента искусственные нейронные сети начинают стремительно развиваться и с каждым годом появляются всё новые улучшения [24-27]. СНС для распознавания изображений в основном обучаются и тестируются на большой базе данных ImageNet [23], содержащей изначально 1000 классов и 14 миллионов изображений и впоследствии расширенной до более 20 тысяч классов. Было показано, что, используя процедуру дообучения (fine-tuning), можно перенести визуальные знания о мире из больших наборов данных в СНС, обучающиеся на относительно небольших данных [28].

При обучении на небольших наборах изображений нейронные сети склонны к переобучению. Это частично компенсируется применением дообучения, но не полностью. Для решения проблемы переобучения предложено множество методов обогащения (аугментации) обучающей выборки. В их числе случайные вырезания подизображений, случайный поворот на небольшой угол, случайное отражение относительно вертикали, варьирование оттенка, добавление шума, размытие, сжатие Jpeg и прочее. В различных задачах также применяются Random Erasing [29] и Cutout [30] и более замысловатые методы MixUp [31] и CutMix [32].

В связи с тем что СНС требовательны к вычислительным ресурсам, **методы повышения быстродействия нейронных сетей** в последнее время бурно развиваются. В данной работе затрагивается повышение скорости объединённого метода на основе СНС только на этапе применения моделей обнаружения лиц и распознавания лежащих детей. В настоящее время используются каскады классификаторов [10], вычисления на GPU [9], квантизация [33] и дистилляция [34] нейронных сетей. Применение большинства из этих методов требует большего времени, дополнительных средств или выполнения некоторых условий, в то время как в данной работе предлагается простой подход на основе использования специфики тестовой выборки.



Рисунок 1. Пример изображений из классов, которые нужно распознать:

а) правильная ориентация; б) неправильная ориентация; в) лежащий ребёнок. Designed by Freepik

### Постановка задачи

Глобально рассматривается задача автоматического определения ориентации фотографии, содержащей человека. Мы делаем предположение, что на фотографии видно лицо, то есть человек повернут к камере на угол не более  $90^\circ$  относительно вертикали. Поскольку в наших интересах найти только фотографии с неправильной ориентацией, а исправлением ошибки уже занимается человек, для себя мы ставим задачу только распознавания фотографий с неправильной ориентацией, то есть повернутых на один из углов ( $-90^\circ$ ,  $+90^\circ$  или  $180^\circ$ ).

При решении задачи возникла необходимость повышения скорости работы метода распознавания неправильной ориентации. Изначально ставилась задача сократить время проверки базы до менее одной минуты без использования GPU, но в итоге нам удалось разработать ещё более быстрый метод.

В связи со спецификой данных ставится дополнительная задача – распознавание лежащих детей среди фотографий с горизонтальным лицом. Для этой задачи особых требований к скорости работы нет, поскольку количество ложных срабатываний очень мало по сравнению с общим объёмом изображений.



## Описание методов

### 1. Метод распознавания неправильной ориентации фотографии на основе детектора лиц

Изначально алгоритм был следующим: применить детектор лиц; если на фотографии найдено лицо, то выдать результат «правильная ориентация»; иначе – «неправильная ориентация».

Первоначально на обработку 35'000 фотографий уходило 7 часов<sup>1</sup>. Впоследствии мы воспользовались библиотекой torch-MTCNN [11] и перенесли запуск кода на машину с графическим процессором, и время сократилось до 12 минут. При этом проведение остальных проверок при загрузке на сайт занимало менее полминуты. На компьютере администратора не было GPU, поэтому данная проверка не очень сочеталась с существующими процессами.

Для сокращения времени работы изначального алгоритма мы предположили, что можно обрезать края изображения, а также уменьшить размер изображения. В редких случаях, когда это приведёт к потере детекции, обнаружение лиц необходимо повторить на исходном изображении.

На большинстве фотографий лицо ребёнка находится близко к центру изображения, поэтому мы решили оценить область обрезки. Мы нашли лица на 1000 случайных изображениях из базы и сохранили координаты прямоугольников лиц в массивы left, top, right, bottom.

Прямоугольник задаётся координатами левого верхнего (left, top) и правого нижнего (right, bottom) углов с осью X, направленной вправо, и осью Y, направленной вниз. Далее мы вычислили нормализованные координаты прямоугольников:

$$\text{left\_n}[i] = \text{left}[i]/w[i], \text{top\_n}[i] = \text{top}[i]/h[i], \text{right\_n}[i] = \text{right}[i]/w[i], \text{bottom\_n}[i] = \text{bottom}[i]/h[i],$$

где  $w[i]$  и  $h[i]$  – ширина и высота изображения.

Когда на фотографии было найдено несколько лиц, выбирался наибольший по площади прямоугольник лица – обычно в базе присутствует только лицо одного ребёнка. Чтобы оценить область, в которой в большинстве случаев находится лицо, по массивам left\_n и top\_n мы посчитали 5-й квантиль, по массивам right\_n и bottom\_n – 95-й квантиль. Получились такие значения квантилей:

$$q\_left = 0.19, q\_top = 0.10, q\_right = 0.80, q\_bottom = 0.75.$$

Это означает, что более  $0.95^4 \approx 81\%$  лиц находится в области площадью  $0.61 \cdot 0.65 \approx 0.40$  от исходного изображения.

Далее мы решили проверить, можно ли обнаруживать лица на уменьшенном изображении, чтобы сократить время работы детектора. Для каждого изображения из 1000 случайных мы рассмотрели масштабированные версии изображения с коэффициентом от 0.05 до 1.0 и шагом 0.025 и определили, на каком минимальном масштабе обнаруживается лицо (обрезка в данном случае не делалась). Далее мы взяли 97-й квантиль массива минимальных масштабов и получили масштаб 0.25, что примерно соответствует изображению размера 90x110. Площадь изображения при этом уменьшается в 16 раз, что существенно увеличивает скорость детекции лиц.

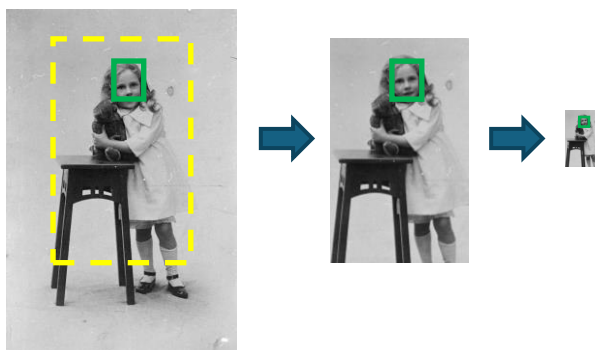


Рисунок 2. Иллюстрация уменьшения области применения детектора лиц за счёт уменьшения области поиска и уменьшения масштаба. Область обрезки такая большая, поскольку на фотографиях встречаются лица крупным планом. Площадь изображения сокращается в 40 раз. Photo by Museums Victoria on Unsplash

<sup>1</sup> Предполагаемая причина очень медленной работы – использование неэффективной реализации метода MTCNN.

Замеры показали, что 96.6 % лиц было найдено на уменьшенном изображении с небольшими временными затратами. Если же лицо не найдено, мы применяем детектор к оригинальному изображению.

Время работы метода составило 10 минут на 35'000 изображений без использования GPU. Замер показал, что по сравнению с изначальным алгоритмом программа ускорилась в 5 раз. Впоследствии процесс проверки данных был усовершенствован за счёт исключения повторной обработки одних и тех же фотографий. Дело в том, что фотографии детей обновляются в большинстве случаев не чаще одного раза в год, в среднем между двумя последовательными проверками обновляется порядка 300 фотографий. Достаточно проверять каждую фотографию только один раз, поэтому мы реализовали хранение md5-сумм проверенных фотографий. По сравнению с обнаружением лиц md5-сумма файла вычисляется практически мгновенно, примерно за 0.1 миллисекунду. Такой подход помог сократить время проверки в 100 раз.

Изначально мы предполагали, что на фотографиях с неправильной ориентацией лицо не будет найдено. В ходе исследований выяснилось, что детектор лиц при применении на фотографии с лицом, повернутым на  $90^\circ$ , в 22 % случаев выдаёт положительный результат обнаружения. Это связано с тем, что при обучении детектора использовались фотографии, снятые с произвольных ракурсов и, соответственно, в нём встречались лица, ориентированные в произвольном направлении. На рисунке 3 показана оценка вероятности обнаружения лица детектором torch-MTCNN [11] в зависимости от угла поворота.

Для уменьшения числа ошибок первого рода мы добавили дополнительную проверку на положение ключевых точек. Если угол между вектором правый-левый глаз ребёнка и осью X, направленной вправо, составляет более  $45^\circ$ , то выводится результат «неправильная ориентация». Даже с учётом такого улучшения ошибка первого рода составляла 11.8 % при замере на 2000 изображениях и поворотах  $\pm 90^\circ$ . Дополнительный анализ показал, что на лицах повернутых на  $90^\circ$ , ключевые точки лица находятся неточно – часто с поворотом около  $45^\circ$ . Поэтому после первого нахождения ключевых точек мы выравниваем положение глаз с горизонталью с помощью поворота изображения и запускаем поиск ключевых точек повторно для уточнения их положения.

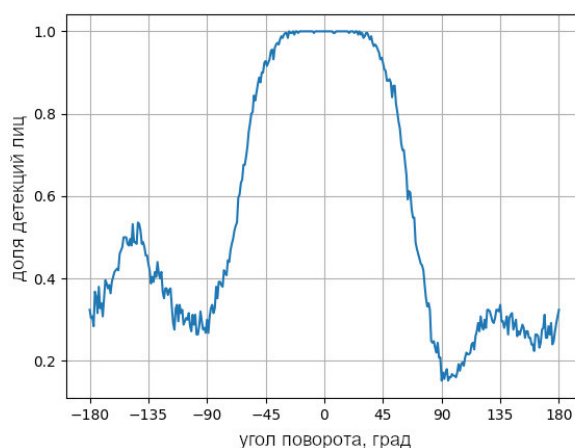


Рисунок 3. Статистика обнаружения лица в зависимости от угла ориентации лица. Направление горизонтальной оси соответствует повороту против часовой стрелки. Тест проведён на 250 фотографиях из базы детей с практически вертикальной ориентацией лица

По сравнению с методом [5] в изначальной версии метода требуется применение детектора лиц только в одной ориентации, что ускоряет обработку в 4 раза. Однако в некоторых библиотеках есть возможность отключения определения ключевых точек, что может компенсировать эти временные затраты. Далее мы приводим сравнение с методом [5], к которому также применим предложенный выше способ повышения скорости. Более точно в данном случае, если ни в одной ориентации не было найдено лицо, осуществляется поиск на изображении исходного размера.

Финальная версия алгоритма выглядит следующим образом:

1. Уменьшить изображение до площади  $10^4$ 000 пикселей и обрезать поля относительного размера  $\text{margin\_left}=0.19$ ,  $\text{margin\_top}=0.10$ ,  $\text{margin\_right}=0.20$ ,  $\text{margin\_bottom}=0.25$ ;
2. Найти ориентацию с наибольшей уверенностью детектора YuNet [14] (метод [5]);
3. Если лица найдены и наилучшая ориентация  $0^\circ$ , то вернуть «правильная ориентация»;
4. Иначе применить метод [5] с детектором YuNet [14] к исходному изображению.

## 2. Метод распознавания лежащих детей

Как выяснилось на практике, алгоритм на основе обнаружения лиц на наших данных выдаёт много фотографий лежащих детей-инвалидов, у которых лицо ориентировано горизонтально. По согласованию с заказчиком было решено такие фотографии считать правильно ориентированными. Мы решили автоматически отсеивать случаи ложных срабатываний из числа потенциально неверно ориентированных фотографий.

Задача распознавания изображений в последние годы обычно решается с помощью нейронных сетей [9, 28]. В случае небольшого количества данных применяется дообучение [28] свёрточной нейронной сети (СНС), предобученной на базе ImageNet [23]. Дообучение предполагает инициализацию СНС с помощью весов, полученных в результате предобучения, и минимизацию функции потерь с более низким темпом обучения. Это позволяет сохранить некоторое количество априорных знаний о природе данных (изображений) и увеличивает точность распознавания. В отличие от [28] мы обновляем веса *всех* слоёв при обучении СНС, подобно тому, как это делают в методах распознавания кровного родства [35].

В качестве базовой модели (СНС) мы используем ResNet-50 [26] с весами IMAGENET1K\_V2. Модель состоит из 50 слоёв. Слои объединяются в базовые блоки, состоящие из трёх слоёв: свёртка  $1 \times 1$ , свёртка  $3 \times 3$  и свёртка  $1 \times 1$ . При приближении к последним слоям количество каналов свёртки увеличивается. Каждый базовый блок дублируется сквозной связью в обход этого блока.

Модель ResNet-50 предполагает входное изображение размером  $224 \times 224$ . Для приведения изображений к такому виду мы сначала масштабируем их к размеру  $232 \times 232$ . В качестве обогащения выборки используется отражение относительно вертикали с вероятностью 0.5 и вырезание случайного квадрата размером  $224 \times 224$ . Поскольку базовая модель решает задачу многоклассовой классификации, мы заменяем её последний полносвязный слой с 1000 выходами на полносвязный слой с одним выходом. В качестве функции потерь используются бинарные кросс-энтропийные потери (см. ниже), для её минимизации используется алгоритм Adam со стандартными параметрами в pytorch и темпом обучения  $10^{-5}$ . Как это общепринято, мы перемешиваем обучающие примеры перед каждой эпохой.

Вручную разметили изображения на три класса: обычные (0), лежащие дети с вертикальным лицом (1) и лежащие дети с горизонтальным лицом (2). Был выбран протокол разметки, при котором дети, сидящие в коляске или на диване или с поднятой головой, не считались лежащими. Все обычные фотографии, которые имели неправильную ориентацию, были приведены к правильной ориентации. При обучении мы поворачиваем фотографии из классов 0 и 1 на  $90^\circ$  влево или вправо, чтобы смоделировать ошибочную ориентацию. Фотографии из класса 2 оставляем как есть, так как их поворот на  $180^\circ$  градусов часто является неправдоподобным изображением.

Поскольку часть изображений из класса 1 также содержит детей-инвалидов, на этом этапе мы решили объединить классы 1 и 2, чтобы модель также отсеивала повернутые изображения из класса 1. Таких изображений примерно в 60 раз меньше, чем в классе 0, поэтому это не должно привести к большому количеству ошибок первого рода. Объединённый класс мы приняли за *positive class*.

В данной формулировке задача является сильно несбалансированной: в обучающей выборке приблизительно в 54 раза больше обычных фотографий, чем лежащих детей. При этом на этапе применения модели у нас в несколько раз больше лежащих детей, чем обычных. Поскольку допускается наличие небольшого количества ложных срабатываний, мы стремимся к точке на ROC-кривой, в которой ошибки первого и второго рода происходят одинаково часто:  $\text{TPR} \approx \text{TNR}$  ( $\text{true positive rate} \approx \text{true negative rate}$ ), но мы немного упростили задачу и сразу выбрали целевое значение  $\text{TNR} = 0.97$ . Соответственно, при валидации модели мы используем значение сбалансированной точности  $\text{BA} = 0.5 \cdot (\text{TPR} + \text{TNR})$  в точке  $\text{TNR} \approx 0.97$ . Однако, при минимизации обычной кросс-

энтропийной функции потерь как функции от весов модели  $\theta$  точность работы модели близка к 100 %, но модель всё время выдаёт класс 0, что приводит к низкому значению ВА. Для корректировки дисбаланса в функцию потерь необходимо добавить вес положительного класса  $w$ , равный 54. Такая сбалансированная бинарная кросс-энтропийная функция потерь задаётся формулой

$$\text{BinaryCrossEntropyLoss}(x, y, \theta) = -(w \cdot y \cdot \log f(x, \theta) + (1 - y) \cdot \log(1 - f(x, \theta))) ,$$

где  $x$  – входное изображение;  $y$  – метка класса (0 или 1);  $f(\cdot)$  – функция, которую задаёт нейронная сеть. На этапе применения модель выдаёт результат «обычная фотография» или «лежащий ребёнок» (нужно отсеять).

Распространённым методом регуляризации нейронных сетей является ранний останов, когда при обучении точность модели время от времени контролируется на валидационной выборке, и в качестве финальных выбираются веса с наилучшим качеством при валидации. Мы делаем контроль каждые 50 итераций при размере мини-батча в 32 изображения. Обучение длится не более 10 эпох, но мы останавливаем обучение, если нет улучшения качества на валидационной выборке в течение 20 раундов.

## Результаты экспериментов

### 1. Распознавание неправильной ориентации фотографии на основе детектора лиц

Для сравнения методов определения ориентации фотографии на основе обнаружения лиц мы используем выборку из 2000 случайных фотографий детей<sup>1</sup>. Эти фотографии были проверены на правильность ориентации и отсутствие лежащих детей. 84 % фотографий имеют размер от 250x400 до 450x450, все фотографии ограничены размером 468x445. Далее были сделаны три копии данной выборки с поворотами на +90° (влево), на -90° (вправо) и на 180°. В таблице 1 показано сравнение точности и скорости различных методов. Сравнение проводилось на ноутбуке с процессором Intel Core i7-12700H с 14 физическими и 20 логическими ядрами с операционной системой Windows 11. Скорость работы замерялась на изображениях с поворотом 0°. В таблице 1 мы исследуем только внутренний параллелизм алгоритмов. Как показано в таблице 2, выполнение алгоритма в нескольких процессах (внешний параллелизм) даёт более существенный прирост скорости. Эксперимент проводился на 800 изображениях, приведены среднее время на обработку одного изображения и стандартное отклонение по трём запускам.

Таблица 1. Сравнение скорости и точности различных методов (без использования GPU)

Название метода	Сред. время обработки одного изображения, с			Точность срабатывания метода			
	1 ядро	2 ядра	4 ядра	0°	+90°	-90°	180°
torch-MTCNN	0.056	0.040	0.033	<b>99.90</b>	81.05	72.05	69.40
+ обрезка/уменьшение	<b>0.013</b>	<b>0.009</b>	<b>0.007</b>	<b>99.95</b>	79.10	70.20	65.95
torch-MTCNN + учёт положения точек	0.112	0.077	0.064	99.75	97.90	96.45	79.35
+ обрезка/уменьшение	0.070	0.049	0.041	99.75	97.70	96.25	78.40
RetinaFace + учёт положения точек	3.720	2.130	1.355	<b>99.90</b>	<b>99.95</b>	<b>100.00</b>	99.40
+ обрезка/уменьшение	2.120	1.220	0.775	<b>99.90</b>	<b>99.95</b>	<b>100.00</b>	99.40
Алгоритм [5] - torch-MTCNN x 4	0.208	0.146	0.121	99.60	<b>99.95</b>	99.75	<b>99.90</b>
+ обрезка/уменьшение	0.038	0.027	0.023	99.55	<b>99.90</b>	<b>99.90</b>	<b>99.90</b>
Алгоритм [5] - onnx-YuNet x 4	0.041	0.042	0.034	<b>99.95</b>	<b>100.00</b>	<b>100.00</b>	<b>99.95</b>
+ обрезка/уменьшение	<u>0.016</u>	<u>0.017</u>	<u>0.015</u>	<b>99.95</b>	<b>99.95</b>	<b>99.95</b>	<b>99.95</b>
Дообучение ResNet-50	0.101	0.053	0.035	<b>99.95</b>	<b>99.95</b>	<b>100.00</b>	<b>99.90</b>
Дообучение ResNet-34	0.076	0.041	0.027	<b>99.90</b>	<b>100.00</b>	<b>100.00</b>	<b>99.95</b>

<sup>1</sup> Среднее количество новых фотографий на одну проверку около 300, среди них в среднем около 0.03 % фотографий с неправильной ориентацией и 2 % фотографий лежащих детей. Фотографии с поворотом 180° достаточно редки.

Таблица 2. Сравнение скорости алгоритмов при запуске в параллельных процессах

Метод	Среднее время обработки одного изображения, с				Опер. пам. на 1 процесс, Мб
	1 процесс	2 процесса	4 процесса	8 процессов	
[5] - torch-MTCNN + обр./уменьш.	0.047 ± 0.007	0.023 ± 0.003	0.014 ± 0.001	0.015 ± 0.001	312
[5] - onnx-YuNet + обр./уменьш.	0.019 ± 0.004	0.009 ± 0.000	0.011 ± 0.000	0.005 ± 0.001	38

Поскольку точность для лидирующих алгоритмов оказалась очень близка к 100 %, мы дополнительно протестировали лучшие алгоритмы на 30253 изображениях с правильной ориентацией. Результаты приведены в таблице 3. Было замечено, что при применении обрезки и уменьшения, метод onnx-YuNet начинает выдавать больше ложных срабатываний на угле 0°. Поэтому мы применили следующее улучшение: если для уменьшенного изображения ориентация определена как неправильная, то проводится дополнительное тестирование на исходном изображении. Это сокращает количество ложных срабатываний на изображениях с правильной ориентацией, но практически не увеличивает время работы, поскольку на большинстве изображений в базе лицо ориентировано правильно. Более точно в тесте с 8 процессами алгоритм обрабатывает одно изображение в среднем за 5 мс.

Таблица 3. Измерение точности работы методов на большой выборке

Название метода	Точность срабатывания метода			
	0°	+90°	-90°	180°
Алгоритм [5] - torch-MTCNN x 4	99.63	99.93	99.82	99.88
+ обрезка/уменьшение	99.69	99.90	99.88	99.79
Алгоритм [5] - onnx-YuNet x 4	<b>99.88</b>	<b>99.98</b>	<b>99.98</b>	<b>99.94</b>
+ обрезка/уменьшение	99.74	<b>99.96</b>	<b>99.96</b>	99.84
+ повторная проверка при угле ≠ 0°	<b>99.89</b>	<b>99.96</b>	<b>99.96</b>	99.84

Таблица 4. Качество методов распознавания лежащих детей

Метод	Валидационная выборка			Тестовая выборка		
	BA, %	TNR, %	TPR, %	BA, %	TNR, %	TPR, %
ResNet50, дообучение, weight=54	97.40	97.01	97.80	96.57	97.00	96.14
+ оба поворота при обучении	<b>97.95</b>	97.01	98.90	<b>96.64</b>	97.14	96.14
+ случайный поворот на ± 10°	<b>97.95</b>	97.01	98.90			

## 2. Распознавание лежащих детей

Мы взяли набор фотографий детей из 31000 изображений (из открытого источника) и разбили его на обучающую, валидационную и тестовую выборку в отношении 10:3:10.

В таблице 4 показаны результаты экспериментов. Мы проводили тестирование на тестовой выборке только в случае, если сбалансированная точность (BA) превышает качество, достигнутое в предыдущих экспериментах. Было замечено, что использование обоих поворотов в обучающей выборке немного улучшает точность. Более точно мы дублируем обучающую выборку и каждый второй раз выбираем поворот на 90°, противоположный предыдущему (в случае классов 0 и 1).

## Обсуждение результатов

### 1. Распознавание неправильной ориентации фотографии на основе детектора лиц

Эксперименты показали, что хотя первая версия алгоритма достаточно точно срабатывает на вертикальных лицах, количество ошибок на фотографиях с неправильной ориентацией довольно большое – 20-30 %. Значительное улучшение точности было достигнуто за счёт применения метода [5]. Для повышения скорости алгоритма был использован детектор YuNet [14]. Для сокращения числа ложных срабатываний предложено проводить дополнительную проверку на изображении исходного размера в случае, если распознана неправильная ориентация.

Несмотря на то что метод RetinaFace [13] показал достаточно высокую точность, а также автоматический внутренний параллелизм, он работает примерно в 70 раз медленнее, и его применение без GPU не подходит под требования задачи.

Алгоритмы распознавания неверной ориентации на основе ResNet также показывают высокую точность. Однако, согласно замерам, на 4 ядрах скорость алгоритма на основе YuNet в 2.5 раза превышает скорость алгоритма на основе ResNet, даже при использовании легковесной сети ResNet-34. Поэтому в задачах, где допускается некоторое количество ложных срабатываний, разметка данных затруднительна и есть требования к скорости работы, метод на основе YuNet является предпочтительным. Дальнейшее улучшение точности возможно в случае применения нейронных сетей для отсеивания ложных срабатываний.

## 2. Распознавание лежащих детей-инвалидов

Метод показал способность отсеивать 96.14 % лежащих детей при отсеивании неверно ориентированной фотографии в 2.86 % случаев. Такая точность достаточна в нашем сценарии использования.

Поскольку мы распознавали всех лежащих детей, а не только детей-инвалидов, как подлежащих отсеиванию, это влияет на количество пропусков, но таких изображений в базе немного. Замер качества работы объединённого метода на независимой выборке из 31000 изображений показал ошибку первого рода 10 % и ошибку второго рода 0,05 %.

В продолжении данного исследования нам удалось сбалансировать ошибку первого и второго родов за счёт прямого предсказания правильности ориентации. Была получена ошибка первого рода 0 % и ошибка второго рода 0,05 %.

## Выводы

В данной работе представлены два метода. Первый метод быстро распознаёт неверно повернутые изображения с помощью детектора лиц, затрачивая на это 5 мс на одно изображение без использования GPU. Второй метод отсеивает ложные срабатывания первого метода, так что итоговые ошибки первого и второго родов составляют 10 % и 0.05 % соответственно.

## Список литературы

1. Vailaya, A., Zhang, H., Yang, C., et al. Automatic image orientation detection / A. Vailaya, H. Zhang, C. Yang, et al. *IEEE Transactions on Image Processing*, 2002, 11(7), 746-755.
2. Joshi, U., & Guerzhoy, M. Automatic photo orientation detection with convolutional neural networks. *14th IEEE Conference on Computer and Robot Vision (CRV)*, 2017, pp. 103-108.
3. Tolstaya, E. Content-based image orientation recognition. *GraphiCon 2007*, pp. 158-161.
4. Shima, Y., Nakashima, Y., & Yasuda, M. Detecting orientation of in-plain rotated face images based on category classification by deep learning. *TENCON 2017 - 2017 IEEE Region 10 Conference*, pp. 127-132.
5. US Patent US7565030B2
6. Rowley, H. A., Baluja, S., & Kanade, T. Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, 1998, 20(1), 23-38.
7. Viola, P., & Jones, M. J. Robust real-time face detection. *Int. journal of computer vision*, 2004, 57(2), 137-154.
8. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
9. Krizhevsky, A., Sutskever, I., & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012, 25.
10. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal processing letters*, 2016, 23(10), 1499-1503.
11. Tim Esler. Facenet-pytorch library. <https://github.com/timesler/facenet-pytorch>
12. Iván de Paz Centeno. ipazc/mtcnn: v1.0.0. *Zenodo*, 2024, doi: 10.5281/zenodo.13901378
13. Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. Retinaface: Single-shot multi-level face localisation in the wild. *IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2020, pp. 5203-5212.
14. Wu, W., Peng, H., & Yu, S. Yunet: A tiny millisecond-level face detector. *Machine Intelligence Research*, 2023, 20(5), 656-665.
15. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 1989, 2.

16. Dalal, N., & Triggs, B. Histograms of oriented gradients for human detection. *2005 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 886-893.
17. Csurka, G., Dance, C., Fan, L., Willamowski, J., & Bray, C. Visual categorization with bags of keypoints. *Workshop on statistical learning in computer vision, ECCV 2004*, vol. 1, no. 1-22, pp. 1-2.
18. Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, 60(2), 91-110.
19. Perronnin, F., & Dance, C. Fisher kernels on visual vocabularies for image categorization. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8.
20. Lazebnik, S., Schmid, C., & Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *2006 IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 2169-2178.
21. Krapac, J., Verbeek, J., & Jurie, F. Modeling spatial layout with fisher vectors for image categorization. *2011 IEEE International Conference on Computer Vision*, pp. 1487-1494.
22. Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. The Pascal Visual Object Classes challenge: A retrospective. *International Journal of Computer Vision*, 2015, 111(1), 98-136.
23. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255.
24. Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *2014 arXiv preprint. arXiv:1409.1556*.
25. Ioffe, S., & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *2015 International Conference on Machine Learning*, pp. 448-456.
26. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.
27. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. *2019 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4690-4699.
28. Oquab, M., Bottou, L., Laptev, I., & Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1717-1724.
29. Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. Random Erasing Data Augmentation. *2020 AAAI Conference on Artificial Intelligence*, Vol. 34, No. 07, pp. 13001-13008.
30. DeVries, T., & Taylor, G. W. Improved regularization of convolutional neural networks with cutout. *2017 arXiv preprint. arXiv:1708.04552*.
31. Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. mixup: Beyond empirical risk minimization. *2017 arXiv preprint. arXiv:1710.09412*.
32. Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. *2019 IEEE International Conference on Computer Vision*, pp. 6023-6032.
33. Gholami, A., Kim, S., Dong, Z., Yao, Z., Mahoney, M. W., & Keutzer, K. A survey of quantization methods for efficient neural network inference. In: *Low-power computer vision*, 2022, pp. 291-326.
34. Hinton, G., Vinyals, O., & Dean, J. Distilling the knowledge in a neural network. 2015, *arXiv:1503.02531*.
35. Shadrikov, A. Achieving better kinship recognition through better baseline. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pp. 872-876.



## Распознавание рукописных графических элементов и мультязычного текста: обзор современных подходов, базовый метод и перспективы развития

Н. Ю. Скопин, М. С. Мосалев, Д. А. Королев, А. М. Пискунова

Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

**Аннотация.** В данной статье представлен обзор современных методов решения задачи распознавания рукописного контента, включающего мультязычный текст (русский и английский) и графические элементы, такие как блок-схемы из прямоугольников и стрелок, с акцентом на распознавание изображений маркерных досок. Для мультязычного рукописного текста рассмотрены архитектуры, способные обрабатывать несколько языков, а для блок-схем — методы распознавания фигур и структуры. Кроме того, в работе предложена базовая модель, в которой для распознавания текста используются решения EasyOCR и TrOCR, а для анализа графических элементов — алгоритмы на базе OpenCV, выделяющие контуры фигур. Цель этой модели — оценить сложность решения поставленной задачи с использованием общедоступных инструментов при минимальной подготовке. Базовая модель ориентирована на специфическую задачу распознавания контента на маркерной доске, а результаты формируются в формате XML для последующей обработки, например визуализации. Система протестирована на специально собранном датасете изображений маркерных досок. В статье также обсуждаются основные сложности задачи распознавания и обозначаются перспективные направления для повышения точности и качества решения.

**Ключевые слова:** распознавание рукописного контента, мультязычный текст, блок-схемы, выделение контуров фигур, распознавание диаграмм, обнаружение объектов, распознавание изображений, маркерная доска

## Recognition of handwritten graphic elements and multilingual text: a review of modern approaches, a basic method and development prospects

N. Y. Skopin, M. S. Mosalev, D. A. Korolev, A. M. Piskunova

National Research University Higher School of Economics, Moscow, Russia

**Abstract.** This article presents an overview of modern methods for solving the problem of recognizing handwritten content, including multilingual text (Russian and English) and graphical elements such as block diagrams consisting of rectangles and arrows, with a focus on the recognition of whiteboard images. For multilingual handwritten text, architectures capable of processing multiple languages are considered, and for block diagrams — methods for recognizing shapes and structure. In addition, the paper proposes a baseline model in which EasyOCR and TrOCR are used for text recognition, and OpenCV-based algorithms that extract shape contours are used for analyzing graphical elements. The purpose of this model is to assess the complexity of solving the given problem using publicly available tools with minimal preparation. The baseline model is focused on the specific task of recognizing content on a whiteboard, and the results are generated in XML format for further processing (e.g., visualization). The system was tested on a specially collected dataset of whiteboard images. The article also discusses the main challenges of the recognition task and outlines promising directions for improving the accuracy and quality of the solution.

**Keywords:** handwriting recognition, multilingual text, flowcharts, shape outline detection, diagram recognition, object detection, image recognition, whiteboard

### Введение

Маркерные доски являются неотъемлемой частью современных образовательных и профессиональных сред, выступая важным инструментом для обсуждений, обучения и совместной работы. Контент на таких досках часто представляет собой сочетание рукописного текста на нескольких языках, таких как русский и английский, и графических элементов, таких как блок-схемы, диаграммы и стрелки.

В образовательной сфере автоматизация извлечения графических схем способствует оцифровке лекционных материалов и упрощает поиск информации в архивах. В научных и инженерных лабораториях такие инструменты позволяют структурированно документировать результаты проектных и экспериментальных работ. В бизнес-контексте распознавание блок-схем и графиков ускоряет фиксацию результатов мозговых штурмов и совещаний. Для достижения полной автоматизации обработки содержимого доски необходимо также преобразовывать полученные данные в машиночитаемые форматы.

### 1. Обзор существующих подходов

В работе Schäfer B. и др. [1] задача распознавания рукописных диаграмм разделяется на два этапа: локальное распознавание, локализация символов и восстановление глобальной структуры. Авторы рассматривают диаграммы как совокупность блоков, стрелок и поясняющего текста — подход, близкий к нашему.

Предлагается модификация ускоренной сверточной нейронной сети на основе регионов (Faster Region-based Convolutional Neural Network, Faster R-CNN) с добавлением предсказания ключевых точек стрелок и пост-обработки для воссоздания общей структуры. Это позволяет интегрировать детекцию объектов и анализ топологии диаграммы. В архитектуру включены пирамида признаков (Feature Pyramid Network) и модуль регрессии ключевых точек стрелки, что улучшает распознавание стрелок и текста. Ключевые точки стрелок используются для определения связей между элементами. Особое внимание уделено аугментации: применялся геометрический пайплайн и специальные преобразования для снижения путаницы между стрелками и текстом. В рамках анализа статьи выделены следующие преимущества описанной модели:

- одновременное распознавание символов и структуры,
- эффективность на малых и крупных датасетах благодаря целевой аугментации,
- высокая скорость — менее 100 мс на диаграмму.

Кроме того, есть недостатки и перспективы:

- проблемы с пересекающимися ограничивающими прямоугольниками (bounding box) стрелок;
- возможность внедрения графовых мер схожести для оценки.

В статье Bluche T. и др. [2] предлагается архитектура со сверточным кодировщиком для извлечения признаков из изображений и особой двунаправленной рекуррентной сетью с долгой краткосрочной памятью (Long short-term memory) в качестве декодировщика для генерации текстовой последовательности. Основная идея — создание универсального кодировщика, обученного однажды на разноразличных данных, который можно использовать без дообучения для распознавания новых языков.

Несмотря на прогресс в задаче распознавания рукописного текста (Handwritten Text Recognition, HTR), например, с использованием многомерной рекуррентной нейронной сети (multidimensional Recurrent Neural Network), ключевые ограничения — построчная обработка и отсутствие поддержки мультязычности. Авторы решают это, обучая модель на семи языках (включая русский и английский), чтобы кодировщик научился извлекать обобщенные признаки, инвариантные к языку. Декодировщик при этом адаптируется под конкретную задачу.

Модель обучалась на 133 тыс. изображений на английском, французском, немецком, испанском, итальянском, португальском и русском языках. Результаты показали высокую эффективность даже без тонкой настройки (fine-tuning): модель способна распознавать текст неизвестного заранее языка, что открывает путь к обработке мультязычных и смешанных текстов.

В работе Omasa T. и др. [3] описано, как сделать так, чтобы языковая модель с возможностью обработки визуальных данных (Visual Language Model) могла понять для себя структуру диаграммы, то есть увидеть, где стрелки, как они соединяют блоки и другие детали, и дальше смогла бы ответить на вопросы по ней. Упоминаются модели, учитывающие стрелки, и конкретная архитектура, обзор по статье с данной архитектурой есть в предыдущем пункте.

Вопрос, который ставят себе авторы: как можно совместить детектор, обращающий свое внимание на стрелки (flowchart-aware detector), и языковую модель через составление с помощью первой модели запроса, который содержит геометрическую информацию для второго?

Основной подход, описанный в статье, состоит из семиэтапного конвейера, который авторы подробно описали в своей работе.

### 2. Основные проблемы распознавания

Распознавание контента с маркерных досок сопряжено с рядом технических трудностей, обусловленных особенностями среды и способа записи информации [4, 5]. Системы распознавания сталкиваются со следующими основными проблемами:

- неравномерное освещение и блики: источники света создают блики на глянцевой поверхности доски и тени, что приводит к неравномерной яркости изображения и затрудняет определение границ символов;
- качество маркеров: маркеры различаются по толщине линий и насыщенности цвета, что влияет на контрастность и читаемость надписей;
- вариативность почерка: индивидуальные различия в стилях письма требуют от системы способности адаптироваться к разнообразным почеркам и способам отрисовки графических элементов;
- использование нескольких цветов: многоцветные надписи могут усложнять распознавание, особенно если цвета плохо контрастируют с фоном;
- частичные стирания и переписывания: следы предыдущих записей создают шум, который может быть ошибочно принят за актуальный контент;
- качество изображения и искажение перспективы: фотографии, сделанные под углом, приводят к геометрическим искажениям, требующим коррекции. Качественное распознавание зависит от параметров сделанного изображения доски (разрешение, фокус, экспозиция);
- фоновые помехи: посторонние объекты в кадре необходимо исключить при обработке изображения;
- перекрытие текста и графики: текст и графические элементы могут пересекаться или быть близкими друг к другу, что затрудняет их разделение и распознавание, также необходимо учитывать случай вложенности одного графического элемента в другой, например, прямоугольник в прямоугольнике.

Эти проблемы с качеством изображения напрямую влияют на точность распознавания, так как затрудняют корректную сегментацию текста и графических элементов, а также классификацию символов. Для преодоления этих вызовов необходимо применять специализированные методы предобработки изображений и использовать продвинутые алгоритмы машинного обучения, способные справляться с шумами и искажениями.

Распознавание смешанных русско-английских рукописных текстов предъявляет особые требования к алгоритмам распознавания рукописного текста. Ниже выделены проблемы, возникающие при разработке мультязычных решений [6–9]. Они препятствуют созданию универсальных моделей, способных работать над смешанными текстами с одинаковой точностью. Список проблем в данном случае следующий:

1. Идентификация языка. При использовании гибридного подхода (раздельное применение одноязычных моделей) возникает задача надежной классификации языка фрагмента на уровне строки или даже слова (Q-сети, D-триггер). Ошибки на этом этапе приводят к каскадному ухудшению качества распознавания.
2. Детектирование строк и слов. Также при таком гибридном проходе возникает задача детектирования слов и строк. Смешанный текст может менять направление или стиль штриха. Это приводит к ошибкам выделения строк и отдельных слов: символы разных алфавитов могут сливаться в одном связном компоненте или, наоборот, разделяться на несколько.
3. Вариативность почерка и артефакты. Индивидуальные особенности письма (разная ширина штриха, наклон, связность букв) в сочетании с шумом и пятнами на доске затрудняют извлечение стабильных признаков.
4. Мультязычная модель. Разработка единой модели для распознавания сразу двух алфавитов требует особых архитектурных решений: мультязычная токенизация, расширенный словарь токенов, схема объединения выходов нейронной сети для работы с последовательностями переменной длины (Connectionist Temporal Classification) и языковой модели. Подобные решения обычно сложнее в обучении и требуют мультязычных датасетов.
5. Различия в наборе символов. Кириллица и латиница имеют перекрывающиеся визуальные символы (А/А, Е/Е, В/В), что порождает неоднозначности при классификации без учета лингвистического контекста. Для их разрешения необходима интеграция языковых моделей и контекстных алгоритмов постобработки.

6. Доменно-специфичные особенности. Технические термины, аббревиатуры и имена собственные часто выходят за пределы базового словаря. Модели нужно дообучать, если они обучались на определенных корпусах, и расширять лексикон, что усложняет процесс обучения и поддержки системы.

### Постановка задачи

Задача данной работы – обзор современных нейросетевых методов распознавания рукописных русско-английских текстов и методов для распознавания фигур, создание базовой модели, сочетающей готовые решения и комбинированный подход для распознавания текста, и оценка работы базовой модели на собранном и размеченном датасете фотографий маркерных досок.

Важно отметить, что задача исследования включает две взаимосвязанные подзадачи: мультязычное распознавание рукописного текста (на русском и английском языках) и детекцию графических элементов (прямоугольников и стрелок). Основные вызовы заключаются в точной сегментации и идентификации языка каждого текстового фрагмента, а также в отделении графических элементов от фонового шума и текстовых меток и в формировании итогового файла в машиночитаемом формате с результатами распознавания. Распознавание рукописного текста сталкивается с вариативностью индивидуальных стилей письма. Кроме того, наличие нескольких языков усложняет процесс из-за смешения алфавитов и необходимости точной идентификации языка. Одновременно распознавание графических элементов требует применения соответствующих методов для анализа их структуры и значения.

### Теория

В данном разделе описана предлагаемая базовая модель для распознавания содержимого маркерной доски (рукописных блок-схем и мультязычного текста). Модель включает несколько ключевых этапов: предобработку изображения, выделение и распознавание графических элементов через контуры, распознавание мультязычного текста и генерацию XML-файла для структурированного представления результатов. В данной работе под мультязычностью понимается поддержка двух языков – русского и английского.

#### 1. Предобработка изображения

Первым шагом в обработке изображения является его предобработка, направленная на улучшение качества, включающая коррекцию контраста и яркости, и подготовку к последующим этапам анализа. Этот процесс начинается с преобразования исходного цветного изображения в оттенки серого, что упрощает цветовое пространство и позволяет сосредоточиться на интенсивности пикселей, игнорируя цветовые вариации. Далее применяется гауссово размытие, которое сглаживает изображение, уменьшая влияние высокочастотных шумов, сохраняя при этом значимые границы объектов.

Для удаления фоновых теней и слабых пикселей, не относящихся к полезной информации, применяется пороговая операция: из цветного изображения извлекается компонент серого, после чего все пиксели с интенсивностью большей 70 переводятся в чисто белый цвет. Сам порог определен эмпирически по результатам экспериментов. Такая операция точки белого позволяет избавиться от бледных пятен и сгладить неравномерную засветку, обеспечивая более четкий контраст между линиями маркера и фоновым покрытием доски. В результате выполнения всех шагов первого этапа получается изображение с повышенной контрастностью, унифицированным разрешением и минимальным уровнем фонового шума.

Для выделения контуров используется алгоритм детекции краев Кэнни, известный своей способностью находить границы объектов с высокой точностью благодаря двухпороговой фильтрации и отслеживанию связности [10]. После этого применяются морфологические операции: дилатация для соединения разрывов в контурах и эрозия для удаления мелких шумовых артефактов. Эти шаги формируют надежную основу для последующего обнаружения контуров, что критически важно для выделения графических элементов диаграммы.

## 2. Выделение и распознавание графических элементов

Выделение графических элементов, таких как прямоугольники и стрелки, осуществляется с использованием методов анализа контуров. После предобработки из изображения извлекаются контуры, представляющие замкнутые области. Эти контуры аппроксимируются до полигонов для упрощения их геометрической структуры с сохранением ключевых характеристик. Классификация форм основана на анализе числа сторон полученного полигона: четырехугольники идентифицируются как прямоугольники, тогда как стрелки распознаются по более сложным критериям.

Для обнаружения стрелок используется анализ выпуклой оболочки полигона. Вычисляется количество вершин выпуклой оболочки. Если оно лежит в диапазоне от 4 до 5 (дополнительная сторона в случае, если стрела имеет плоское основание) и если форма стрелки имеет ровно две дополнительные точки, которых нет в выпуклой оболочке (рис. 1), то контур рассматривается как потенциальная стрелка и ищется наконечник и основание [11]. Точка, не принадлежащая выпуклой оболочке, являющаяся самой выпуклой острой вершиной и наиболее выступающая из общего контура, интерпретируются как наконечник стрелки, а наиболее удаленная точка от наконечника стрелки на оболочке определяется как основание. Этот подход позволяет надежно различать направление стрелок, что важно для интерпретации структуры диаграммы.



Рисунок 1. Контур стрелки

Результаты распознавания контуров сохраняются в виде двух списков для дальнейшей обработки. Для прямоугольников каждый элемент — это список из четырёх пар координат (x,y), соответствующих вершинам прямоугольника в порядке обхода. Для стрелок — это кортеж из двух элементов, где первый элемент — координаты наконечника стрелки, а второй элемент — координаты основания.

Чтобы обеспечить устойчивость метода к вариациям качества изображения и стилей рисования, применяется итеративный подход к настройке параметров аппроксимации контуров. Модель тестирует различные пороговые значения, выбирая конфигурацию, которая максимизирует количество распознанных элементов. Такой адаптивный механизм повышает работоспособность системы, как было выявлено на этапе тестирования, и делает её применимой к более широкому спектру входных данных.

Результат такой предобработки изображения и выделения контуров изображен на рисунке 2. Красным цветом визуализированы распознанные прямоугольники, зеленым — стрелки.

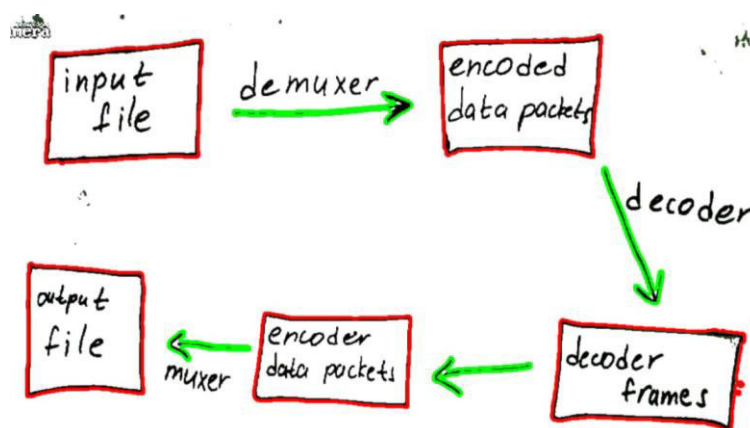


Рисунок 2. Результат распознавания контуров

### 3. Методы распознавания мультязычного текста

Распознавание текста в базовой модели реализовано с использованием гибридного подхода, сочетающего детекцию слов, классификацию языка и последующее распознавание, а также готовых инструментов. Особое внимание было уделено точности распознавания, скорости работы и способности адаптироваться к смешанному тексту. Мы протестировали такие инструменты, как EasyOCR [12], Shiftlab [13], Tesseract [14] и TrOCR [15, 16], для распознавания рукописного текста. В итоговую реализацию базовой модели взяли модели на основе TrOCR для решения задачи распознавания текста и EasyOCR для решения задачи детекции слов. Ниже перечислены гибридные методы, которые были протестированы для базовой модели распознавания мультязычного текста (русский и английский языки).

Детекция слов и формирование строк. Первым этапом в обработке текстового контента стало выделение отдельных слов и формирование строк, так как модель TrOCR принимает на вход только строки рукописного текста или отдельные слова. Данная модель не приспособлена для распознавания многострочного контента. Для детекции слов использовался инструмент EasyOCR, который предоставляет координаты ограничивающего прямоугольника (bounding box) для каждого распознанного слова. На основе этих данных был реализован алгоритм группировки слов по строкам: объекты объединялись в группы по вертикальной координате  $Y$  правого нижнего угла ограничивающих их прямоугольников с допустимой погрешностью  $d$ . Таким образом, каждая группа представляла собой потенциальную строку текста. После группировки производилось формирование единого ограничивающего прямоугольника для каждой строки путем объединения отдельных ограничивающих прямоугольников слов, входящих в эту группу. Полученные области использовались далее для последовательного применения моделей HTR к целым строкам или отдельным словам.

На основе анализа существующих технологий и результатов предварительных экспериментов были протестированы три подхода для базовой модели к распознаванию смешанного русско-английского текста.

Подход 1: гибридный метод с классификацией языка. Данный подход основан на комбинации EasyOCR, двух моделей TrOCR, обученных для русского и английского языков соответственно, и классификатора языка. После детекции слов с помощью EasyOCR для каждого слова осуществляется предсказание языка с использованием дообученной модели ResNet [12]. Затем применяется соответствующая модель TrOCR для распознавания текста. Недостатком этого подхода является необходимость запуска нескольких моделей, что увеличивает общее время обработки и снижает производительность, а также накопление ошибки при неправильной детекции слова или классификации языка, что может отразиться на итоговом распознавании. Важно отметить, что классификатор на основе ResNet срабатывал не идеально, тем не менее было принято решение протестировать данный подход более подробно.

Подход 2: выбор наиболее вероятного результата между двумя языками. В данном случае также используются EasyOCR для детекции и две модели TrOCR, но вместо предварительной классификации языка текст каждого слова распознаётся обеими моделями. Итоговый результат выбирается на основе сравнения нормализованных значений логитов. Однако, как показали эксперименты, значения логитов разных моделей не всегда сравнимы между собой даже после нормализации, что может приводить к ошибкам выбора языка. Это связано с различиями в процессе обучения исходных моделей, которые не были специально адаптированы под такую задачу.

Подход 3: единая мультязычная модель TrOCR. Третий подход заключается в применении уже обученной модели TrOCR [16], поддерживающей распознавание текста сразу на двух языках — русском и английском. Такой подход позволяет обрабатывать текст за один проход, что положительно влияет на скорость работы. Однако, как показали первоначальные тесты, доступные предобученные модели не обеспечивают достаточного качества распознавания на нашем наборе данных, особенно в случае сложных технических терминов и аббревиатур.

Все три рассмотренных подхода имеют свои преимущества и ограничения. Первый обеспечивает гибкий выбор модели в зависимости от языка, но требует дополнительных вычислений. Второй

предлагает более простую логику выбора, однако сталкивается с проблемой интерпретации логитов. Третий требует дообучения или тонкой настройки модели под специфику данных.

В рамках разработанной базовой модели был выбран третий подход, поскольку он позволил достичь наилучшего баланса между точностью и надежностью распознавания.

Текст внутри и вне прямоугольников блок-схемы распознается отдельно, чтобы затем можно было точно привязать текст к соответствующему прямоугольнику при генерации XML файла. Для этого после распознавания расположения прямоугольников эти области вырезаются и отдельно подаются в модель для распознавания текста. Затем области с прямоугольниками исключаются из изображения, чтобы избежать повторного учета текста при распознавании текста вне прямоугольников на изображении целиком.

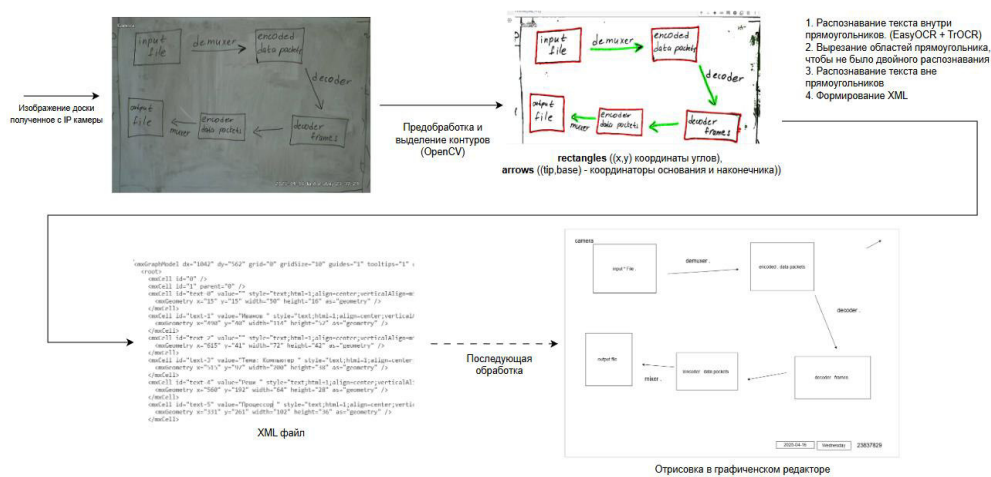
#### 4. Формирование XML-файла

Итоговым результатом работы модели является генерация файла в формате XML, который структурирует распознанные элементы в машиночитаемом формате. На вход функции генерации подаются координаты и значения всех распознанных элементов. В XML-файле указываются стили, размер шрифта и толщина линий. При генерации выделяются три основных типа данных.

- Прямоугольники представлены как вершины с указанием их позиций (координаты левого верхнего угла) и размеров (длина и ширина). В этом же объекте сохраняется текст, если он был написан внутри соответствующего прямоугольника.
- Стрелки – координаты, содержащие исходные (sourcePoint) и конечные точки (targetPoint), по которым отрисовывается направление стрелки.
- Текст – самостоятельные блоки с текстом вне прямоугольников. Указывается координата левого верхнего угла, ширина и высота блока с текстом, а также значение самого текста.

Структура XML обеспечивает точное воспроизведение пространственных отношений между элементами и текстом, что позволяет использовать результат для дальнейшей обработки, анализа или визуализации (например, отрисовки в графическом редакторе).

Общая структура предлагаемой базовой модели изображена на рисунке 3.





правильной детекции прямоугольников, стрелок и текста. Для решения задачи распознавания текста был сформирован отдельный датасет, состоящий из изображений строк и их расшифровок. Количество объектов в нем равно 21.

Условия эксперимента:

- константные параметры: положение IP-камеры, угол обзора, разрешение изображений, рабочая область доски;
- изменяющиеся параметры: освещение (естественное, искусственное, смешанное), почерк и способ изображения графических элементов (индивидуальные особенности письма), цвет и насыщенность маркера, размеры изображения.

Эти параметры отражают реальные условия использования и позволяют оценить устойчивость алгоритмов к внешним факторам.

## 2. Метрики для оценки

Для анализа эффективности модели выбраны метрики, обеспечивающие простоту разметки и вычисления.

Распознавание текста:

- уровень ошибок в символах (Character Error Rate, CER): доля ошибок на уровне символов между предсказанным и истинным текстом;
- уровень ошибок в словах (Word Error Rate, WER): доля ошибок на уровне слов.

Эти метрики стандартны для задач распознавания текста и подходят для оценки мультязычного текста.

Идеальным значением для метрик является 0. Это означает, что распознавание прошло идеально — без ошибок, то есть метрика уровня ошибок в символах показывает, насколько точно система распознает отдельные символы, а метрика уровня ошибок в словах важна для оценки понятности текста, она отражает смысловые ошибки.

Выделение и распознавание графических элементов:

- пересечение по блокам (Intersection over Union, IoU): степень пересечения предсказанных и истинных ограничивающих прямоугольников (bounding box) для оценки точности локализации прямоугольников и стрелок;
- F1-мера: гармоническое среднее точности (precision) и полноты (recall) для оценки качества обнаружения графических элементов;
- точность (accuracy): доля правильно распознанных объектов среди всех объектов

Метрики разделены по задачам, что позволяет детально оценить производительность модели.

Диаграмма считается правильно распознанной, если:

- каждый объект (блок-прямоугольник или текст) правильно классифицирован и локализован с значением метрики пересечения по блокам (IoU) не менее 70 %, то есть пересечение истинного ограничивающего прямоугольника и полученного с помощью нашей базовой модели - 0,7;
- для каждой стрелки правильно идентифицированы узлы основания и наконечника. Это определяется через оценку расстояния между точками основания и наконечника предсказанного и размеченного объекта. Если евклидовы расстояния между этими точками меньше 100, то считается, что стрелка распознана правильно.

## 3. Результаты базовой модели

В данном разделе представлены результаты количественной оценки производительности базовой модели по четырем ключевым задачам: детекция прямоугольников, локализация текстовых строк, распознавание стрелок и распознавание мультязычного текста на основе собственного датасета диаграмм и текстов преимущественно на белой маркерной доске. Для оценки использовались стандартные метрики, которые были ранее упомянуты в статье. При оценке локализации строк текста выделялся весь текст на изображении, а не только текст вне блоков, для большего количества примеров.

Оценка детекции прямоугольников, локализации строк текста и детекции стрелок представлена в таблице 1, оценка качества распознавания мультязычного текста – в таблице 2.

Таблица 1. Оценка детекции объектов

Метрика	Оценка детекции прямоугольников	Оценка локализации строк текста	Оценка детекции стрелок
IoU	0.5003	0.6316	-
precision	0.5729	0.6433	0.693
recall	0.5359	0.6040	0.522
accuracy	0.5359	0.6040	0.539
F1 score	0.5538	0.6231	0.595

Таблица 2. Оценка мультязычного распознавания текста

Подход	Подход 1 (с ResNet)	Подход 2 (анализ логитов)	Подход 3 (мультязычная TrOCR)
CER	0.85	0.51	0.43
WER	1.67	0.89	0.84

Как упоминалось ранее при предобработке изображения для удаления фоновых теней и слабых пикселей, не относящихся к полезной информации, применяется пороговая операция: из цветного изображения извлекается компонент серого, после чего все пиксели с интенсивностью большей 70 переводятся в чисто белый цвет. В таблице 3 представлены результаты распознавания при разных показателях этого порога.

Таблица 3. Метрики при разном пороге предобработки

Метрика	Попор (threshold)				
	50	60	70	80	90
IoU rectangle	0.4212	0.4647	0.5003	<b>0.5043</b>	0.4910
F1 score rectangle	0.5117	0.5415	0.5538	0.5710	<b>0.5747</b>
Accuracy rectangle	0.4524	0.4995	0.5359	<b>0.5421</b>	0.5315
IoU text	0.5876	0.6085	<b>0.6316</b>	0.6250	0.6215
F1 score text	0.5193	0.5538	<b>0.6231</b>	0.6084	0.6062
Accuracy text	0.5028	0.5336	<b>0.6040</b>	0.5832	0.5867
F1 score arrow	0.5924	0.5836	<b>0.5952</b>	0.5830	0.5410
Accuracy arrow	<b>0.5543</b>	0.5472	0.5388	0.5531	0.5200

### Обсуждение результатов

Несмотря на полученные результаты и выявленный потенциал базового метода, эксперименты продемонстрировали существенные ограничения текущей реализации, которые обуславливают низкую точность и нестабильность работы в реальных условиях. В связи с этим целесообразно рассмотреть векторы работы дальнейших исследований, направленных на преодоление выявленных проблем и повышение качества распознавания. Ниже представлены ключевые перспективы развития системы, базирующиеся на анализе выявленных недостатков и современных подходах в обработке рукописного контента.

#### 1. Улучшение качества изображения через попиксельные (Pixel-to-Pixel) преобразования

Одним из перспективных направлений является применение генеративных моделей с попиксельными преобразованиями для перевода изображений маркерной доски в более структурированный вид, близкий к цифровым чертежам. Такой подход позволяет устранить неравномерное освещение, блики, фоновые помехи и повысить четкость линий и символов. Это особенно важно при работе с изображениями, сделанными под углом, с пятнами и разводами или при плохом освещении. Предварительная обработка с использованием таких моделей может значительно улучшить качество входных данных для последующих этапов распознавания.

Кроме того, использование моделей с попиксельным преобразованием может способствовать унификации почерка и стиля рисования графических элементов, что положительно скажется на стабильности работы классификаторов форм и моделей распознавания рукописного текста.

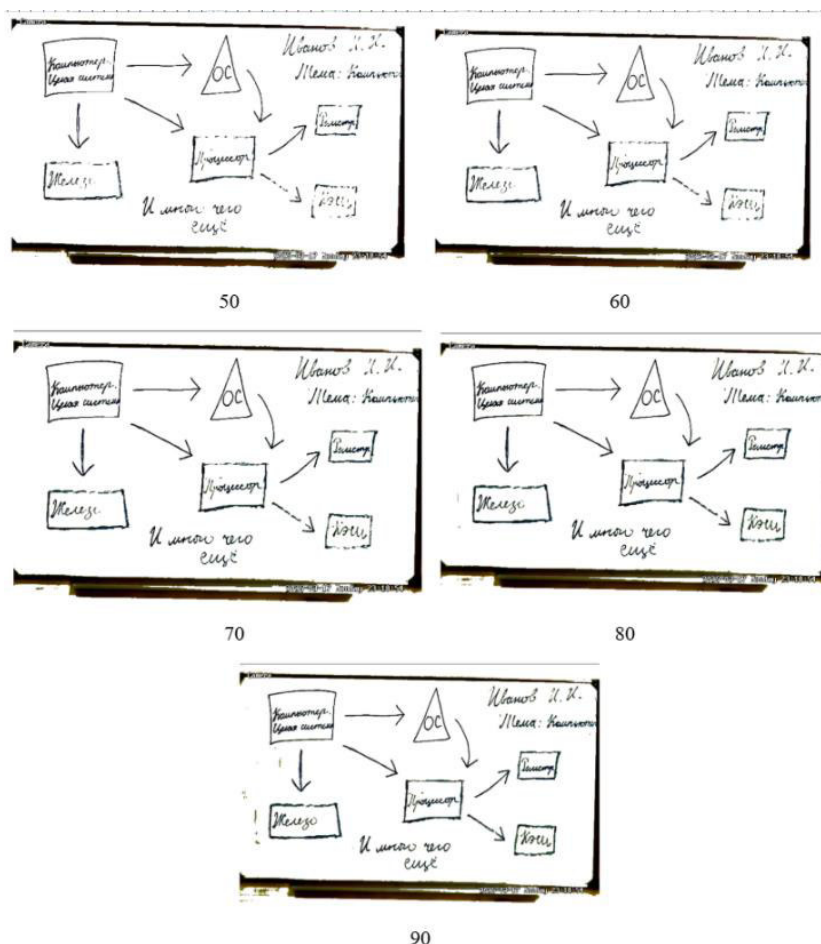


Рисунок 4. Изображения после преобработки с разным порогом белого

## 2. Совместное распознавание текста и графики: монолитные архитектуры(end-to-end)

Следующим важным направлением является переход от модульной архитектуры к монолитной модели, которая будет одновременно обрабатывать как текстовые, так и графические элементы. Такие модели могут использовать архитектуры типа трансформер или их модификации, учитывающие пространственное расположение объектов.

Интеграция информации о взаимном расположении текста и графики позволит улучшить интерпретируемость результата и повысить точность определения принадлежности текстовых блоков к графическим элементам, например, к стрелкам или прямоугольникам. Также это поможет в автоматическом восстановлении структуры диаграммы.

## 3. Дообучение мультязычных моделей распознавания рукописного текста (HTR) на специализированных датасетах

Основным недостатком гибридного подхода остается необходимость запуска нескольких моделей, что увеличивает вычислительную нагрузку. Поэтому перспективным направлением является разработка и дообучение универсальной мультязычной модели распознавания рукописного текста на датасетах, содержащих смешанные русско-английские рукописные образцы.

Такой подход позволит повысить стабильность распознавания смешанных фраз и аббревиатур.

## 4. Расширение функциональности: поддержка новых типов графических элементов

На данном этапе модель поддерживает распознавание только прямоугольников и стрелок. Для более широкого применения в образовательной и профессиональной среде целесообразно расширить спектр распознаваемых графических примитивов, включив окружности, ромбы, параллелограммы и другие элементы блок-схем. Это потребует как доработки модуля выделения контуров, так и расширения XML-структуры выходного файла.

## Закключение

В данной работе представлен комплексный подход к распознаванию рукописного мультязычного текста и графических элементов на маркерных досках. Предложена базовая модель, объединяющая методы распознавания текста (EasyOCR и TrOCR) и анализ контуров через OpenCV для выделения фигур. Результаты представлены в структурированном формате, что позволяет использовать их в системах автоматической обработки данных. Основные результаты включают решение ключевых проблем: детекции строк в смешанном тексте, классификации языка, отделения графики от фонового шума. В дальнейшем планируется улучшение качества распознавания через попиксельные преобразования, разработку монолитных моделей для совместного распознавания текста и графики. Полученные результаты открывают возможности для преобразования диаграмм в машиночитаемые форматы, что актуально для образовательных и профессиональных сред.

Ноутбуки для расчета метрик, отрисовки результата и собранный датасет с разметкой (директория Research), а также Телеграмм-бот как интерфейс взаимодействия для распознавания базовой моделью расположены в репозитории GitLab – Режим доступа: <https://git.miem.hse.ru/msmosalev1/board-recognizer>.

## Список литературы

1. Schäfer B., Keuper M., Stuckenschmidt H. Arrow R-CNN for handwritten diagram recognition // International Journal on Document Analysis and Recognition (IJDAR). 2021. Режим доступа: <https://link.springer.com/article/10.1007/s10032-020-00361-1>
2. Bluche T., Messina R. Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition // Proceedings of the 14th International Conference on Document Analysis and Recognition (ICDAR). 2017. Режим доступа: [https://tbluche.com/files/icdar17\\_gnn.pdf](https://tbluche.com/files/icdar17_gnn.pdf)
3. Omasa T., Koshihara R., Morishige M. Arrow-Guided VLM: Enhancing Flowchart Understanding via Arrow Direction Encoding // arXiv. 2025. arXiv:2505.07864. Режим доступа: <https://arxiv.org/pdf/2505.07864>
4. DrawnNet: Offline Hand-Drawn Diagram Recognition Based on Keypoint Detection // PMC. 2022. Режим доступа: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8947756/>
5. Sabeghi Saroui B. Recognition of Mathematical Handwriting on Whiteboards // University of Birmingham. 2015. Режим доступа: <https://etheses.bham.ac.uk/6251/1/SabeghiSaroui15PhD.pdf>
6. Ali S. H., Abdulrazzaq M. B. A Comprehensive Overview of Handwritten Recognition Techniques: A Survey // Journal of Computer Science. 2023. Vol. 19, no. 5. P. 569–587. Режим доступа: <https://doi.org/10.3844/jcssp.2023.569.587>.
7. Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey / W. AlKendi, F. Gechter, L. Heyberger, C. Guyeux // Journal of Imaging. 2024. Vol. 10, iss. 1. Art. 18. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10817575/>
8. Hamad, H. Handwritten Recognition Techniques: A Comprehensive Review / H. Hamad, S. T. Al-Dhaher, M. S. Al-Hakeem // Symmetry. 2024. Vol. 16, Iss. 6. Art. 681. URL: <https://www.mdpi.com/2073-8994/16/6/681>
9. OCR рукописного текста: почему это так важно и трудно // Яндекс Образование : [сайт]. 2023. 17 ноября. URL: <https://education.yandex.ru/journal/kak-raspoznat-rukopisnyi-tekst>
10. Canny J. A computational approach to edge detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1986. Vol. 8, no. 6. P. 679–698.
11. How to detect different types of arrows in image // Stack Overflow. Режим доступа: <https://stackoverflow.com/questions/66718462/how-to-detect-different-types-of-arrows-in-image>
12. A Novel Technique for Handwritten Text Recognition Using Easy OCR / B. K. Pattanayak, A. K. Biswal, S. R. Laha [и др.] // 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) : Proceedings. IEEE, 2023. P. 1115–1119. Режим доступа: [https://theGrenze.com/pages/servej.php?fn=640\\_1.pdf&name=Efficient+Text+Extraction+and+Summarization+usingEasyocr+and+GPT-3&id=2886&association=GRENZE&journal=GIJET&year=2024&volume=10&issue=2](https://theGrenze.com/pages/servej.php?fn=640_1.pdf&name=Efficient+Text+Extraction+and+Summarization+usingEasyocr+and+GPT-3&id=2886&association=GRENZE&journal=GIJET&year=2024&volume=10&issue=2)
13. shiftlab\_ocr: A python OCR library to read and generation // GitHub repository. Режим доступа: [https://github.com/konverner/shiftlab\\_ocr](https://github.com/konverner/shiftlab_ocr) (дата обращения: 20.05.2025).
14. Ракшит С., Басу С. Recognition of Handwritten Roman Script Using Tesseract Open source OCR Engine // Proc. National Conference on NAQC. 2008. P. 141–145. Режим доступа: <https://arxiv.org/pdf/1003.5891.pdf>
15. A Comprehensive Evaluation of TrOCR with Varying Image Effects // The National High School Journal of Science. 2024. Режим доступа: <https://nhsjs.com/2024/a-comprehensive-evaluation-of-trocr-with-varying-image-effects/>
16. Sergak0. text-recognition // Репозиторий GitHub. Режим доступа: <https://github.com/sergak0/text-recognition>

УДК 004.932

DOI: 10.25686/978-5-8158-2474-4-2025-670-674

## Обнаружение клубней картофеля на фоне грунта

А. В. Левитин, Е. П. Мудров

Рязанский государственный радиотехнический университет имени В.Ф. Уткина, Рязань, Россия

**Аннотация.** В статье рассматривается задача сегментации изображений картофельных клубней на фоне грунта. Для классификации пикселей в качестве признаков применялись размытые цветовые каналы, а также два алгоритма машинного обучения: случайный лес (Random Forest) и линейный дискриминантный анализ (Linear Discriminant Analysis, LDA). Экспериментальные результаты подтвердили эффективность предложенного решения.

**Ключевые слова:** сегментация изображений, классификация, линейный дискриминантный анализ, случайный лес, обнаружение клубней.

## Detection of potato tubers against the background of soil

A. V. Levitin, E. P. Mudrov

Ryazan State Radio Engineering University named after V.F.Utkin, Ryazan, Russia

**Abstract.** The article considers the problem of segmentation of images of potato tubers on the background of the soil. To classify pixels as features, blurred color channels were used, as well as two machine learning algorithms: Random Forest and Linear Discriminant Analysis (LDA). Experimental results have confirmed the effectiveness of the proposed solution.

**Keywords:** image segmentation, classification, linear discriminant analysis, random forest, tuber detection.

### Введение

Задача обнаружения клубней картофеля на фоне грунта является ключевой при создании системы технического зрения для интеллектуального картофелеуборочного комбайна. Целью данной работы является разработка алгоритмов для СТЗ, способных решать эту задачу [1, 2].

### Постановка задачи

Система технического зрения должна выполнять бинарную сегментацию получаемых изображений с разрешением 640×480 пикселей, классифицируя каждый элемент как «клубень» или «не клубень» на основе обученной модели. Сложность решения данной задачи заключается в том, что клубни часто бывают частично скрыты слоем грунта. Это требует анализа не отдельного пикселя, а его окрестности для формирования нелокальных признаков, что является критически важным для точной классификации.

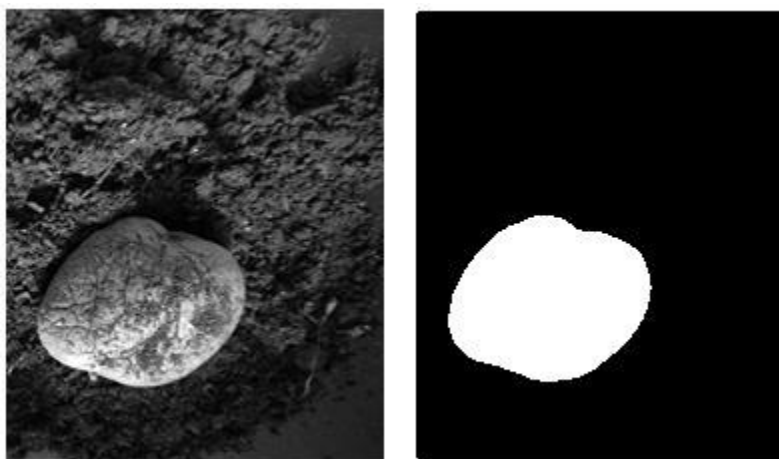


Рисунок 1. Пример исходного изображения и его маски

## Теория

### 1. Признаки элементов изображения

Для классификации каждого элемента изображения (пикселя) использовался вектор признаков  $\mathbf{z}_{ij} \in R^3$ , компоненты которого вычисляются по формуле

$$z_{ij}^k = \sum_{n,m} a_{nm} q_{i+n,j+m}^k, \quad (1)$$

где  $q_{ij}^k$  – значение  $k$ -го канала  $ij$ -го элемента;  $a_{nm}$  – коэффициенты, подбираемые так, чтобы значение признака  $z_{ij}^k$  определялось значениями канала в некоторой окрестности  $ij$ -го элемента, обеспечивая тем самым нелокальность многомерного признака  $\mathbf{z}_{ij}$ .

Коэффициенты в составе соотношения (1) принимали значения

$$a_{nm} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{n^2 + m^2}{2\sigma^2}\right), \quad (2)$$

где параметр  $\sigma$ , оптимальное значение которого подлежит экспериментальному определению, определяет глубину размытия изображения.

Размытие «скрывает» мелкие фрагменты грунта на поверхности клубня, сглаживая цветовые переходы и упрощая последующее разделение классов на основе обобщенной информации.

Под «каналами», используемыми для формирования вектора признаков, подразумеваются компоненты стандартных цветовых моделей. В различных экспериментах использовались: RGB или HSL. Выбор модели обусловлен необходимостью оценить, какое цветовое пространство наиболее эффективно для выделения клубней картофеля на фоне грунта, в условиях частичного перекрытия почвой.

### 2. Методы классификации элементов изображения

Для задачи бинарной классификации пикселей были применены и сравнены два алгоритма: линейный дискриминантный анализ [3] и случайный лес [4].

LDA предполагает, что векторы признаков  $\mathbf{z}_{ij}$  для обоих классов («клубень» и «грунт») имеют многомерное нормальное распределение с разными средними значениями, но с одинаковой ковариационной матрицей. Введем обозначения:  $x_{ij} \in T_0$  – элемент изображения  $x_{ij}$  расположен в области  $T_0$  с грунтом;  $x_{ij} \in T_1$  – элемент  $x_{ij}$  расположен в области  $T_1$  с клубнем. Тогда, в соответствии с выше сделанными предположениями LDA, справедливо соотношение

$$\ln p(x_{ij} \in T_k | \mathbf{z}_{ij}) = \mathbf{v}_k^T \mathbf{z}_{ij} + \eta_k + c, \quad (3)$$

где  $p(x_{ij} \in T_k | \mathbf{z}_{ij})$  – оценка апостериорной вероятности того, что  $x_{ij} \in T_k$ ;  $\mathbf{v}_k = \Sigma^{-1} \mathbf{m}_k$ ;

$\eta_k = -\frac{1}{2} \mathbf{m}_k^T \Sigma^{-1} \mathbf{m}_k + \ln p(x_{ij} \in T_k)$ ;  $\mathbf{m}_k$  – выборочная оценка математического ожидания  $\mathbf{z}_{ij}$  при  $x_{ij} \in T_k$ ;  $\Sigma$  – выборочная оценка ковариационной матрицы  $\mathbf{z}_{ij}$ ;  $p(x_{ij} \in T_k)$  – оценка априорной вероятности того, что  $x_{ij} \in T_k$ ;  $c$  – нормирующая величина, обеспечивающая равенство

$$\sum_{k=0}^1 p(x_{ij} \in T_k | \mathbf{z}_{ij}) = 1.$$

Случайный лес — это ансамблевый метод, который строит множество решающих деревьев и объединяет их результаты для повышения общей точности и устойчивости модели. Каждое дерево обучается на своей собственной бутстрэп-выборке (случайной подвыборке данных с повторениями). При построении каждого узла дерева рассматривается не весь набор признаков, а лишь их случайное подмножество. Это предотвращает переобучение. После обучения каждое дерево в ансамбле «голосует» за отнесение объекта к тому или иному классу. Итоговый прогноз модели определяется по результату большинства голосов.

### Результаты экспериментов

На рисунках 2 и 3 представлены результаты сегментации тестового RGB-изображения для Случайного леса и LDA соответственно. Вероятности принадлежности пикселей к классу "клубни" получены с помощью метода `predict_proba()`. Значения метрик составляют:  $\text{precision} = 0.94$ ;  $\text{recall} = 0.95$  при  $\sigma = 14$  для Случайного леса с параметрами максимальной глубины равной 5 и числа деревьев равного 15 (данные параметры были подобраны с помощью поиска по сетке) и  $\text{precision} = 0.84$ ;  $\text{recall} = 1.0$  при  $\sigma = 23$  для LDA.



Рисунок 2. Результат сегментации тестового RGB-изображения с использованием Random Forest



Рисунок 3. Результат сегментации тестового RGB-изображения с использованием LDA

На рисунке 4 показаны зависимости точности ( $\text{precision}$ ) и полноты ( $\text{recall}$ ) обнаружения пикселей клубней от параметра глубины размытия для методов LDA и RF.

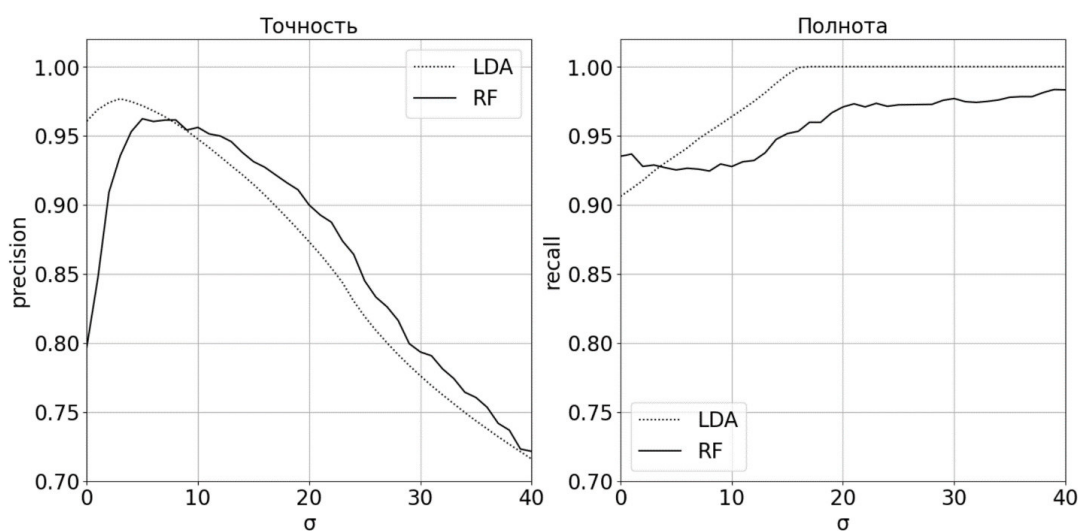


Рисунок 4. Зависимости характеристик обнаружения клубней от параметра глубины размытия

На рисунке 5 представлены кривые  $\text{precision-recall}$  и зависимости площади под этими кривыми (AUC-ROC) от параметра глубины размытия.



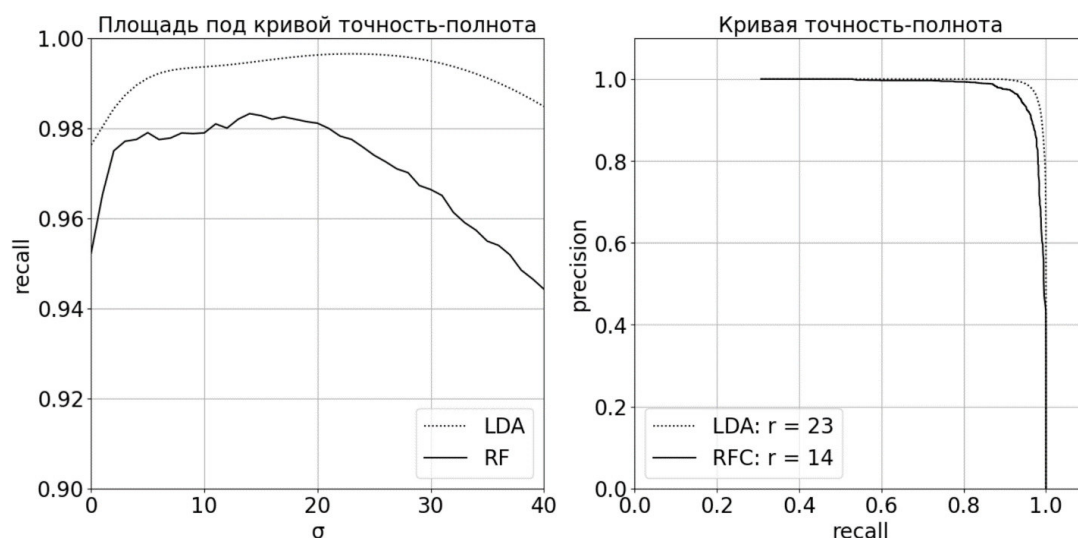


Рисунок 5. Кривые precision–recall и зависимость площади под кривой от параметра глубины размытия

На рисунках 6 и 7 представлены результаты сегментации тестового HSL-изображения для Случайного леса и LDA соответственно. Значения метрик составляют: precision = 0.93; recall = 0.86 при  $\sigma = 0$  для Случайного леса с параметрами максимальной глубины равной 9 и числа деревьев равного 20 и precision = 0.92; recall = 0.85 при  $\sigma = 3$  для LDA.



Рисунок 6. Результат сегментации тестового HSL-изображения с использованием Random Forest



Рисунок 7. Результат сегментации тестового HSL-изображения с использованием LDA

На рисунке 8 показаны зависимости точности (precision) и полноты (recall) обнаружения пикселей клубней от параметра глубины размытия для методов LDA и RF.

На рисунке 9 представлены кривые precision–recall и зависимости площади под этими кривыми (AUC-ROC) от параметра глубины размытия.

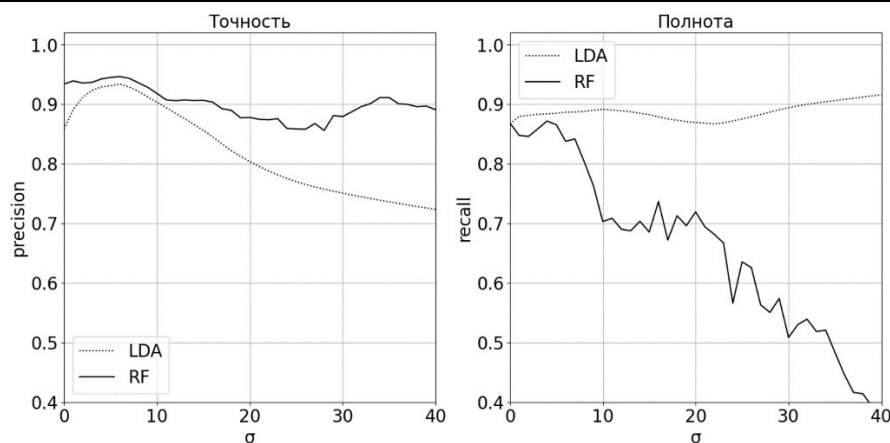


Рисунок 8. Зависимости характеристик обнаружения клубней от параметра глубины размытия

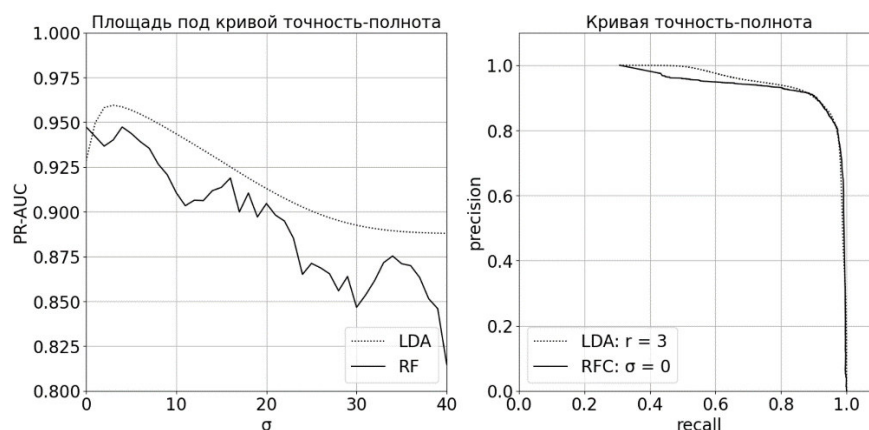


Рисунок 9. Кривые precision–recall и зависимость площади под кривой от параметра глубины размытия

## Выводы

Результаты исследования показали, что предложенные подходы демонстрируют разную эффективность в зависимости от используемого цветового пространства.

В пространстве RGB оба метода показали хорошие результаты, при этом метод LDA продемонстрировал значительно более высокое качество классификации. Также преимуществом LDA стала значительно меньшая вычислительная сложность по сравнению с Random Forest.

В отличие от RGB, в пространстве HSL оба метода показали неудовлетворительные результаты классификации, что свидетельствует о меньшей пригодности этого цветового пространства для решения данной конкретной задачи сегментации.

Оптимальное значение параметра глубины размытия  $\sigma$  может быть найдено путем максимизации площади под кривой точность–полнота для обучающего изображения.

## Список литературы

1. Левитин А. В., Мудров Е. П. Обнаружение клубней картофеля на фоне грунта // Новые информационные технологии в научных исследованиях: материалы XXIX Всероссийской научно-технической конференции студентов, молодых ученых и специалистов, Рязань, 27–29 ноября 2024 года / Рязанский государственный радиотехнический университет им. В.Ф. Уткина. 2024. Т. 2. С. 118.
2. Левитин А. В., Мудров Е. П. Обнаружение клубней картофеля на фоне грунта // Современные технологии в науке и образовании – СТНО-2025: сборник трудов VIII международного научно-технического форума, Рязань, 4–6 марта 2025 года / Рязанский государственный радиотехнический университет им. В.Ф. Уткина, 2025 Т. 5. С. 142–147.
3. Duda R. O., Hart P. E., Stork D. G. Pattern Classification. 2nd ed. New York: John Wiley & Sons, 2001. Section 2.6.2.
4. Breiman L. Random Forests // Machine Learning. 2001. Vol. 45, no. 1. P. 5–32.

## Сегментация строк в изображениях рукописных документов

Л. М. Местецкий<sup>1,2</sup>, В. С. Смирнова<sup>1</sup>

<sup>1</sup>Московский государственный университет имени М. В. Ломоносова, Москва, Россия

<sup>2</sup>Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

**Аннотация.** В работе предложен оригинальный метод построчной сегментации рукописных документов, основанный на построении векторного поля локальных направлений текста и формировании осевых линий строк, аппроксимирующих интегральные кривые данного поля. Разработан и реализован алгоритм, робастно работающий с реальными изображениями архивных документов, включая сложные случаи изгибающихся строк, неравномерных межстрочных интервалов и вариативности почерка. Экспериментальные результаты демонстрируют высокую эффективность (F1-score = 0.92).

**Ключевые слова:** сегментация строк, рукописные документы, триангуляция Делоне

## Line-by-line segmentation of handwritten documents

L. Mestetskiy<sup>1,2</sup>, V. Smirnova<sup>1</sup>

<sup>1</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> National Research University Higher School of Economics, Moscow, Russia

**Abstract.** We propose an original method for line-by-line segmentation of handwritten documents based on constructing a vector field of local text orientations and extracting line centerlines that approximate the field's integral curves. We develop and implement an algorithm that robustly handles real images of archival documents, including challenging cases with curved lines, non-uniform interline spacing, and handwriting variability. Experimental results demonstrate high effectiveness (F1-score = 0.92).

**Keywords:** text line segmentation, handwritten documents, Delaunay triangulation

### Введение

Исторические рукописные документы, такие как письма, дневники и заметки, содержат важную информацию о культурных, социальных и научных аспектах прошлого. Для автоматизированного распознавания текста в рукописных документах важно обеспечить точную построчную сегментацию, поскольку большинство современных OCR-систем выполняют распознавание именно на уровне отдельных строк.

Процесс выделения строк в рукописных документах осложняется вариативностью почерка, плотным расположением и нерегулярностью наклона строк, частыми перекрытиями выносных элементов букв, а также фоновым шумом в виде пятен и потертостей на бумаге, неоднородным освещением и углом съемки страниц документов.

Существуют четыре основных подхода к задаче сегментации строк. Проекционные методы строят профиль плотности пикселей по горизонтали и определяют границы строк по локальным минимумам, что хорошо работает на страницах с горизонтальными строками, имеющими статичный интервал, но слабо приспособлено к изогнутым строкам и перекрытиям [1, 2]. Морфологические алгоритмы, такие как RLSA (Run-Length Smoothing Algorithm), объединяют близко расположенные компоненты текста посредством морфологического замыкания, однако выбор оптимального структурирующего элемента остаётся нетривиальной задачей [3–5]. Графовые модели представляют документ в виде графа, где узлы — это связанные компоненты текста, а рёбра отражают их пространственные отношения; сегментация достигается решением задачи минимального графового разреза, обеспечивая гибкость при разделении тесно переплетённых строк [6]. Нейросетевые подходы (FCN, U-Net, RNN/LSTM) демонстрируют наивысшую точность на больших размеченных датасетах, но их применение ограничено высокими вычислительными затратами и необходимостью разметки больших объёмов данных [7].

Существует потребность в методе, который сочетал бы интерпретируемость и робастность графовых моделей с возможностью локального анализа без дорогостоящего этапа обучения. В рамках данной работы предлагается оригинальный алгоритм построчной сегментации, основанный на

построении нерегулярной решётки узлов в центрах локальной плотности текстовых пикселей, вычислении в каждой точке решётки вектора направления текста как минимального градиента функции плотности и аппроксимации осевых линий строк с помощью триангуляции Делоне. Такой подход обеспечивает высокую точность выделения строк, робастность к шуму и кривизне линий, а также исключает необходимость предварительного обучения на больших размеченных выборках.

Стоит отметить, что существующие подходы к сегментации строк без обучения (например, минимально-энергетические «швы» в *constrained seam carving* [8], голосование в пространстве Хафа по блокам связных компонент [9]) имеют ограничения: высокую чувствительность к предобработке и бинаризации; геометрические допущения, нарушаемые сильно искривленными строками, неравномерным межстрочным интервалом, пересечениями и слияниями штрихов, что приводит к множественным локальным минимумам и фрагментации/слиянию линий; выраженная масштабная зависимость (высота строки, толщина штриха, интервал), требующая ручной калибровки окон, фильтров и порогов и ухудшающая перенос между коллекциями без ретюнинга.

Цель исследования — разработать и экспериментально оценить предложенный метод построения сегментации рукописных документов с математически обоснованной структурой, обеспечивающей интерпретируемость и универсальность при обработке разнообразных архивных материалов.

### Постановка задачи

Исходными данными задачи являются изображения страниц исторических рукописных документов, таких как письма, дневники и заметки, представленные в виде растровых изображений. Для каждого исходного изображения требуется автоматически выполнить сегментацию текстовых строк путём построения множества осевых линий строк. Каждая осевая линия должна быть представлена упорядоченной ломаной, которая аппроксимирует центральную траекторию расположения компонент текста, отражая как глобальную кривизну строки, так и её локальный наклон. Выходом предложенного алгоритма является набор таких ломаных линий, предназначенный для последующего использования в системах оптического распознавания текста и других приложениях анализа рукописных документов.

Основные сложности задачи обусловлены вариативностью почерка авторов, неравномерностью межстрочных интервалов и изменчивостью угла наклона строк. Для формализации задачи приняты следующие предположения:

1. Каждая строка может быть аппроксимирована ломаной с углами перегиба, не превышающими  $45^\circ$ ;
2. Преобладающий наклон соседних строк считается постоянным;
3. В пределах одной строки локальные участки текста в малой окрестности имеют схожие углы наклона;
4. Строки преимущественно имеют горизонтальный наклон, то есть угол наклона строк к горизонтали, для которых возможна корректная сегментация, не превышает  $45^\circ$ .

На основании описанных выше допущений задача сегментации разбивается на две основные подзадачи.

**1. Построение строковой решётки.** Рукописная строка представляется как область с повышенной плотностью чёрных пикселей, по сравнению с межстрочными промежутками. Центры таких скоплений позволяют определить *узловые точки*, по которым восстанавливается траектория строки. В их окрестности текстовые пиксели имеют близкие направления, что даёт возможность оценить *локальный наклон строки*.

На основе узлов формируется *строковая решётка*  $G$ , и для каждой точки  $g_i \in G$  задаётся *вектор локального направления*  $d(g_i)$ , формирующий поле направлений  $d: G \rightarrow S$ . Далее формируются осевые линии строк как линии, аппроксимирующие интегральные кривые векторного поля  $d$ , проходящие через узловые точки  $g_i$  с касательным направлением, совпадающим с вектором  $d(g_i)$ . Это обеспечивает согласованность траекторий с локальной геометрией и кривизной строк.

**2. Формирование осевых линий.** На основе строковой решётки  $G$  строится граф потенциальных связей  $N = (G, E_0)$ , где рёбра  $E_0$  соединяют узловые точки. Отбирается подмножество рёбер  $E_S$ , согласованных с локальными направлениями поля  $d$ , что даёт подграф  $N_S = (G_{\text{ext}}, E_S)$ . Его связные

компоненты соответствуют фрагментам осевых линий, а также на нём определены две основные операции:

- 1) склейка рёбер, имеющих общие вершины и близкие направления – отбор фрагментов ломанной линии при их соединении концами;
- 2) склейка фрагментов – объединение связанных цепочек рёбер в строки.

В итоге задача заключается в построении целостных линий, согласованных по направлению, то есть выделенных связанных цепей в графе для восстановления полной геометрии строк.

### Теория

Задача сегментации строк сводится к автоматическому выделению осевых линий текста на бинаризованном изображении страницы  $D$ . Для точной формализации вводятся следующие понятия.

**Определение 1 (Узловая точка).** Пусть  $D \in \mathbb{Z}^2$  — множество координат текстовых пикселей. Точка  $d \in D$  называется *узловой*, если существует окрестность  $U_\epsilon(g)$  такая, что плотность текстовых пикселей в ней максимальна по сравнению с соседними областями.

**Определение 2 (Строковая решётка).** Множество всех узловых точек

$$G = \{g_1, g_2, \dots, g_N\} \subset D$$

называется *строковой решёткой*. Решётка обеспечивает дискретное представление страницы через ключевые точки, по которым восстанавливаются осевые линии строк.

**Определение 3 (Поле направлений строк).**

$$d: G \rightarrow S, \quad S = \{v \in \mathbb{R}^2 : \|v\| = 1\},$$

где  $d(g_i)$  – единичный вектор локального направления строки в окрестности  $g_i$ .

**Определение 4 (Осевая линия строки).** Непрерывная упорядоченная ломаная

$$l = (v_1, \dots, v_k), \quad v_k \in G,$$

такая, что для каждого сегмента  $(v_k, v_{k+1})$  вектор  $v_{k+1} - v_k$  примерно совпадает с  $d(v_k)$ , называется *осевой линией*.

### Метод решения

#### Преобразование изображения

Входные данные – .jpg-изображения страниц архивных рукописных дневников. Сначала проводится *медианное сглаживание* для подавления шумов при сохранении структуры текста. Затем выполняется *адаптивная бинаризация* с локальным порогом

$$T(x, y) = \frac{1}{\sum_{(i,j) \in \mathcal{N}} w(i, j)} \sum_{(i,j) \in \mathcal{N}} I(i, j) w(i, j) - C, \quad \mathcal{N} = 9 \times 9, \quad C = 6,$$

где  $w(i, j)$  – веса гауссова ядра.

Пиксель  $(x, y)$  считается чёрным, если  $I(x, y) > T(x, y)$ . Наконец, для удаления мелких артефактов применяются последовательно морфологические преобразования: сначала открытие, затем закрытие (композиции эрозии и дилатации) с квадратным структурирующим элементом  $3 \times 3$ .

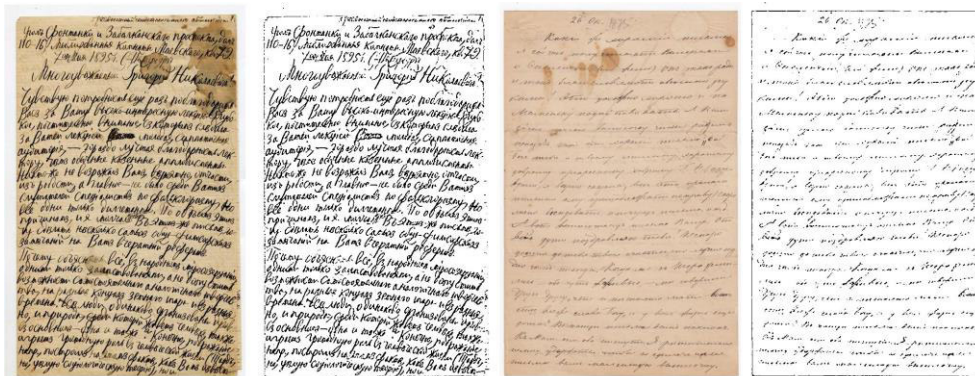


Рисунок 1. Примеры бинаризации: справа сходное изображение, слева – бинаризованное

### Построение строковой решётки

В основе метода лежит гипотеза о том, что в пределах ограниченной области угол наклона текстовой строки статичен, то есть на выбранном фрагменте изображения направление строки практически не меняется. Более того, чем уже рассматриваемая область, тем точнее выполняется это предположение, поскольку локальная структура строки слабо подвержена искажениям и кривизне на малом масштабе.

В соответствии с введенным предположением для перехода к дискретному представлению страницы область изображения  $D$  разбивается на  $N$  вертикальных полос одинаковой ширины (далее  $N = 8$ ):

$$S_j = \{(x, y) \in D \mid x \in [jW/N, (j+1)W/N)\}, \quad j = 0, \dots, N-1,$$

где  $W$  – ширина изображения.

В каждой полосе строится вертикальный переходный профиль

$$r_j(y) = \sum_{x:(x,x+1) \in S_j} |D(x+1, y) - D(x, y)|,$$

где  $D(x, y) \in \{0, 1\}$ .

Локальный максимум  $r_j(y_{j,k})$  принимается, если выполняются следующие условия:

$$r_j(y_{j,k-1}) < r_j(y_{j,k}),$$

$$r_j(y_{j,k}) > r_j(y_{j,k+1}),$$

$$r_j(y_{j,k}) \geq \tau_j, \quad \tau_j = 1,5 \text{ median}(r_j(y)).$$

Здесь  $k$  – индекс дискретной вертикальной позиции в строке  $j$ .

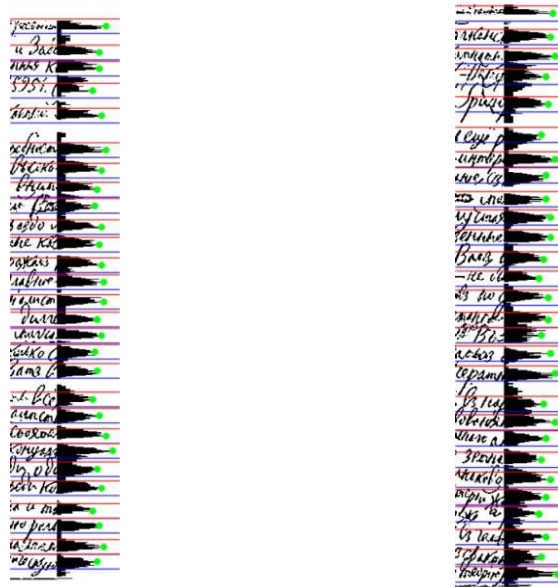


Рисунок 2. Фрагменты полученных полос и соответствующих гистограмм строчных профилей. Зеленые точки – локальные максимумы, красные/синие линии – верхняя/нижняя границы описанных фреймов

Максимумы дают уровни  $y_{j,k}$  с наибольшей плотностью переходов «фон–текст» и служат опорой для определения центров тяжести компонент строки. Порог отсекает шумовые пики и нетекстовые детали.

Для каждого  $y_{j,k}$  вычисляется центр тяжести

$$x_{j,k} = \text{median}\{x \mid D(x, y_{j,k}) = 1, (x, y_{j,k}) \in S_j\}$$

и формируется прямоугольный фрейм

$$R_{j,k} = [x_{j,k} - \frac{w_j}{2}, x_{j,k} + \frac{w_j}{2}] \times [y_{j,k} - \frac{h_j}{2}, y_{j,k} + \frac{h_j}{2}],$$

где  $w_j = \lfloor W/N \rfloor$ ,  $h_j = \frac{2}{3}w_j$ .



Такой фрейм охватывает достаточное число пикселей для надёжной локальной оценки направления, минимально пересекается с соседними полосами и, так как высота по построению меньше ширины ( $h_j < w_j$ ), повышает точность оценки вдоль строки:

$$P_{j,k} = \{p \in R_{j,k} \mid D(p) = 1\}.$$

### Оценка локального направления методом главных компонент

Для каждого облака точек  $P_{j,k}$  направление строки и её центр тяжести оцениваются методом главных компонент, который робастен к шуму, имеет низкую вычислительную сложность и согласуется с допущением о локальной стационарности наклона, а экспериментально подобранное соотношение сторон фрейма  $h_j : w_j = 2:3$  обеспечивает надёжность оценки локального направления при сильной кривизне строки. Сначала вычисляется центр масс

$$\bar{p} = \frac{1}{|P_{j,k}|} \sum_{p \in P_{j,k}} p,$$

затем ковариационная матрица

$$\Sigma_{j,k} = \frac{1}{|P_{j,k}|} \sum_{p \in P_{j,k}} (p - \bar{p})(p - \bar{p})^T,$$

и её старший собственный вектор  $v_1$  дает оценку нормированного локального направления

$$d(g_{j,k}) = v_1, \quad \|d(g_{j,k})\| = 1.$$

Пара  $(g_{j,k}, d(g_{j,k}))$ , где  $g_{j,k} = \bar{p}$  формирует узел строковой решётки и локальное направление. Совокупность

$$(G, d) = \{(g_{j,k}, d(g_{j,k}))\}$$

является компактным множеством, робастным к артефактам письма и шумам, и служит основой для дальнейшей реконструкции непрерывных осевых линий.

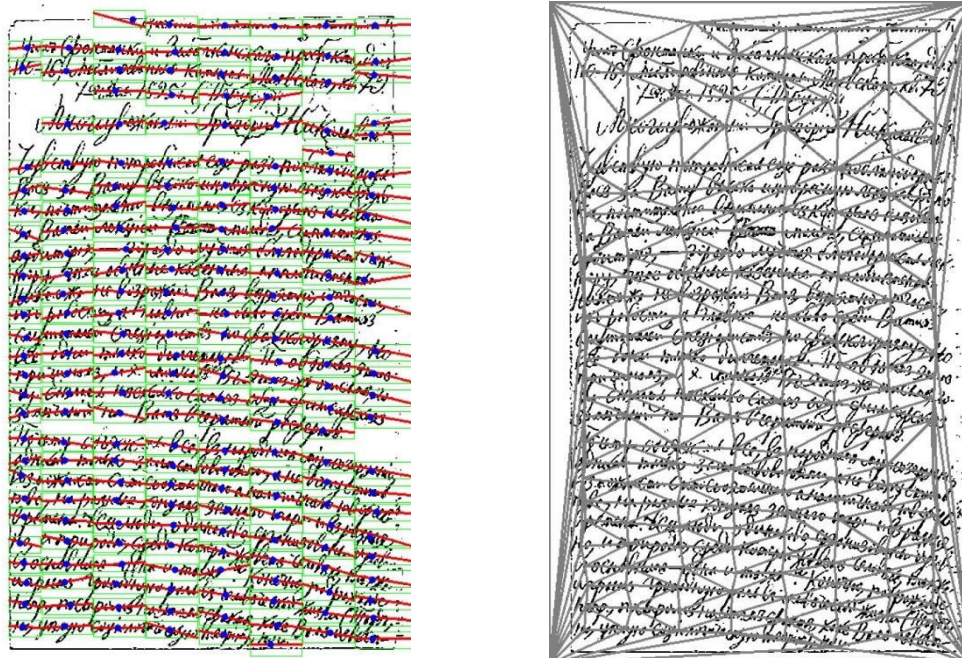


Рисунок 3. Пример страницы с выделенными узловыми точками (синие), векторами главных компонент (красные). Зеленым обозначены описанные фреймы, серым – построенная триангуляция Делоне

### Построение осевых линий строк

Чтобы восстановить траекторию каждой строки, к узлам строковой решётки  $G$  добавляются четыре вершины для корректного построения модели страницы, совпадающие с углами изображения



$C = \{(0, 0), (W - 1, 0), (W - 1, H - 1), (0, H - 1)\}$ , что даёт расширенное множество  $G_{\text{ext}} = G \cup C$ . По этому множеству строится триангуляция Делоне, то есть такое разбиение плоскости на треугольники, в описанную окружность каждого из которых не попадает ни одна другая точка  $G_{\text{ext}}$ . Поскольку узлы  $G$  лежат в центрах текстовых кластеров, их ближайшие соседи в этой триангуляции, как правило, принадлежат той же строке, а свойство Делоне минимизирует узкие и длинные треугольники, благодаря чему большинство рёбер ориентировано почти горизонтально, а построение занимает  $O(|G_{\text{ext}}| \log |G_{\text{ext}}|)$  времени. Обозначим полный граф триангуляции  $N_0 = (G_{\text{ext}}, E_0)$ .

Далее каждое ребро  $e = (g, g') \in E_0$  проверяется на согласованность с локальным направлением, определённым в узлах методом главных компонент. Пусть  $w = g' - g$  и пусть  $d(g), d(g')$  – нормированные векторы, показывающие ориентацию строки в точках  $g$  и  $g'$ . Тогда

$$\theta(g) = \arccos \frac{w \cdot d(g)}{\|w\|}, \quad \theta(g') = \arccos \frac{w \cdot d(g')}{\|w\|},$$

а меру несогласованности задает

$$\delta(e) = \max\{\min(\theta(g), 180^\circ - \theta(g)), \min(\theta(g'), 180^\circ - \theta(g'))\}.$$

Ребро считается подходящим, если  $\delta(e) \leq \varphi_{\text{max}}$ , где  $\varphi_{\text{max}} = 7^\circ$  допускает умеренную кривизну строк. Рёбра, удовлетворяющие этому условию, образуют подграф  $N_s = (G_{\text{ext}}, E_s)$ ,  $E_s = \{e \in E_0 \mid \delta(e) \leq \varphi_{\text{max}}\}$ ; каждая его связная компонента, как правило, соответствует одной строке.

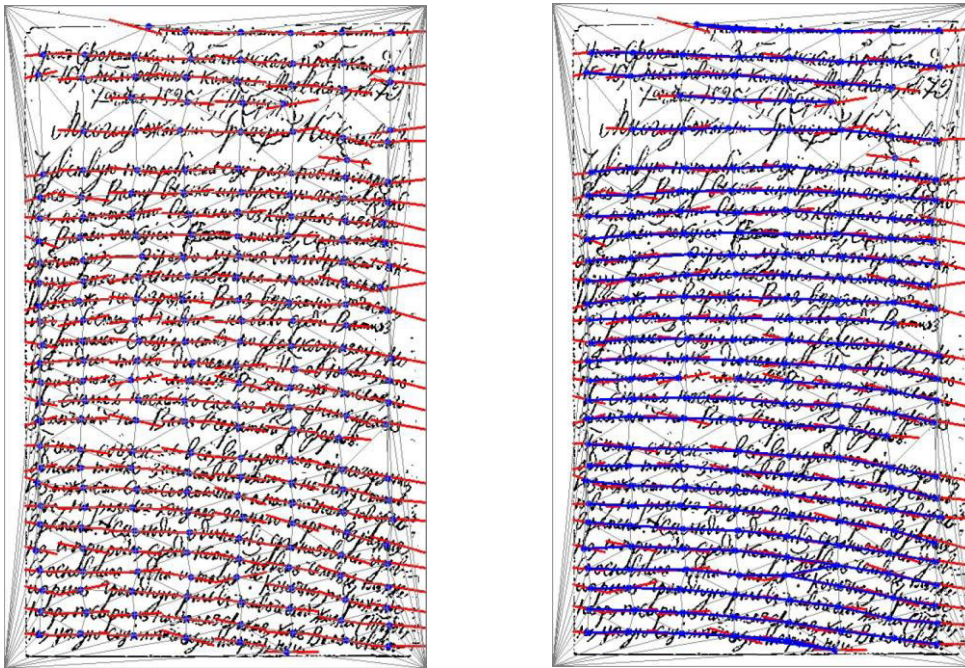


Рисунок 4. Пример страницы с построенной триангуляцией (серый) и векторами направлений (красный).

Справа синим выделены отобранные по критерию ребра триангуляции

Наконец, рёбра с близкими направлениями, исходящие из общей вершины, склеиваются в ломаные, а соседние ломаные, совпадающие по ориентации и расположению, соединяются в непрерывные полилинии, которые аппроксимируют интегральные кривые поля направлений  $d$  и, таким образом, задают искомые осевые линии строк.

#### Постобработка осевых линий

После отбора рёбер получен граф, где каждая компонента уже аппроксимирует строку, но могут оставаться два типа локальных артефактов:

- разрывы цепочек: между некоторыми последовательными вершинами отсутствует ребро;
- шумовые развилки: вершина графа имеет степень более двух, образуя неоднозначные ветвления.

### Устранение разрывов цепочек

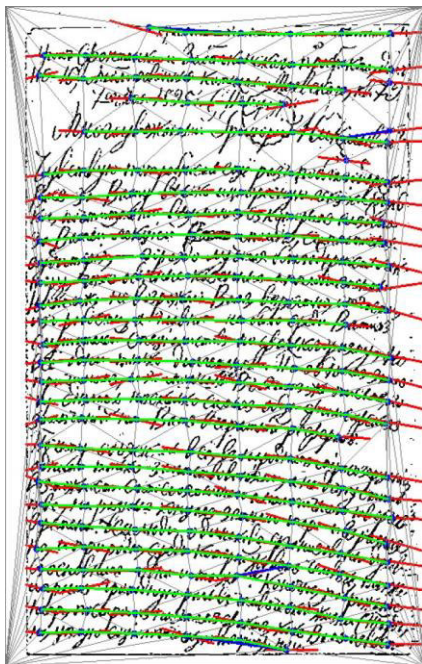
Пусть  $u$  и  $w$  – правый и левый концы двух компонент, для которых  $\|u - w\|$  мала. Ищем путь  $P$  внутри узкой вертикальной полосы и выбираем

$$P^* = \arg \min_P \sum_{(p,q) \in P} (\lambda \|q - p\| + (1 - \lambda) [1 - \frac{(q-p)d(p)}{\|q-p\|}]), \quad 0 < \lambda < 1,$$

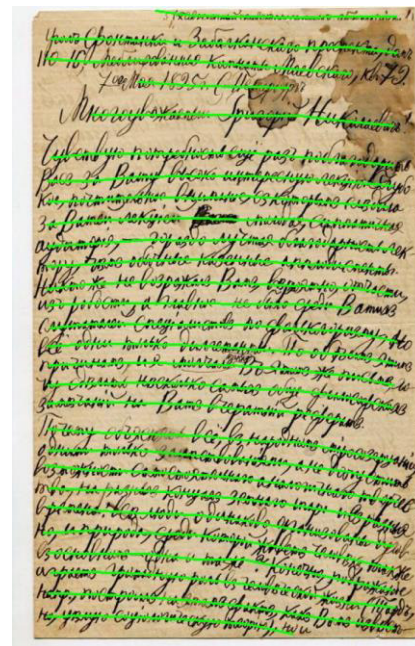
что одновременно минимизирует длину вставки и её отклонение от поля направлений  $d$ , тем самым восстанавливая непрерывность линии.

**Фильтр шумовых развилок.** Для вершины  $n$  с  $\deg(n) > 2$  оставляем два ребра с максимальным  $\frac{(q-n)d(n)}{\|q-n\|}$ , остальные удаляем. Эта фильтрация убирает ложные ответвления, оставляя ребра, наиболее согласованные с локальным направлением.

После обеих процедур каждая компонента превращается в единственную ломаную; упорядочив её по  $x$ , получаем гладкую осевую линию строки.



Граф после постобработки



Итоговая сегментация строк

Рисунок 5. Постобработка: зелёным – итоговые рёбра, синим – рёбра до фильтрации, красным – локальные направления

### Результаты экспериментов

Для тестирования алгоритма была использована коллекция фотографий архивных писем XIX века. Письма написаны разными авторами, имеют разную структуру бумаги, интенсивность чернил, размер написания, отсутствуют линии написания текста, что помогает оценить разработанный алгоритм на робастность к различным видам данных.

Из полного архива, включающего более 2600 изображений, вручную были размечены и отобраны 120 наиболее сложных для сегментации экземпляров. Эти изображения содержат типичные проблемные случаи: сильно искривлённые строки, малый межстрочный интервал, фоновый шум и нерегулярность написания. Такой выбор обеспечивает тщательную проверку надежности работы алгоритма на реальных задачах анализа исторических рукописей. Оценка качества сегментации проводилась по результатам сравнения сегментированных изображений с эталонной разметкой.

Введены следующие обозначения:

- **True Positives (TP)** – число эталонных строк, для которых алгоритм выдал хотя бы один фрагмент. Это отражает, что строка найдена, даже если разбита на несколько частей;



- **False Positives (FP)** – число предсказанных фрагментов, не соответствующих ни одной эталонной строке, а именно:
  - ложно выделенные фрагменты;
  - лишние части при дроблении строки на фрагменты;
  - объединённые (слипшиеся) несколько строк в один фрагмент, который уже не может быть отнесён к отдельной строке.

Таблица 1. Оценки результатов экспериментов

Всего строк	TP	FP	FN	Precision	Recall	$F_1$
1546	1482	181	64	0.89	0.96	0.92

- **False Negatives (FN)** – число эталонных строк, для которых алгоритм не выдал ни одного фрагмента, а именно:
  - полностью пропущенные строки;
  - строки «потерянные» в результате слияния нескольких эталонных строк в один фрагменты.

На основе этих значений определяются два ключевых показателя качества:

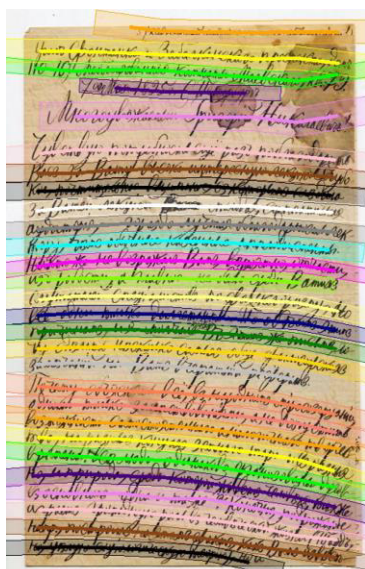
$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}.$$

Первая оценка показывает то, насколько предложенный метод вообще способен детектировать строки, а вторая – насколько точно он это делает. Также рассмотрена величина  $F_1$ -score, которая объединяет точность (Precision) и полноту (Recall):

$$F_1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

### Обсуждение результатов

Предложенный метод обладает высокой полнотой (*recall*) и точностью (*precision*) и, соответственно, величиной  $F_1$ -score. Значение  $F_1 = 0,92$  свидетельствует о высоком качестве сегментации, поскольку гармоническое среднее не позволяет одной из компонент доминировать над другой и обеспечивает сбалансированную оценку результатов при обработке случаев, имеющих шумовые артефакты, различную толщину строк и междустрочных интервалов.



Пример корректной сегментации (вариант 1)



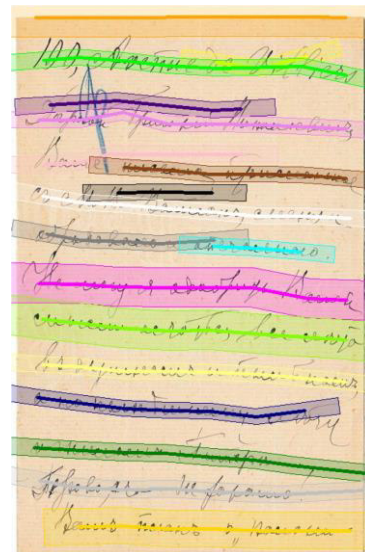
Пример корректной сегментации (вариант 2)

Рисунок 6. Все строки сегментированы корректно

Большинство ошибок, входящих в число *FP*, связано с тем, что метод корректно локализует строки, но внутри одной осевой линии компоненты оказываются разорванными, что создаёт избыточные фрагменты. Ошибки *FN* зачастую обусловлены объединением нескольких истинных строк в один сегмент в тех случаях, где межстрочный интервал сравнительно мал. При этом алгоритм допускает минимальное число полностью ложных выделений и пропусков.



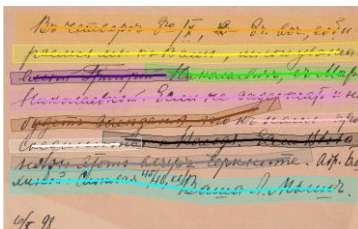
Слипшиеся фрагменты строк:  
серая линия объединяет две  
строки, вызывая  
дополнительные подразделения



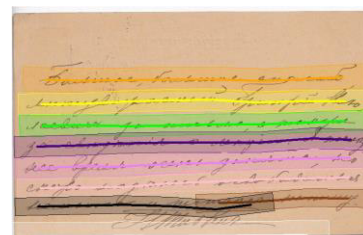
Ложные фрагменты:  
оранжевая, фиолетовая (2-я  
строка) и чёрная (4-я) линии

Рисунок 7. Ошибки False Positives

Кроме того, тестовый набор включал наиболее сложные страницы (с сильными изгибами строк, нерегулярным межстрочным интервалом и фоновыми шумами), тогда как на документах со стандартной структурой метод работает ещё более робастно.



Не выделена последняя строка,  
у предпоследней возможно  
слияние



Не выделена последняя строка

Рисунок 8. Ошибки False Negatives

Большинство ошибок появляется там, где почерк имеет крупный размер, и может быть связано с фиксированными размерами фреймов, относительно которых вычисляются локальные направления строк.

### Заключение

В работе предложен метод построчной сегментации рукописных документов, основывающийся на анализе локального распределения текстовых пикселей и построении дискретного векторного поля направлений. Для реализации метода был разработан программный комплекс, а также было произведено тестирование алгоритма на обширном архиве изображений исторических писем. Эксперименты подтвердили робастность метода к фоновому шуму, вариативности почерка и искривлённым строкам: достигнуты высокие значения полноты ( $\text{recall} = 0,96$ ) и точности

(precision = 0,89), а  $F_1$ -score = 0,92 свидетельствует о сбалансированном сочетании этих показателей – метод последовательно находит все релевантные строки и минимизирует число ложных срабатываний. Дополнительно показано, что алгоритм эффективно аппроксимирует и корректно сегментирует даже сильно изогнутые или наклонённые текстовые строки.

Таким образом, основные полученные результаты таковы:

1. Разработан метод построчной сегментации изображений рукописных документов на основе модели интегральных кривых векторного поля;
2. Реализован подход определения связей узловых точек, согласованных с векторным полем локальных направлений текста;
3. Предложена процедура восстановления интегральных кривых векторного поля как подграфов в триангуляции Делоне.

Основными направлениями дальнейшей работы являются преобразования всех имеющихся фиксированных порогов в адаптивные на основе структуры каждой страницы для более точного её восстановления, а также улучшение этапа постобработки для предотвращения наличия артефактов.

#### Источник финансирования

Разработанный алгоритм применялся при выполнении проекта Российского научного фонда «Культурное наследие России: интеллектуальный анализ и тематическое моделирование корпуса рукописных текстов» (грант РНФ № 22-68-00066).

#### Список литературы

1. R. Ptak, B. Zygadlo, O. Unold, Projection-based text line segmentation with a variable threshold, *International Journal of Applied Mathematics and Computer Science* 27 (2017) 195-206.
2. S. Arivazhagan, D. J. Hemanth, Statistical projection methods for text line segmentation, in: *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR)*, 2007, pp. 1029-1033.
3. K. Y. Wong, R. G. Casey, F. M. Wahl, Document analysis system, *IBM Journal of Research and Development* 26 (1982) 647-656.
4. N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos, N. Papamarkos, Segmentation of historical machine-printed documents using adaptive run length smoothing and skeleton segmentation paths, *Image and Vision Computing* 28 (2010) 590-604.
5. V. Papavassiliou, T. Stafylakis, V. Katsouros, G. Carayannis, Handwritten document image segmentation into text lines and words, *Pattern Recognition* 43 (2010) 369-377.
6. D. J. Kennard, W. A. Barrett, Separating lines of text in free-form handwritten historical documents, in: *Proceedings of the 2nd International Conference on Document Image Analysis for Libraries (DIAL'06)*, 2006, pp. 12-23.
7. G. Renton, C. Chatelain, S. Adam, C. Kermorvant, T. Paquet, Handwritten text line segmentation using fully convolutional network, in: *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, pp. 321-326.
8. X. Zhang, C. L. Tan, Text line segmentation for handwritten documents using constrained seam carving, in: *14th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, IEEE, Crete, Greece, 2014, pp. 98-103. doi:10.1109/ICFHR.2014.24.
9. G. Louloudis, B. Gatos, C. Halatsis, Text line detection in unconstrained handwritten documents using a block-based hough transform approach, in: *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR)*, IEEE Computer Society, Curitiba, Brazil, 2007, pp. 599-603. URL: [https://users.iit.demokritos.gr/~bgat/Icdar2007\\_TextLineDetection.pdf](https://users.iit.demokritos.gr/~bgat/Icdar2007_TextLineDetection.pdf).

## Моделирование, распознавание и анализ на основе компьютерного зрения движущихся в газе твердых частиц различных масштабов

И. А. Дорощенко, И. А. Знаменская

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

**Аннотация.** В работе представлен комплексный подход к моделированию, распознаванию и анализу твердых частиц различных размеров, движущихся в газовой среде, с использованием компьютерного зрения и методов машинного обучения. Исследование основано на лабораторных экспериментах, проведенных в ударной трубе при скоростях потока до 900 м/с. Визуализация движения частиц выполнено методом высокоскоростной теневой съемки. Разработаны алгоритмы обработки изображений для обнаружения и отслеживания частиц размерами от 1 мкм до 5 мм. Последовательность обработки включает перевод изображений в оттенки серого, пороговую фильтрацию, морфологические операции, выделение контуров и отслеживание центроидов. Кроме того, обучены модели глубокого обучения на основе архитектуры YOLO на размеченном наборе изображений частиц для обеспечения обнаружения и локализации отдельных частиц в реальном времени в сложных течениях.

Анализ также включал автоматическую оценку яркости частиц и интенсивности фона, что даёт дополнительный параметр для идентификации и классификации частиц. Количественные результаты получены путём реконструкции траекторий частиц на основе их трекинга, построения  $x-t$  диаграмм.

Предложенные методы демонстрируют высокую эффективность при обнаружении и анализе быстро движущихся твердых частиц в газовых потоках и применимы к широкому кругу задач, включая лабораторное моделирование многофазных потоков.

**Ключевые слова:** компьютерное зрение; глубокое обучение; детекция и трекинг объектов; сверточные нейросети; YOLO; сверхзвуковой поток; высокоскоростная теневая съемка

## Modeling, recognition, and analysis of solid particles of various scales moving in a gas using computer vision

I. A. Doroshchenko, I. A. Znamenskaya

Lomonosov Moscow State University, Moscow, Russia

**Abstract.** This work presents a comprehensive approach to modeling, recognition, and analysis of solid particles of various sizes moving in a gaseous medium using computer vision and machine learning techniques. The study is based on laboratory experiments conducted in a shock tube facility with flow velocities up to 900 m/s. Visualization of particle motion was carried out using high-speed shadowgraph imaging. Custom image processing algorithms were developed to detect and track particles with sizes ranging from 1  $\mu\text{m}$  to 5 mm. The processing pipeline includes grayscale conversion, adaptive thresholding, morphological operations, contour detection, and centroid tracking. In addition, deep learning models based on the YOLO architecture were trained on a labeled dataset of particle images to enable real-time detection and localization of individual particles within complex flow fields.

The analysis also included automatic estimation of particle brightness and background intensity, which provide an additional metric for particle identification and classification. Quantitative results were obtained through tracking-based reconstruction of particle trajectories,  $x-t$  diagrams, and velocity-time profiles.

The proposed methods demonstrate high efficiency in the detection and analysis of fast-moving solid particles in gas flows and are applicable to a range of problems, including laboratory modeling of multiphase flows.

**Keywords:** computer vision; deep learning; object detection and tracking; convolutional neural networks; YOLO; supersonic flow; high-speed shadowgraphy

### Введение

Отслеживание движущихся твёрдых частиц имеет критическое значение во многих областях науки и техники. В области ситуационной осведомлённости в космосе возможность мониторинга спутников и орбитального мусора необходима для предотвращения столкновений и обеспечения устойчивости космических операций. В работе [1] был предложен простой алгоритм компьютерного зрения для автоматического обнаружения траекторий движения спутников на основе преобразования Хафа. В работе [2] авторы разработали систему обнаружения объектов на основе сверточной нейросети YOLOv8 для отслеживания космического мусора на оптических изображениях, обеспечив значительное повышение точности по сравнению с методами классического компьютерного зрения.

Помимо космических приложений, системы на основе машинного обучения применяются для отслеживания частиц в многофазных потоках [3]. Для обучения моделей глубокого обучения для задач детекции объектов требуется разметка данных. Обычно это делается либо с использованием реальных экспериментальных или натурных данных, либо путём генерации синтетических наборов данных с использованием средств моделирования [4]. Однако модели, обученные исключительно на синтетических данных, часто сталкиваются с проблемами обобщения, поскольку такие наборы данных упрощают взаимодействие среды и частиц и условия освещения. С другой стороны, натурные данные имеют ограничения, такие как неконтролируемое освещение, низкая частота кадров и несогласованность разметки.

Таким образом, лабораторные исследования представляют собой «золотую середину», совмещающая физический реализм и контролируемость условий проведения эксперимента. Ударные трубы, способные создавать сверхзвуковые потоки газа в сочетании с такими методами визуализации течений, как высокоскоростная теневая съемка [5], PIV (англ. Particle Image Velocimetry), предоставляют визуальные данные в высоком разрешении, хорошо подходящими для их обработки методами компьютерного зрения [6, 7].

В последние годы методы компьютерного зрения и машинного обучения всё активнее используются в сочетании с высокоскоростной визуализацией для автоматизации распознавания динамических объектов в потоках. Например, обнаружение пузырьков в жидкой среде было реализовано с использованием свёрточных нейросетей, метода выделения границ Канни и алгоритма ORB [8]. Для идентификации пересечений ударных волн в газодинамических экспериментах применялся метод детекции углов Харриса [9]. В одном из исследований [10] классификация режимов течения для оценки скорости потока осуществлялась путём анализа положений ударных волн с использованием машины опорных векторов (англ. Support Vector Machine, SVM). В другой работе [11] с помощью сверточной нейросети Mask R-CNN на шпирен-изображениях распознавалась зона детонации топливной смеси в двигателе внутреннего сгорания. Волновые структуры на границе газ–жидкость успешно регистрировались с применением нейросети YOLOv7, при этом обучающая выборка формировалась на основе объединения нескольких теневых изображений для повышения достоверности распознавания [12]. Активно используются различные алгоритмы отслеживания точечных частиц в потоках [13].

В наших предыдущих работах [14, 15] были разработаны методы компьютерного зрения и глубокого обучения для детекции различных структур на кадрах высокоскоростной теневой съемки течений. Данная работа дополняет ранее полученные результаты. Реализованы автоматизированное отслеживание, анализ динамики непрозрачных частиц в потоках и измерение их параметров, яркости.

### **Экспериментальная установка**

Экспериментальная установка представляет собой однодиафрагменную ударную трубу, позволяющую создавать ударные волны с числом Маха  $M=4$ . Изучались течения со скоростями спутного потока до 900 м/с. Камера высокого давления заполнялась гелием, камера низкого давления – воздухом. Таким образом, воздух использовался в качестве рабочего газа. Ударная труба прямоугольного сечения размером 48 мм × 24 мм. В камеру низкого давления встроена тестовая секция, боковые стенки которой выполнены из кварцевого стекла для оптического доступа. Синхронизация течения и высокоскоростной камеры осуществлялась с помощью сигналов с пьезодатчиков, встроенных в камеру низкого давления.

Производился засев потока в ударной трубе непрозрачными твердыми частицами различных размеров, которые использовались для отработки методов компьютерного зрения.

Для высокоскоростной теневой съемки в качестве подсветки использовался стационарный лазер, длина волны 532 нм. Системой линз создавался параллельный пучок света, который направлялся в тестовую секцию, а затем изображение собиралось на матрице высокоскоростной камеры с помощью объектива. Использовалась высокоскоростная камера Photron Fastcam SA5. Скорость съемки составляла до 150 000 кадров в секунду, экспозиция – 1 мкс.



## Методы

### 1. Алгоритмы компьютерного зрения на основе выделения контуров

Для отслеживания и трекинга твердых частиц на последовательностях теневых изображений был разработан алгоритм компьютерного зрения на основе детекции контуров и трекинга центроидов объекта. Основные шаги проиллюстрированы на рисунке 1. Алгоритм написан на языке программирования Python. Пайплайн обработки изображений включает в себя преобразование каждого кадра видео в монохромный, размытие изображения с помощью функции Гаусса. Далее применяется динамическая пороговая фильтрация. Пороговое значение яркости рассчитывается на основе средней яркости кадра, позволяя алгоритму подстроиться под освещение. Далее изображение переводится в бинарное и цвета инвертируются. Применяется операция морфологического открытия. Морфологическое открытие полезно для удаления маленьких объектов на изображении при сохранении формы и размера больших объектов. Далее производится детекция контуров с помощью метода `findContours()` из библиотеки OpenCV и их фильтрация по длине периметра для удаления слишком мелких объектов и шумов.

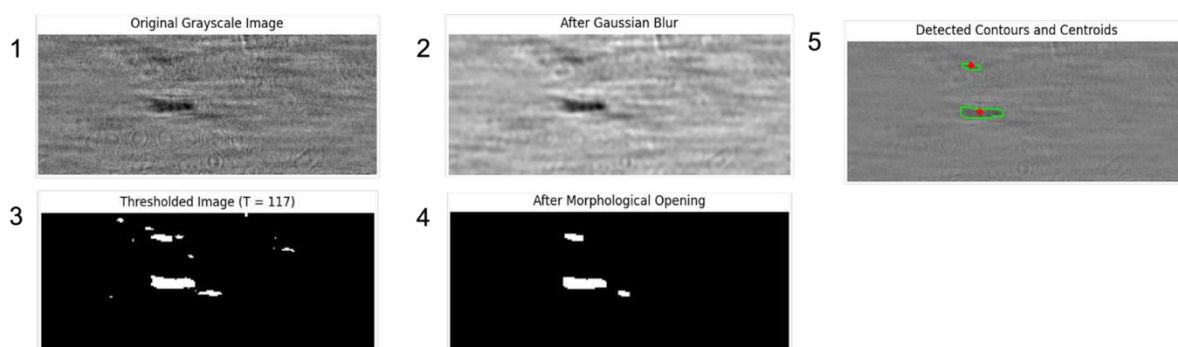


Рисунок 1. Пошаговая визуализация алгоритма детекции движущихся частиц в потоке на примере одного теневого кадра

Для каждого валидного контура вычисляется положение геометрического центроида, координаты которого принимаются за положение частицы. Для трекинга траекторий частиц между последовательными кадрами видео использовалась библиотека `Trackpy` для Python, которая идентифицировала одну и ту же частицу между кадрами с помощью оценки расстояний и смещений. Каждой частице присваивался уникальный идентификатор, а кратковременные исчезновения частиц из поля зрения митигировались с помощью параметра памяти.

Координаты центроидов конвертировались в пиксели, строились x-t диаграммы движения распознанных частиц. Разработанная программа сохраняла новую версию видеозаписи с выделенными контурами и отмеченными центроидами для распознанных частиц. Также на каждом кадре указывалось его время. Примеры полученных в результате теневых кадров приведены на рисунке 2.

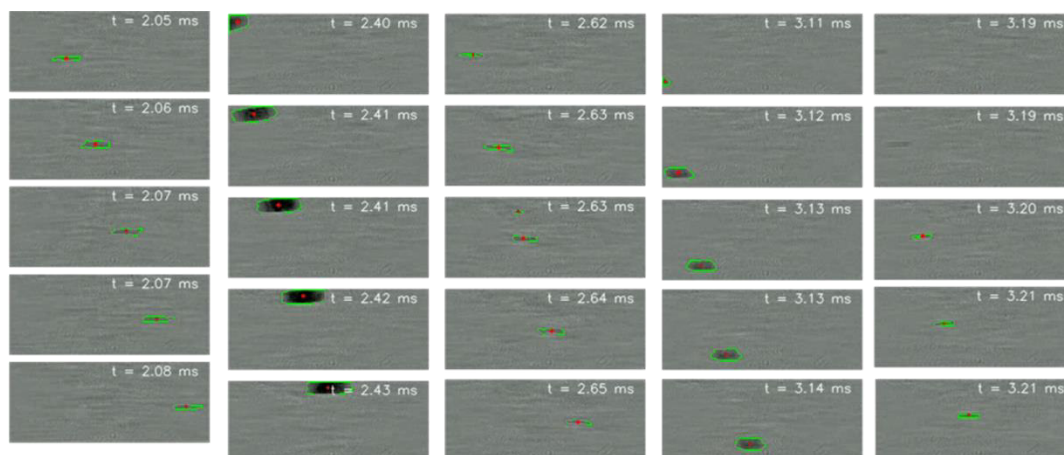


Рисунок 2. Последовательность отслеживаемых на теневых кадрах частиц. На рисунке отмечены контуры распознанных объектов и положения центроидов

## 2. Модель глубокого обучения

Распознавание частиц-трассеров в потоке также осуществлялось с помощью более продвинутого метода на основе глубокого обучения – сверточной нейросети на архитектуре YOLO (англ. (You Only Look Once)). Была обучена модель машинного обучения, позволяющая распознавать различные структуры течений на изображениях. В данной работе речь пойдет только об одном типе объектов – частице в потоке.

Модель YOLOv11 была выбрана в связи с сочетанием высокой точности и скорости работы. Обучение производилось на размеченном наборе данных, содержащем теневые кадры различных течений в ударной трубе. Размечались следующие классы объектов: головная ударная волна (bow shock), ударная волна (shock), контактная поверхность (plume) и частица-трассер (particle). В настоящей работе нас интересовало только распознавание объектов, относящихся к классу частиц-трассеров. Размеченный набор данных дополнялся аугментациями – копиями исходных изображений, полученными с помощью применения таких операций как добавление искусственного шума, функции размытия Гаусса, изменения яркости, контрастности и других параметров. Полученный набор данных для обучения содержал 3235 изображений. Подробная информация о наборе данных и обучении модели опубликована в одной из прошлых работ [15].

Матрица ошибок модели, характеризующая ее точность, приведена на рисунке 3. Точность распознавания частиц-трассеров достигла 92 %. Средняя точность (англ. Average Precision, AP) превысила 90 % при пороговом значении IoU (англ. Intersection over Union) равном 0.5.

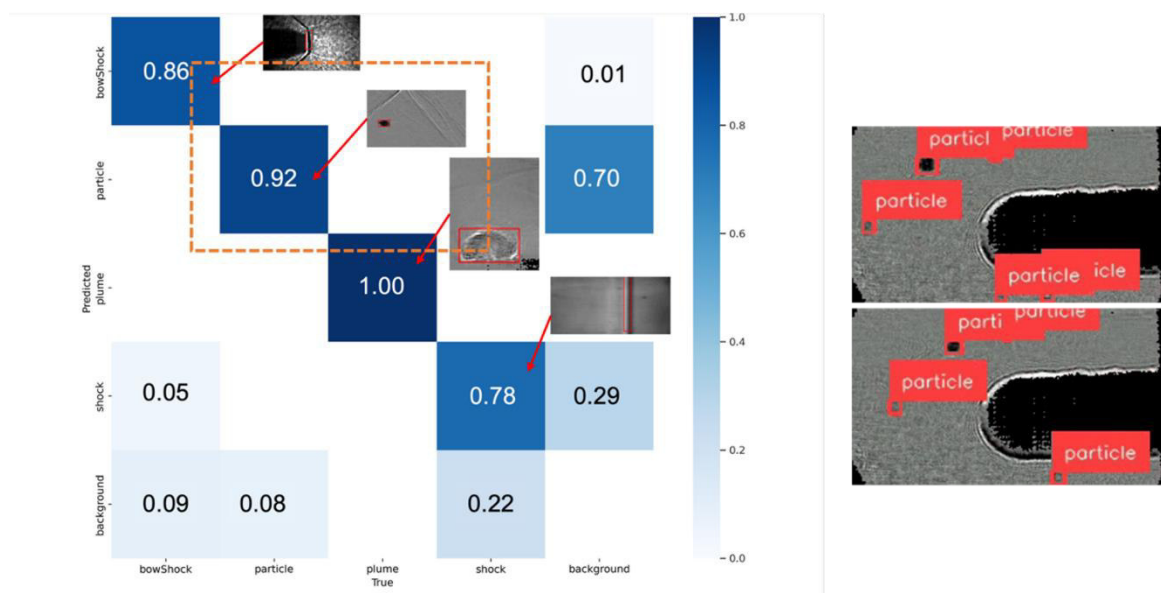


Рисунок 3. Матрица ошибок обученной модели YOLO (слева), примеры детекции частиц на теневых кадрах в присутствии аэродинамического тела в потоке (справа)

## Обсуждение результатов

С помощью алгоритма компьютерного зрения на основе выделения контуров проведен анализ геометрических параметров частиц-трассеров в потоке и их яркость по сравнению с фоновой. На рисунке 4 приведено распределение измеренной ширины и высоты частиц для одного из экспериментов. Большинство объектов имеет ширину от 2 до 6 мм и высоту от 1 до 3 мм.

На рисунке 5 приведено распределение усредненной по каждой частице яркости, которая оказалась значительно меньше фоновой яркости. Фоновая яркость отмечена вертикальной линией на рисунке. Средняя яркость частиц составила 80–115 единиц (на шкале 0–255 единиц), тогда как средняя яркость фона составила 125 единиц. Данное распределение показывает, что частицы являются достаточно контрастными для их точного распознавания методами компьютерного зрения и позволяют оценить качество настройки теневого метода в эксперименте.

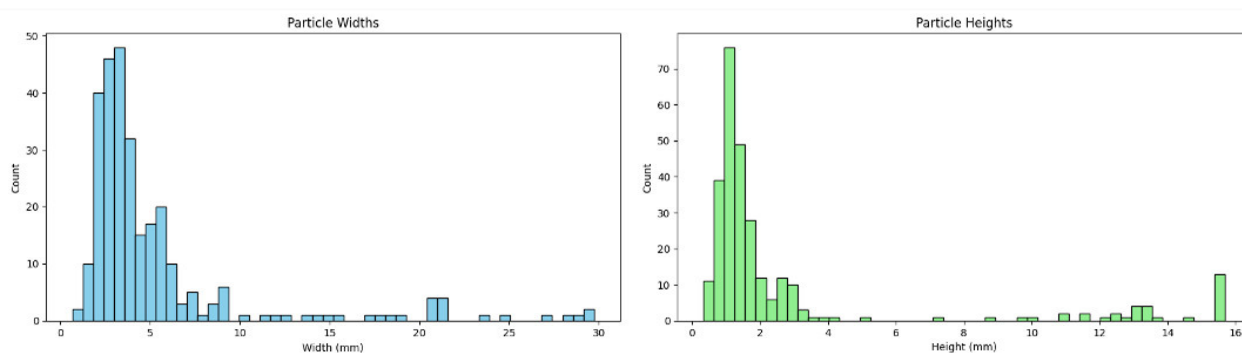


Рисунок 4. Распределение параметров распознанных частиц-трассеров в потоке: *слева* – горизонтальные размеры, *справа* – вертикальные размеры

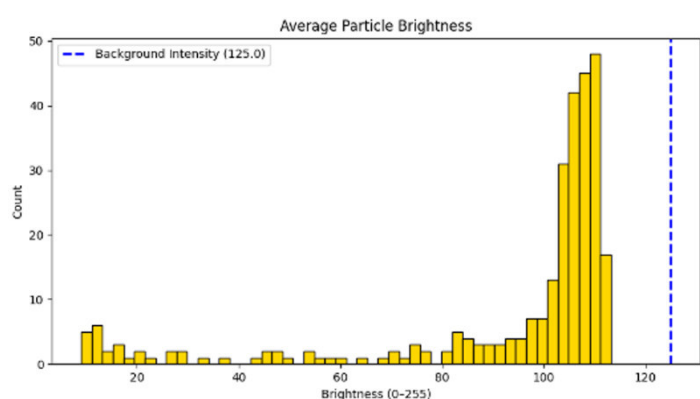


Рисунок 5. Распределение усредненной яркости частиц, сравнение со средней яркостью фона

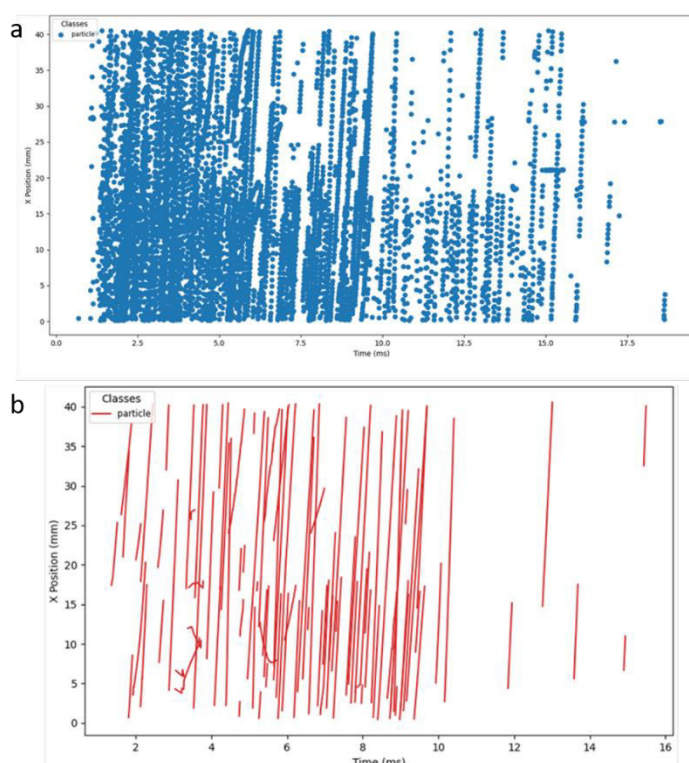


Рисунок 6. X-t диаграммы движения распознанных частиц для одного из экспериментов: а – координаты частиц; б – непрерывные траектории, полученные в режиме трекинга

Модель глубокого обучения использовалась для автоматического распознавания частиц и траекторий их движения. Соответствующие x-t диаграммы приведены на рисунке 6. Полученные

диаграммы позволяют оценить концентрацию частиц в области наблюдения для различных промежутков времени, а также измерить скорости по наклону  $x$ - $t$  диаграмм для каждой частицы. Измеренные скорости частиц составили от 150 м/с до 300 м/с в зависимости от стадии течения и размера частиц. Чем меньше частица, тем ее скорость ближе к реальной скорости течения. Возможность автоматического анализа больших данных видеозаписей позволяет значительно ускорить получение новой физической информации о течениях.

### **Заключение**

Разработан и протестирован метод детекции и анализа твердых частиц в течении газа, визуализируемого тeneвым методом. Метод включает в себя высокоскоростную тeneвую съемку течения, применение методов как классического компьютерного зрения, так и современной модели глубокого обучения. Эксперименты проводились в ударной трубе, позволяющей создавать ударные волны с числами Маха от 2.5 до 3.5 и скоростью потока до 900 м/с.

Используя алгоритм компьютерного зрения на основе выделения контуров, были обнаружены частицы размером от 0.1 до 30 мм в диаметре. Были извлечены такие параметры, как распределение высоты и ширины частиц, их средней яркости. Измеренная яркость сравнивалась с усредненной яркостью фона. Наиболее часто встречающаяся ширина частиц составила от 2 до 6 мм, высота от 1 до 3 мм, яркость – от 80 до 115 единиц (на шкале от 0 до 255 единиц). При этом средняя фоновая яркость составила 125 единиц. Разница между средней яркостью частиц и фона позволяет оценить качество настройки оптики в тeneвом методе и эффективность распознавания частиц методами компьютерного зрения.

Обучена модель глубокого обучения для детекции частиц на основе архитектуры YOLOv11. Точность детекции составила 92 %, средняя точность (англ. Average Precision, AP) превысила 90 % при пороговом значении IoU равном 0.5. С помощью разработанного кода автоматически строились  $x$ - $t$  диаграммы движения распознанных объектов для каждой видеозаписи, содержащей тысячи кадров и точек данных. Видеозаписи включали в себя 20 мс течения в ударной трубе, регистрируемого со скоростью 150 000 кадров в секунду. Измерялись скорости частиц, они составили 150–300 м/с в зависимости от размера частицы и стадии течения.

Таким образом, сочетание высокоскоростной тeneвой съемки и методов компьютерного зрения позволило значительно ускорить обработку экспериментальных данных течений и получить новую физическую информацию об исследуемых потоках. Метод подходит не только для течений в ударных трубах, но и для любых течений, визуализируемых тeneвым или шпирен методом, в том числе в аэродинамических трубах, в натурных экспериментах.

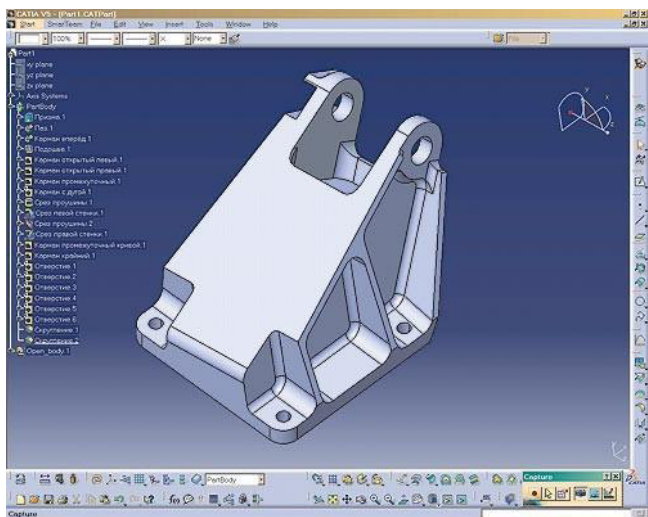
### **Источник финансирования**

Исследование проводилось в рамках государственного задания Московского государственного университета имени М.В. Ломоносова.

### **Список литературы**

1. Prototype of Automatic Satellite Streak Detection, Identification and Initial Orbit Determination Pipeline from Optical Observation / M. D. Danarianto, A. M. Maharani, B. M. Falah, F. Rohmah // *Journal of Physics: Conference Series*. 2022. Vol. 2214, no. 1. P. 1-6.
2. Enhanced YOLOv8-Based Method for Space Debris Detection Using Cross-Scale Feature Fusion / Y. Guo, X. Yin, Y. Xiao, et al. // *Discover Applied Sciences*. 2025. Vol. 7, no. 2. P. 95.
3. Ghedia N., Vithalani C., Kothari A. M., Thanki R. M. Moving Objects Detection Using Machine Learning // Cham: Springer International Publishing, 2022. P. 85.
4. Detection & Identification of On-Orbit Objects Using Machine Learning, Detection & Identification of On-Orbit Objects Using Machine Learning / M. D. Perez, M. A. Musallam, A. Garcia, et al. // 8th European Conference on Space Debris, ESA Space Debris Office. 2021. P. 1-10.
5. Settles G. S., Hargather M. J. A Review of Recent Developments in Schlieren and Shadowgraph Techniques // *Measurement Science and Technology*. 2017. Vol. 28, no. 4. P. 1-26.
6. Supersonic Combustion Flow Visualization at Hypersonic Flow / T. V. C. Marcos, D. R. Pinto, G. S. Moura, et al. // 28th International Symposium on Shock Waves. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. P. 1041–47.

7. Znamenskaya I. A., Doroshchenko I. A. Edge detection and machine learning for automatic flow structures detection and tracking on schlieren and shadowgraph images // *Journal of Flow Visualization and Image Processing*. 2021. Vol. 28, no. 4. P. 1–26
8. Bubble Trajectory Tracking Based on ORB / S. Wang, S. Lu, J. Liu, et al. // *Algorithm, Image Analysis & Stereology*. 2023. Vol. 42, no. 1. P. 17–23
9. Smith N. T., Lewis M. J., Chellappa R. Detection, Localization, and Tracking of Shock Contour Salient Points in Schlieren Sequences // *AIAA Journal*. 2014. Vol. 52, no. 6. P. 1–16.
10. Manshadi M. D., Vahdat-Nejad H., Kazemi-Esfeh M. Speed Detection in Wind-Tunnels by Processing Schlieren Images // *International Journal of Engineering*. 2016. Vol. 29, no. 7. P. 1–6.
11. Detecting the Flame Front Evolution in Spark-Ignition Engine under Lean Condition Using the Mask R-CNN Approach / L. Petrucci, F. Ricci, R. Martinelli, F. Mariani // *Vehicles*. 2022. Vol. 4, no. 4. P. 1–18.
12. Fast Detection of Bag-Breakups in Pulsating and Steady Airflow Using Video Analysis and Deep Learning / D. Morita, B. Raytchev, A. Elhanashi, et al. // *Journal of Real-Time Image Processing*. 2023. Vol. 20, no. 6. P. 114.
13. Xuezhong X., Ali S. M. Visual tracking with methodologies – a literature survey // *Journal of Flow Visualization and Image Processing*. 2016. Vol. 23, no. 3–4. P. 275–321.
14. Simulation of Supersonic Jet Flow Past a Blunt Body in a Laboratory Experiment Using Computer Vision / I. Doroshchenko, I. Znamenskaya, N. Sysoev, A. Lutsii // *Acta Astronautica*. 2024. Vol. 215. P. 69–78.
15. Doroshchenko I. A. Analysis of the experimental flow shadowgraph images by computer vision methods // *Numerical Methods and Programming (Vychislitel'nye Metody i Programirovanie)*. 2023. No. 2. P. 231–42.



Секция 4

# ***СИСТЕМЫ АВТОМАТИЗАЦИИ ПРОЕКТИРОВАНИЯ – РАЗРАБОТКА, ВНЕДРЕНИЕ И ПРИОРИТЕТЫ РАЗВИТИЯ***

---

DEVELOPMENT, IMPLEMENTATION  
AND ADVANCEMENT OF COMPUTER-AIDED  
DESIGN SYSTEMS



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия

## Автоматизация движения мобильного робота за стеной Turtlebro с помощью ПИД-регулятора ROS

Зеар Аунг<sup>1</sup>, Шидловский Станислав Викторович<sup>2</sup>, Ну Ну Вар<sup>3</sup>

<sup>1,2</sup>Томский государственный университет, Томск, Россия, zayaraung53@gmail.com, shidlovskiysv@mail.ru

<sup>3</sup>Университет корпоративного управления, Сагаинг (UCMS)  
Сагаинг, Мьянма, dr.nuwar@gmail.com

**Аннотация.** В статье представлена реализация траектории алгоритмы отслеживания и обхода препятствий в Turtlebro3 Burger, использующие операционную систему ROS 2. Для отслеживания лемнискатных и круговых траекторий используется пропорциональное управление с компенсацией. Кроме того, для обхода препятствий в динамичных условиях используется алгоритм потенциального искусственного поля (APF). При таком подходе робот моделируется как частица, погруженная в потенциальное поле, которая использует притягивающие функции для достижения пункта назначения и отталкивающие функции для обхода препятствий. Для проверки работоспособности алгоритмов были проведены имитационные тесты в программном обеспечении Gazebo и экспериментальные испытания. Результаты показали, что пройденный маршрут соответствует требованиям задачи, подтверждая эффективность предложенного подхода к планированию маршрута.

**Ключевые слова:** ROS 2, искусственное потенциальное поле, Turtlebro, планирование пути, объезд препятствий.

## Automated wall following of Turtlebro using PID control ROS

Zayar Aung<sup>1</sup>, Stanislav Viktorovich Shidlovsky<sup>2</sup>, Nu Nu War<sup>3</sup>

<sup>1,2</sup>Tomsk State University, Tomsk, Russian Federation, zayaraung53@gmail.com, shidlovskiysv@mail.ru

<sup>3</sup>University of Co-operative and Management of Sagaing (UCMS), Sagaing, Myanmar, dr.nuwar@gmail.com

**Abstract.** The paper presents the implementation of trajectory tracking and obstacle avoidance algorithms on a Turtlebro Burger using the ROS 2 operating system. A proportional control with compensation is used for lemniscate and circle trajectory tracking. In addition, a potential artificial field (APF) algorithm is incorporated for obstacle avoidance in dynamic environments. In this approach, the robot is modeled as a particle immersed in a potential field that uses attractive functions to reach a destination point and repulsive functions to go around obstacles. Simulated tests in Gazebo software and experimental tests are performed to validate the performance of the algorithms. The results showed that the route followed meets the requirements of the problem, confirming that the proposed route planning approach is efficient.

**Keywords:** ROS 2, Artificial potential field, Turtlebro, path planning, obstacle avoidance.

### Введение

Одной из главных проблем мобильной робототехники является автономная навигация, при которой робот должен двигаться к цели и одновременно избегать столкновений с препятствиями окружающей среды [1]. Эта задача известна как планирование и отслеживание пути, она учитывает распределение препятствий, присутствующих в окружающей среде статически и динамически, чтобы рассчитать непрерывный маршрут и достичь целевой точки из исходного положения с эффективным и быстрым обходом препятствий [2].

Алгоритмы планирования траектории в мобильных роботах должны соответствовать требованиям безопасной маршрутизации, скорости вычислений и надежности. Алгоритм искусственного потенциального поля (APF) – это эвристический подход, который обеспечивает простое и эффективное планирование траектории, что может гарантировать обработку в режиме реального времени. Этот подход использует потенциальные поля отталкивания вокруг препятствий, которые отталкивают робота, и потенциальное поле притяжения в целевой точке, которое притягивает робота к ней [3]. Сумма двух потенциалов заставляет робота испытывать обобщенную силу, равную отрицательному градиенту общего потенциала. Эта сила поддерживает направление и величину, которые перемещают робота в поисках точки с наименьшим потреблением энергии. Эффективная управляемость этого алгоритма в режиме реального времени позволяет обрабатывать однородные траектории, поэтому он широко используется в автономной навигации. Например, способ управления формированием двух



мобильных роботов, обеспечивающий эффективное преодоление препятствий с использованием алгоритма APF, представлен в [4]. Авторы работы [5] используют алгоритм APF, чтобы заставить робота достичь желаемого положения, избегая препятствий, добавляя локальные аттракторы, чтобы избежать проблемы локальных минимумов. Исходя из этого, в работе [6] предложен усовершенствованный метод APF для обнаружения и обхода фиксированных препятствий, который позволяет роботу достигать цели по оптимальной траектории, решая проблему, связанную с тем, что робот может попасть в ловушку локального минимума. Новое применение этой методики представлено в работе [7], в которой разрабатывается динамическое планирование траектории для отслеживания движущихся целей с использованием алгоритма APF, применяемого к беспилотному летательному аппарату. Недавно в [8] был представлен алгоритм навигации для мобильных роботов на базе операционной системы ROS, который позволяет осуществлять автономную навигацию в любой среде, используя Turtlebot Burger в качестве испытательного стенда, выполняя обработку данных с датчиков, установленных на роботе, для обработки алгоритмом APF и планирования маршрута на основе сенсорной информации.

В этой статье предлагается пропорциональный контроллер для отслеживания траектории движения и обхода препятствий мобильными роботами с дифференциальным приводом. Внутренняя одометрия генерируется с использованием алгоритма искусственного потенциального поля, обработки данных в режиме реального времени, лидара и датчиков абсолютного кодирования. Зависимость от удаленного компьютера устранена благодаря реализации алгоритмов навигации на Raspberry pi4 с ROS 2. Все эксперименты проводились на роботе Turtlebot Burger Robot. Результаты позволяют следовать установленным маршрутом с минимальными ошибками, что демонстрирует эффективность метода.

Статья организована следующим образом: сначала представлен анализ кинематической модели дифференциального робота с учетом смещенной контрольной точки, а также модели одометрии, основанной на данных датчика, для оценки его относительного положения в плоскости. В конце этого раздела разработан алгоритм отслеживания траектории с использованием пропорционального закона управления с компенсацией. Затем алгоритм обхода препятствий разработан в соответствии с подходом APF для определения наиболее оптимального пути достижения роботом цели. Далее, наконец, сгенерированные алгоритмы тестируются с помощью моделирования и эксперимента.

## Предварительные приготовления

### А. Кинематическая модель со смещенной точкой

Кинематическая модель описывала скорости и взаимосвязь между управляющими входами и поведением системы, учитывая представление системы в пространстве состояний. Предполагается, что транспортное средство движется в горизонтальной плоскости (X,Y) относительно смещенной точки P(x,y), как показано на рисунке 1 [9]. Параметры приведены в таблице 1. Рассматриваются две системы отсчета: локальная система {P}, которая размещается в контрольной точке и перемещается вместе с роботом; глобальная система {G} представляет собой фиксированную систему координат, в которой робот описывает траекторию. Позиция робота может быть вычислена следующим образом:

$$P = \begin{pmatrix} x_r(t) + a \cos(\theta_r(t)) \\ y_r(t) + a \sin(\theta_r(t)) \end{pmatrix}. \quad (1)$$

Исходя из этих положений, уравнения линейной и угловой скоростей выглядят следующим образом:

$$\dot{P} = \begin{pmatrix} v(t) \cos(\theta_r(t)) - a\omega \sin(\theta_r(t)) \\ v(t) \sin(\theta_r(t)) + a\omega \cos(\theta_r(t)) \end{pmatrix}. \quad (2)$$

Перепишав (2) в матричной форме, кинематическая модель в перемещенной точке P задаем формулой

$$\begin{bmatrix} \dot{P}_x \\ \dot{P}_y \end{bmatrix} = \begin{bmatrix} \cos(\theta_r) & -a \sin(\theta_r) \\ \sin(\theta_r) & a \cos(\theta_r) \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix} = B(\theta_r)U. \quad (3)$$

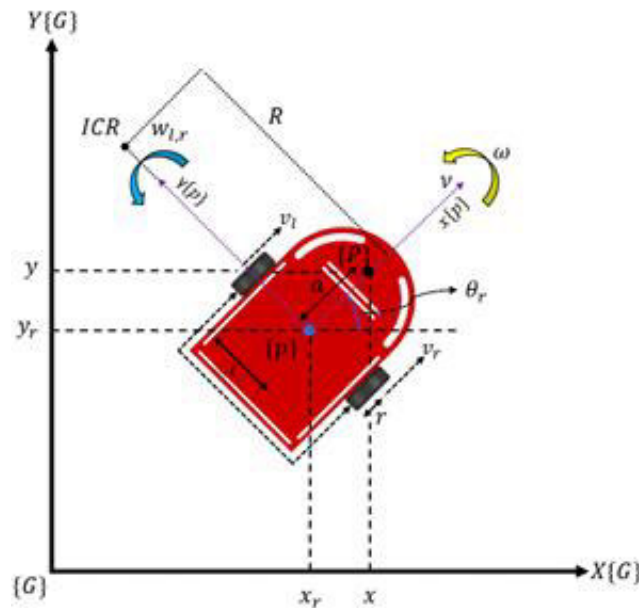


Рис. 1. DCL со смещенной точкой

Таблица 1. Номенклатура для дифференциального анализа роботов

Номенклатура
G – Глобальная справочная система
P – Локальная справочная система
P – Центральная точка робота
ICR – Мгновенный радиус поворота
R – Расстояние от центральной точки до ICR
$x_r, y_r$ – Положение робота в системе отсчета G
$\theta_r$ – Ориентация робота относительно оси $X_G$
$x, y$ – Положение смещенной контрольной точки
$a$ – Расстояние от центра робота до смещенной точки
$v, v_r, v_l$ – Линейные скорости робота
$\omega$ – Угловая скорость робота
$l$ – Расстояние от центра робота до колес
$r$ – Радиус колеса

## В. Одометрия

Turtlebot3burger оснащен динамическим микшером XL430-W250-Tm с абсолютным датчиком, который необходим для расчета одометрии. Расстояние, пройденное роботом в плоскости, определяется импульсами или отсчетами, генерируемыми датчиками, общим количеством импульсов за оборот (PPR) и диаметром колес.

**1) Пройденное расстояние.** Расстояние, пройденное каждым из колес  $d_i$ , с учетом информации от датчиков определяется по формуле

$$d_l = \frac{(S_l t - S_{l(t-1)})}{q} C = \frac{\Delta S_l}{q} C, \quad (4)$$

$$d_r = \frac{(S_r t - S_{r(t-1)})}{q} C = \frac{\Delta S_r}{q} C. \quad (5)$$

**2) Одометрическая модель.** На рисунке 2а показана диаграмма, соответствующая перемещению и ориентации робота относительно контрольной точки. В таблице 2 используются соответствующие параметры.

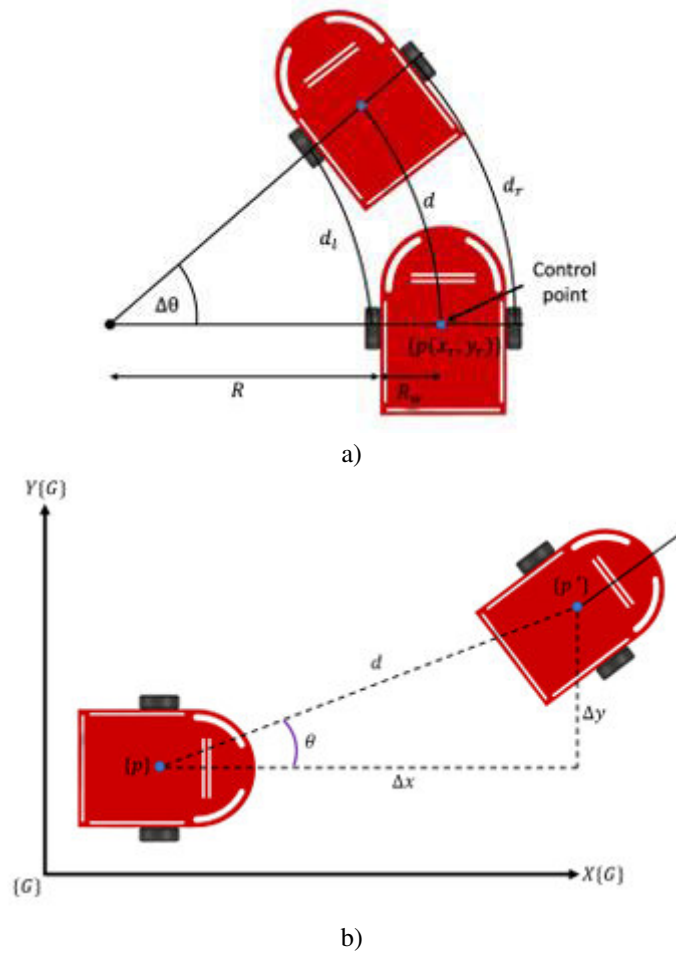


Рис. 2. Диаграммы перемещения и вращения робота

Как видно из рисунка 2а, где приведены формулы для вычисления дуги окружности,  $d_l$ ,  $d_r$  и  $d$  определяются с помощью следующих уравнений:

$$d_l = R \Delta \theta; \quad (6)$$

$$d_r = (R + 2R_\omega) \Delta \theta; \quad (7)$$

$$d = (R + R_\omega) \Delta \theta, \quad (8)$$

где  $d_l$  и  $d_r$  – известные значения, поскольку они могут быть определены с помощью уравнений (4) и (5). Тогда  $\theta$  и  $d$  – неизвестные значения, соответствующие изменению ориентации и общему пройденному расстоянию. Полное смещение и изменение ориентации определяется алгебраическими манипуляциями с уравнениями (6), (7) и (8) следующим образом:

$$\Delta \theta = \frac{d_r - R \Delta \theta}{2R_\omega} = \frac{d_r - d_l}{2R_\omega}, \quad (9)$$

$$d = (R + R_\omega) \Delta \theta = d_l + \frac{d_r - d_l}{2} = \frac{d_l + d_r}{2}. \quad (10)$$

Таблица 2. Номенклатура, используемая для расчета с помощью термометра

Номенклатура
$d_l, d_r$ – Пройденные расстояния на каждом колесе
D – Расстояние, пройденное от центральной точки
P – Центральная точка робота
$\Delta \theta$ – Увеличенный угол поворота
R – Расстояние от центральной точки до центра левого края
$R_\omega$ – Расстояние от центральной точки до центра ободьев
$\Delta x, \Delta y$ – Составляющие пройденного расстояния

С помощью уравнения (10) вычисляется изменение глобальных координат, как показано на рисунке 2b, где расстояние  $d$  разложено на составляющие  $x$  и представлено в виде приращений, поскольку расчет производится для каждого момента времени, например,

$$\begin{aligned}\Delta x(t) &= d \cos(\theta_r(t)); \\ \Delta y(t) &= d \sin(\theta_r(t)).\end{aligned}\quad (11)$$

**3) Одноплатный компьютер (SBC).** Для выполнения расчетов одометрии на ROS 2 необходимо понимать взаимодействие между Raspberry Pi 4 (RP4) и компонентами, составляющими TurtleBot в ROS.RP4 работает под управлением операционной системы ROS2 Foxy Fitzroy. Узлы связи, показанные на рисунке 3, инициализируют узел, отвечающий за взаимодействие компонентов робота. Эта сеть обеспечивает поток данных через темы и сообщения.

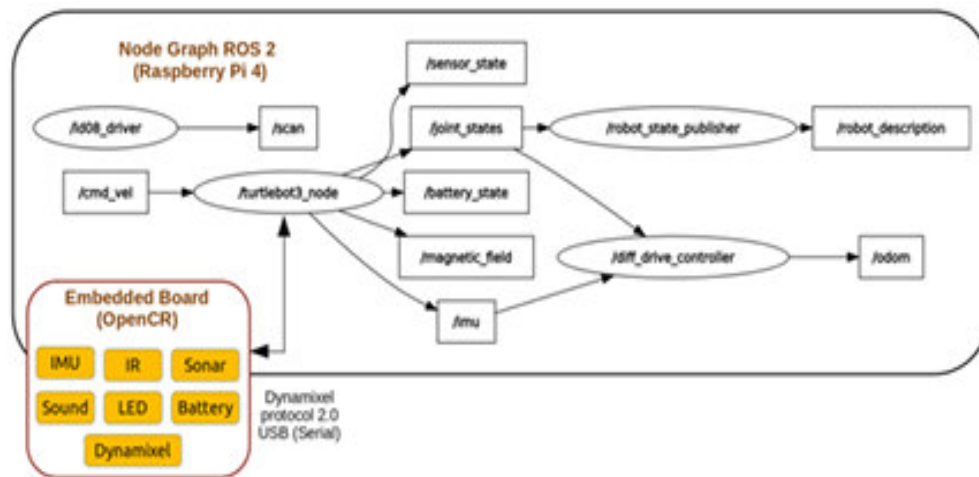


Рис. 3. Узловая сеть и поток данных

На рисунке 3 овалы обозначают узлы, а прямоугольники – темы. Стрелки указывают на передачу данных. Начиная с узла, направление стрелки указывает на то, что узел публикует информацию по одной или нескольким темам. И наоборот, когда стрелка указывает на узел, это означает, что узел подписан на одну или несколько тем. Ниже мы опишем каждую тему, которая содержит конкретную информацию о компоненте робота.

**/состояние датчика:** в этом разделе представлена информация от датчиков в виде счетчика и проверяется, включен ли крутящий момент на двигателях.

**/состояние батареи:** в этом разделе представлена вся информация, относящаяся к батарее, такая как напряжение, процент заряда и напряжение в каждом элементе.

**/imu:** содержит информацию о линейном и угловом ускорении, относящемся к локальной системе отсчета.

**/магнитное поле:** обеспечивает ориентацию робота и силу магнитного поля в навигационной среде.

**/состояние соединения:** предоставляет информацию об угловом положении каждого колеса и его угловых скоростях.

**/сканирование:** содержит информацию о расстоянии до объектов, обнаруживаемых вокруг робота каждым лучом лидарного датчика.

**/odom:** в этом разделе указывается положение относительно исходной точки, заданное при инициализации anode.

**/cmdvel:** записывает команды линейной и угловой скорости, отправляемые роботу в качестве управляющих воздействий, для расчета одометрии.

**/состояние датчика:** раздел используется для отображения количества тиков, генерируемых каждым двигателем при движении робота, с использованием сообщений левого и правого датчиков.

### Алгоритм отслеживания траектории

Учитывая, что  $B(\theta_r)$  не является сингулярным, разрабатывается контроллер обратной связи по состоянию, такой как

$$U = B(\theta_r)^{-1}u, \quad (12)$$

где  $u$  представляет собой вспомогательный управляющий вход. Подставляя (12) в (3), получаем линеаризованную модель дифференциального робота в виде

$$\dot{p} = u. \quad (13)$$

Затем, как показано в работе [10], предлагается пропорциональный регулятор с компенсацией, который регулирует линейную и угловую скорости в соответствии с генерируемой погрешностью.

Что касается предыдущих зависимостей, то мы получили

$$u = \begin{bmatrix} \dot{x}_{dr} \\ \dot{y}_{dr} \end{bmatrix} + K_p e(t), \quad (14)$$

где  $k_p \in \mathbb{R}^+$ . Ошибка отслеживания  $e(t)$  определяется как разница между желаемой траекторией  $(x_{dr}, y_{dr})$  и расчетной  $(p_x, p_y)$  относительно глобальной системы отсчета  $\{G\}$ , что означает

$$e(t) = \begin{bmatrix} e_x(t) \\ e_y(t) \end{bmatrix} = \begin{bmatrix} x_{dr} - p_x \\ y_{dr} - p_y \end{bmatrix}. \quad (15)$$

Подставляя (13) в (14), получим

$$\begin{bmatrix} \dot{p}_x \\ \dot{p}_y \end{bmatrix} = \begin{bmatrix} \dot{x}_{dr} \\ \dot{y}_{dr} \end{bmatrix} + K_p e(t) = \begin{bmatrix} \dot{x}_{dr} - \dot{p}_x \\ \dot{y}_{dr} - \dot{p}_y \end{bmatrix} + K_p e(t) = 0. \quad (16)$$

Следовательно,

$$\dot{e}(t) + K_p e(t) = 0. \quad (17)$$

Таким образом, управляющие входы задаются с помощью:

$$U = \begin{bmatrix} \cos(\theta_r) & -a \sin(\theta_r) \\ \sin(\theta_r) & -a \cos(\theta_r) \end{bmatrix}^{-1} \begin{bmatrix} \dot{x}_{dr} \\ \dot{y}_{dr} \end{bmatrix} + K_p e(t). \quad (18)$$

### Обнаружение препятствий и их объезд с помощью APF

С помощью алгоритма APF робот-черепаха моделируется как частица, находящаяся под воздействием двух потенциальных полей: притягивающего и отталкивающего [3]. Потенциальное поле определяется функциями притяжения и отталкивания, как показано на рисунке 4. Оба они спроектированы так, чтобы робот притягивался к месту назначения и удалялся от препятствий, как частица, стремящаяся к точке с минимальной энергией [11]:

$$U(q, q_d, q_{obs}) = U_{attr}(q, q_d) + U_{rep}(q, q_{obs}), \quad (19)$$

поскольку  $q$  – относительные расстояния робота,  $q_d$  – желаемые расстояния, а  $q_{obs}$  – расстояния до препятствий.

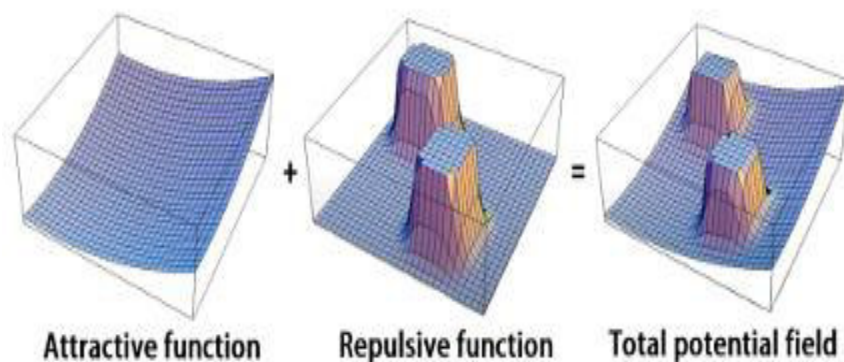


Рис. 4. Представление алгоритма APF

1) Функция притяжения. Считается, что параболическая функция конического притяжения определяет пороговое значение  $Q_d$ , соответствующее расстоянию, на котором функция изменяется с параболической на коническую, и постоянную параболической функции  $Q_d$ , получая:

$$U_{attr}(q, q_d) = \begin{cases} \frac{1}{2} \epsilon_q d^2 & si\ d < Q_d \\ Q_d \epsilon_q d - \epsilon_d \frac{Q_d^2}{2} & si\ d \geq Q_d \end{cases} \quad (20)$$

2) Отталкивающая функция. Гиперболическая отталкивающая функция определяется как

$$U_{rep}(q, q_{obs}) = \begin{cases} 0 & si\ Q^* \leq d \\ \epsilon_r \left( \frac{1}{d} - \frac{1}{Q^*} \right) & si\ Q_* \leq d \leq Q^* , \\ \epsilon_r \left( \frac{1}{Q^*} - \frac{1}{Q_*} \right) & si\ d \leq Q_* \end{cases} \quad (21)$$

где  $Q^* = Q_{inf}$  и  $Q_* = Q_{min}$ , где  $Q_{min} \cdot Q_{inf}$  представляет собой расстояние, на которое препятствие не оказывает влияния, а  $Q_{min}$  – минимальное безопасное расстояние, которое должно быть окружено. Параметр  $\epsilon_r$  – это показатель, который показывает, насколько функция сможет отразить удар робота.

3) Общее потенциальное поле. После вычисления функций притяжения и отталкивания общее потенциальное поле с учетом  $n$  препятствий может быть вычислено следующим образом:

$$U_{(q, q_d, q_{obs})} = U_{attr}(q, q_d) + \sum_{i=1}^n U_{rep}(q, q_{obs}^i). \quad (22)$$

Чтобы спланировать траекторию, необходимо следовать направлению, противоположному градиенту, который представляет собой вектор сил, определяемый суммой градиентов функций притяжения и отталкивания, величина и направление которых будут зависеть от конфигурации робота, пункта назначения и препятствий. Предполагается, что отрицательный градиент будет указывать на направление с минимальной энергией, то есть направление, которое минимизирует потенциальное воздействие на робота. Отрицательный градиент силы притяжения определяется формулой

$$\nabla U_{attr}(q, q_d) = \begin{cases} \begin{bmatrix} \epsilon_q d_d (x - x_d) \\ \epsilon_q d_d (y - y_d) \end{bmatrix} & si\ d_d \leq Q_d \\ \begin{bmatrix} \frac{Q_d \epsilon_q (x - x_d)}{d_d} \\ \frac{Q_d \epsilon_q (y - y_d)}{d_d} \end{bmatrix} & si\ d_d \geq Q_d \end{cases}, \quad (23)$$

где  $d(d, d_q) = \sqrt{(x - x_d)^2 + (y - y_d)^2}$ ,  $(x, y)$  представляют относительное положение робота и  $(x_d, y_d)$  желаемое положение. Аналогично сила отталкивания, полученная в результате суммарного отталкивающего воздействия всех препятствий при подходе APF, представляет собой отрицательный градиент, заданный формулой

$$\nabla U_{rep}(l) = \sum_i \begin{cases} \begin{bmatrix} \frac{\epsilon_r l_{x,i}}{d_i^3} \\ \frac{\epsilon_r l_{y,i}}{d_i^3} \end{bmatrix} & si\ Q_* \leq d \leq Q^* , \\ [0\ 0]^T & Otherwise \end{cases} \quad (24)$$

где  $(l_x, l_y)$  – компоненты расстояния каждого датчика до препятствия.  $d_i = \sqrt{l_{x,i}^2 + l_{y,i}^2}$  и  $\epsilon_r$  являются параметрами отталкивания. Сумма значений указывает на то, что при расчете учитываются все расстояния, определяемые каждым лучом лидарного датчика. Эта информация служит основой для закона управления, позволяющего безопасно доставить робота TurtleBro к желаемому месту назначения.

4) Закон управления. Исходя из диаграммы, показанной на рисунке 1, мы знаем относительное положение робота в смещенной контрольной точке, кроме того, матрица  $B(\theta_r)$  из уравнения (3) обратима. Следовательно, мы получаем следующее уравнение, умножая обе части на обратную матрицу, которую мы будем называть  $J$ , и очищая линейную и угловую скорости

$$\begin{bmatrix} \omega \\ v \end{bmatrix} = \begin{bmatrix} \cos(\theta_r) & \sin(\theta_r) \\ -a \sin(\theta_r) & a \cos(\theta_r) \end{bmatrix} \begin{bmatrix} \dot{p}_x \\ \dot{p}_y \end{bmatrix}, \quad (25)$$

где  $v$  и  $\omega$  – линейная и угловая скорости. Следовательно, с учетом формулы равномерного кругового движения угловые и линейные скорости, относящиеся к угловым скоростям, вычисляются по формуле

$$\omega(t) = \frac{r(\omega_r - \omega_l)}{2l}, \quad (26)$$

$$v(t) = \frac{r(\omega_l + \omega_r)}{2}. \quad (27)$$

Подставляя уравнения (26) и (27) в (2), получим

$$\begin{aligned} \dot{p}_x &= \frac{r}{2}(\omega_l + \omega_r)\cos(\theta_r) - \frac{r}{2l}(\omega_r - \omega_l)a \sin(\theta_r); \\ \dot{p}_y &= \frac{r}{2}(\omega_l + \omega_r)\sin(\theta_r) + \frac{r}{2l}(\omega_r - \omega_l)a \cos(\theta_r). \end{aligned} \quad (28)$$

Поскольку представляет интерес модулировать угловые скорости на каждом колесе, то после алгебраических манипуляций их можно переписать в виде выражений

$$\begin{aligned} \dot{p}_x &= \frac{\omega_r r}{2}(\cos(\theta_r) - \frac{a}{l}\sin(\theta_r) + \frac{\omega_l r}{2}(\cos(\theta_r) + \frac{a}{l}\sin(\theta_r) \\ \dot{p}_y &= \frac{\omega_r r}{2}(\sin(\theta_r) - \frac{a}{l}\cos(\theta_r) + \frac{\omega_l r}{2}(\sin(\theta_r) - \frac{a}{l}\cos(\theta_r)) \end{aligned}, \quad (29)$$

которые могут быть перестроены следующим образом:

$$\begin{bmatrix} \dot{p}_x \\ \dot{p}_y \end{bmatrix} = \frac{r}{2} \begin{bmatrix} \cos(\theta_r) + \frac{a}{l}\sin(\theta_r) & \cos(\theta_r) - \frac{a}{l}\sin(\theta_r) \\ \sin(\theta_r) - \frac{a}{l}\cos(\theta_r) & \sin(\theta_r) + \frac{a}{l}\cos(\theta_r) \end{bmatrix} \begin{bmatrix} \omega_l \\ \omega_r \end{bmatrix} = j(q)h, \quad (30)$$

где  $q = [x, y, \theta_r]^T$  и  $h = [\omega_l, \omega_r]^T$ . Таким образом, умножая обе части на обратную матрицу  $J$ , получаем следующее:

$$h = j^{-1}(q)[\dot{p}_x, \dot{p}_y]^T \quad (31)$$

$$j^{-1}q = \frac{1}{r} \begin{bmatrix} \cos(\theta_r) + \frac{l}{q}\sin(\theta_r) & \sin(\theta_r) - \frac{a}{l}\cos(\theta_r) \\ \cos(\theta_r) - \frac{q}{a}\sin(\theta_r) & \sin(\theta_r) + \frac{q}{a}\cos(\theta_r) \end{bmatrix}.$$

Основываясь на предыдущем, предлагаем следующий закон регулирования:

$$h = j^{-1}(q)\dot{\eta}^{ref}(q, q_d, l), \quad (32)$$

где  $\eta^{ref}$  – произведение линейной опорной скорости на суммарные силы, рассчитанные по отрицательному градиенту суммарного потенциального поля как  $\eta^{ref} = v_{ref} \begin{bmatrix} F_x \\ F_y \end{bmatrix} =$

$$= v_{ref} \begin{bmatrix} \cos(\theta_r) & -\sin(\theta_r) \\ \sin(\theta_r) & \cos(\theta_r) \end{bmatrix} F_{rep}.$$

## Результаты

В этом разделе представлены результаты моделирования и экспериментов по отслеживанию траектории полета Turtleburger и схеме обхода препятствий.

### А. Следование по пути

Для отслеживания траектории был реализован контроллер, указанный в уравнении (14). Были протестированы два типа траекторий: лемниската 1 и круговая траектория. Мы описали эти траектории, используя следующие уравнения:

$$\text{лемниската: } [x_d, y_d] = [x_0 + a \sin(\omega t) \quad y_0 + b \sin(2\omega t)];$$

$$\text{круг: } [x_d, y_d] = [r \cos(\omega t) \quad r \sin(\omega t)].$$

Параметры, используемые для характеристики алгоритма, приведены в таблице 3.



Таблица 3. Предлагаемые параметры траекторий

Параметр	Переменная	Ценность
Размер (по оси Y)	a	0.6 m
Размер (ось X)	b	0.48 m
Центр лемнискаты	$(x_0, y_0)$	(0,0) m
Угловая частота (лемниската)	$\omega$	0.157 rad/s
Радиус окружности	r	0.4 m
Угловая частота (окружность)	$\omega_c$	0.1 rad/s

Эффективность отслеживания траектории была эвристически скорректирована путем проведения многочисленных имитационных и экспериментальных испытаний. Следующие значения были признаны оптимальными для каждого случая:  $k_{sim} = 4$ ,  $k_{exp} = 1,6$ . На рисунке 5 показаны результаты моделирования, полученные при заданной траектории, по которой бургер Turtlebro следует непрерывной линией, а также ошибка отслеживания, полученная при траектории в форме лемнискаты с использованием Gazebo, которая обеспечивает среда 3D-моделирования, имитирующая реальный мир. Кроме того, можно импортировать модель робота и протестировать его работу в сложных имитируемых условиях. На графике показана максимальная погрешность в пределах  $\pm 0,3$  см, соответствующая отклонению желаемого маршрута от реального, что является хорошим приближением. В то же время результаты экспериментов показаны пунктирной линией с учетом тех же условий эксплуатации. Мы получаем, что погрешность отслеживания находится в диапазоне от  $\pm 2,5$  см по оси y до  $\pm 2$  см по оси x.

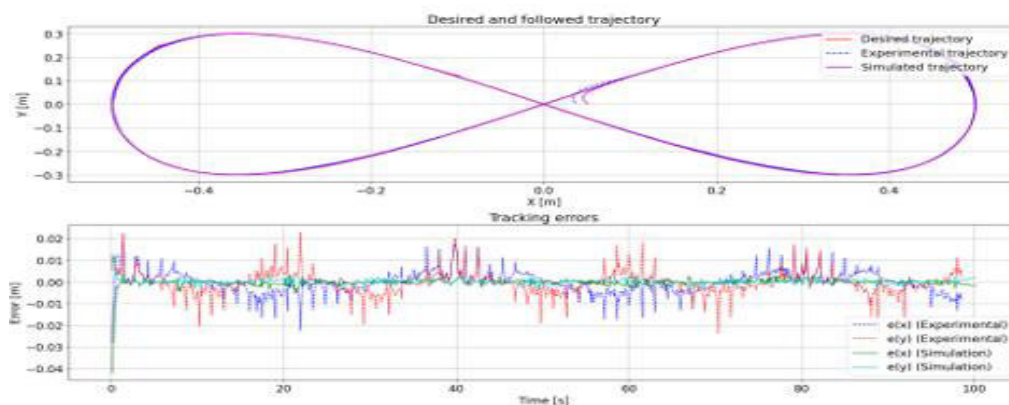


Рис. 5. Следующая траектория в форме лемнискаты

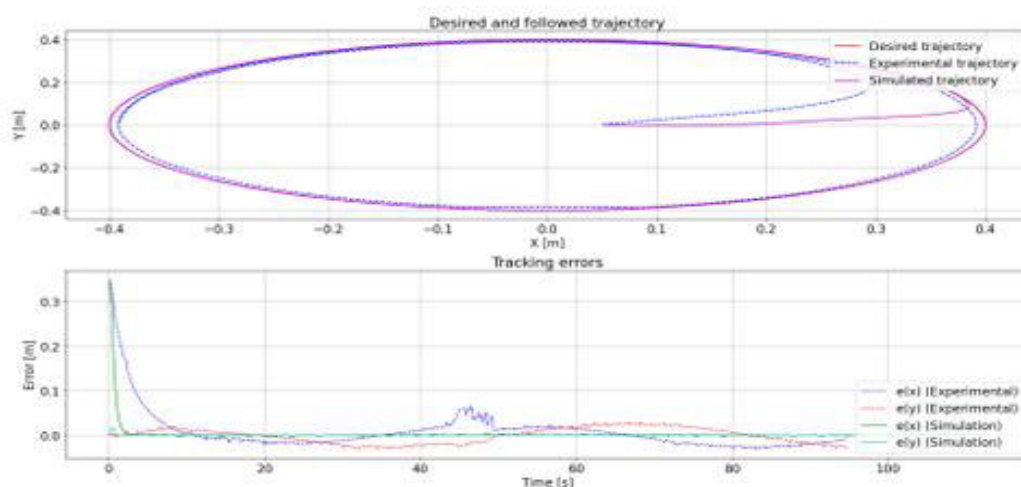


Рис. 6. Следующая траектория в форме круга

Впоследствии были проведены тесты с использованием круговой траектории для наблюдения за работой контроллера. Результаты, полученные по желаемой и пройденной траекториям, а также

ошибки отслеживания при моделировании показаны на рисунке 6 непрерывной линией, где наблюдается практически незначительная ошибка, которая асимптотически быстро сводится к нулю. Кроме того, в экспериментальной части пунктирными линиями показана погрешность отслеживания, составляющая приблизительно  $\pm 5$  см по обеим осям, что отражено на графике траектории.

### В. Избегание препятствий и следование по пути

В следующем разделе показаны результаты реализации алгоритма APF с использованием закона управления, описанного в уравнении (32). В моделируемой и экспериментальной среде тестирования одни и те же препятствия расположены в следующих позициях в метрах и ориентированы в радианах:  $[P_{Obs1}, O_{Obs1}] = [0.45, -0.6, 0]$ ,  $[P_{Obs2}, O_{Obs2}] = [1.4, -0.6, -0.3952]$  и  $[P_{Obs3}, O_{Obs3}] = [1, -1.4, 0]$ . Параметрами, использованными для моделирования и экспериментальных испытаний, были радиус колеса  $r = 0.033$  м, расстояние  $l = 0.08$  м, расстояние от центра до смещенной точки  $a = 0.04$  м, скорость  $v = 0.15$  м/с, которая обеспечивала наилучшие результаты при отслеживании и объезде препятствий, минимальное расстояние  $Q_{inf} = 0.3$  м, безопасное расстояние  $Q_{safe} = 0.16$  м,  $\epsilon_q = 2.2$  и  $\epsilon_r = 0.000845$ , причем последний параметр является наиболее важным для оценки.

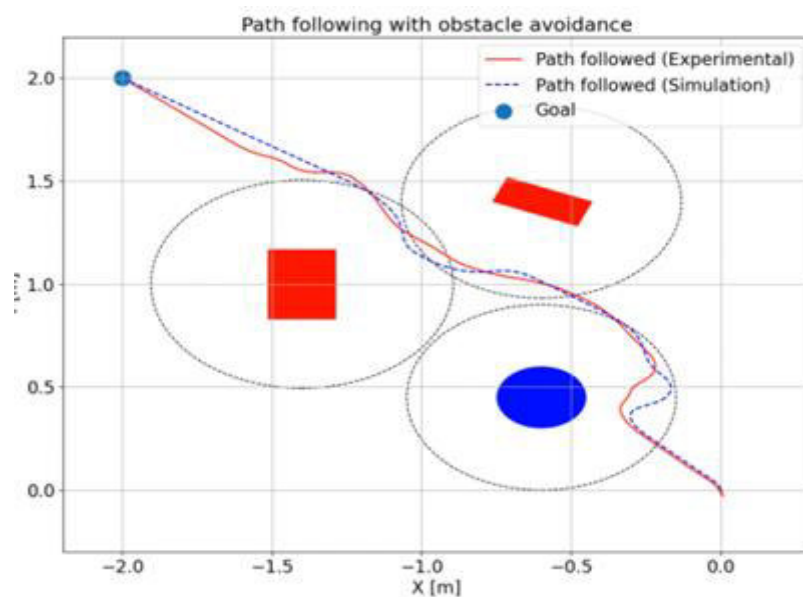


Рис. 7. Обход препятствия с помощью APF

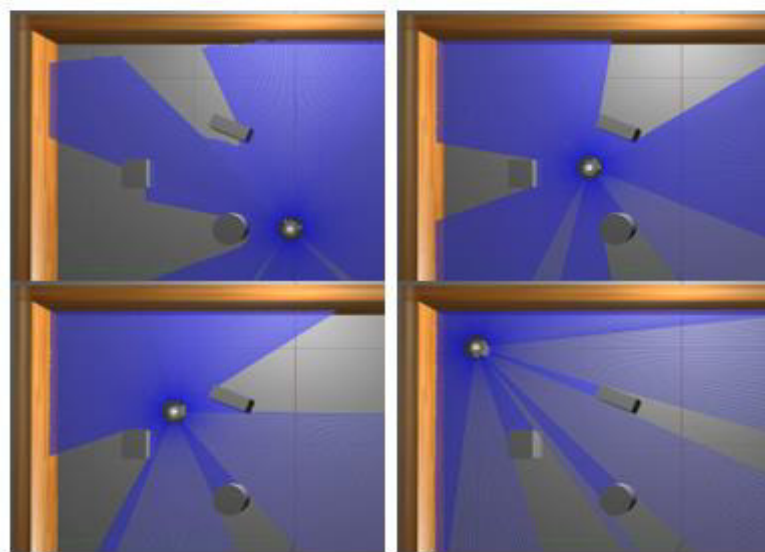


Рис. 8. Последовательность траекторий, пройденных в Gazebo

На рисунке 7 показаны результаты, полученные при моделировании Gazebo, при соблюдении траектории, описанной красным цветом, для достижения цели обхода препятствий в навигационном

пространстве. Результаты эксперимента показаны зеленым цветом. Для сравнения эффективности метода с теми же параметрами была создана точная копия имитируемой среды, которая показана на рисунке 8 в виде последовательности кадров, идущих к цели. Препятствия окружены по окружности пунктирной линией, которая указывает минимальную область, где робот обнаружит наличие блока. Как видно из обоих случаев, имитационного и экспериментального, результаты демонстрируют эффективность предложенного метода. Цель достигается за 26 секунд при моделировании и за 28 секунд при реальном применении.

### Заключение

В данной статье рассмотрена проблема автономной навигации в мобильной робототехнике путем реализации алгоритмов управления, соответствующих отслеживанию траектории и обходу препятствий, для дифференциального робота Turtlebot 3, запрограммированного в ROS 2. Представлен подход, основанный на алгоритме APF, для расчета непрерывного маршрута, который позволяет роботу эффективно достигать пункта назначения, избегая препятствий. Особое внимание уделяется использованию одометрии с абсолютными датчиками и лидаром для обработки сенсорных данных в режиме реального времени. Эффективность этих методов подтверждается результатами, которые демонстрируют способность робота следовать установленным маршрутам с минимальными ошибками. Кроме того, предлагаемый подход обеспечивает гибкость, предназначен для практического применения при планировании маршрутов в статических и динамических средах, что подчеркивает его способность обрабатывать данные в режиме реального времени и устранять внешние зависимости. В дальнейшем мы планируем повысить точность отслеживания маршрута, используя подход model predictive control (MPC), который оценивает будущую переменную и для оптимизации будущего поведения выходных данных установки у [12–14], алгоритмы для определения наиболее быстрого маршрута в динамических навигационных средах, таких как D\*, а также интеграцию дополнительных датчиков для получения более точных результатов за счет повышения эффективности вычислений.

### Список литературы (References)

1. R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, Introduction to autonomous mobile robots. MIT press, 2011.
2. J. A. Oroko and G. N. Nyakoe, "Obstacle avoidance and path planning schemes for autonomous navigation of a mobile robot: A review."
3. J. A. Oroko and G. Nyakoe, "Obstacle avoidance and path planning schemes for autonomous navigation of a mobile robot: a review," in: Proceedings of the Sustainable Research and Innovation Conference, 2022, pp. 314–318.
4. J. Lagunas-Avila, R. Castro-Linares, and J. Alvarez-Gallegos, "Obstacle avoidance in leader-follower formation using artificial potential field algorithm," 2021, pp. 1–6.
5. M. Melchiorre, L. S. Scimmi, L. Salamina, S. Mauro, S. Pastorelli et al., "Robot collision avoidance based on artificial potential field with local attractors," 2022.
6. S. M. H. Rostami, A. K. Sangaiah, J. Wang, and X. Liu, "Obstacle avoidance of mobile robots using modified artificial potential field algorithm," EURASIP Journal on Wireless Communications and Networking, vol. 2019, p. 70, 2019. Available: <https://doi.org/10.1186/s13638-019-1396-2>
7. H. M. Jayaweera and S. Hanoun, "A dynamic artificial potential field (d apf) uav path planning technique for following ground moving targets," IEEE Access, vol. 8, pp. 192760–192776, 2020.
8. M. Panagoda, M. Lokuliyana, A. Senarath, N. K. V. M. N. Nisansala, R. W. A. D. U. Rajapaksha, U. U. S. Rajapaksha, and C. Jayawardena, "Moving robots in unknown environments using potential field graphs," 2022, pp. 96–101.
9. F. I. Heredia-Moreno, M. E. J'ativa-Brito, C. G. Merino-S'anchez, and A. V. Mac'ias-Espinales, "Simulaci' on del modelo matem'atico de un robot m'ovil diferencial con control de posici' on sin orientaci' on basado en ley de lyapunov," Polo del Conocimiento, vol. 5, no. 1, pp. 880–899, 2020.
10. A. Escamilla and C. Mauricio, "Esquema de formaci' on de robots m'oviles con evasi' on de obst' aculos est' aticos y din' amicos," 2022.
11. O. Khatib, "Real-time obstacle avoidance system for manipulators and mobile robots," in Proceedings of the 1985 IEEE International Conference on Robotics and Automation, St. Louis, MO, USA, 1985, pp. 25–28.
12. G. Valencia-Palomo and J. A. Rossiter, "Auto-tuned predictive control based on minimal plant information," IFAC Proceedings Volumes, vol. 42, no. 11, pp. 554–559, 2009.
13. G. Valencia-Palomo, K. R. Hilton, and J. A. Rossiter, "Predictive control implementation in a plc using the iec 1131.3 programming standard," in 2009 European Control Conference (ECC). IEEE, 2009, pp. 1317–1322.
14. S. Chitraganti, S. Aberkane, C. Aubrun, G. Valencia-Palomo, and V. Dragan, "On control of discrete-time state-dependent jump linear systems with probabilistic constraints: A receding horizon approach," Systems & Control Letters, vol. 74, pp. 81–89, 2014.

УДК 087.2

DOI: 10.25686/978-5-8158-2474-4-2025-704-714

## Аналитическое сравнение выполнения сборок разной сложности в СПЖЦ «САРУС»

И. Ю. Аберясева, Д. П. Волков, Н. А. Денисова

Саровский физико-технический институт – филиал Национального исследовательского ядерного университета «МИФИ» (СарФТИ НИЯУ МИФИ), Саров, Россия

**Аннотация.** Авторы принимают участие в версионном функциональном тестировании вновь разрабатываемого комплекса программ «Система полного жизненного цикла изделий «Цифровое предприятие» («САРУС»). Цель исследования – проведение аналитического сравнения выполнения сборок разной сложности средствами программного модуля «Система конструкторского проектирования» (ПМ CAD). В CAD-проектировании создание сборочных моделей является одной из ключевых задач. В результате проделанной работы установлено, что ПМ CAD имеет схожие механизмы получения моделей для проектирования сборок с системами аналогичного класса, например, КОМПАС-3D. Однако были выявлены факты несрабатывания функций при использовании импортированных моделей, связанных с отсутствием возможности прямого импорта проприетарных форматов, а именно изменение параметров и атрибутов модели. Авторами работы выдвинуты предложения по развитию механизмов создания сборочных моделей средствами ПМ CAD на основе импортированных моделей.

**Ключевые слова:** конструирование в машиностроении, программный модуль CAD, сборочная модель, 3D-сборка, оптимизация алгоритма сборки.

## Analytical comparison of the assemblies of different complexity in the FLCS «SARUS»

I. Y. Aberyaseva, D. P. Volkov, N. A. Denisova

Sarov Institute of Physics and Technology – branch of the National Research Nuclear University "MEPhI" (SarPhTI NRNU MEPhI), Sarov, Russia

**Abstract.** The authors are participating in versioning functional testing of the newly developed software package "Digital Enterprise Product Lifecycle System" ("SARUS"). The purpose of the study is to conduct an analytical comparison of the performance of assemblies of varying complexity using the software module "Design Engineering System" (PM CAD). In CAD design, creating assembly models is one of the key tasks. As a result of the work done, it was found that PM CAD has similar mechanisms for obtaining models for assembly design, with systems of a similar class, for example, COMPAS-3D. However, the facts of non-functioning of functions, when using imported models were revealed, due to the lack of the possibility of direct import of proprietary formats, namely, changing the parameters and attributes of the model. The authors of the work put forward proposals for the development of mechanisms for creating assembly models using PM CAD based on imported models.

**Keywords:** design in mechanical engineering, CAD software module, 3D modeling, assembly model, 3D assembly, optimization of the assembly algorithm

### Введение

Практическим опытом организации современного производства доказано, что в производстве и проектировании системы автоматизированного проектирования (САПР) играют ключевую роль. Их важность обусловлена множеством факторов, таких как ускорение процессов проектирования и производства, повышение точности и качества продукции, снижение производственных затрат, поддержание инноваций и конкурентоспособности, обеспечение гибкости и адаптивности в условиях меняющихся требований рынка. Без этих технологий современное производство и инженерия были бы гораздо менее эффективными и конкурентоспособными.

В условиях высокой конкуренции требуется разработка и совершенствование новых систем автоматизированного проектирования (CAD) и активное их внедрение на предприятиях машиностроения. CAD-системы предназначены для моделирования технических объектов и разработки проектно-конструкторской документации, соответствующей стандартам и нормативным требованиям. Ключевым элементом такой документации являются сборочные модели, которые обеспечивают полное понимание структуры и состава конечного изделия.

Кафедра технологии специального машиностроения на базе кафедры цифровых технологий СарФТИ НИЯУ МИФИ является одной из экспериментальных групп по изучению, обучению и

внедрению в пользование нового САПР – СПЖЦ «САРУС» [1, 2] и проводит практико-ориентированное исследование по теме *«Количественная оптимизация элементов сборки в изделиях разной сложности в ПМ САД ПО «САРУС»*, которое проходит на обновленной базовой версии продукта 2024.2.1.1-VNF, утверждённой в 2024 году.

Документацией ПО «САРУС» [3] регламентируется, что *сборка изделия* является одним из основных технологических процессов в машиностроении. Поэтому поддержка процесса моделирования сборок является важной задачей программного модуля САД. Задача усложняется высокой размерностью, так как во многих изделиях число деталей, узлов, подборок может достигать сотен тысяч и миллионов. Вместе с тем встречаются модели с относительно небольшим числом компонентов – от нескольких десятков до тысяч. Руководством оператора определены операции и функции, которые поддерживает программный модуль САД «САРУС».

Из опыта работы в аналогичных САД, например, КОМПАС-3D, известно, что в сопряжениях могут участвовать координатные плоскости и оси, начала координат, грани, ребра, вершины тел и поверхностей (в том числе построенных в сборке), точки, вершины кривых, сегменты ломаных, графические объекты в эскизах, а также вспомогательные оси и плоскости. В КОМПАС-3D задаются сопряжения следующих типов: совпадение, касание, соосность, параллельность, перпендикулярность, расположение элементов на заданном расстоянии, Расположение элементов под заданным углом. Любое сопряжение можно удалить или отредактировать [4].

Исходя из практического опыта была определена область нашего исследования ПО «САРУС»:

- сравнить возможность задания соединений в двух САПР;
- за основу создания соединений в ПМ САД возьмем *форму сопрягаемых поверхностей и метод образования соединений*.

Таким образом, основная задача данной работы – исследовать реальность выполнения базовых операций сборки: возможность вставки компонента (детали или под сборки) в сборку; сопряжение компонент друг с другом по системам координат и сопряжениям; возможность имитации движения модели изделия по степеням свободы компонент; возможность движения отдельных компонент как единого целого, так и с учётом степеней свободы, содержащихся в самом компоненте.

Конкретно в данной работе рассмотрено аналитическое сравнение выполнения сборочных моделей разной сложности в программном обеспечении СПЖЦ «САРУС», базовая версия которого находится в активном тестировании с целью внедрения в конструкторско-технологические структурные подразделения РФЯЦ-ВНИИЭФ.

### Постановка задачи

Нас будут интересовать оригинальные и заимствованные детали, которые необходимо проектировать в ПО САД. Именно сложность проектирования сборок будет нашим объектом исследования, причем не имеет значения, оригинальные они или заимствованные. Необходимо проследить возможность выполнения операций сборки в ПМ САД СПЖЦ «САРУС» при создании сборок разной сложности.

Определим *цель* текущей работы: провести аналитическое сравнение выполнения сборок разной сложности.

Исходя из цели исследования, за основу возьмем по большей части заимствованные сборки, состоящие из деталей различной сложности [5].

Основные *задачи* работы:

- 1) выбрать две сборочные модели, на примере процесса сборки которых будут проведен сравнительный анализ;
- 2) выбрать систему проектирования, осуществимую в СПЖЦ «САРУС»;
- 3) создать сборочные модели в «САРУС» и произвести замер времени, затрачиваемого на процесс сборки;
- 4) создать сборочные модели в КОМПАС-3D и произвести замер времени, затрачиваемого на процесс сборки;
- 5) сравнить полученные результаты замеров времени, затрачиваемого на процесс создания сборочных моделей в сравниваемых системах.

## Теория

### 1. Краткое описание характеристики и назначения сборок

Сборочная модель – это цифровая 3D-модель, которая представляет собой сборку из нескольких отдельных компонентов или деталей в единое изделие. Она используется в системах CAD для визуализации, анализа и документирования взаимодействия различных частей изделия между собой. Современные CAD-системы должны обеспечивать баланс между функциональной полнотой и удобством работы, предлагая конструкторам эффективные инструменты для создания сложных сборок при минимальных временных затратах. Особое значение приобретает оптимизация процессов позиционирования и соединения компонентов, а также продуманная организация пользовательского интерфейса.

Рассмотрим основные характеристики сборочной модели.

#### 1) Составные элементы:

- Модель включает все детали, узлы и компоненты, из которых состоит изделие.
- Каждая деталь может быть создана отдельно, а затем интегрирована в сборку.

#### 2) Взаимное расположение:

- Детали размещаются в соответствии с их реальным положением в изделии.
- Учитываются соединения, зазоры, крепления и другие параметры.

3) Иерархия: сборочная модель может быть многоуровневой, включая подузлы (например, двигатель как подузел в автомобиле).

4) Параметризация: изменение одной детали может автоматически отражаться на всей сборке, если модель параметризована.

5) Документирование: на основе сборочной модели создаются чертежи, спецификации и другие документы, необходимые для производства.

### 2. Описание процесса создания сборочной модели

Из рассмотренных выше характеристик очевидно, что создание сборочной модели – многоэтапный процесс, который требует тщательной проработки каждого компонента и их взаимодействия. Мы основывались на основных этапах создания сборочной модели в системах CAD [6]:

#### 1. Подготовка и планирование

- Определение цели: четкое понимание, для чего создается сборочная модель (проектирование, производство, презентация и т.д.).
- Анализ требований: изучение технических условий, стандартов и нормативов, которые должны быть соблюдены.
- Разработка структуры сборки: определение иерархии компонентов (узлы, подузлы, детали) и их взаимосвязей.

#### 2. Создание отдельных деталей

- Проектирование компонентов: каждая деталь создается отдельно в CAD-системе. Это может быть как простая деталь (например, болт), так и сложный узел (например, корпус устройства).
- Параметризация: если необходимо, детали создаются с использованием параметров, чтобы их можно было легко изменять в будущем.
- Проверка геометрии: убедиться, что каждая деталь соответствует техническим требованиям.

#### 3. Сборка компонентов

- Импорт деталей: все созданные детали загружаются в файл сборки.
- Позиционирование: детали размещаются в пространстве сборки с использованием ограничений (constraints) и сопряжений (mates). Например, совмещение осей, фиксация расстояний или углов.
- Проверка взаимодействия: убедиться, что детали правильно соединены и нет пересечений.

## Результаты экспериментов

### 1. Описание выбранных сборочных моделей

Для проведения данной работы были выбраны две сборочные единицы: «Кривошип» и «Кондуктор перекидной». Рассмотрим каждую сборку.



Кривошип (рис. 1) – вращающееся звено кривошипного механизма, служащего для передачи движения от шатуна к валу. При этом поступательное движение поршня паровой машины преобразуется во вращательное движение вала.

Кривошип состоит из эксцентрично расположенных пальца и вала, соединенных посредством плеча. Способ соединения вала и пальца с плечом могут быть различными. Их выбирают в зависимости от условий, в которых будет работать данный механизм. В нашем случае соединение реализовано с помощью шпоночного соединения и винтового крепления.

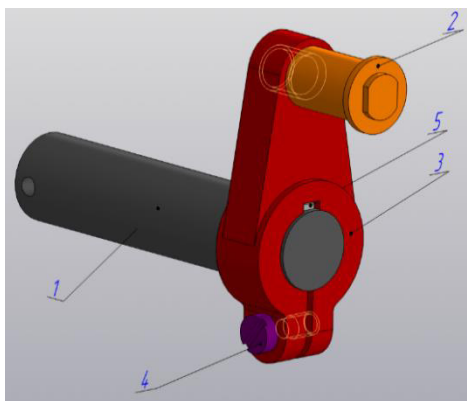


Рис. 1. Сборочная модель «Кривошип»

Сборочная модель «Кривошип» состоит из следующих компонентов: 1 – вал; 2 – палец; 3 – плечо; 4 – винт М6-6gx16 (ГОСТ 1491-80); 5 – шпонка А6х4х14 (ГОСТ 29175-91).

Вторая сборочная единица «Кондуктор перекидной» (рис. 2). Кондуктор перекидной, как универсальный механизм может использоваться в различных областях техники и промышленности. Его основное назначение — изменение направления движения или передача усилия между элементами под углом. Кондуктор перекидной имеет широкий спектр применения в различных отраслях: от машиностроения до бытовой техники. Его универсальность заключается в способности изменять направление движения или усилия, что делает его незаменимым в сложных механизмах.

Перекидной кондуктор представляет собой приспособление для сверления отверстий через втулки (поз. 5 и 6) в двух противоположных фланцах детали, устанавливаемой на базу (поз. 2) и закрепляемой быстросъемной шайбой (поз. 4) и гайкой (поз. 15).

В состав кондуктора входят 20 деталей, восемь из которых стандартные изделия:

- поз. 13 – болт М8-8g x 60.36 ГОСТ 3033-79 (1 шт.);
- поз. 14 – винт АМ6-8g x 12.58 ГОСТ 1491-80 (6 шт.);
- поз. 15 – гайка М10-7Н.5 ГОСТ 5915-70 (1 шт.);
- поз. 16 – гайка М10-7Н.5 ГОСТ 5916-70 (1 шт.);
- поз. 17 – шпилька М10-8g x 40.58 ГОСТ 22038-76 (1 шт.);
- поз. 18 – штифт 4n6 x 18 ГОСТ 3128-70 (4 шт.);
- поз. 19 – штифт 6n6 x 40 ГОСТ 3128-70 (1 шт.);
- поз. 20 – штифт 8n6 x 50 ГОСТ 3128-70 (1 шт.).

На корпусе 1 винтами 14 укреплена база 2, на фланец которой устанавливается и выступающей частью Ø25 центрируется обрабатываемая деталь. Плита 3 при этом откинута. В базу 2 ввернута шпилька 17, застопоренная низкой гайкой 16. На другой конец шпильки наворачивается гайка 15. С помощью быстросъемной шайбы 4 и гайки 15 обрабатываемую деталь жестко закрепляют в кондукторе, надвинув на верхний торец детали шайбу 4. После установки и закрепления детали кондукторная плита 3 поворачивается вокруг штифта 20 до горизонтального положения. Запрессованные в плиту 3 опоры 7 должны при этом накладываться на такие же опоры в корпусе 1. Откидной болт 13 с гайкой 10, установленный на штифте 19, накидывается на плиту 3. С помощью гайки 10 плита 3 закрепляется неподвижно. Через запрессованные в плиту втулки 5 производят сверление пяти отверстий Ø4 в верхнем фланце детали. Для сверления через втулки 6 четырех



отверстий  $\varnothing 5$  в противоположном фланце детали необходимо, взяв кондуктор за ручки (11, 12, 18), перекантовать его вместе с деталью на  $180^\circ$  и поставить на пяты 8. Просверлив отверстия, кондуктор снова поворачивают на  $180^\circ$  в прежнее положение, открепляют и откидывают плиту 3, ослабляют гайку 15, снимают шайбу 4 и вынимают готовую деталь.

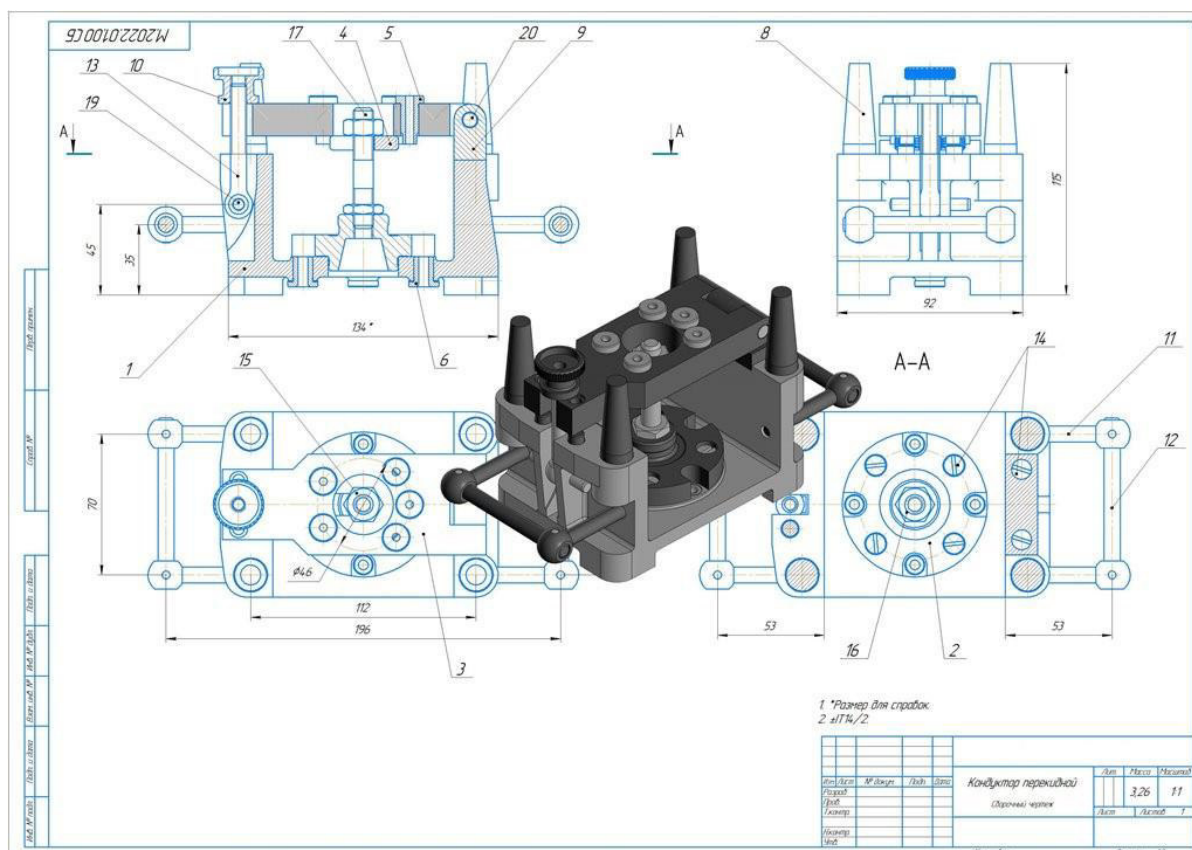


Рис. 2. Кондуктор перекидной

## 2. Сравнимые системы проектирования

СПЖЦ «САРУС» (далее – САРУС) – технологически независимая система управления полным жизненным циклом изделий для промышленных предприятий. Система представляет собой программное обеспечение (ПО) PLM-класса (Product Lifecycle Management – управление жизненным циклом изделий), входит в число основных классов индустриального ПО, разрабатываемого в соответствии с Дорожной картой развития высокотехнологичных областей «Новые производственные технологии», утвержденной Правительством РФ в 2021 году [1].

Система САРУС разработана предприятием Госкорпорации «Росатом» РФЯЦ-ВНИИЭФ. Брендовое название цифрового продукта созвучно месту его создания – городу Сарову Нижегородской области. Система САРУС создавалась с 2018 года в интересах атомной отрасли и оборонно-промышленного комплекса (ОПК) как моноплатформенное решение, полностью независимое от зарубежных технологий и элементов. Подтверждена её кроссплатформенность – продукт может работать как на операционной системе Windows, так и на российской операционной системе Astra Linux. В числе преимуществ также стоит отметить применение суверенного российского геометрического ядра RGK (Russian Geometric Kernel) для работы с 3D-объектами. Полная импортонезависимость и защищенное исполнение являются ключевыми преимуществами САРУС сегодня.

Система САРУС может применяться на предприятиях, использующих различные типы производства. Осуществленная апробация САРУС на предприятиях Росатома позволяет уверенно предлагать её заказчикам из других отраслей машиностроения в качестве цифрового продукта, имеющего опыт практического применения в условиях высоких требований атомной индустрии [1].

Как было сказано выше, САРУС находится на стадии разработки и местом его создания является Саров. Кроме этого, Саровский физико-технический институт выступает в качестве тестировщика данного ПО [2]. Поэтому оптимизация создания сборочных моделей путем выдвижения предложений по введению новых команд или доработки интерфейса будет производиться в программном обеспечении (ПО) САРУС в версии 2024.2.1.1-VNF.

Для объективного выдвижения предложений по введению новых команд или доработки интерфейса рационально будет сравнить САРУС с ПО, которое уже представлено на рынке. На основе этого сравнения и будут выдвинуты предложения для оптимизации создания сборок. Студенты в СарФТИ НИЯУ МИФИ среди множества САД-систем наиболее подробно изучают российскую систему трехмерного проектирования КОМПАС-3D, которая была выпущена на рынок в 1989 году. Поэтому САРУС будет сравниваться с КОМПАС-3D в версии V22.

КОМПАС-3D разработан российской компанией «АСКОН» и предназначен для проектирования изделий основного и вспомогательного производств в таких отраслях промышленности, как машиностроение (транспортное, сельскохозяйственное, энергетическое, нефтегазовое, химическое и т.д.), приборостроение, авиастроение, судостроение, станкостроение, вагоностроение, металлургия, промышленное и гражданское строительство, товары народного потребления и т. д. В основе КОМПАС-3D лежит российское геометрическое ядро С3D (создано С3D Labs, дочерней компанией АСКОН) и собственные программные технологии. Но для установки системы применяется автономный установщик Microsoft.NET Framework 4.8, являющийся разработкой американской компании Microsoft. Поэтому КОМПАС-3D считается отечественной разработкой, но с применением зарубежных составляющих [4].

### 3. Создание сборочных моделей в ПО КОМПАС-3D

Процесс создания сборки в КОМПАС-3D включает в себя следующие общие шаги:

1. Подготовительный этап. Перед началом работы необходимо тщательно изучить конструкторскую документацию (КД), на основе которой будет создаваться сборка.
2. Создание файла сборки. В главном меню выбрать: «Файл» → «Создать» → «Сборка». После создания файла активируется панель инструментов «Сборка» для дальнейшего редактирования.
3. Добавление базовой детали. Использовать команду «Вставить компонент из файла» в разделе «Компоненты». В качестве первой детали рекомендуется выбирать элемент, являющийся основой для большинства других компонентов (например, «Корпус» для сборки «Нагнетатель»).
4. Позиционирование детали. Задать пространственное положение базовой детали. Подтвердить выполнение команды добавления компонента.
5. Добавление последующих компонентов. Повторить процедуру добавления для каждой следующей детали (аналогично п. 3-4).
6. Создание сопряжений. Для корректного взаимного расположения деталей используются различные типы сопряжений: геометрические: совпадение, соосность, параллельность, перпендикулярность; позиционные: заданное расстояние, определенный угол; специальные: касание, симметрия, зависимое положение.
7. Завершение сборки. Последовательно повторить этапы 5-6 для всех оставшихся компонентов до полного формирования сборки.

Время, за которое удалось создать сборочные модели «Кривошип» и «Кондуктор перекидной» в результате использования вышеописанных шагов и применения массивов при данной конфигурации ПО КОМПАС-3D, представлено в таблице 1. Замер времени производился 3 раза.

Таблица 1. Замер времени построения в КОМПАС-3D

Элемент \ Замер времени	1	2	3	Среднее время
«Кривошип»	6,43 мин.	6,28 мин.	6,16 мин.	6,29 мин.
«Кондуктор перекидной»	28,44 мин.	27,54 мин.	27,33 мин.	28,17 мин.

В итоге среднее время создания сборочных моделей «Кривошип» в ПО КОМПАС-3D равно 6,29 мин., а «Кондуктор перекидной» – 28,17 мин.

#### 4. Создание сборочных моделей в СПЖЦ «САРУС»

Во время тестирования разработан полный пользовательский сценарий создания обеих сборок.

Приведем краткое его описание аналогично этапам, показанным выше, исключая очевидные операции.

1. Подготовительный этап выполнен.
2. Начало создания файла сборки. В окне «Приветствие» выбираем документ «Сборочная модель».

В ленте выбираем функцию «Вставить компонент» В строке «Файл» выбираем команду «Появляется окно с деталями, которые были выполнены раньше (рис. 3). Выбираем деталь «Вал».

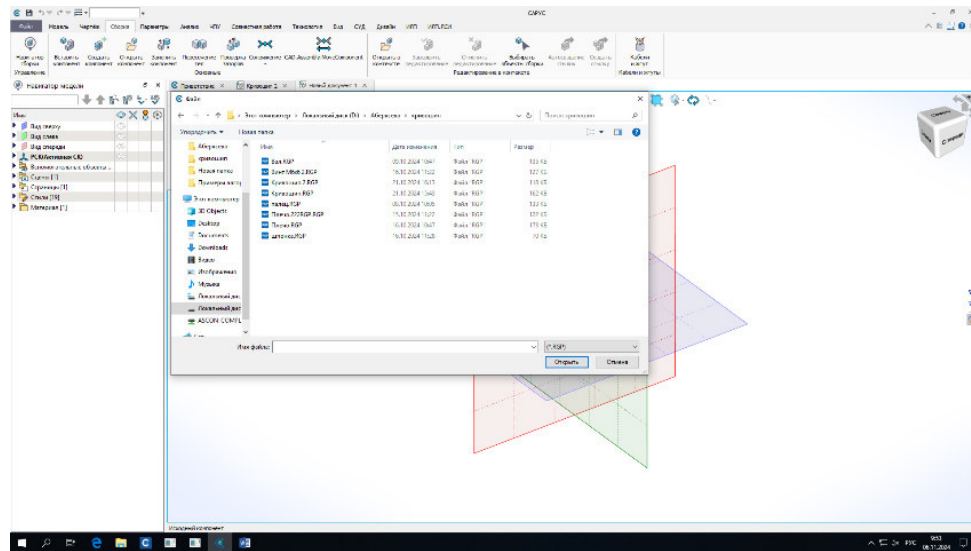



Рис. 3. Диалоговое окно для выбора компонентов сборки

3. Добавление базовой детали. В окне «Компонент» ничего не изменяем и выбираем функцию . Затем воспользуемся командой «Сопряжение». В открытом окне выбираем команду «Фиксация», даем согласие нажатием иконки с зеленой галочкой (при согласии выполнения операции этим приемом всегда обязательно пользуемся). После фиксации повторяем использование команды «Вставить компонент», но теперь выбираем деталь «Шпонка» (рис. 4).

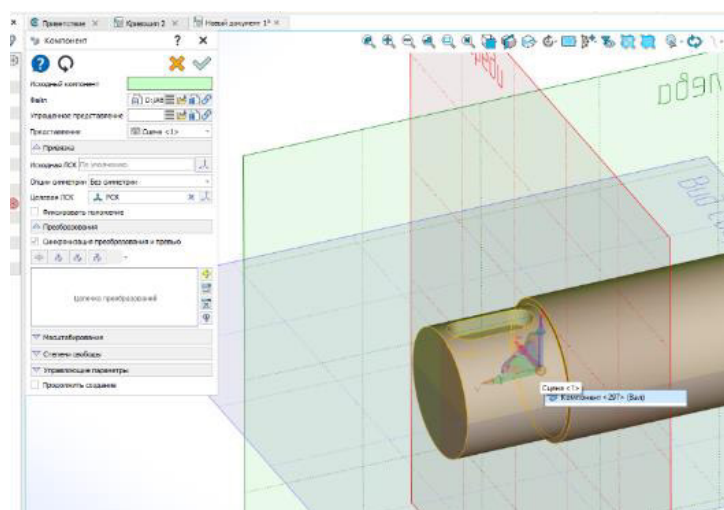


Рис. 4. Функция добавления компонента

4. Позиционирование детали. Курсором мышь наводим на центр системы координат (СК) детали «Шпонка» и перемещаем ее немного влево под углом, для удобства дальнейшего сопряжения. После выполнения данного действия используем зеленую галочку. Для наглядности процесса можно

воспользоваться функцией «Полупрозрачное изображение». Выбираем команду «Совпадение». Выбираем грани, как показано на рисунке 5, а. Видим, что деталь «Шпонка» расположена неверно (рис. 5, б). Курсором мыши наводим на последнее сопряжение и правой клавишей мыши открываем окно, где выбираем команду «Редактировать». В появившемся окне выбираем команду «Обратное направление». Если ничего не изменилось, команду необходимо повторить. На рисунке 5, в размещение детали «Шпонка» завершено.

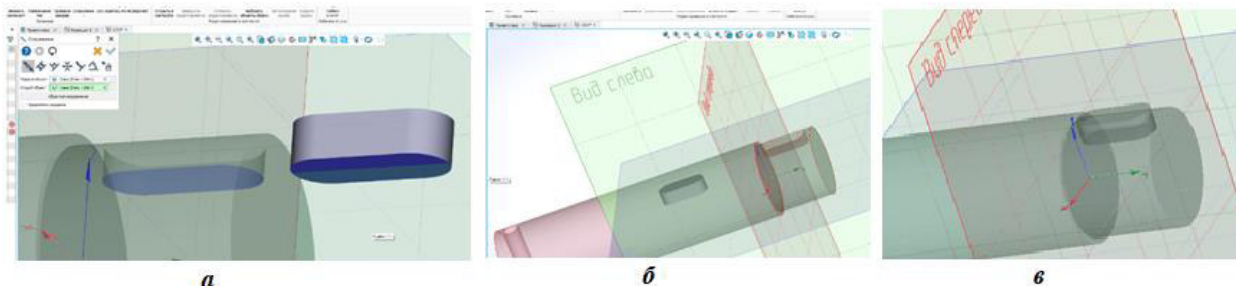


Рис. 5. Настройка правильного позиционирования компонентов сборки

5. Добавление последующих компонентов. Повторить процедуру добавления для каждой следующей детали (аналогично п. 3-4), например, повторяем использование команды «Вставить компонент», но теперь выбираем деталь «Плечо».

6. Создание сопряжений. Курсором мышь наводим на центр СК детали «Плечо» и перемещаем ее немного вправо, для удобства дальнейшего сопряжения (рис. 6, а). После выполнения данного действия используем ✓. В ленте выбираем команду «Сопряжение». В появившемся окне выбираем команду «Соосность» (рис. 6, б). Выбираем цилиндрические поверхности, после чего используем ✓ (рис. 6, в). В команде «Сопряжение» выбираем «Параллельность». Выбираем поверхность, как показано на рисунке 7, а, и плоскость «Вид слева». Воспользуемся командой «Совпадение». Выбираем поверхность и грань, показанные на рисунке 7, б. Размещение детали «Плечо» завершено (рис. 7, в).

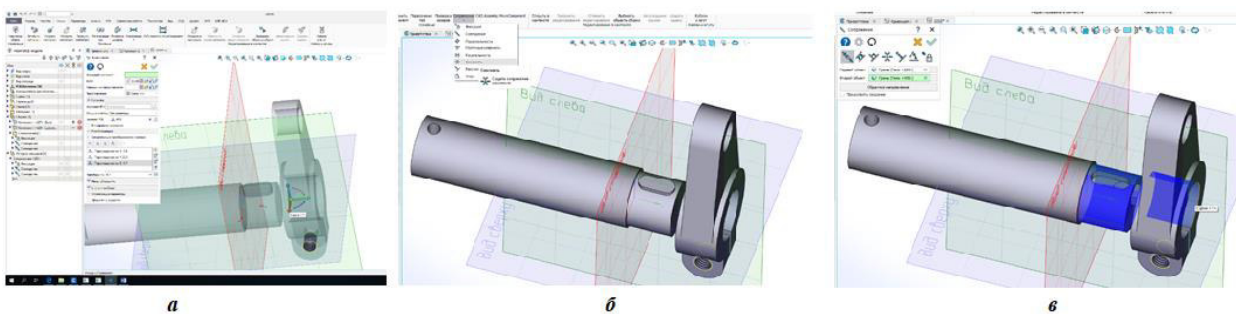


Рис. 6. Создание сопряжения компонентов сборки

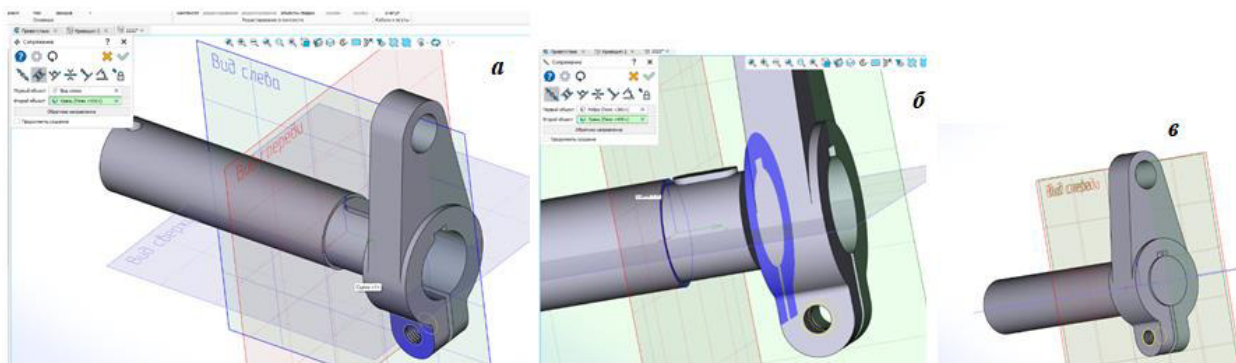


Рис. 7. Завершение сопряжения компонентов сборки



7. Завершение сборки. Последовательно повторить этапы 5-6 для всех оставшихся компонентов до полного формирования сборки: деталь «Палец» (рис. 8, а); последнюю деталь «Винт». Желательно ее расположить так, как это указано в чертеже (рис. 8, б). Если детали расположены неверно, используем команду «Редактировать». Сборка завершена (рис. 8, в).

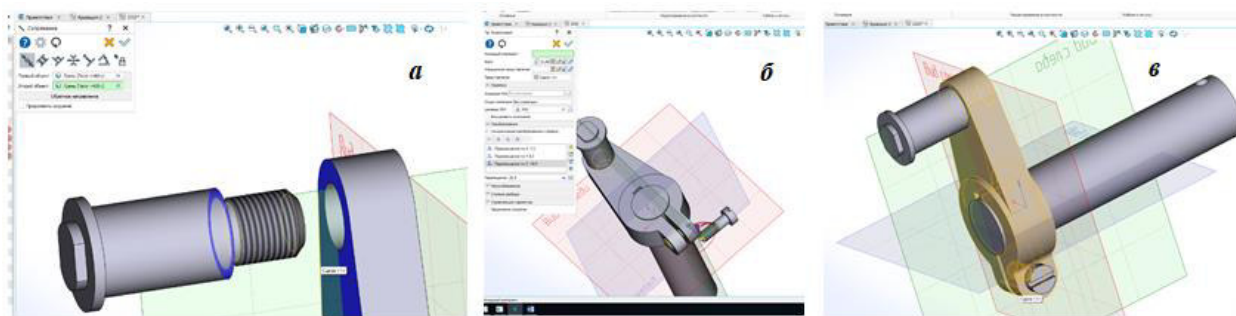


Рис. 8. Завершение создания сборки изделия «Кривошип»

В СПЖЦ «САРУС» есть и другой метод построения сборочной модели – построение сборки через локальную систему координат (ЛСК). Но на данной версии этот метод выдает ошибку, поэтому воспользоваться и продемонстрировать ее не удалось, так как новая ЛСК располагается не в центре отверстия, где мы изначально выбираем (рис. 9).

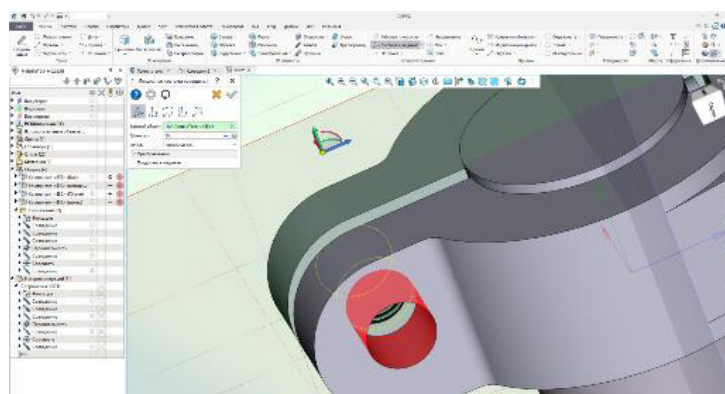


Рис. 9. Попытка построения сборки методом ЛСК

Сборку изделия «Кондуктор перекидной» выполняем аналогично. По результатам тестирования представлен пользовательский сценарий.

Время, за которое удалось создать сборочные модели «Кривошип» и «Кондуктор перекидной» в результате использования вышеописанных шагов ПО САРУС, представлено в таблице 2. Замер времени производился 3 раза.

Таблица 2. Замер времени построения в СПЖЦ «САРУС»

Элемент \ Замер времени	1	2	3	Среднее время
«Кривошип»	8,13 мин.	8,32 мин.	8,47 мин.	8,3 мин.
«Кондуктор перекидной»	32,45 мин.	31,33 мин.	30,4 мин.	31,39 мин.

В итоге среднее время создания сборочных моделей «Кривошип» в СПЖЦ САРУС равно 8,3 мин. и «Кондуктор перекидной» – 31,39 мин.

### Обсуждение результатов

#### 1. Сравнение процесса создания сборочных моделей в ПО КОМПАС-3D и СПЖЦ САРУС

Для наглядного представления полученных результатов сведем их в таблицу в виде разницы времени, затрачиваемого на создание сборочных моделей в ПО САРУС и КОМПАС-3D.

Таблица 3. Разница времени, затраченного на создание сборочных моделей «Кривошип» и «Кондуктор перекидной» в КОМПАС-3D и СПЖЦ «САРУС»

Элемент \ Замер времени	1	2	3	Среднее время
«Кривошип»	1,7 мин.	2,04 мин.	2,3 мин.	2,01 мин.
«Кондуктор перекидной»	4,01 мин.	3,79 мин.	3,07 мин.	3,22 мин.

Отсюда следует: процесс создания сборки в САРУС затрачивает больше времени на 2,01 минуты на создание «Кривошип» и 3,22 минуты на построение «Кондуктор перекидной», чем в КОМПАС-3D.

Полученный результат вызван, в первую очередь, задержкой отклика при вызове команды. Данная проблема связана с несколькими факторами:

1) аппаратные ограничения – слабые характеристики ПК: недостаток оперативной памяти; низкая производительность процессора;

2) оптимизация работы программы – большое количество компонентов: каждая деталь и сопряжение требуют вычислительных ресурсов; фоновые процессы: другие программы (антивирус, браузер) могут загружать систему.

В ПО САРУС команда «Вставить компонент» объединяет вставку компонента из файла, библиотеки и исходного компонента. Поэтому в диалоговом окне команды приходится выбирать необходимый источник файла и соответственно вызывать его. На это также затрачивается часть времени за счет излишних передвижений курсора и процесса выбора.

При необходимости добавить в модель несколько одинаковых деталей в САРУС за один сеанс команды «Вставить компонент» возможно добавить только одну деталь. Для добавления остальных таких же деталей необходимо снова использовать команду «Вставить компонент», но в этот раз в рабочем окне в поле «Исходный компонент» необходимо выбрать нужную деталь и только после этого возможно ее скопировать. На это действие также затрачивается некоторое количество времени.

Кроме этого, часть времени затрачивается на поиск или вызов необходимых команд. Примером служат массивы. Большая часть команд, необходимых для создания сборки, находятся на инструментальной панели во вкладке «Сборка», но массивы расположены во вкладке «Модель». Поэтому для вызова массивов приходится переходить из вкладки «Сборка» во вкладку «Модель», вызывать необходимый вид массива и после завершения работы с массивами возвращаться обратно во вкладку «Сборка» для продолжения дальнейшей работы.

Единственным существенным минусом на данный момент является отсутствие библиотек стандартных изделий в СПЖЦ «САРУС». В КОМПАС-3D в версии v22 учебной версии данная библиотека тоже отсутствовала.

На данный момент в ПО САРУС команда «Локальная система координат» работает некорректно. Также была замечена ошибка при использовании команды «Совпадение». После выполнения данной функции цилиндрические детали (чаще всего) меняют направление на 180°. Чтобы расположить деталь верно, требуется использование дополнительной команды. Иногда при изменении в сборке цилиндрические детали, в которых было изменено направление, принимают исходное положение, что требует повторного использования команд «Редактировать» и «Обратное направление». В некоторых случаях использование команд «Редактировать» и «Обратное направление» требуется неоднократно для расположения одной детали, на что тоже затрачивается время.

В КОМПАС-3D удастся сэкономить часть времени за счет возможности быстрого вызова некоторых команд, например, «Вставить компонент», с помощью нажатия правой клавиши мыши и открытия дополнительной миниатюрной панели инструментов, тем самым минимизировать передвижение курсора.

### Заключение

Проведенное исследование производительности систем автоматизированного проектирования выявило существенную разницу во временных затратах при создании сборочных моделей. На примере сборочных моделей «Кривошип» и «Кондуктор перекидной» было установлено, что САРУС демонстрирует более низкую скорость работы по сравнению с КОМПАС-3D. Этот результат требует детального рассмотрения и системного подхода к оптимизации.

Ключевые **выводы** исследования:

1. Факторы, влияющие на производительность:
  - архитектурные особенности систем;
  - алгоритмы обработки сопряжений;
  - эффективность использования аппаратных ресурсов;
  - удобство интерфейса и рабочих инструментов.
2. Предложения по оптимизации САРУС:
  - Технические улучшения:
  - ✓ ускорение алгоритмов обработки сборок;
  - ✓ оптимизация механизма создания сопряжений;
  - ✓ улучшение работы с аппаратными ресурсами.
3. Интерфейсные доработки:
  - упрощение последовательности операций;
  - введение "горячих клавиш" для частых действий;
  - улучшение визуализации в режиме реального времени.
4. Функциональные расширения:
  - введение шаблонов типовых соединений;
  - автоматизация повторяющихся операций;
  - улучшение системы предварительного просмотра.

Проведенное исследование подтвердило необходимость оптимизации процессов создания сборок в САРУС. Реализация предложенных улучшений позволит сократить временные затраты и повысить эффективность работы конструкторов, приблизив показатели системы к уровню КОМПАС-3D. Дальнейшие исследования могут быть направлены на детальный анализ конкретных "узких мест" в алгоритмах работы системы.

#### Источник финансирования

Договор от 25 марта 2024 г. № 96-2024/244 на выполнение научно-исследовательской работы по теме: «Проведение научных исследований в интересах ФГУП «РФЯЦ-ВНИИЭФ» с использованием результатов для актуализации образовательных программ подготовки специалистов с высшим профессиональным образованием для ФГУП «РФЯЦ-ВНИИЭФ», Задача 10 «Исследование программно-аппаратных решений в области цифровых технологий, возможностей их отраслевого применения, разработка и версионное тестирование компонентов полностью защищенной системы полного жизненного цикла среднего класса СПЖЦ V3 «САРУС».

**Благодарности.** Авторы благодарят кафедру цифровых технологий СарФТИ-НИЯУ МИФИ под руководством О.В. Кривошеева, к.т.н., зав. кафедрой, зам. директора ФГУП «РФЯЦ-ВНИИЭФ» по технологиям полного жизненного цикла – директор института цифровых технологий (ИТЦ) – главный конструктор СПЖЦ.

#### Список литературы

1. Об утверждении стратегии развития аддитивных технологий в Российской Федерации на период до 2030 года: Распоряжение Правительства РФ от 14.07.2021 № 1913-р (ред. от 21.10.2024).
2. Денисова Н.А., Федоренко Г.А. Применение в учебном процессе вуза тестирования программного модуля САД вновь создаваемого САПР «САРУС» // GraphiCon 2023: труды 33-й Междунар. конф. по компьютерной графике и машинному зрению (Москва, 19–21 сент. 2023 г.). М.: Институт прикладной математики имени М.В. Келдыша РАН, 2023. С. 848-859.
3. Комплекс программ в защищенном исполнении «система полного жизненного цикла изделий «Цифровое предприятие». Основная версия программного модуля. «Система конструкторского проектирования» (версия 2)/Руководство оператора 07623615.00423-06 34 01. 2021. 338 с.
4. КОМПАС-3D. Система трехмерного моделирования. Сайт «ascon.ru». URL: <https://ascon.ru/products/kompas-3d/>
5. Отчет о научно-исследовательской работе исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖЦ V3 (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. № № НО/38-278.10-2023-О от 15.11.2023. С. 8-27.
6. Выполнение сборочной модели в комплексе программ в защищенном исполнении «Система полного жизненного цикла изделий «Цифровое предприятие» / С. Е. Наседкина, Д. С. Коротков, Е. В. Множинская, К.О. Махов // Математика и математическое моделирование: сборник материалов XVIII Всероссийской молодежной научно-инновационной школы. Саров: СарФТИ, Интерконтакт, 2024. С. 537.



## Автоматизация расчета несущей способности антенно-мачтовых сооружений

Ю. Д. Маркина

Нижегородский государственный архитектурно-строительный университет, Нижний Новгород, Россия

**Аннотация.** Представлены результаты разработки и внедрения элементов автоматизированной инженерной системы расчёта несущей способности антенно-мачтовых сооружений на базе отечественного программного обеспечения. Актуальность работы определяется массовой эксплуатацией и модернизацией антенно-мачтовых сооружений, а также ростом требований к инженерной экспертизе при реализации проектов цифровизации отрасли связи. Проведён комплексный численный анализ влияния направлений ветровой нагрузки и конфигурации оборудования на напряжённо-деформированное состояние конструкций, подтверждающий необходимость учёта нестандартных сценариев ветрового воздействия и пространственного расположения оборудования. Разработаны специализированные программные модули с графическим интерфейсом, реализующие автоматизированный сбор исходных данных, расчёт аэродинамических характеристик, задание ветровых нагрузок и генерацию отчётной документации в соответствии с требованиями нормативных документов. Создана интегрированная база данных оборудования, включающая более 1100 наименований антенно-фидерных устройств, что позволяет существенно сократить трудоёмкость инженерных расчётов и снизить вероятность ошибок. Выполненные работы подтвердили эффективность предлагаемого подхода к комплексной автоматизации расчёта несущей способности антенно-мачтовых сооружений, обеспечивая сокращение времени подготовки расчётных моделей и повышение точности инженерного анализа. Полученные результаты являются основой для дальнейшего создания единой инженерной системы расчёта, сопровождения и анализа антенно-мачтовых сооружений.

**Ключевые слова:** антенно-мачтовые сооружения, несущая способность, автоматизация расчёта, ветровая нагрузка, цифровая модель, аэродинамика, программные комплексы, инженерный анализ.

## Automation of load-bearing capacity calculation for antenna mast structures

Y. D. Markina

Nizhny Novgorod State University of Architecture and Civil Engineering, Nizhny Novgorod, Moscow, Russia

**Abstract.** The paper presents the results of the development and implementation of components of an automated engineering system for calculating the load-bearing capacity of antenna mast structures based on domestic software. The relevance of the study is determined by the widespread operation and modernization of antenna mast structures, as well as the increasing requirements for engineering expertise within the framework of digitalization projects in the telecommunications sector. A comprehensive numerical analysis was conducted to assess the influence of wind load directions and equipment configurations on the stress-strain state of antenna mast structures, confirming the necessity of considering non-standard wind load scenarios and the spatial arrangement of equipment. Specialized software modules with graphical interfaces have been developed to enable automated collection of input data, calculation of aerodynamic characteristics, assignment of wind loads, and generation of reporting documentation in accordance with regulatory requirements. An integrated equipment database containing over 1,100 items of antenna-feeder devices has been created, significantly reducing the labor intensity of engineering calculations and minimizing the risk of errors. The completed work has confirmed the effectiveness of the proposed approach to comprehensive automation of antenna mast structures load-bearing capacity calculations, ensuring a reduction in the time required to prepare calculation models and improving the accuracy of engineering analysis. The obtained results provide a foundation for the further development of a unified engineering system for the calculation, support, and analysis of antenna mast structures.

**Keywords:** antenna mast structures, load-bearing capacity, calculation automation, wind load, digital model, aerodynamics, software packages, engineering analysis.

### Введение

Антенно-мачтовые сооружения (АМС) являются важнейшим элементом современной телекоммуникационной и технологической инфраструктуры. По данным Минцифры России и отраслевых аналитиков, на территории Российской Федерации эксплуатируется более 100 тысяч антенно-мачтовых сооружений различного назначения, включая объекты сотовой связи, телерадиовещания и промышленных систем [1, 2], при этом ежегодно вводятся в эксплуатацию тысячи новых объектов. Значительная часть АМС подвергается ежегодной модернизации и переоборудованию, связанным с расширением состава установленного оборудования [3].

Дополнительным фактором, усиливающим потребность в инженерных проверках и расчётах, является реализация «Проекта развития отрасли связи Российской Федерации на период до 2035 года» [4], предусматривающего развитие сетей 5G, подготовку инфраструктуры к технологиям 6G и значительное расширение сети связи. Это создаёт предпосылки к росту интенсивности эксплуатации и необходимости проведения расчётов несущей способности АМС в рамках переоборудования.

Одним из критических аспектов расчёта несущей способности АМС является учёт ветрового воздействия, включая влияние расположения оборудования при выборе критического направления ветра, что в настоящее время не регламентировано действующей нормативной базой [5]. Это усложняет инженерные расчёты и требует выполнения альтернативных расчётных сценариев для комплексного анализа конструкций.

Современные отечественные программные комплексы (SCAD Office, ЛИРА-САПР, RFEM) обеспечивают высокую точность инженерного анализа, включая моделирование сценариев ветрового воздействия, проверку устойчивости и прочности конструкций [6, 7]. Однако они не реализуют сквозной процесс расчёта и сопровождаются значительными трудозатратами на подготовку данных и формирование отчётности.

В связи с этим возникает объективная необходимость в создании специализированных программных модулей, обеспечивающих полный цикл моделирования, анализа и документирования расчётов несущей способности АМС, соответствующих отраслевым требованиям и интегрируемых в существующую цифровую инфраструктуру предприятий.

В данной статье приведены обобщенные результаты отдельных комплексных исследований автора по автоматизации расчета несущей способности АМС. Основные положения и расширенные численные данные в настоящее время приняты к публикации либо находятся на стадии подготовки к публикации и будут представлены в специализированных научных изданиях.

### **Постановка задачи**

Итоговой целью исследования является разработка единой инженерной системы расчёта на базе отечественного программного обеспечения, предназначенной для комплексного анализа несущей способности антенно-мачтовых сооружений (АМС), их инженерного сопровождения и подготовки технической документации в рамках единого цифрового цикла.

Для достижения указанной цели планируется последовательно решить следующие ключевые задачи:

1. Разработка стандартизированных правил построения цифровых моделей АМС.
2. Формирование и актуализация базы данных оборудования с возможностью оперативного обновления при появлении новых технических решений.
3. Автоматизация процессов подготовки исходных данных для расчёта.
4. Интеграция отечественных программных комплексов в единый инженерный контур.
5. Создание специализированных расчётных модулей, учитывающих конструктивные и эксплуатационные особенности АМС.
6. Автоматизация процедур экспертизы и подготовки нормативно-технической документации.

Настоящая статья содержит промежуточные результаты выполненных на данном этапе работ по каждому из указанных направлений.

### **Теория**

В рамках выполненной работы проведён комплексный анализ особенностей расчёта несущей способности АМС с учётом реальных конструктивных и эксплуатационных факторов. Объектами исследования выступили функционирующие АМС, расположенные в различных регионах Российской Федерации. АМС1 высотой 72 м расположена в Рязанской области; АМС2 высотой 70 м расположена в Тульской области; АМС3 высотой 67,5 м, АМС4 высотой 72,5 м и АМС5 высотой 40 м расположены в Калужской области. Особое внимание было уделено влиянию пространственного расположения технологического оборудования и направлений ветрового воздействия на напряжённо-деформированное состояние конструкций.

На основе численных исследований установлено, что традиционные подходы к выбору расчётных направлений ветра трехгранных АМС, ограниченные стандартным перпендикулярным направлением относительно одной из граней конструкции, не обеспечивают надлежащей полноты инженерного анализа. Моделирование показало, что при изменении направления ветра относительно конструкции, особенно при воздействии вдоль плоскостей граней, значительно возрастают усилия в отдельных элементах АМС. При этом локальные перераспределения нагрузок и пространственные деформации оказывают критическое влияние на общую работу сооружения. Численные исследования, выполненные на основании детальных моделей пяти реально эксплуатируемых АМС, подтвердили, что изменение направления ветрового воздействия, в частности при атаке воздушного потока вдоль одной из граней башни, приводит к перераспределению усилий между элементами конструкций и возникновению дополнительных напряжённых зон. Для расчёта процентного прироста сжимающих усилий за основу (100 %) принимались усилия в элементах АМС, полученные при ветровой нагрузке, направленной перпендикулярно грани, что соответствует методике определения ветровой нагрузки по [5]. Наибольшие отклонения по максимальным продольным усилиям  $N_{max}$  зафиксированы в элементах решётки — раскосах и распорках, где прирост продольных сжимающих усилий может превышать 20 % по сравнению с нормативными расчётами, а поведение распорок демонстрирует нестабильность и высокую чувствительность к углу атаки потока (рис. 1).

Кроме того, моделирование выявило характерное пространственное закручивание конструкции при воздействии ветра вдоль грани, что подтверждает необходимость учёта трёхмерного поведения сооружений при инженерных расчётах. Этот эффект обусловлен перераспределением ветровой нагрузки и возникновением крутящих моментов, что особенно актуально для гибких решётчатых конструкций и башенных сооружений. Современные экспериментальные и численные исследования подтверждают, что игнорирование пространственного закручивания может привести к существенным ошибкам в оценке напряжённо-деформированного состояния конструкции. Так, в [8] с использованием аэрогибкой модели стальной решётчатой башни показано, что при воздействии ураганного ветра возникает выраженный торсионный отклик конструкции. В работе [9] численным путём установлено, что угол атаки воздушного потока существенно влияет на коэффициент вибрационного отклика башен, что связано с перераспределением нагрузок и пространственным деформированием. В исследовании [10] экспериментально подтверждено наличие трёхмерных динамических нагрузок, а в [11] лабораторные испытания показали значительное возрастание моментов скручивания при косом ветре, аналогичном воздействию вдоль грани.

Отдельным направлением работы стал анализ влияния технологического оборудования, устанавливаемого на АМС, на аэродинамическое поведение конструкции. Выполненные моделирования с учётом реального размещения оборудования показали, что его расположение, габариты и парусность способны существенно изменять общую ветровую нагрузку, а следовательно, влиять на распределение внутренних усилий. Все расчеты выполнялись в ПК SCAD Office с использованием метода конечных элементов, что позволило учесть пространственную работу конструкции при различных направлениях ветровой нагрузки.

Таким образом, доказано, что при оценке несущей способности трехгранных АМС критическое значение имеют не только ветровое воздействие, действующее перпендикулярно наветренной грани [5], но и ветровое воздействие вдоль одной из граней башни, которое может приводить к повышенным усилиям в отдельных элементах. Также установлено, что для корректного выбора расчётного направления ветра необходимо учитывать как геометрию конструкции, так и фактическое расположение оборудования. Стандартного анализа по 1–2 направлениям ветра, предусмотренного [5], недостаточно. Количество и ориентация расчётных схем должны определяться индивидуально на основе инженерного анализа.

Сторонние исследования подтверждают, что для обеспечения надёжности АМС необходимо учитывать влияние нестандартных направлений ветрового воздействия, поскольку их игнорирование может приводить к недооценке усилий в отдельных элементах конструкций [12, 13]. Одновременно наблюдается активное развитие концепций цифровых двойников и информационного моделирования зданий [14, 15]. Однако, несмотря на прогресс в отдельных направлениях, комплексного решения,

объединяющего процессы сбора данных, численного моделирования, расчёта и автоматизированного документирования в единую систему, до настоящего времени не создано.



Рис. 1. Прирост  $N_{max}$  элементов решетки: а) AMC1; б) AMC2; в) AMC3-5

Выполнен комплекс работ, направленных на автоматизацию подготовки исходных данных для расчёта несущей способности АМС, что является критически важным этапом инженерного анализа подобных объектов. Традиционно подготовка исходных данных, особенно для сложных

пространственных конструкций с большим количеством оборудования и вспомогательных элементов, требует значительных временных затрат и сопряжена с высоким риском ошибок.

В ходе исследования разработаны и внедрены программные модули, обеспечивающие автоматизацию следующих ключевых процессов:

- сбор статической составляющей ветровой нагрузки на элементы основной решётчатой конструкции башни. Программа формирует перечень конструктивных элементов, определяет их геометрические характеристики и автоматически рассчитывает нормативные ветровые нагрузки в соответствии с действующими стандартами;
- сбор статической составляющей ветровой нагрузки на элементы кабельроста, лестниц и кабельных трасс, размещённых на АМС;
- сбор ветровой нагрузки на рабочие и технологические площадки, расположенные на конструкциях АМС;
- задание рассчитанных нагрузок непосредственно на расчётную схему в ПК SCAD Office. За счёт интеграции с расчётной моделью полностью исключается необходимость ручного ввода данных, что повышает точность и воспроизводимость результатов;
- подготовка стандартизированной отчётной документации по результатам всех этапов сбора и задания ветровых нагрузок. Программные модули формируют отчёты, соответствующие требованиям нормативных документов и технического задания, что существенно снижает трудоёмкость подготовки технической документации;
- экспертиза несущей способности элементов решётчатой башни на основании результатов расчёта, включая автоматическую генерацию отчётных форм для представления в составе технической документации по результатам инженерного анализа несущей способности АМС.

Особенностью разработанного комплекса является интеграция с расчётными моделями, созданными в ПК SCAD Office. Геометрические характеристики конструктивных элементов, необходимых для расчёта ветровой нагрузки, и значения усилий, необходимые для последующего анализа, автоматически импортируются из SCAD Office, что практически полностью исключает необходимость ручного ввода данных и минимизирует влияние человеческого фактора. Комплекс масштабируем и позволяет добавлять новые типы оборудования и новые расчётные сценарии, планируется интеграция с расчётными моделями, создаваемыми в различных ПК.

Реализация описанных подходов позволила существенно сократить трудоёмкость подготовки исходных данных, повысить точность результатов инженерного анализа и стандартизировать процедуру расчёта несущей способности АМС, что особенно актуально в условиях массового переоборудования и интенсификации эксплуатации подобных конструкций.

Также в рамках проекта разрабатывается комплекс специализированных программных средств с графическим пользовательским интерфейсом, предназначенный для автоматизации ключевых этапов подготовки исходных данных, необходимых для расчёта несущей способности АМС с учётом конструктивных и эксплуатационных особенностей. На данном этапе разработано три элемента комплекса.

Первым элементом является программа инженерного расчета нагрузок от антенно-фидерных устройств на антенно-мачтовые сооружения (рис. 2а). Это приложение, предназначенное для автоматизированного расчёта ветровой и весовой нагрузки от оборудования, размещаемого на АМС. Программа обеспечивает вычисление нормативных нагрузок на основе геометрических, конструктивных и аэродинамических характеристик оборудования, с последующим учётом их влияния при инженерном анализе общей пространственной модели АМС. Существенным преимуществом данного модуля является интеграция с расширяемой базой данных оборудования, что исключает необходимость ручного ввода информации и повышает точность расчётов. Как и в предыдущем модуле, реализована функция автоматической генерации отчётов, полностью соответствующих требованиям расчёта несущей способности.

База данных оборудования представляет собой структурированное справочное хранилище, включающее сведения о более чем 1100 наименованиях типового оборудования, устанавливаемого на антенно-мачтовых сооружениях: секториальных антеннах, радиомодулях, блоках питания, системных

модулях, климатических шкафов и прочем. Разработана программа для ведения и обновления базы данных антенно-фидерных устройств. Пользовательский интерфейс программы (рис. 2б) предусматривает возможность быстрого поиска, просмотра технических характеристик оборудования, а также интерактивного расширения содержимого. Данный функционал позволяет поддерживать актуальность справочной информации, своевременно учитывать новые модели оборудования, поступающие на рынок, и оперативно обновлять данные расчётных моделей.

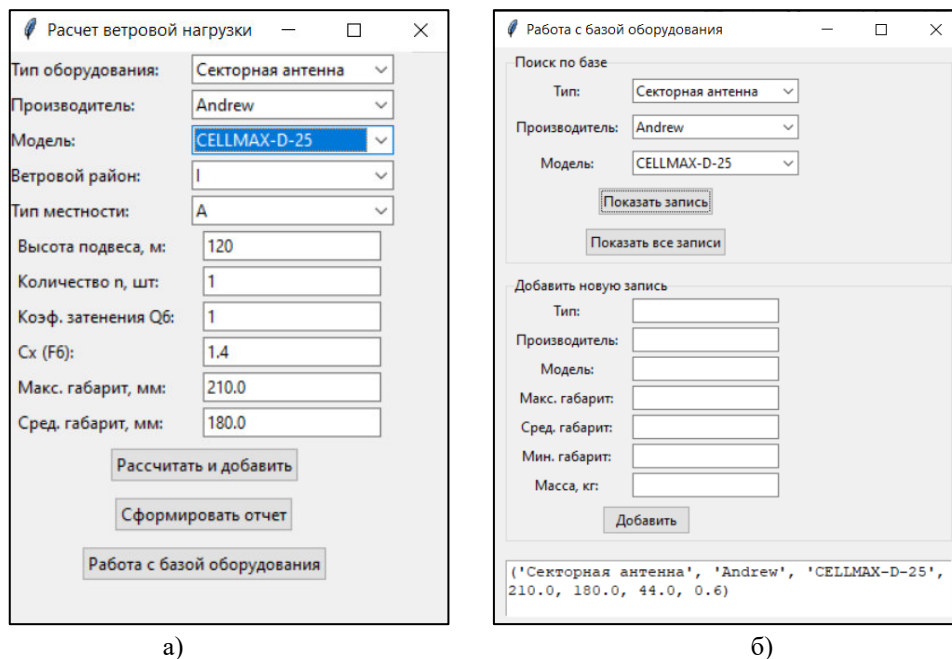
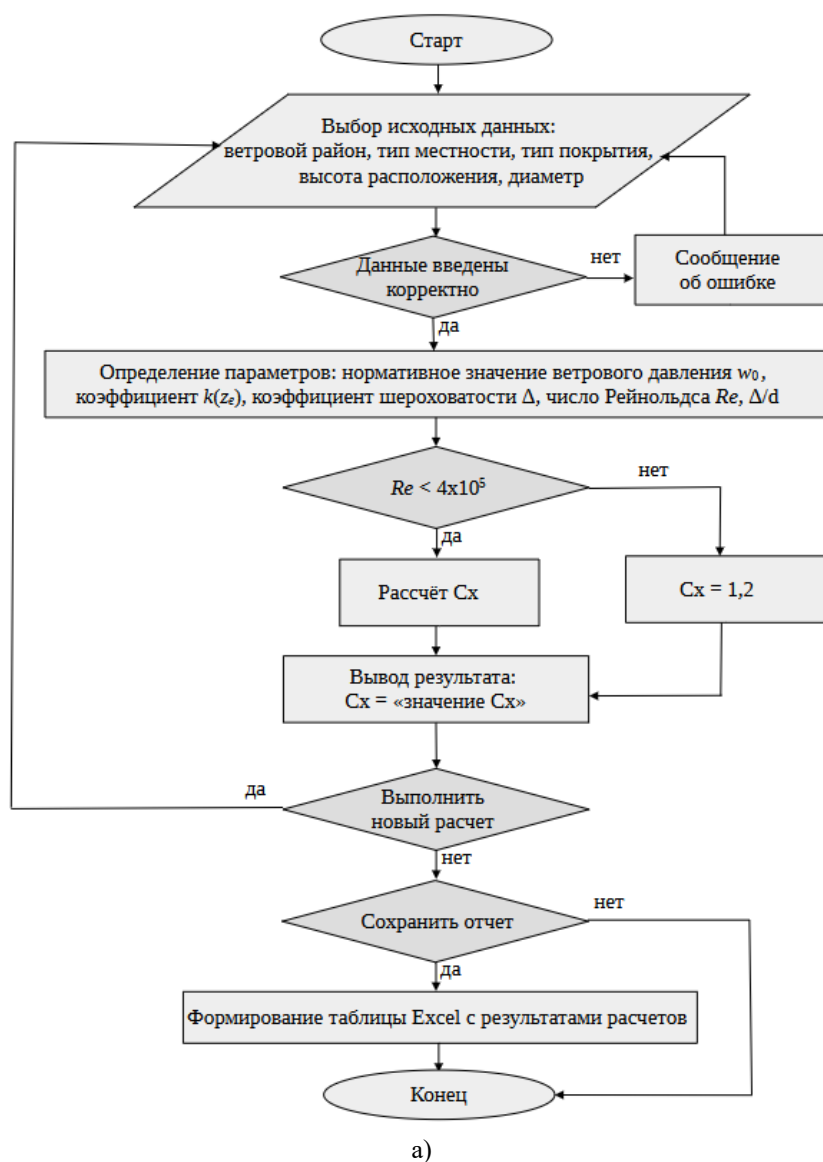


Рис. 2. Пользовательский интерфейс программ: а) инженерного расчета нагрузок от антенно-фидерных устройств на антенно-мачтовые сооружения; б) для ведения и обновления базы данных антенно-фидерных устройств

Третьим компонентом комплекса является программный модуль, предназначенный для оперативного определения аэродинамических коэффициентов сопротивления  $S_x$  элементов конструкции с круговой цилиндрической поверхностью (рис. 3). Указанный модуль реализует алгоритмы расчёта  $S_x$  на основании геометрических характеристик элемента (диаметра, шероховатости поверхности) и параметров воздушного потока. Реализация пользовательского интерфейса обеспечивает интуитивно понятный процесс ввода исходных данных и генерации отчётных материалов. В модуле предусмотрена возможность многократного пакетного расчёта для групп элементов конструкции, а также автоматического формирования итоговых отчётов, выполненных в соответствии с требованиями, предъявляемыми к документации по расчёту несущей способности АМС.

Совокупность разработанных программных инструментов обеспечивает комплексное повышение эффективности инженерного анализа АМС за счёт значительного сокращения трудоёмкости подготовки исходных данных, снижения вероятности ошибок, связанных с ручным вводом информации, стандартизации отчётной документации, повышения точности и повторяемости расчётных процедур.

Реализация предложенного комплекса программных средств позволяет создать основу для дальнейшей интеграции в единую инженерную систему расчёта, анализа и сопровождения антенно-мачтовых сооружений.



б)

Р...

Ветровой район  
|

Тип местности  
В

Покрытие  
Распыленная краска

$z_e$ , м  
39.0

$d$ , мм  
159.0

Рассчитать

Сохранить отчёт

График Cx

в)

Рис. 3. Программа инженерного расчета аэродинамического коэффициента элементов с круговой цилиндрической поверхностью: а) блок-схема; б) иконка; в) пользовательский интерфейс программы

## Результаты

В соответствии с ключевыми задачами, сформулированными ранее, выполнены следующие работы:

### 1. Разработка стандартизированных правил построения цифровых моделей АМС:

- Проведены численные исследования влияния нестандартных направлений ветровой нагрузки на напряженно-деформированное состояние конструкций.
- Подтверждена необходимость индивидуального выбора направлений ветра и построения нескольких расчетных моделей АМС.

### 2. Формирование базы данных оборудования и аэродинамических характеристик:

- Сформирована база данных из 11339 наименований оборудования, устанавливаемого на АМС.
- База данных интегрирована с разработанными программными модулями.

### 3. Автоматизация подготовки исходных данных:

- Разработан алгоритм автоматизированного сбора и задания статической составляющей ветровой нагрузки на элементы конструкции и оборудование.
- Реализована генерация отчетов по результатам автоматического сбора данных в стандартизированной форме (Excel).



**4. Интеграция существующих российских программных комплексов в общий инженерный контур:**

- Автоматизирован экспорт исходных данных из Excel в SCAD Office и результатов расчёта из SCAD Office в Excel.

**5. Разработка специализированных расчётных модулей для учёта особенностей АМС:**

- Разработаны прототипы расчётных сценариев с учётом нестандартных направлений ветра и расположения оборудования.
- Подготовлены программные модули с графическим интерфейсом для расчёта аэродинамического коэффициента  $C_x$  элементов с круговой цилиндрической поверхностью и ветровой и весовой нагрузки от оборудования.

**6. Автоматизация экспертизы и подготовки отчётности:**

- Разработан прототип системы автоматизированной экспертизы несущей способности элементов решетчатых опор.
- Реализована генерация наиболее трудоемкой части отчётной документации на основе результатов расчётов без необходимости ручного ввода.
- Сформированы стандартизированные шаблоны отчётов, готовые к подаче в экспертные и надзорные организации.

**Обсуждение результатов**

Выполненные работы продемонстрировали практическую применимость предложенных подходов. База данных оборудования, автоматизация подготовки исходных данных и экспертизы снижает трудоёмкость инженерных расчётов и вероятность ошибок. Учёт нестандартных направлений ветра позволяет повысить достоверность оценки несущей способности конструкций. Интеграция существующих отечественных программных комплексов с разработанными модулями создаёт основу для формирования единого инженерного контура.

Вместе с тем для достижения итоговой цели требуется дальнейшее развитие системы, включая расширенные аэродинамические исследования, полную интеграцию всех компонентов, расширение функциональности и повышение степени автоматизации.

**Заключение**

Выполнены начальные этапы подготовки к созданию единой инженерной системы расчёта несущей способности АМС. Результаты подтверждают перспективность комплексного подхода к автоматизации расчёта АМС. Дальнейшее развитие работ будет направлено на создание полноценной отечественной инженерной системы, обеспечивающей расчёт, анализ, сопровождение и подготовку технической документации в едином цифровом цикле.

**Список литературы**

1. Минцифры России. Единый реестр антенно-мачтовых сооружений связи. Режим доступа: <https://digital.gov.ru>
2. Ассоциация операторов связи «Телеком-Союз». Состояние и перспективы развития АМС в РФ. 2023. Режим доступа: <https://telecomunion.ru>
3. GSMA. Россия: Обзор рынка мобильной связи и инфраструктуры. 2023. Режим доступа: <https://www.gsma.com/russia>
4. Проект развития отрасли связи Российской Федерации на период до 2035 года. Режим доступа: <https://regulation.gov.ru/projects>
5. СП 20.13330.2016. Нагрузки и воздействия. Актуализированная редакция СНиП 2.01.07-85\*. М.: Минстрой России, 2016.
6. SCAD Soft. Программный комплекс SCAD Office: функциональные возможности и примеры применения. Режим доступа: <https://www.scadsoft.com>
7. ЛИРА-САПР. Системы автоматизированного проектирования и расчёта строительных конструкций. Режим доступа: <https://www.liraland.ru>

8. Zhang H., Huang X., Peng S. Aeroelastic modeling to study the wind-induced response of a self-supported steel lattice tower subjected to simulated hurricane winds // *Engineering Structures*. 2021. Vol. 244. 112759. DOI: 10.1016/j.engstruct.2021.112759
9. Zhao S., Zhang C., Yue J., Gu M. Wind-induced vibration coefficient of landscape tower with curved and twisted columns based on wind tunnel test and numerical analysis // *Buildings*. 2022. Vol. 12, Issue 10. 1635. DOI: 10.3390/buildings12101635
10. Aly A. Investigation of 3-D dynamic wind loads on lattice towers // *The Arabian Journal for Science and Engineering*. 2014. Vol. 39, iss. 8. P. 5863–5874. DOI: 10.1007/s13369-014-1173-5
11. Bashor R., Kareem A. Wind tunnel study of wind-induced torques on L-shaped tall buildings // *Journal of Wind Engineering and Industrial Aerodynamics*. 2009. Vol. 97, iss. 3–4. P. 131–147. DOI: 10.1016/j.jweia.2009.01.004
12. Хайбулина Р. З., Никулина О. Особенности определения ветровой нагрузки на металлические башни связи по РФ и Еврокодам // *Журнал строительных конструкций*. 2020. № 4. С. 95–108. DOI: 10.12345/jsc.2020.4.0095
13. Wang Z., Xu Y., Zhao J. Wind-induced responses and safety evaluation of communication towers using digital twin approach // *Journal of Wind Engineering and Industrial Aerodynamics*. 2023. Vol. 230. 105019. DOI: 10.1016/j.jweia.2023.105019
14. Stadtman F., Rasheed A., Kvamsdal T. Digital Twins in Wind Energy: Emerging Technologies and Future Directions // *Renewable Energy Digitalization*. 2023. Vol. 15. P. 45–65. DOI: 10.1016/j.red.2023.03.002
15. Kapteyn M. G., Willcox K. E. From Physics-Based Models to Predictive Digital Twins via Interpretable Machine Learning // *Journal of Computational Mechanics*. 2020. Vol. 58, no. 3. P. 211–225. DOI: 10.1002/jcm.31578

УДК 087.2

DOI: 10.25686/978-5-8158-2474-4-2025-724-733

## Оптимизация выполнения криволинейных поверхностей в САД «САРУС» путём параметризации

А. С. Костина, И. И. Макаров, Н. А. Денисова

Саровский физико-технический институт – филиал Национального исследовательского ядерного университета «МИФИ» (СарФТИ НИЯУ МИФИ), Саров, Россия

**Аннотация.** В рамках исследования программного модуля «Система конструкторского проектирования» (ПМ САД) комплекса программ «Система полного жизненного цикла изделий «Цифровое предприятие» («САРУС»), внедряемого на цифровом предприятии РФЯЦ-ВНИИЭФ, построение криволинейных поверхностей является достаточно актуальной задачей, так как большинство проектируемых деталей разной степени сложности имеют криволинейные поверхности. В ходе проведенного практико-ориентированного исследования срабатываемости функций создания моделей в модуле САД системы «САРУС» была успешно продемонстрирована эффективность параметризации как метода оптимизации построения криволинейных поверхностей. Использование параметризации значительно упростило процесс проектирования, обеспечив гибкость и точность моделирования. Благодаря наложению ограничений и корректному выбору типов дуг удалось избежать ошибок построения и добиться полной управляемости модели. Это особенно важно при работе со сложными фасонными формами, где традиционные методы могут оказаться недостаточно эффективными.

**Ключевые слова:** программный модуль САД, 3D-моделирование, криволинейная поверхность, параметризация, наложение ограничений.

## Optimization of curved surfaces in CAD "SARUS" by parameterization

A. S. Kostina, I. I. Makarov, N. A. Denisova

Sarov Institute of Physics and Technology – branch of the National Research Nuclear University "MEPhI" (SarPhTI NRNU MEPhI), Sarov, Russia

**Abstract.** As part of the study of the software module “Design System” (PM CAD) of the software package "Digital Enterprise Product Lifecycle System “Digital enterprise” (SARUS), implemented at the RFNC-VNIIEF digital enterprise, the construction of curved surfaces is a fairly urgent task, since most of the designed parts of varying degrees of complexity have curved surfaces. In the course of a practice-oriented study of the actuation of model creation functions in the CAD module of the SARUS system, the effectiveness of parameterization as a method for optimizing the construction of curved surfaces was demonstrated successfully. The use of parameterization has greatly simplified the design process, providing flexibility and accuracy of modeling. Due to the imposition of restrictions and the correct choice of arc types, it was possible to avoid construction errors and achieve full controllability of the model. This is especially important when working with complex shapes, where traditional methods may not be effective enough.

**Keywords:** CAD software module, 3D modeling, curved surface, parameterization, constraint imposition.

### Введение

В результате тестирования нового САПР «САРУС» (далее по тексту – САРУС), внедряемого в работу на производственных площадках ФГУП «РФЯЦ-ВНИИЭФ», студентами кафедры ТСМ СарФТИ выполнен большой объём проблемно-ориентированного исследования, направленного на развитие системы в области программного модуля «Система конструкторского проектирования» (ПМ САД), который является компонентом СПЖЦ «САРУС» и представляет собой САД-систему [1].

На описываемом этапе практико-ориентированного исследования методом тестирования нами продолжено изучение функциональных возможностей и алгоритмов срабатывания команд программного модуля САД.

На предыдущих этапах исследования [2, 3] были выявлены возможности применения операций, предлагаемых интерфейсом ПМ САД по построению криволинейных поверхностей в 2D- и 3D-режимах простых и средней сложности деталей машиностроения. Выявлено, что в представленных версиях программного комплекса СПЖЦ «САРУС» 1.0.17.9 и 2023.1.0.1-RC16 при проверке работоспособности ряда функций «Скругление» в СПЖЦ «САРУС» установлено, что все алгоритмы, кроме управления опциями перекрытия скруглений – взаимной обрезки или сопряжения скруглений с наращиванием материала, оказались работоспособными. Также была продемонстрирована работа

данных алгоритмов на деталях с усложненными элементами формы в СПЖЦ «САРУС». Здесь все алгоритмы работали исправно, даже в местах пересечений скруглений, чего, кстати, не удалось сделать на детали простейшей формы (I группы конструктивной сложности) [4].

Было отмечено, что разработчик проводил тестирование алгоритма работы функций, но не на деталях, а на поверхностях или примитивных конструкциях [5]. При незначительном усложнении сочетания поверхностей при построении 3D-моделей возникали трудности и сбои программы. Поэтому была поставлена цель исследования возможности применения операций, предлагаемых интерфейсом ПМ CAD, входящему в систему «САРУС», по построению криволинейных поверхностей в 2D- и 3D-режимах в деталях *простой формы* с теоретическим анализом предлагаемых функций построения.

Был проведен сравнительный анализ с известным САПР «КОМПАС-3D». Здесь у СПЖЦ «САРУС» были свои достоинства и недостатки. К первым относятся более корректное построение скругления с переменным радиусом, а также возможность построения скругления с перекрытием отверстия (скругление повторяет форму отверстия), что в «КОМПАС-3D» не получилось сделать. Из недостатков СПЖЦ «САРУС» по отношению к «КОМПАС-3D» можно отметить отсутствие некоторого функционала, который есть в «КОМПАС-3D», а именно: полное скругление, остановка скругления в точке и обход с гладкой стыковкой на пересечении двух ребер. Также «КОМПАС-3D» корректно выполнил алгоритм управления опциями перекрытия скруглений – взаимной обрезки или сопряжения скруглений с наращиванием материала. Однако и в той, и в другой САПР есть факты несрабатывания функций.

Был сделан вывод, что работать с функциями «Скругление» в СПЖЦ «САРУС» возможно на базовом уровне программного обеспечения CAD.

В Техническом задании для реализации третьего этапа научной работы запланированы следующие действия:

- разработка алгоритмов создания 3D-моделей **деталей средней сложности** с разным типом криволинейных поверхностей с применением разных 3D-операций;
- формулирование аналитического вывода рациональности построений;
- выявление недостатков построения криволинейных поверхностей в ПО «СПЖЦ ЦП»; формулирование рекомендаций для доработки ПМ.

### Постановка задачи

Обновлённая версия «САРУС» 2024.2.1.1-VNF позволила сократить обучающую часть работы студентов в связи с более совершенным интерфейсом и удобным уровнем его применения. Студенты выполняют построения 3D-моделей деталей с поверхностями достаточной сложности.

Поэтому **целью тестирования** на данном этапе мы ставим исследование возможности применения операций, предлагаемых интерфейсом ПМ CAD, входящему в систему «САРУС», по построению криволинейных поверхностей в 2D- и 3D-режимах в **деталях средней сложности с элементами сложной формы** с теоретическим анализом предлагаемых функций построения. Конструктивную сложность в соответствии с классификацией, предложенной в отчёте по выполнению второго этапа исследования [3], рассмотрим для деталей геометрических форм, которые требуют специальных расчетов (стаканы, петли, линзы и др.).

### Задачи:

- тестирование ПМ CAD по срабатыванию выполнения функций, операций и команд построения, рекомендуемых разработчиком при построении 3D-моделей, исходя из обозначенной в цели конструктивной сложности деталей и с использованием ранее сформулированных рекомендаций исследования применения операций: ЛСК, булевы операции, скругления;
- описание логики выбора и разработки оптимальных алгоритмов построения криволинейных поверхностей в ПМ CAD «САРУС» в 3D-моделях деталей средней сложности с элементами сложной формы:
  - 1) проанализировать построение основания детали «Фланец» с применением операции «Круговой массив»;
  - 2) проанализировать построение эскиза с применением операций группы «Параметризация».

- анализ работоспособности алгоритмов построения криволинейных поверхностей при построении 3D-моделей;
- подготовка методических материалов по обучению работы с ПМ CAD «САРУС» студентов по направлению подготовки 15.03.04 – Конструкторско-технологическое обеспечение машиностроительных производств, а также специалистов, имеющих опыт работы с рядом САПР в направлении конструкторской подготовки.

### Теория

В выводах по начальному этапу исследования, исходя из того, что студенты тестировали ПО без каких-либо предлагаемых алгоритмов построения, используя только свои навыки работы в известных ПО, были заданы следующие критерии анализа работоспособности тестируемого ПМ сравнительного характера:

- 1) функциональное назначение;
- 2) удобство и понятность интерфейса, интуитивный уровень его освоения;
- 3) производительность – быстрота построения одинаковых деталей.

На следующем этапе у нас уже появились алгоритмы построения 3D-моделей, выполненных и описанных по результатам этапа 1, критерии слегка скорректировали:

- 1) функциональное назначение конкретных операций и исследование двух алгоритмов построения в САРУС и сравнение их с алгоритмом построения в КОМПАС-3D;
- 2) удобство и понятность интерфейса, интуитивный уровень его освоения;
- 3) производительность – быстрота построения одинаковых деталей.

Аналитическое сравнение в данном случае также имеет качественный характер. Даже производительность зависит не только от количества операций, необходимых для построения 3D-модели. Здесь большую роль играют опыт конструктора, умение применить оптимальный алгоритм построения, ведь одну и ту же модель можно создать, используя разные операции [2].

Этап 3 отличается уже рефлексивным пониманием ПМ CAD «САРУС»: имея операции и функции построения, идентичные большинству САПР в области 3D-моделирования, эта САПР имеет отличную от, например, известного и хорошо освоенного КОМПАС-3D, логику построения.

При попытке освоить программный модуль (ПМ) CAD интуитивным способом пользователи встречаются с непониманием несрабатывания функций, когда, казалось бы, все действия и приёмы выполняются. Практическая работа в ПМ показала, что при построении 3D-моделей даже сложных деталей они действительно срабатывают, только надо правильно спроектировать и задать во всплывающем окне все необходимые действия.

Предлагаем несколько рекомендаций при начальном освоении ПМ.

1. Построить 3D-модель здесь можно двумя способами: по примитивам и по эскизу. Третий способ – комбинированный, тоже возможен и часто оптимален.

2. Большую роль здесь играет система координат (СК или ЛСК (локальная). При построении разных элементов детали её необходимо перемещать по осям в соответствии с размерами и расположением элемента детали в пространстве, особенно при построении по примитивам. Если при построении по эскизу существует новая плоскость, ЛСК перемещается автоматически при указании этой плоскости [6].

3. Булевы операции рекомендуются к постоянному вниманию: если булева операция не задана или отмечена неправильно, построение будет невозможно [7].

4. При выборе операции всплывает окно. Его необходимо внимательно изучить и выбрать те функции, которые необходимы в Вашем случае. Обратите внимание на галочку ОК во всплывающем окне. Если она зелёная, операция выполнена полностью, по ней можно кликать для завершения операции.

При построении активно пользуемся документами [5, 8, 9, 10].

Перед началом работы необходимо хорошо ознакомиться с «Лентой» документа «Модель», где выбираются операции и осуществляются функции построений заданных поверхностей (рис. 1). Здесь есть возможность выбора способа построения «По эскизу» с геометрическим построением в 2D (a) или

«Примитиву» с автоматизированным построением простейших геометрических фигур (б); выбрать 3D-операции для построения объёмной поверхности (в); построить элементы модели (г); а также использовать вспомогательные (д) и дополнительные (е) функции построения и т.д.



Рис. 1. Панель «Лента» в программном модуле CAD «САРУС»

## Результаты экспериментов

### 1. Анализ построения основания детали «Фланец» с применением операции «Круговой массив».

#### 1.1. Анализ детали «Фланец»

Особенностью детали «Фланец» (рис. 2) можно определить наличие дискового элемента с волнистой гранью. Все остальные элементы детали имеют достаточно простую форму и понятный алгоритм построения 3D-модели.

- Рассмотрим построение основания в виде диска с волнистой поверхностью.
- Проведём построение с показом истории операций детали «Фланец».

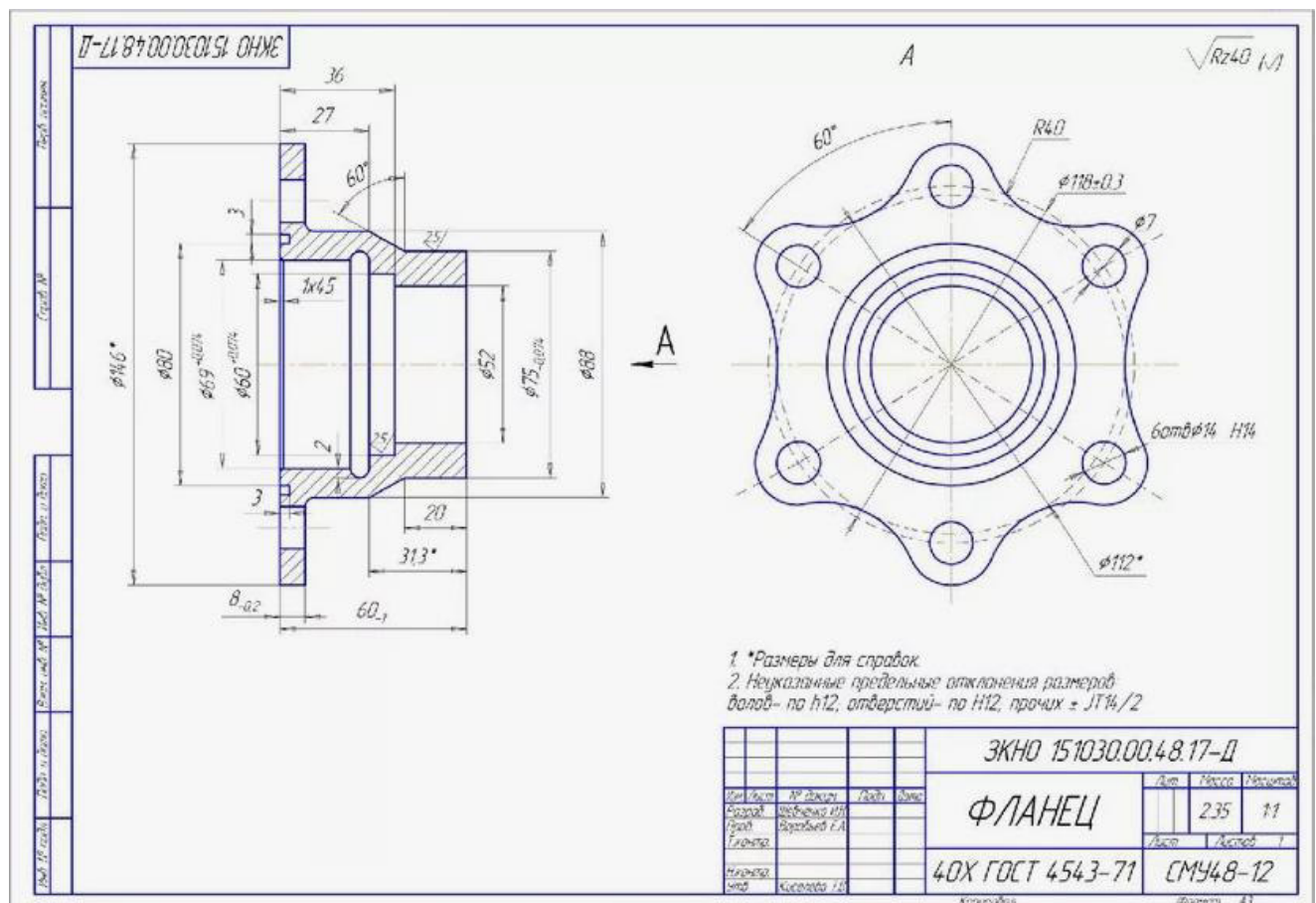


Рис. 2. Чертеж детали «Фланец»

#### 1.2. Построение основания по эскизу с применением операции «Круговой массив».

- Эскиз представляет собой 6 отв.  $\varnothing 14$ , расположенных на окружности  $\varnothing 118$  и углом  $60^\circ$  между ними. Для точного построения строим одну окружность и с помощью команды «круговой массив», копируем окружности с выставлением нужного угла между ними.

- С помощью центра одной из построенных окружностей строим дугу R15. Повторяем операцию «круговой массив», размножая построенную дугу (рис. 3).

Сразу замечаем сбой построения (отмечен красным эллипсом на рис. 3). Забегая вперёд, заметим, что не удалось этот элемент построения исправить при реализации данного алгоритма (рис. 4).

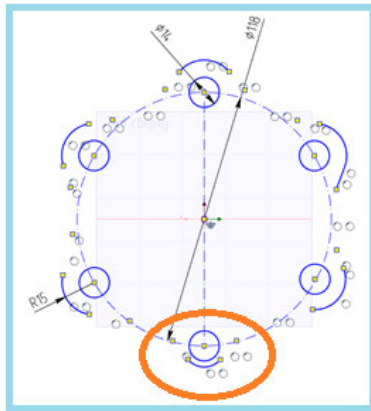


Рис. 3. Эскиз построения с применением операций «круговой массив» со сбоем построения

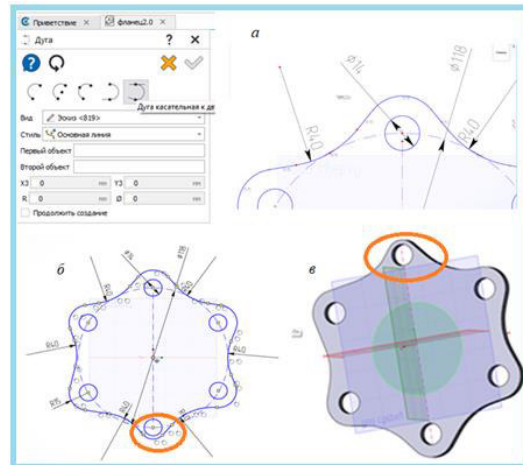


Рис. 4. Построение рассматриваемого элемента по эскизу: а – с помощью команды «дуга» под тип «дуга, касательная к двум прямым» – создание объекта «дуга» между двумя кривыми; б – эскиз со сбоем; в – 3D-модель элемента, построенная операцией «Вытягивание»

Мы проводили практико-ориентированное исследование операции «Скругление» при построении моделей на протяжении тестирования ПО «САРУС». Эта операция достаточно сложна, всегда вызывала моменты несрабатывания. Однако в последней версии показала достаточно хороший результат.

В данном случае мы пытаемся построить 2D-скругления. Выполнялось большое количество перестроений разными способами, но с одной логикой, которые не принесли положительного результата. Анализируя результат построения, можно заключить, что, видимо, большое число элементов вызывает сбой программы и применение множества криволинейных поверхностей в одной операции в данной версии программы ограничено.

**Вывод:** необходимо искать другой способ построения рассматриваемого элемента в 2D-формате.

## 2. Анализ построения эскиза основания с применением операций группы «Параметризация»

2.1. Параметризация является мощным инструментом в современном конструкторском проектировании, и ее внедрение в рабочие процессы способствует повышению качества и эффективности разработки сложных деталей. Необходимо пояснить, что процесс, если не имеешь опыта, достаточно трудоёмкий, построения должны быть аккуратными и точными. Зато здесь при построении не имеют значения размеры – они устанавливаются для завершения этапов построения.

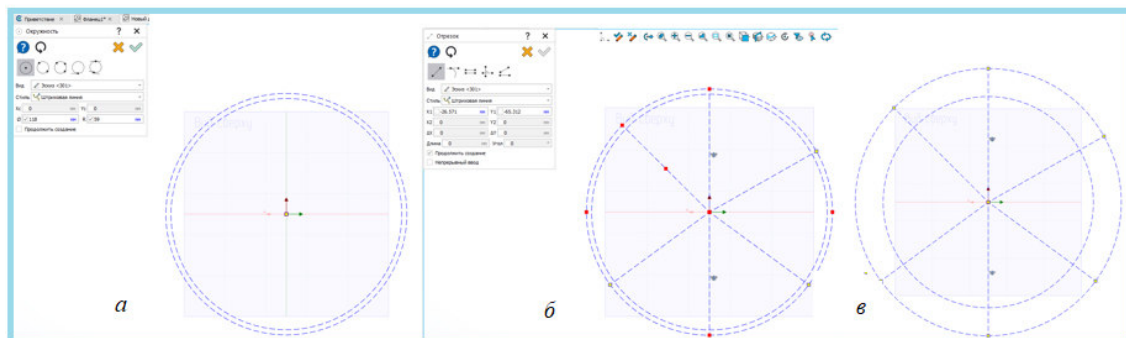


Рис. 5. Подготовка эскиза к параметризации



а) На выбранной плоскости строим две соосные окружности диаметром 118 мм и 112 мм штриховой линией. Эта линия играет роль вспомогательной при построении, но является образующей при параметризации (рис. 5, а). Проводим из центра окружностей шесть отрезков тоже штриховой линией, соблюдая примерное их расположение. Отрезки привязываем к окружности диаметром 118 мм (рис. 5, б). Привязку можно проверить, потянув за окружность, – при изменении диаметра она тянет за собой отрезки (рис. 5, в).

б) В группе «Параметризация» выбираем операцию «Ограничение» (рис. 6). Каждое ограничение на эскизе отобразится специальным значком (рис. 6, а). Фиксируем размеры окружностей: они стали черного цвета, что говорит о срабатывании ограничений параметризации.

Ставим ограничения на расположение отрезков:

- ставим один угловой размер  $60^\circ$  (рис. 6, б);
- позицию остальных линий можно ограничить ещё рядом угловых размеров. Но они будут загромождать эскиз. Выбираем симметрию;
- если эскиз правильно параметризован, он становится черным (рис. 6, в);
- эскиз может быть переопределён или сформирован с ошибкой. Тогда он принимает оранжево-желтый цвет (рис. 6, г).

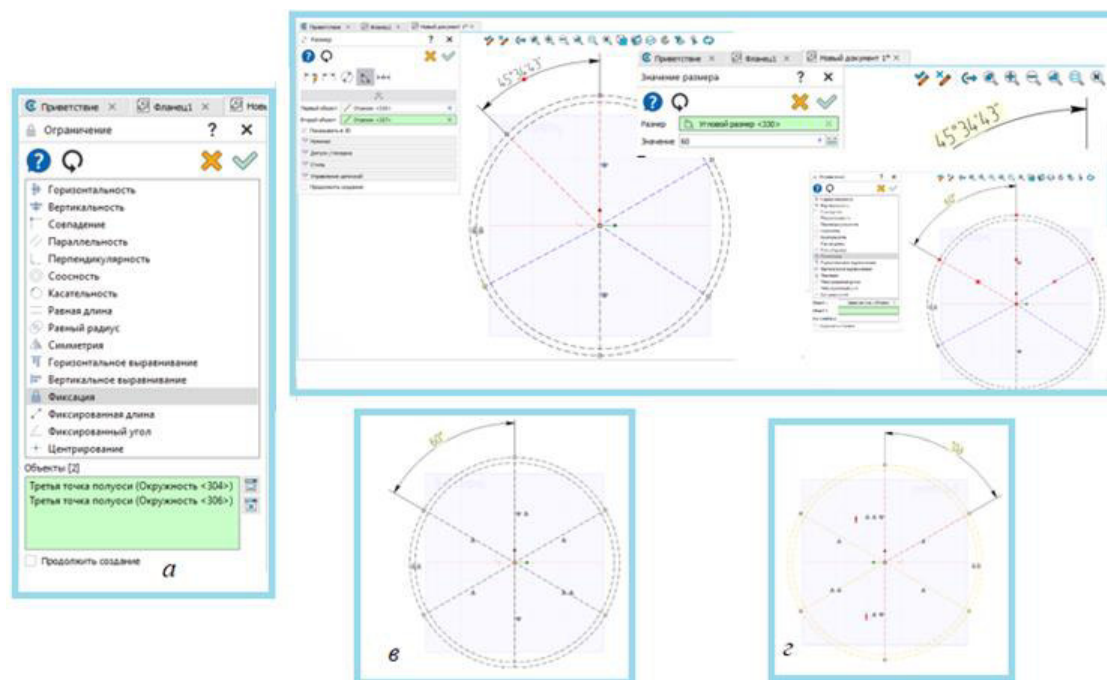


Рис. 6. Этап параметризации эскиза: наложение ограничений

в) Формируем криволинейную поверхность. Для этого выбираем операцию «Дуга» (рис. 7):

- для выпуклых дуги выбираем тип дуги «по центру, точке и углу». Строим без размеров, но небольшим радиусом, сопоставимым с размерами чертежом детали (рис. 7, а);
- вогнутые дуги строим по типу «дуга, касательная к двум прямым». Точки, определяющие прямую, выбираем на выпуклой дуге, иначе получаем «абракадабру» - сбой построения, как, например, на рисунке 8. Нижнюю точку дуги определяем, как она строится – впоследствии сделаем ограничение по касательности к окружности диаметром 112 мм, согласно чертежу (рис. 7, б);
- ставим ограничения ко всем дугам – по равенству диаметров, предварительно поставив радиус вогнутой дуги 40 мм. Ограничиваем вогнутые дуги касательной к диаметру 112 мм;
- лишние кривые обрезаем, найдя в группе «Линии» операцию «Кривые: обрезка/удлинение кривой»;
- полностью параметризованный эскиз постепенно приобретёт черный цвет. Если что-то останется синим, надо найти ограничение, которого не хватает (рис. 7, в).

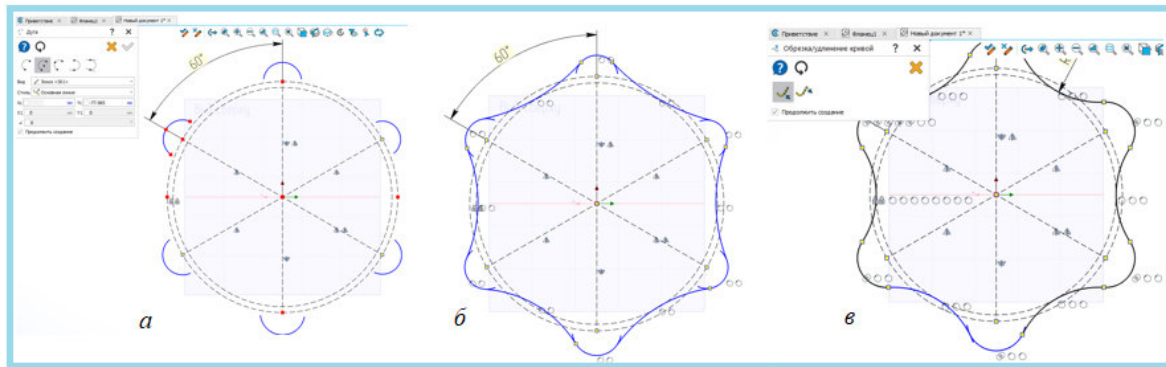


Рис. 7. Построение и параметризация дуг

г) Эскиз завершаем построением шести окружностей диаметром 14 мм (рис.8, а). Аккуратность построения соблюдаем, выделяя только те кривые, ограничения по которым выполняем. Иначе получаем обещанную ранее «абракадабру» (рис. 8, б). Отверстия также можно ограничить равенством диаметров и найти недостающее ограничение, например, проставить размер (рис. 8, в). Но для дальнейшего построения их включать в параметризированный эскиз необязательно. 3D-модель, выполненная операций «Вытягивание» на расстояние 8 мм, создаётся (рис. 8, г).

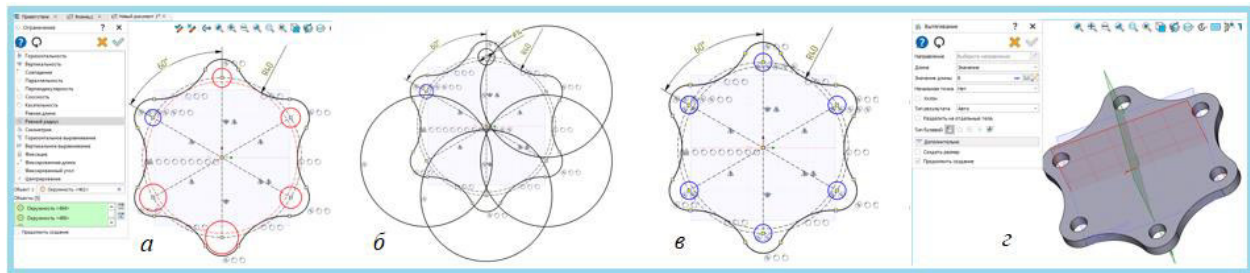


Рис. 8. Результат построения 3D-модели с криволинейной поверхностью с применением группы операций «Параметризация»

### 3. Завершение создания 3D-модели детали «Фланец»

3.1. Построение 3D-модели детали «Фланец» продолжим с применением ранее изученных операций. 3D-модель детали «Фланец» показана на рисунке 9, навигатор построения модели – на рисунке 10.

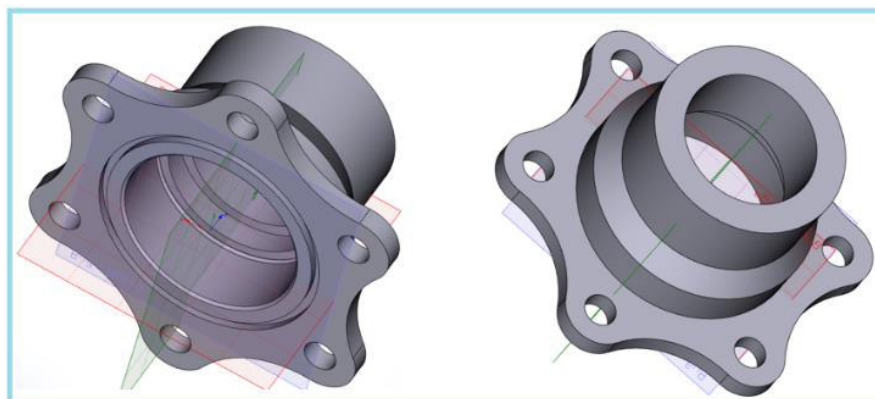


Рис. 9. 3D-модель детали «Фланец», представленная в разной ориентации

### Обсуждение результатов

В данной версии ПО «САРУС» 2024.2.1.1-VNF возможность построения сложных деталей, вплоть до IV конструктивной группы сложности, определяемых как «особо сложные детали (преимущественно отливки) с пространственными расположениями поверхностей» [3].

Как и в любом ПО, здесь необходимо нарабатывать практический опыт, определяемый не столько интуитивностью пользования CAD, сколько рефлексивным пониманием логики работы CAD «САРУС». Логика вырабатывается пользователем, некоторые основные её положения мы попытались отразить в данном методическом руководстве.

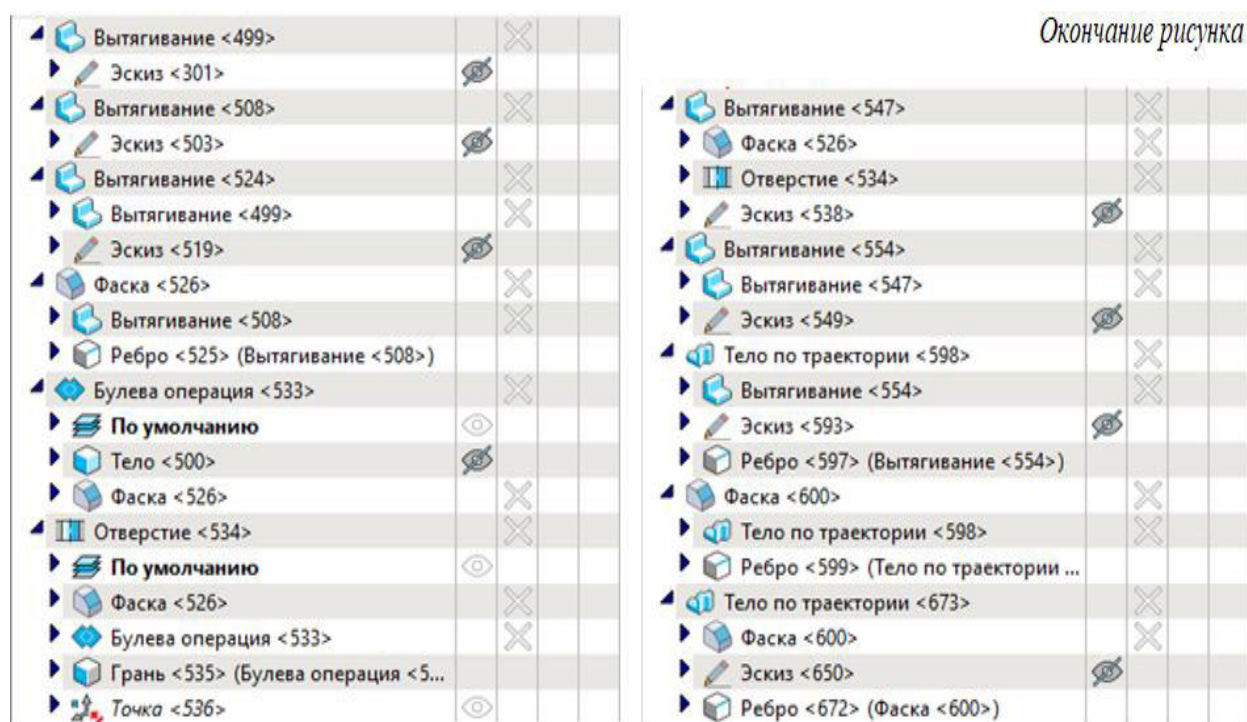


Рис. 10. История операций построения детали «Фланец»

Во избежание ошибок построения при большом количестве линий в эскизе, рационально использовать операции группы «Параметризация». При этом важно обратить внимание на то, что эскиз приобретает черный цвет. Это значит, что он определен. В Навигаторе модели в Истории построения эскиз не должен быть подсвечен красным знаком (!). Проверить, определён ли эскиз, можно, изменяя размеры.

Для успешной работы в CAD рекомендуется при пробных работах изучить как можно больше операций и функций для выбора оптимального алгоритма построения. Сделаем заключение по функциям и операциям, применяемым нами при создании 3D-модели.

Следует иметь в виду, что в данной CAD на сегодняшний момент трудоёмкость построения выше, чем в других, давно используемых CAD, например, «Компас-3D». Требуется частые расчёты размеров и координат, поэтому рекомендуется выбрать способ информационного обмена, удобный для пользователя (можно расчёты «держать в голове», можно записывать и т.п.). Однако эта трудоёмкость часто обусловлена правилами работы с платформой ПО. Так, выставленные в 3D-модели координаты не изменяются или корректируются незначительно в САМ «САРУС» при разработке управляющих программ [6].

Булева операция здесь имеет свою логику применения. «Сверхзадачей» булева алгоритма является способность работать и выдавать адекватный результат на произвольных пользовательских данных. Булева операция дает инженеру полную свободу в формообразовании, не привлекая каких-либо кинематических соображений (как, например, операция «заметания» вдоль траектории). Платой за гибкость является повышенная сложность подготовки операндов, так как булев алгоритм не является предметно-ориентированным, т.е. формируя конструктивный элемент посредством БО, инженер должен сам позаботиться о соблюдении всевозможных ограничений моделирования» [7].

При построениях выявлено некорректное выполнение некоторых функций:

- «САРУС» не имеет функцию отображать степени свободы, что немаловажно при построении параметризованного эскиза;

- «САРУС» не ловит оси начала координат при попытке создания кругового массива;
- «САРУС» не позволяет отменить выделенную линию путем повторного нажатия на нее;
- «САРУС» не ловит привязки к центру созданной ранее окружности;
- «САРУС» не позволяет вырезать в двух направлениях сразу, в каждом направлении на свою глубину, т.е. не создает отдельные параметры для каждого направления:
- «САРУС» некорректно отображает операцию «круговой массив»;
- в ПО «САРУС» некорректно работает строка отката в дереве;
- «САРУС» некорректно производит операцию «усечение кривой»;
- «САРУС» на данный момент не имеет рабочей функции «создать исполнение». Она есть, но только в замысле;
- «САРУС» начинает зависать и вылетать, если делать много операций за один подход; также при сохранении в абсолютно случайном порядке программа зависает и вылетает.

Однако, надо отметить, что эти недостатки выявлены из-за недостаточного опыта работы в САПР.

### **Заключение**

В ходе проведенного исследования была успешно продемонстрирована эффективность параметризации как метода оптимизации построения криволинейных поверхностей в модуле CAD системы «САРУС». На примере детали «Фланец» проанализированы ключевые этапы работы, включая применение операций «Круговой массив» и группы «Параметризация», что позволило выявить преимущества и особенности данного подхода.

Использование параметризации значительно упростило процесс проектирования, обеспечив гибкость и точность моделирования. Благодаря наложению ограничений и корректному выбору типов дуг удалось избежать ошибок построения и добиться полной управляемости модели. Это особенно важно при работе со сложными фасонными формами, где традиционные методы могут оказаться недостаточно эффективными.

Результаты работы подтвердили, что параметризация не только ускоряет процесс создания 3D-моделей, но и повышает их точность и адаптивность к изменениям. Это открывает новые возможности для проектирования в рамках системы «САРУС» и других CAD-решений.

Доработок в СПЖЦ «САРУС» требуют многие функции. Наиболее явные из них – удаление и редактирование отрезков и линий эскиза, доработка в целом команды «Обрезка/удлинение прямой», «Разделение кривой», потому что в некоторых случаях эти команды работают некорректно. При выполнении возврата предыдущей операции велика вероятность перестраивания эскиза в абсолютно случайном порядке и с непредсказуемым исходом, впоследствии восстановление наработки является труднореализуемой задачей.

В основном нам нравится логика построения деталей в СПЖЦ «САРУС». Программа отличается своим функционалом от других CAD, CAM-систем тем, что в ней есть полный жизненный цикл изделия. В моделировании присутствуют новые команды, которые еще не доводилось применять, но если они будут так же полезны, как примитивы, то СПЖЦ «САРУС» станет намного интереснее и полезнее многих CAD, CAM-систем.

По результатам исследования разработано методическое пособие «СПЖЦ «САРУС». Пользовательские сценарии создания 3D-моделей деталей средней сложности и сложных с пояснениями и иллюстрациями; методическое руководство по выполнению практических работ по дисциплине «Система полного жизненного цикла изделий "Сарус" RPLM V3» студентами, обучающимися по направлению 15.04.05 – Конструкторско-технологическое обеспечение машиностроительного производства.

### **Источник финансирования**

Договор от 25 марта 2024 г. № 96-2024/244 на выполнение научно-исследовательской работы по теме: «Проведение научных исследований в интересах ФГУП «РФЯЦ-ВНИИЭФ» с использованием результатов для актуализации образовательных программ подготовки специалистов с высшим профессиональным образованием для ФГУП «РФЯЦ-ВНИИЭФ», Задача 10 «Исследование программно-аппаратных решений в области цифровых

технологий, возможностей их отраслевого применения, разработка и версионное тестирование компонентов полностью защищенной системы полного жизненного цикла среднего класса СПЖЦ V3 «САРУС».

### Благодарности

Авторы благодарят кафедру цифровых технологий СарФТИ-НИЯУ МИФИ под руководством О.В. Кривошеева, к.т.н., зав. кафедрой, зам. директора ФГУП «РФЯЦ-ВНИИЭФ» по технологиям полного жизненного цикла – директор института цифровых технологий (ИТЦ) – главный конструктор СПЖЦ;

- научно-исследовательский отдел поддержки и внедрения Института цифровых технологий РФЯЦ-ВНИИЭФ, руководитель Г.С. Лазарев.

### Список литературы

1. Денисова Н.А., Федоренко Г.А. Применение в учебном процессе вуза тестирования программного модуля САД вновь создаваемого САПР «САРУС» // GraphiCon 2023: труды 33-й Междунар. конф. по компьютерной графике и машинному зрению (Москва, 19–21 сент. 2023 г.). М.: Институт прикладной математики им. М.В. Келдыша РАН, 2023. С. 848-859.
2. Отчёт о НИР «Исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖЦ V3» (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. номер НО/38-278.10-2023-О от 15.11.23. С. 64-88.
3. Отчет о научно-исследовательской работе исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖЦ V3. (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. № НО/37-278.10-2023-О. С. 8–27.
4. Самарина Е.А., Денисова Н.А. Исследование операции «Скругление» в программном обеспечении «Система полного жизненного цикла «САРУС» // GraphiCon 2024 : материалы 34-й Междунар. конф. по компьютерной графике и машинному зрению (Россия, Омск, 17–19 сент. 2024 г.). Омск : Изд-во ОмГТУ, 2024. С. 766–775.
5. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»: Основная версия программного модуля. «Система конструкторского проектирования» (версия 2)/Руководство оператора 07623615.00423-06 34 01. 2021. 338 с.
6. Прыткова Ю.Б., Денисова Н.А. Роль системы координат в построении 3D-модели в СПЖЦ «САРУС» // GraphiCon 2024 : материалы 34-й Междунар. конф. по компьютерной графике и машинному зрению (Россия, Омск, 17–19 сент. 2024 г.). Омск : Изд-во ОмГТУ, 2024. С. 795-804.
7. Жарков Н.А., Денисова Н.А. Особенности применения булевых операций в программном модуле «Система конструкторского проектирования» // GraphiCon 2024 : материалы 34-й Междунар. конф. по компьютерной графике и машинному зрению (Россия, Омск, 17–19 сент. 2024 г.). Омск : Изд-во ОмГТУ, 2024. С. 805-812.
8. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»: Программный модуль «Система конструкторского проектирования». Пояснительная записка технического проекта- 07623615.00423-12 81 01. Часть 1. 2021. 924 с.
9. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»: Программный модуль «Система конструкторского проектирования». Описание программы 07623615.00423-05 13 01. Часть 2. 2021. 80 с.
10. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»: Программный модуль «Система конструкторского проектирования». ПЗ ТП 07623615.00423-04 81 01-4. Часть 4. 2021. 282 с.



УДК 087.2

DOI: 10.25686/978-5-8158-2474-4-2025-734-743

## Оптимизация построения 3D-модели в СПЖЦ «САРУС» с использованием таблиц переменных

Д. А. Аладкин, И. В. Михайленко, Н. А. Денисова

Саровский физико-технический институт – филиал Национального исследовательского ядерного университета «МИФИ» (СарФТИ НИЯУ МИФИ), Саров, Россия

**Аннотация.** В условиях растущих требований к скорости и качеству моделирования эффективное управление параметрами становится ключевым фактором успеха в инженерном и дизайнерском процессах. Перед студентами кафедры специального машиностроения СарФТИ НИЯУ МИФИ поставлена задача тестирования САПР «Система полного жизненного цикла изделия «Цифровое предприятие» (СПЖЦ «САРУС») в связи с переходом РФЯЦ ВНИИЭФ на использование нового комплекса. Процесс тестирования нового САПР предусматривает в Техническом задании выявление и разработку оптимальных алгоритмов создания в ПМ CAD 3D-моделей различной сложности с целью обучения будущих пользователей.

Опыт работы в различных САПР (например, Компас-3D, Solid Work и др.) показал, что при построении 3D-моделей результативно используются таблицы переменных. Концепция таблиц переменных заключается в использовании структурированных данных для управления параметрами 3D-моделей. Таблицы переменных позволяют централизованно хранить и изменять ключевые параметры, что упрощает процесс моделирования и минимизирует ошибки при внесении изменений.

В статье рассматривается подход к оптимизации процесса создания 3D-модели на примере создания 3D-модели детали «Сфера» в СПЖЦ «САРУС» с использованием таблиц переменных.

**Ключевые слова:** программный модуль CAD, 3D-моделирование, таблицы переменных, параметрическая модель, связывание переменных.

## Optimization of the construction of a 3D model in the SARUS FLCS using tables of variables

D. A. Aladkin, I. V. Mikhailenko, N. A. Denisova

Sarov Institute of Physics and Technology – branch of the National Research Nuclear University "MEPhI" (SarPhTI NRNU MEPhI), Sarov, Russia

**Abstract.** With increasing demands on the speed and quality of modeling, effective parameter management is becoming a key success factor in engineering and design processes. The students of the Department of Special Mechanical Engineering of the SarPhTI of the National Research Nuclear University MEPhI were tasked with testing the CAD "Digital Enterprise Product Full Life Cycle System" (FLCS SARUS) in connection with the transition of the RFNC VNIIEF to the use of a new complex. The process of testing the new CAD provides in the Terms of Reference for the identification and development of optimal algorithms for creating 3D models of various complexity in PM CAD in order to train future users. Experience in various CAD systems (for example, Compass-3D, Solid Work, etc.) has shown that tables of variables are effectively used when building 3D models. The concept of variable tables is to use structured data to control the parameters of 3D models. Variable tables allow you to centrally store and change key parameters, which simplifies the modeling process and minimizes errors when making changes. The article discusses an approach to optimizing the process of creating a 3D model using the example of creating a 3D model of the Sphere part in the FLCS SARUS using tables of variables.

**Keywords:** CAD software module, 3D modeling, tables of variables, parametric model, linking variables.

### Введение

В условиях растущих требований к скорости и качеству моделирования эффективное управление параметрами становится ключевым фактором успеха в инженерном и дизайнерском процессах. Процесс тестирования нового САПР «Система полного жизненного цикла изделия «Цифровое предприятие» (далее – СПЖЦ «САРУС»), разрабатываемого во РФЯЦ-ВНИИЭФ, предусматривает в Техническом задании выявление оптимальных алгоритмов создания в ПМ CAD, входящего в СПЖЦ «САРУС», 3D-моделей различной сложности [1].

Опыт работы в различных САПР (например, Компас-3D, Solid Work и др.) показал, что при построении 3D-моделей результативно используются *таблицы переменных*. Концепция таблиц переменных заключается в использовании структурированных данных для управления параметрами 3D-моделей. Таблицы переменных позволяют централизованно хранить и изменять ключевые параметры, что упрощает процесс моделирования и минимизирует ошибки при внесении изменений.

С целью тестирования программного модуля САД СПЖЦ «САРУС» проведен анализ срабатывания функций создания, как пример, 3D-модели детали «Сфера» с использованием таблицы переменных в операции «Редактор переменных» [2]. **Основное её назначение** – быстрое присвоение значений внешним переменным файла, вставляемого в другой документ. Кроме того, таблицу переменных можно использовать для изменения текущих значений переменных в самом файле, содержащем таблицу. В качестве примера к рассмотрению предлагается создание 3D-модели детали «Сфера», имеющей сложную геометрическую форму по классификации А.П. Соколовского [3; 11-12] (рис. 1).

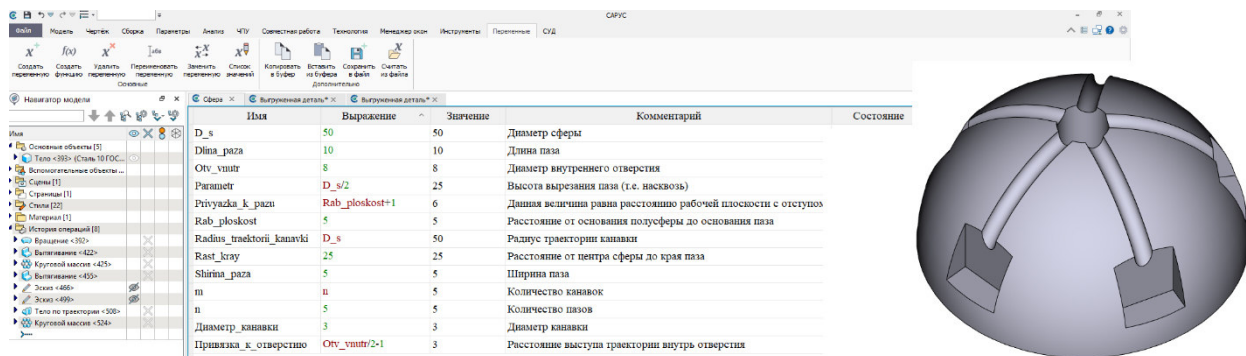


Рис. 1. 3D-модель детали «Сфера», созданная в САПР «САРУС» с использованием таблиц переменных

Одна из особенностей построения 3D-моделей заключается в способе параметризации размеров, которая осуществляется через операции группы «Редактор переменных». На начальном этапе моделирования конструктор базируется на СК, которой следует на протяжении всего процесса [4]. Традиционное построение элементов эскиза всегда сопряжено с изменением численных значений размеров детали. Постоянный пересчёт размеров в процессе построения усложняет задачу и ведёт к трудоёмкости её решения. Для снижения сложности построения в работе проверяется возможность создания эскиза через назначение переменных табличным способом. Это значительно упрощает трудоёмкость моделирования, что можно рассматривать как оптимизацию процесса.

В данной работе мы провели сравнение функции «Таблица переменных» в системе трехмерного моделирования КОМПАС и модуля САД в СПЖЦ «САРУС».

### Постановка задачи

Общее направление практико-ориентированного исследования сформулировано как исследование оптимальных алгоритмов построения криволинейных поверхностей в ПМ САД ПО «СПЖЦ «САРУС». Тестирование началось с изучения алгоритмов построения криволинейных поверхностей в ряде знакомых нам САПР, например, «Компас-3D», и дальнейшего аналитического сравнения сложности и производительности построений с СПЖЦ «САРУС». Далее осуществлялась разработка алгоритмов создания 3D-моделей простых, средней сложности и сложных деталей с разным типом криволинейных поверхностей с применением разных 3D-операций, делался аналитический вывод рациональности построений, выявлялись недостатки построения криволинейных поверхностей в ПО СПЖЦ «САРУС», формулировались рекомендации для доработки ПМ САД [3].

Исходя из общего плана исследования *целью* данной работы мы определили практико-ориентированное исследование возможности снижения сложности построения 3D-моделей сложных деталей, образованных криволинейными поверхностями при создании эскиза через назначение переменных табличным способом, предлагаемых интерфейсом программного модуля САД, входящего в систему СПЖЦ «САРУС».

#### Задачи:

- Анализ практического опыта способов работы с таблицами переменных в программном обеспечении «КОМПАС-3D».
- Сравнительный обзор функции «Таблица переменных» в системе трехмерного моделирования КОМПАС-3D и модуля САД в СПЖЦ «САРУС».
- Формулировка рекомендаций для доработки ПМ САД.



- Подготовка методических материалов по обучению работы с ПМ CAD вышеназванной программы студентов по направлению подготовки 15.03.04 – Конструкторско-технологическое обеспечение машиностроительных производств, а также специалистов, имеющих опыт работы с рядом САПР в направлении конструкторской подготовки.

Объектом исследования укажем ПМ CAD ПО «СПЖЦ «САРУС».

Предмет исследования – возможности предлагаемых операций построения 3D-моделей деталей с элементами криволинейных поверхностей в системе сквозного проектирования СПЖЦ «САРУС».

## Теория

Таблицы переменных (таблицы значений внешних переменных) – инструмент в системах автоматизированного проектирования (САПР), который содержит предопределённые значения внешних переменных файла. Такие таблицы используются:

- для создания исполнений деталей – например, для типовых деталей, отличающихся размерами, которые помещены в таблицу;
- реализации базы стандартных элементов – таблица переменных помогает реализовать базу данных с изменяемыми параметрами;
- работы с вставкой файла (компонента или фрагмента) с внешними переменными в другой документ – из таблицы выбирается строка, каждая ячейка которой содержит значение одной внешней переменной, и эти значения присваиваются внешним переменным вставляемого файла.

Таблицы переменных позволяют централизованно хранить и изменять ключевые параметры, что упрощает процесс моделирования и минимизирует ошибки при внесении изменений.

Структура таблиц обычно включает колонки для названий переменных, их значений и описаний, что обеспечивает ясность и удобство использования.

Например, «Таблица переменных» в «КОМПАС-3D» организована следующим образом (рис. 1):

- первая строка, начиная со второй ячейки, содержит имена переменных — заголовки столбцов таблицы;
- первый столбец, начиная со второй ячейки, содержит комментарии к строкам;
- остальные ячейки содержат значения переменных.

Комментарий	z	d	D	b	d1	a	c	r
Легкая серия, z=6	6	23	26	6	22.1	3.54	0.3	0.2
Легкая серия, z=8	8	32	36	6	30.4	2.71	0.4	0.3
Средняя серия, z=8	8	36	42	7	33.5	1.02	0.4	0.3
Средняя серия, z=10	10	82	92	12	77.1	3	0.5	0.5

Рис. 1. Окно «Таблица переменных» в Системе трехмерного моделирования КОМПАС-3D

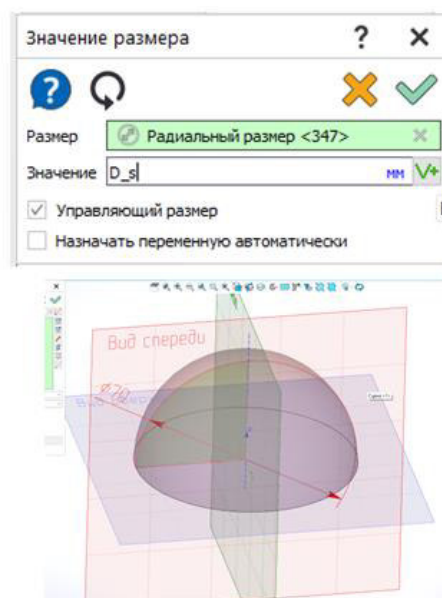


Рис. 2. Пример параметризации образмериванием диаметра окружности в СПЖЦ «САРУС»

Можно выделить особенности работы с «Таблицей переменных» в «КОМПАС-3D»: если в файле уже есть внешние переменные, можно использовать функцию «Читать внешние переменные», тогда в таблице появятся столбцы, соответствующие имеющимся внешним переменным файла.

Можно выбрать способ отображения во время вставки файла в другой документ таблицы «по умолчанию» – отображение всей таблицы или отображение первого столбца (только комментариев к строкам).

Методы параметризации включают:

1. *Параметрические модели* представляют собой подход, при котором модель определяется через набор параметров. Эти параметры могут быть изменены для анализа различных сценариев и оптимизации результатов. Например, инженерные конструкции: параметры, такие как размеры и материалы, могут быть настроены для достижения оптимальных характеристик конструкции (рис. 2).

2. *Связывание переменных* позволяет установить взаимосвязи между различными параметрами модели. Это может помочь в более точном управлении системой и улучшении ее производительности.

3. *Алгоритмы автоматизации* для динамического обновления моделей на основе изменений исходных данных, что повышает эффективность и точность проектирования.

Преимущества использования таблиц переменных в САПР:

- Удобство работы с несколькими исполнениями детали. Таблицы переменных позволяют унифицировать деталь, работать с несколькими вариантами без необходимости создавать отдельные эскизы.

- Возможность изменять геометрию без редактирования эскизов. Переменные позволяют работать с геометрией в параметрическом режиме, что упрощает процесс.

- Использование для поиска оптимального решения в оптимизационных задачах. Значения переменных, полученные в результате решения таких задач, могут применяться для создания проекта анимации.

- Использование для создания анимации. Например, с помощью переменных можно создавать эффект зубчатого зацепления, не прибегая к методу подбора.

Таким образом, таблицы переменных облегчают работу в САПР, позволяя выполнять различные задачи с минимальными усилиями и временем и способны:

- ✓ автоматизировать процесс: быстро обновлять модели при изменении значений в таблице;
- ✓ упрощать управление: легко редактировать и отслеживать параметры;
- ✓ к гибкости: возможности быстро адаптироваться к новым требованиям без необходимости ручного редактирования каждой модели.

Существуют некоторые ограничения таблиц переменных в САПР:

- Требования к формату файла. Например, в «КОМПАС-3D» импортируемая таблица не должна содержать полностью пустых строк или столбцов, а ячейки значений переменных могут содержать только действительные числа.

- Ограничения при импорте данных. Если таблица содержит данные неверного формата, то при импорте они заменяются на «0». Такое же сообщение появляется, если в импортируемой таблице есть пустые или объединённые ячейки.

- Ограничения при работе с ограничениями. Если наложено меньше ограничений, чем требуется, некоторые параметры не могут быть однозначно определены и могут изменяться в некотором диапазоне значений. В таких случаях говорят, что чертёж имеет степени свободы.

- Ограничения при работе с сборками. Переопределённая сборка содержит больше ограничений, чем требуется для взаимного расположения деталей в пространстве. Несовместимая сборка содержит параметрические ограничения, текущие значения которых невозможно удовлетворить.

## Результаты экспериментов

1. *Описание процесса создания 3D-моделей в системе трехмерного моделирования КОМПАС-3D с использованием таблиц переменных.*

Рассмотрим создание 3D-модели поэтапно. Выполнение каждого этапа проиллюстрировано на рисунках 3–8.

На этапе 1 (рис. 3) создаем эскиз (рис. 3, а) и задаем переменную (рис. 3, б).

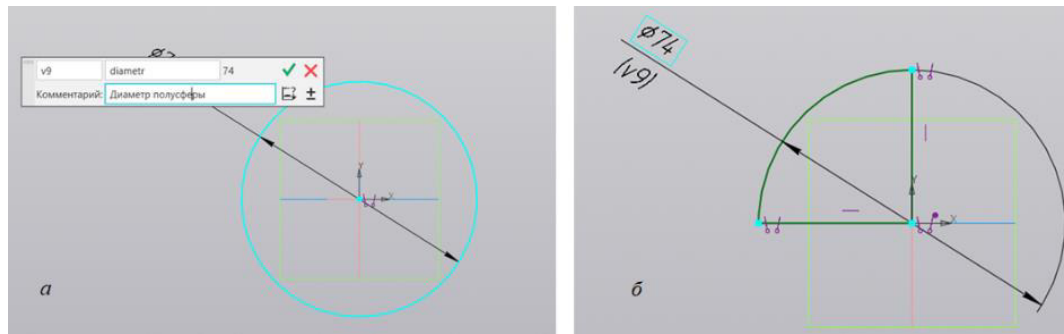


Рис. 3. Этап 1: а – создание эскиза; б – задание переменной

Этап 2 – создаем чертеж полусферы (рис.4, а) и при задании размера укажем переменную «diametr», также можно добавить в столбец «Комментарии» запись «Диаметр полусферы» для памяти и удобства дальнейшего использования таблицы. Выбираем функцию «Элемент вращения» и выбираем ось вращения для создания полусферы. Во вкладке переменные заполнилась первая строка таблицы (рис. 4, б).

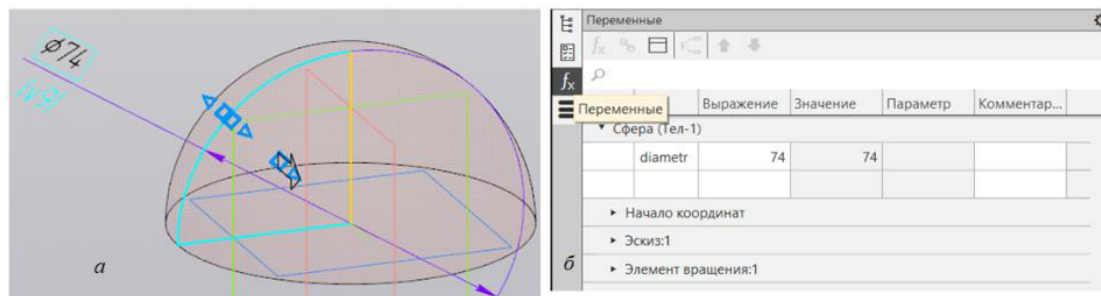


Рис. 4. Этап 2: а – создание чертежа полусферы; б – задание переменной в «Таблице переменных»

Этап 3 – создадим эскиз для сквозного отверстия в центре полусферы (рис. 5, а). Применяем операцию «Вырезать выдавливанием» и указываем высоту равную половине диаметра полусферы, прописывая её как «diametr/2» (рис. 5, б). В таблице переменных появляется новый параметр, численное значение которого мы можем изменять, тем самым меняя численное значение размера модели (рис. 5, в).

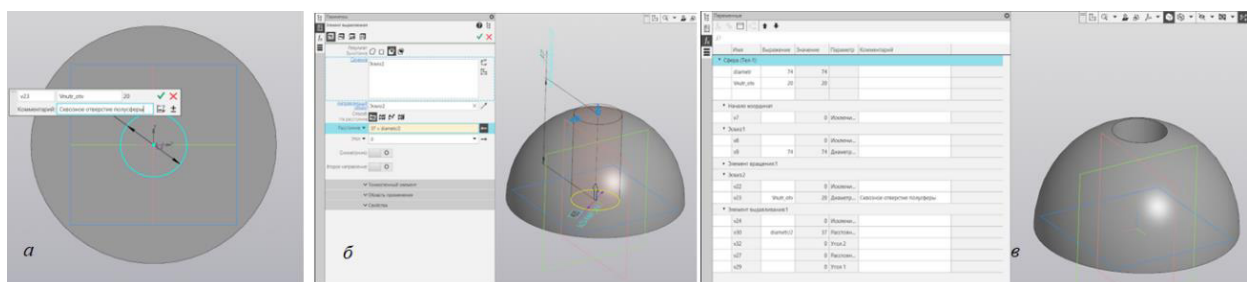


Рис. 5. Создание сквозного отверстия в центре полусферы

Этап 4. Аналогичным образом выполняем дальнейшие операции, при этом параметризируя все переменные. В конечном итоге мы имеем 3D-модель детали со связанной с ней таблицей параметров (рис. 6, а). Для того чтобы таблица переменных работала, нужно каждой переменной задать статус «внешняя» (рис. 6, б). Сохраняем таблицу переменных, привязанную к данной модели, файлом формата «Excel» (рис. 6, в).

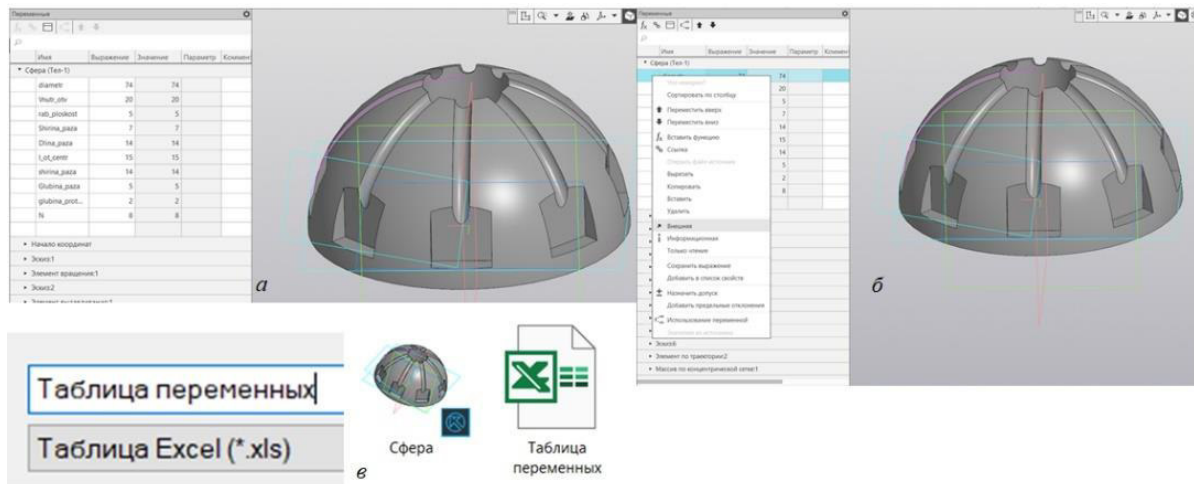


Рис. 6. Итог создания 3D-модели детали «Сфера»:

а – 3D-модель детали «Сфера» со связанной с ней таблицей переменных;

б – присвоение переменным статуса «Внешний»;

в - сохранение таблицы переменных, привязанную к данной модели, файлом формата «Excel»

Этап 5. При изменении численных значений в файле формата «Excel» параметры модели не изменяются сразу. Для этого необходимо открыть модель, перейти во вкладку «Таблица переменных» и кликнуть по функции «Читать из файла» (рис. 5, а).

Далее появится окно, где необходимо подтвердить замену данных таблицы (рис. 7, б).

Затем кликаем на любой столбец из таблицы, появляется кнопка «Присвоить значения переменным», прожимаем её и завершаем операцию, кликнув на функцию «Перестроить модель» (рис. 7, в).

На рисунке 8 показаны модели после изменения численных значений таблицы переменных.

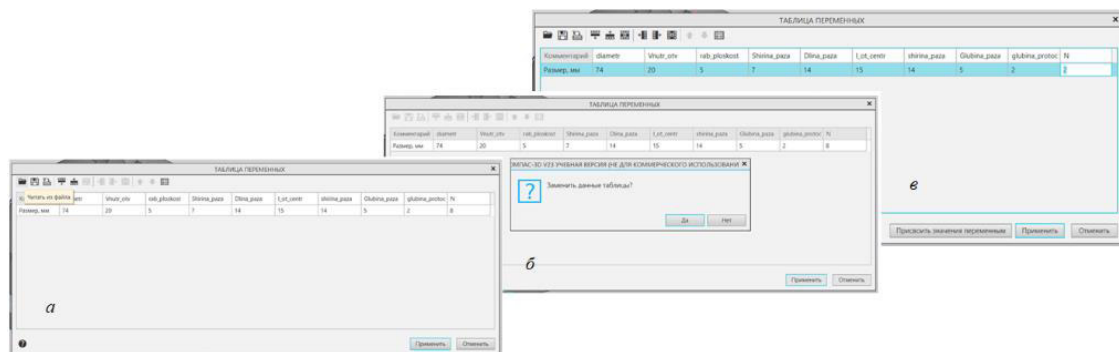


Рис. 7. Этап 5 «Корректировка параметров 3D модели при помощи численных значений таблицы переменных»

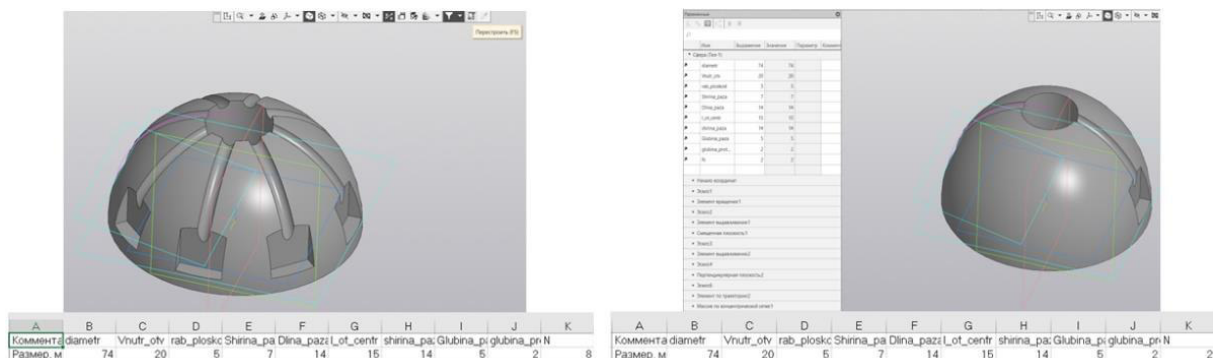


Рис. 8. Демонстрация моделей после изменения численных значений таблицы переменных

Таким образом, мы показали возможность быстрого изменения параметров 3D-модели, изменяя численные значения таблицы переменных в «КОМПАС-3D».

## 2. Исследование процесса создания 3D-моделей в СПЖЦ «САРУС» с использованием таблиц переменных

Процесс создания 3D-модели осуществлялся на обновленной базовой версии продукта 2024.2.1.1-VNF, утверждённой в 2024 году, тоже поэтапно. Выполнение каждого этапа проиллюстрировано на рисунке 9.

Этап 1. Создаём окружность произвольного диаметра и образмериваем её (рис. 9, а). Далее во вкладке «Размер», в графе «Значение», вместо численного значения задаём наименование параметра (по красной стрелке показано увеличенное диалоговое окно с заданием параметров, выделенное красным прямоугольником). В нашем случае наименование параметра «D\_s», его можно задавать как в латинской, так и в русской раскладке (рис. 9, б).

Далее необходимо воспользоваться командой «V+», справа от графы «Значение» (на рис. 9, в показано синей стрелкой, для создания новой переменной). Переходим во вкладку «Параметры» и используем команду «Редактор переменных». После этого будет создан новый документ, содержащий таблицу со всеми параметрами. На данном этапе таблица содержит лишь одну заданную переменную «D\_s». В таблице переменных, в столбце «Выражение» (рис. 9, в по зеленой стрелке), мы можем изменять численное значение размера, тем самым изменяя размер в 3D-модели. В столбце «Комментарий» задаем необходимое описание параметра, например, «Диаметр сферы».

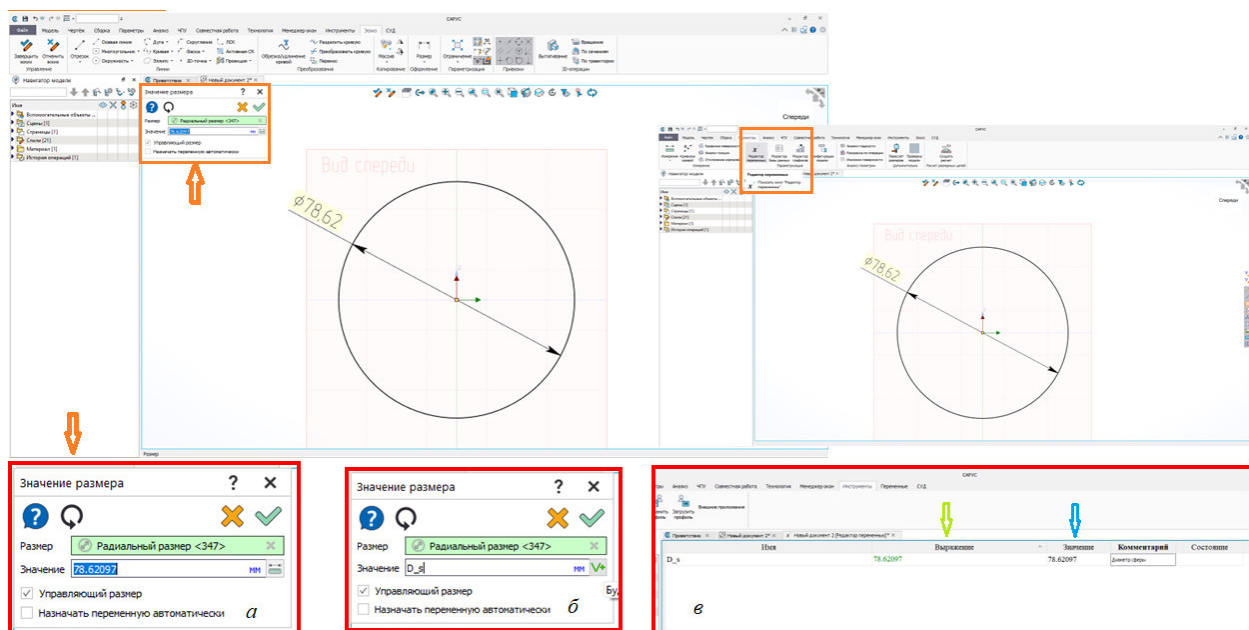


Рис. 9. Этап 1. Создание эскиза

Этап 2. Далее создаем полусферическую деталь с заданным диаметром с помощью команды «Вращение».

Создаем рабочую плоскость, параллельную плоскости вида сверху, и во вкладке плоскости, в графе «Расстояние», вместо численного значения задаем наименование переменной «Rasst\_ot\_osnov». Используем команду «V+» для добавления новой переменной, которая автоматически появится в таблице (рис. 10).

Полусферическую деталь с заданным диаметром создаем с помощью команды «Вращение».



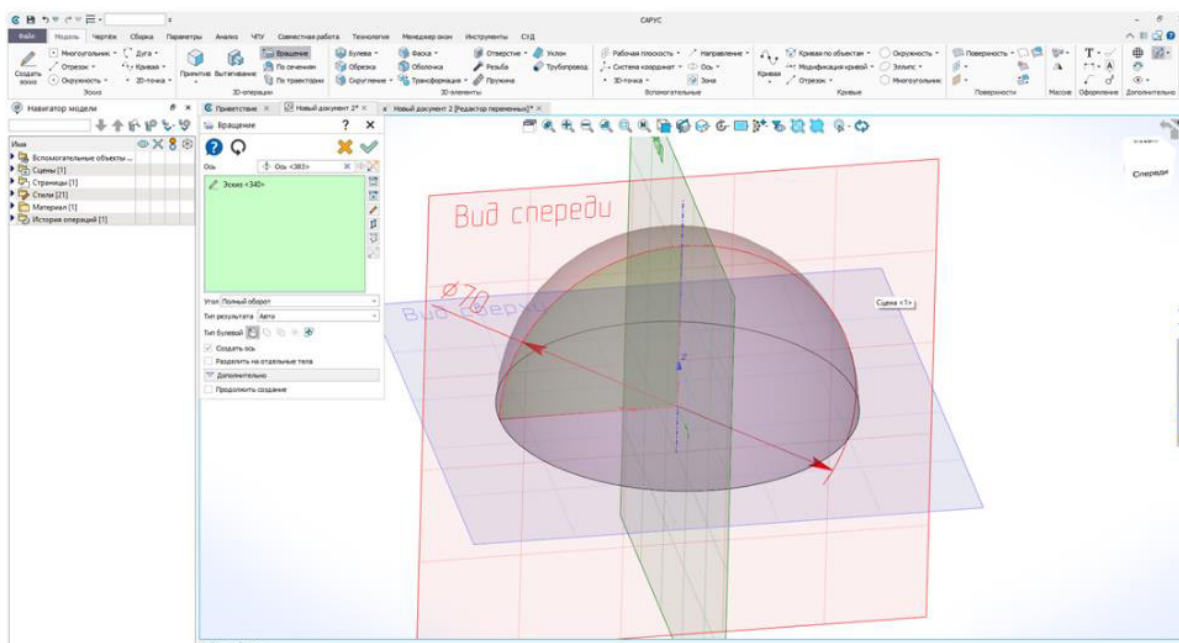


Рис. 10. Этап 2. Создание полусферы

Этап 3. В рабочей плоскости создаем эскиз будущего паза, образмериваем его, с проставлением наименований новых переменных, вместо численных значений (рис. 11).

С помощью операции «Вычитание» вырезаем паз, предварительно заменив в графе «Значение длины» численное значение на наименование переменной «Parametr».

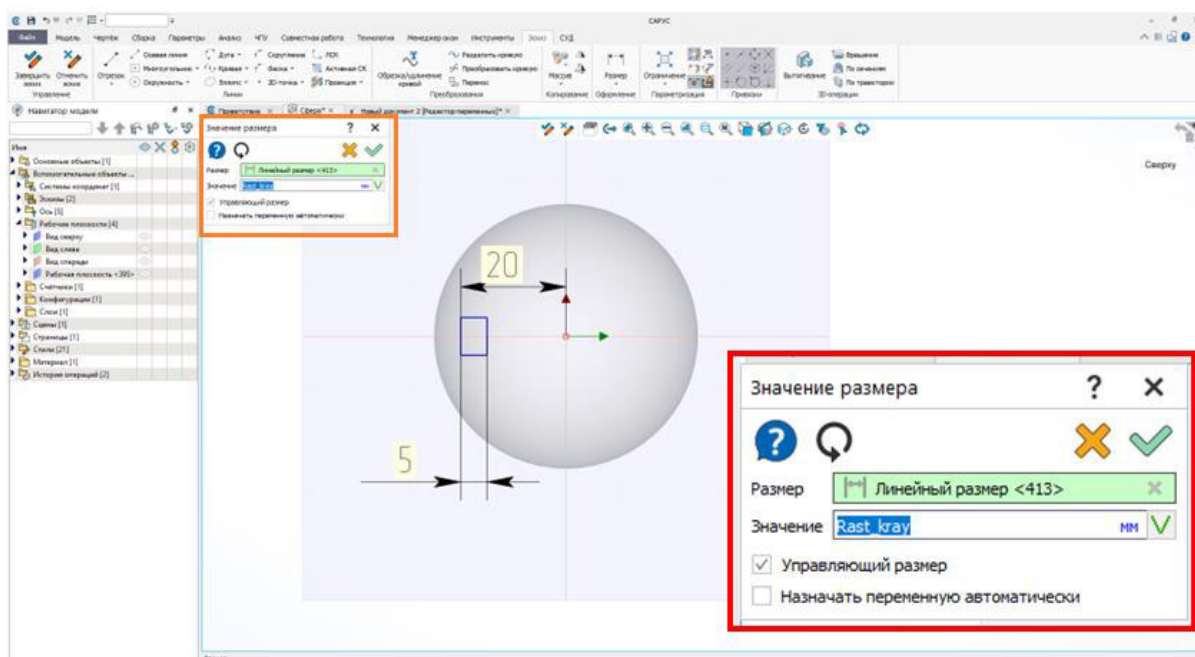


Рис. 11. Этап 3. Задание переменной для требуемой геометрии

Этап 4. Аналогичным образом проделываем дальнейшие операции, при этом параметризуя все переменные. В конечном итоге, мы имеем 3D-модель детали (рис. 12, а1), со связанной с ней таблицей переменных (рис. 12, б1).

Изменяя значения в таблице переменных, мы можем с легкостью редактировать параметры 3D-модели на необходимые нам численные значения (рис. 12, а2, б2).

После завершения работы, можно сохранить таблицу переменных, связанную с файлом, в текстовый документ формата «txt.», содержащий параметры 3D-модели, которую мы также можем

редактировать, изменяя при этом 3D-модель (рис. 12, *в1*, *в2*). Впоследствии при открытии модели для внесения изменений нам необходимо будет воспользоваться командой «Считать из файла».

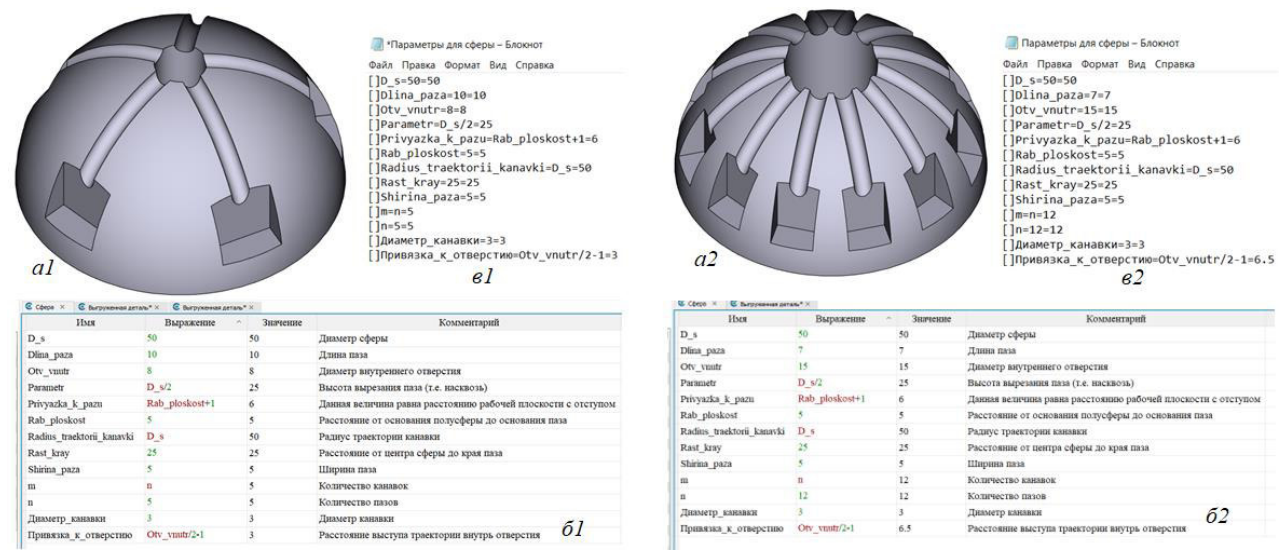


Рис. 12. Этап 4. Демонстрация моделей после изменения численных значений таблицы переменных

**Обсуждение результатов**

**1. Сравнительный анализ функции «Таблица переменных»**

После сравнения алгоритмов построения мы увидели, что результат в виде 3D-моделей в обеих программах создан. Однако в процессе работы есть что обсудить.

В таблицу 1 сведены критерии оценки результативности программ, сформулированные по результатам практического опыта и оцененные по оптимальности и удобству реализации функций.

**Интуитивность интерфейса:** в CAD-модуле СПЖЦ «САРУС» интерфейс понятен на интуитивном уровне, нет необходимости долго искать нужную команду.

**Простота реализации:** в CAD-модуле СПЖЦ «САРУС» для реализации функции «Таблица переменных» необходимо использовать меньшее количество операций.

**3D-моделирование:** в системе трехмерного моделирования КОМПАС-3D построение делается проще и быстрее. Это может быть связано в первую очередь с тем, что программа более привычна и имеется опыт работы с ней.

**Корректировка параметров модели:** применить отредактированные данные из сохраненного файла с таблицей переменных намного проще и удобнее в CAD-модуле СПЖЦ «САРУС», по сравнению с системой трехмерного моделирования КОМПАС-3D.

**Визуализация таблицы:** при добавлении переменных в CAD-модуль СПЖЦ «САРУС» таблица переменных имеет более оптимальную визуализацию.

Таблица 1. Сравнительный анализ функции «Таблица переменных» в ПМ CAD САРУС и ПО «КОМПАС-3D»

Характеристика	CAD модуль СПЖЦ «САРУС»	Система трехмерного моделирования КОМПАС-3D
Интуитивность интерфейса	+	-
Простота реализации	+	-
3D моделирование	-	+
Корректировка параметров модели	+	-
Визуализация таблицы	+	-
Итог	4	1



## **Заключение**

В статье было продемонстрировано создание параметризованных моделей в системе трехмерного моделирования КОМПАС-3D и САД-модуле СПЖЦ «САРУС».

Можно сделать вывод, что функция параметризации размеров и создания таблиц переменных интуитивно понятна и оптимизирована в САД-модуле СПЖЦ «САРУС», также данное программное обеспечение имеет ряд преимуществ для работы на площадках ФГУП РФЯЦ-ВНИИЭФ:

- импортонезависимость программного обеспечения, отсутствие санкционных рисков;
- защищенность данных: продукт сертифицирован на обработку данных, содержащих сведения, составляющие государственную тайну;
- полное соответствие требованиям российского законодательства.

Работа предназначена для специалистов в области 3D-моделирования, инженеров и всех заинтересованных в повышении эффективности работы с трехмерными данными специалистов.

## **Источник финансирования**

Договор от 25 марта 2024 г. № 96-2024/244 на выполнение научно-исследовательской работы по теме: «Проведение научных исследований в интересах ФГУП «РФЯЦ-ВНИИЭФ» с использованием результатов для актуализации образовательных программ подготовки специалистов с высшим профессиональным образованием для ФГУП «РФЯЦ-ВНИИЭФ», Задача 10 «Исследование программно-аппаратных решений в области цифровых технологий, возможностей их отраслевого применения, разработка и версионное тестирование компонентов полностью защищенной системы полного жизненного цикла среднего класса СПЖЦ V3 «САРУС».

## **Благодарности**

Авторы благодарят кафедру цифровых технологий СарФТИ-НИЯУ МИФИ под руководством О.В. Кривошеева, к.т.н., зав. кафедрой, зам. директора ФГУП «РФЯЦ-ВНИИЭФ» по технологиям полного жизненного цикла – директор института цифровых технологий (ИТЦ) – главный конструктор СПЖЦ.

## **Список литературы**

1. Денисова Н.А., Федоренко Г.А. Применение в учебном процессе вуза тестирования программного модуля САД вновь создаваемого САПР «САРУС» // GraphiCon 2023: труды 33-й Междунар. конф. по компьютерной графике и машинному зрению (Москва, 19–21 сент. 2023 г.). М.: Институт прикладной математики им. М.В. Келдыша РАН, 2023. С. 848-859.
2. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие». Основная версия программного модуля. «Система конструкторского проектирования» (версия 2) / Руководство оператора 07623615.00423-06 34 01. 2021. 338 с.
3. Отчёт о научно-исследовательской работе «Исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖЦ V3. (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. № НО/37-278.10-2023-О. С. 8–27.
4. Прыткова Ю.Б., Денисова Н.А. Роль системы координат в построении 3D-модели в СПЖЦ «САРУС». GraphiCon 2024: материалы 34-й Междунар. конф. по компьютерной графике и машинному зрению (Россия, Омск, 17–19 сент. 2024 г.). Омск, 2024. С. 795-804.

УДК 087.2

DOI: 10.25686/978-5-8158-2474-4-2025-744-754

## Исследование срабатывания функций группы «Ограничения» при построении фасонных поверхностей в программном обеспечении «САРУС»

А. С. Быкова, Н. А. Денисова

Саровский физико-технический институт – филиал Национального исследовательского ядерного университета «МИФИ», Саров, Россия

**Аннотация.** В процессе тестирования вновь разрабатываемого в РФЯЦ-ВНИИЭФ САПР «Система полного жизненного цикла изделий «Цифровое предприятие» или СПЖЦ «САРУС» студентами СарФТИ НИЯУ МИФИ выявляются оптимальные алгоритмы построения поверхностей разного рода, характерных при конструировании деталей машиностроения. Построение фасонных поверхностей является важной задачей в проектировании и моделировании в CAD. Результаты трех этапов тестирования программного модуля CAD СПЖЦ «САРУС» показали, что при моделировании объектов сложных форм выявлены проблемы построения поверхностей, имеющих криволинейные образующие, особенно радиусные линии. Поиск оптимальных алгоритмов построения подобных поверхностей привел к необходимости выявления срабатывания функций группы «Ограничения» при построении параметрической модели. В работе исследован метод построения криволинейных поверхностей средствами ПМ CAD, подробно описан оптимальный порядок построения радиусных кривых с учетом выявленных особенностей в работе команд. Показано, что СПЖЦ «САРУС» представляет собой инструмент, позволяющий создавать сложные трехмерные модели с использованием различных функций и операций, а функции группы «Ограничения» играют ключевую роль в срабатывании проектных задач, связанных с созданием фасонных поверхностей.

**Ключевые слова:** программный модуль CAD, 3D-моделирование, фасонная поверхность, радиусные линии, скругления.

## Implementation of testing of the new SARUS CAD system into the educational process of the university

A. S. Bykova, N. A. Denisova

Sarov Institute of Physics and Technology – branch of the National Research Nuclear University "MEPhI", Sarov, Russia

**Abstract.** In the process of testing the newly developed RFNC-VNIIEF CAD "Digital Enterprise Full Life Cycle System" or "SARUS" CAD, students of the SarFTI of the National Research Nuclear University MEPhI identify optimal algorithms for constructing surfaces of various kinds, typical for the design of mechanical engineering parts. The construction of shaped surfaces is an important task in CAD design and modeling. The results of the three stages of testing the CAD software module of the FLCS "SARUS" showed that problems were identified in constructing surfaces with curved shapes, when modeling objects of complex shapes, especially radius lines. The search for optimal algorithms for constructing such surfaces led to the need to identify the operation of the functions of the "Constraints" group when constructing a parametric model. The paper examines the method of constructing curved surfaces using PM CAD, describes in detail the optimal procedure for constructing radius curves, taking into account the identified features in the work of teams. It is shown, that the FLCS "SARUS" is a tool that allows you to create complex three-dimensional models using various functions and operations, and the functions of the "Constraints" group play a key role in triggering design tasks related to the creation of shaped surfaces.

**Keywords:** CAD software module, 3D modeling, shaped surface, radius lines, rounding.

### Введение

В процессе тестирования вновь разрабатываемого в РФЯЦ-ВНИИЭФ САПР «Система полного жизненного цикла изделий «Цифровое предприятие» или СПЖЦ «САРУС» студентами СарФТИ НИЯУ МИФИ выявляются оптимальные алгоритмы построения поверхностей разного рода, характерных при конструировании деталей машиностроения.

На предыдущих этапах исследования при тестировании версий 1.0.17.9 и 2023.1.0.1-RC16 [1, 2] мы обратили внимание на необходимость выявления возможностей применения операций, предлагаемых интерфейсом ПМ CAD по построению криволинейных поверхностей в 2D- и 3D-режимах при создании 3D-моделей деталей машиностроения различной сложности.

Так, например, при попытке построения детали «Шаблон», которую можно представить как объединение простейших геометрических фигур, сопрягаемых радиусными кривыми (рис. 1, а), в версии 1.0.16.23 построение не осуществилось командой «Сопряжения», как в Компас-3D. Необходимо было строить окружности с удалением лишних кривых. При выборе команды «Кривые, удаление и образка кривой» всплывала критическая ошибка или, если удавалось что-то обрезать, эскиз слетал, необходимо было перестроение (рис. 1, б). При повторении построения в более поздней версии (1.0.19.76) 3D-модель детали была построена (рис. 1, в), но замечена нестабильность программы: операции, которые успешно выполнялись при построении модели, могут не выполняться (программа выдаёт критическую ошибку) при построении по тому же алгоритму. После успешного построения детали при ее повторном построении по тому же алгоритму некоторые операции не выполняются. Выявлена некорректная работа размеров.

По оценке исполнителей, трудоемкость построения выше, чем в остальных сравниваемых ПО [1].

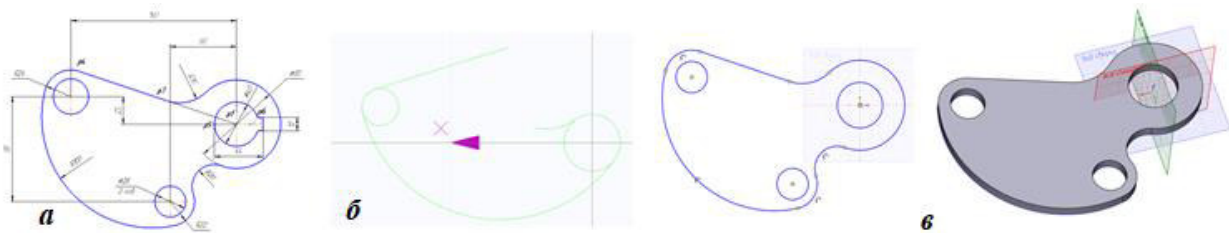


Рисунок 1. Построения эскиза для создания 3D-модели детали «Шаблон»:

а – эскиз детали «Шаблон»;

б – попытка построения 3D-модели «по эскизу» в версии 1.0.16.23;

в – 3D-модель детали, построенный «по эскизу» в версии 1.0.19.76

Ясно, что построение криволинейных поверхностей с фасонной образующей является важной задачей в проектировании и моделировании в CAD.

Исходя из вышеизложенного, в Техническом задании для реализации следующего этапа практико-ориентированного исследования запланированы следующие действия:

- разработка алгоритмов создания 3D-моделей **деталей средней сложности** с различным типом криволинейных поверхностей с применением предлагаемых программным обеспечением 3D-операций;
- формулирование аналитического вывода рациональности построений;
- выявление недостатков построения криволинейных поверхностей в ПО «СПЖЦ ЦП»; формулирование рекомендаций для доработки ПМ.

### Постановка задачи

Целью тестирования на данном этапе является исследование возможности применения операций, предлагаемых интерфейсом ПМ CAD, входящему в систему «САРУС», по построению криволинейных поверхностей в 2D- и 3D-режимах в деталях средней сложности с элементами сложной формы с теоретическим анализом предлагаемых функций построения.

Задачи:

- тестирование ПМ CAD по срабатыванию выполнения функций, операций и команд построения, рекомендуемых разработчиком при построении 3D-модели детали с фасонной образующей, сформированной радиусными линиями;
- описание логики выбора и разработки оптимального алгоритма построения криволинейных поверхностей в ПМ CAD «САРУС» в 3D-модели детали «Крюк чалочный» (рис. 2);
- анализ работоспособности алгоритмов построения скруглений при построении 3D-модели;
- аналитическое сравнение построений исходной детали с ПО «КОМПАС-3D»;
- подготовка методических материалов по обучению работы с ПМ CAD «САРУС» студентов по направлению подготовки 15.03.04 – Конструкторско-технологическое обеспечение

машиностроительных производств, а также специалистов, имеющих опыт работы с рядом САПР в направлении конструкторской подготовки.



Рисунок 2. Пример изделия «Крюк чалочный»

## Теория

### 1. Криволинейная поверхность

Определим понятие криволинейной поверхности как основы детали машиностроения, конструирование которой нас интересует, с использованием терминологического словаря по начертательной геометрии [4].

Согласно терминологии начертательной геометрии, любая деталь машин рассматривается как ГЕОМЕТРИЧЕСКОЕ ТЕЛО – некоторая замкнутая часть пространства, ограниченная плоскими или кривыми поверхностями.

ПОВЕРХНОСТЬ в элементарной геометрии определяется как граница тела или как след движущейся линии (не вдоль себя). Различают поверхности плоские, выпуклые, вогнутые, а в различных разделах математики они имеют более сложную классификацию. В начертательной геометрии пользуются кинематическим способом образования поверхности путем непрерывного перемещения образующей линии в пространстве, причем производящая линия при своем движении может как сохранять свою форму, так и менять ее.

ЛИНИЯ КРИВАЯ определяется как траектория непрерывно движущейся точки в постоянно изменяющемся направлении. Кривая, все точки которой принадлежат одной плоскости, называется плоской. Кривая, все точки которой не могут принадлежать одной плоскости, называется пространственной. Такая линия имеет двоякую кривизну. Кривые линии, как плоские, так и пространственные, могут быть закономерными или случайного вида. Свойства кривых изучаются в аналитической и дифференциальной геометрии, а также в топологии. Единственная кривая, изучаемая в элементарной геометрии, – окружность.

ПОВЕРХНОСТЬ ВРАЩЕНИЯ – это поверхность, образованная вращением какой-либо образующей линии вокруг неподвижной прямой – оси. Производящая (образующая) линия может быть прямой, кривой, ломаной и составной; замкнутой и незамкнутой; плоской и пространственной. Совсем необязательно, чтобы образующая лежала в плоскости, проходящей через ось вращения (например, прямая образующая однополостного гиперболоида вращения). Если начало и конец незамкнутой линии лежат на оси вращения, то поверхность получится замкнутой. Всякая замкнутая поверхность вращения образует тело вращения.

ПОВЕРХНОСТИ ВТОРОГО ПОРЯДКА. В аналитической геометрии так называют поверхности, уравнения которых в прямоугольной системе координат – уравнения второй степени. К ним относятся сфера, эллипсоиды, однополостной и двуполостной гиперболоиды, эллиптический и гиперболический параболоиды, конические и цилиндрические поверхности. Прямая линия пересекает такие поверхности в двух точках.

ПОВЕРХНОСТИ КИНЕМАТИЧЕСКИЕ – это поверхности, образуемые движением какой-либо линии (поверхности) в пространстве, например цилиндрические или конические поверхности. Движущийся элемент называется производящим или образующим. У цилиндра или конуса образующая – прямая линия. Перемещение производящей линии называется поступательным, если все ее точки перемещаются на равные расстояния и по параллельным траекториям.

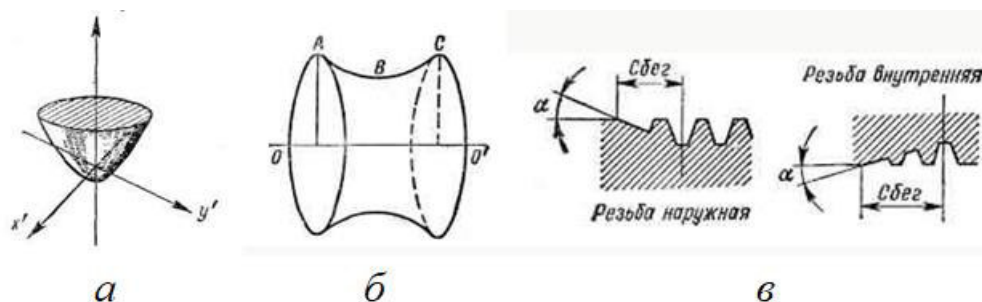


Рисунок 3. Примеры криволинейных поверхностей:  
а – параболоид вращения; б – катеноид; в – резьбовые поверхности

Таким образом, можно заключить, что все детали в машиностроении являются комбинированными геометрическими телами (рис. 3), состоящими из примитивных элементов, представляющими собой простые геометрические тела.

Тогда нас будут интересовать типичные детали и их поверхности: валы ступенчатые, сферические поверхности, резьбовые и зубчатые поверхности, корпусные детали с элементами криволинейных поверхностей – сопряжения плоских и кривых поверхностей.

**СОПРЯЖЕНИЕ ЛИНИЙ** определяется как плавный переход одной линии в другую. Общая для этих линий точка называется точкой сопряжения или точкой перехода. Точка сопряжения двух дуг окружностей лежит на линии их центров. Точка касания прямой и окружности служит точкой их сопряжения.

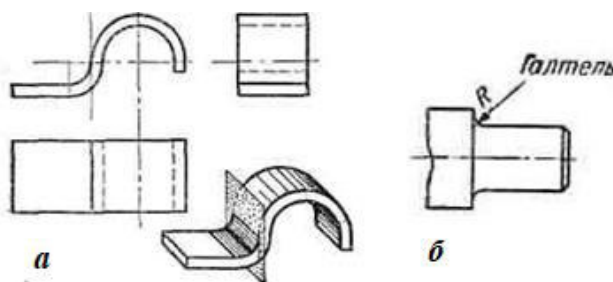


Рисунок 4. Примеры сопряжений:  
а – пример детали с сопряжением поверхностей; б – эскиз ступенчатого вала с галтелью

**СОПРЯЖЕНИЕ ПОВЕРХНОСТЕЙ** есть плавный переход одной поверхности в другую. Линия касания поверхностей называется линией разграничения. Ее показывают только на той проекции, перпендикулярно которой можно провести через линию разграничения поверхность (цилиндр, плоскость), касательную к обеим сопрягаемым поверхностям (рис. 4, а).

**ГАЛТЕЛЬ** (нем. *Hohlkehle* — выкружка) – криволинейная поверхность плавного перехода от меньшего сечения вала к плоской части заплечика или буртика (рис. 4, б).

## 2. Анализ группы «Ограничения» в САПР

Построение фасонных поверхностей является важной задачей в проектировании и моделировании в различных областях, таких как CAD (Computer-Aided Design) и CAM (Computer-Aided Manufacturing). Программное обеспечение «САРУС» представляет собой мощный инструмент, позволяющий создавать сложные трехмерные модели с использованием различных функций и инструментов. В данном исследовании рассматриваются функции группы «Ограничения», которые играют ключевую роль в срабатывании проектных задач, связанных с созданием фасонных поверхностей.

Фасонные поверхности, как описано выше, характеризуются сложной геометрией и могут принимать различные формы, такие как цилиндры, конусы, сферы и более сложные сплайны. Эти поверхности имеют широкий диапазон применения: от автомобилей и самолетов до медицинских устройств и художественных изделий.

Программное обеспечение «САРУС» разработано для эффективного моделирования и анализа, позволяя пользователям работать с большим объемом данных и приложений. Исследование показало, что возможности «САРУС» в контексте фасонных поверхностей значительно расширяются благодаря наличию функции «Ограничения».

Функции группы «Ограничения» в «САРУС» позволяют задавать дополнительные параметры, которые регулируют форму и свойства создаваемых поверхностей. Эти ограничения могут включать, но не ограничиваясь:

- 1) геометрическими – параметрами размеров и форм;
- 2) физическими – свойствами материалов;
- 3) технологическими – условиями производства и обработки.

В 90-е годы XX века началось активное развитие CAD-систем, которые вскоре стали стандартом в индустрии. Первые версии программного обеспечения «САРУС» появились на рынке в начале 2010-х, предоставляя пользователям инструменты для работы с 2D- и 3D-геометрией. В процессе развития системы специалисты начали внедрять функции ограничения, что позволило улучшить качество проектирования и повысить точность моделирования (рис. 5).

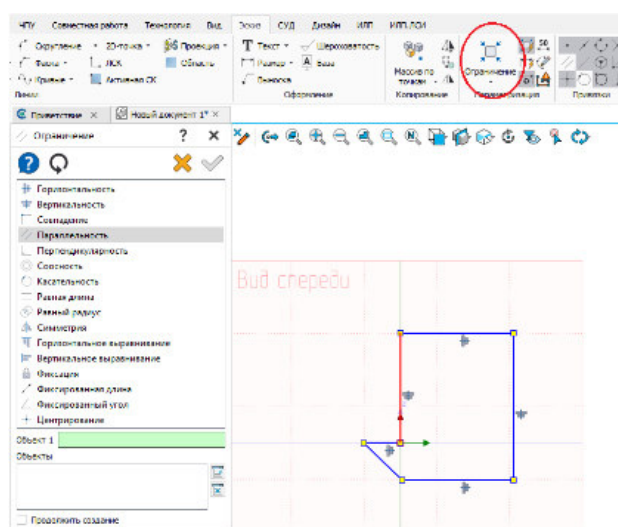


Рисунок 5. Функция «Ограничение» на Ленте CAD «САРУС»

На сегодняшний день функции группы «Ограничения» в «САРУС» продолжают развиваться и актуализироваться. Операторы ограничений позволили значительно упростить процесс создания сложных фасонных поверхностей, улучшив автоматизацию процессов проектирования. Исследования показывают, что применение этих функций снижает ошибки в проектировании на 25-30 %.

## Результаты экспериментов

### 1. Исследование оптимальных алгоритмов построения криволинейных поверхностей в ПМ CAD ПО «СПЖЦ «САРУС»

Объектом исследования стали функции группы «Ограничения», которые интегрированы в программное обеспечение «САРУС». Они представляют собой важный инструмент, позволяющий проектировщикам создавать сложные геометрические формы. Предметом исследования являются свойства и качества этих функций, а также возможность их применения для повышения качества проектирования и точности создания фасонных поверхностей.

**Анализ детали.** Деталь «Крюк чалочный» (рис. 6) можно представить как объединение простейших геометрических фигур, а именно окружностей, сопрягаемых радиусными кривыми.

**Стратегия построения 3D-модели.** Построение можно выполнить операцией «Вытягивание» через создание эскиза. За начало координат принимаем центр окружности диаметром 56 мм. Последовательность построения показана на рисунках 7, 8.

На начальных этапах, согласно размерам эскиза, производим построение окружностей.



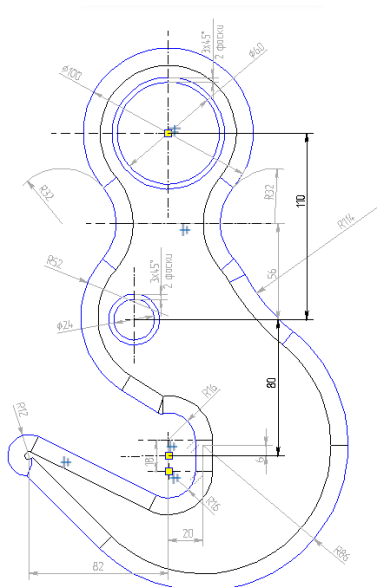


Рисунок 6. Эскиз детали «Крюк чалочный». Эскиз выполнен в ПМ САД «САРУС»

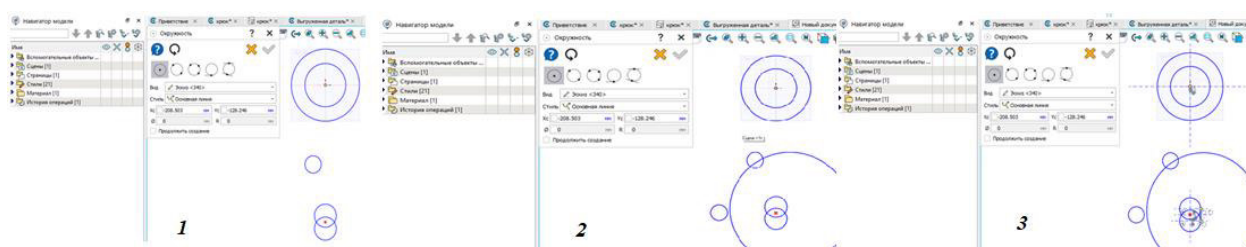


Рисунок 7. Этапы построения эскиза с помощью окружностей

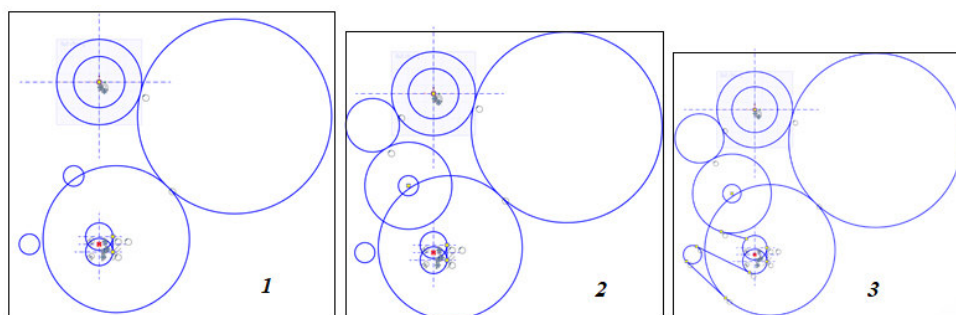


Рисунок 8. Этапы построения эскиза с примерным представлением очертания крюка

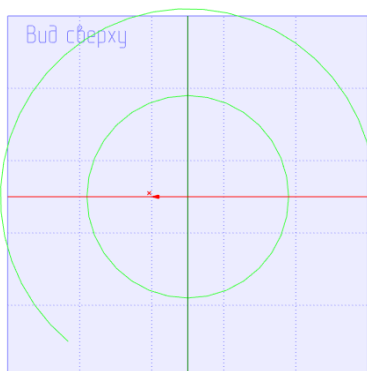


Рисунок 9. Недостроенный эскиз для построения 3D-модели детали «Крюк чалочный»



Построение не может осуществляться командой «Сопряжения», как в ПО «Компас». Необходимо строить окружности с удалением лишних кривых. При выборе команды «Кривые» при удалении и обрезке кривой всплывает критическая ошибка или, если удастся что-то обрезать (рис. 9), эскиз слетает – перестраивается.

К сожалению, на данном этапе проектирования эскиза произошла ошибка (выделена красным овалом на рис. 10), при изменении диаметра одной из окружностей, остальные элементы начинают изменяться автоматически. Такое бывает, когда программа содержит большое количество элементов.

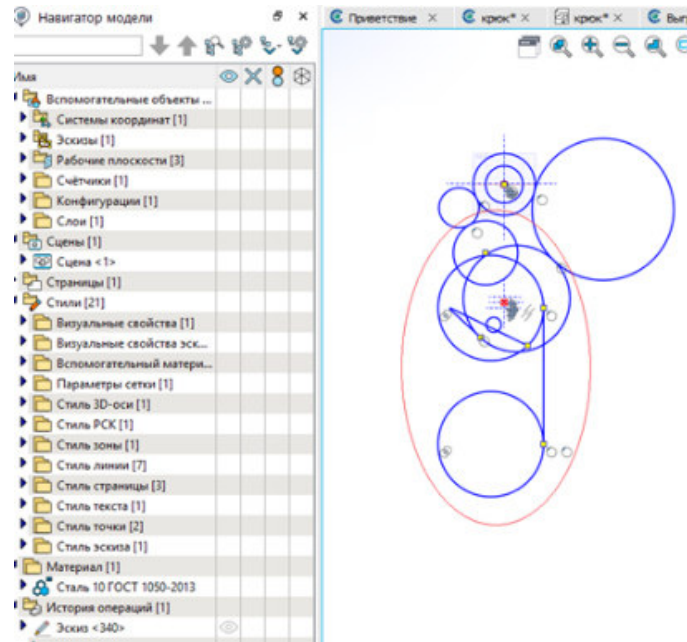


Рисунок 10. Возникновение ошибки при большом количестве элементов

После исправления ошибок эскиз выглядит так, как показано на рисунке 11. Это уже окончательный вид эскиза. На данном этапе (при удалении лишних элементов) ошибок не возникало.

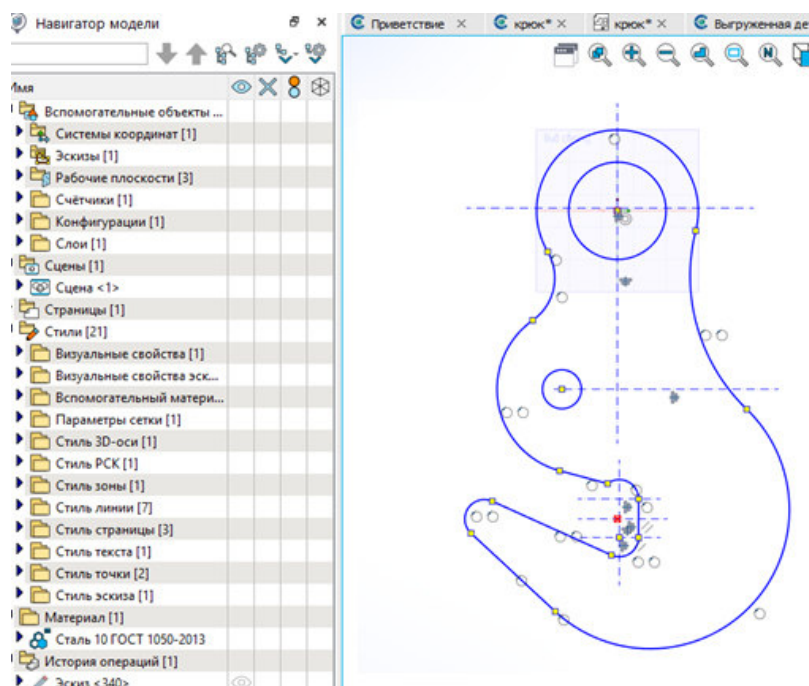


Рисунок 11. Окончательный вид эскиза

Вариант разработки 3D-модели крюка имеет вид, представленный на рисунке 12, но это не окончательный вариант.

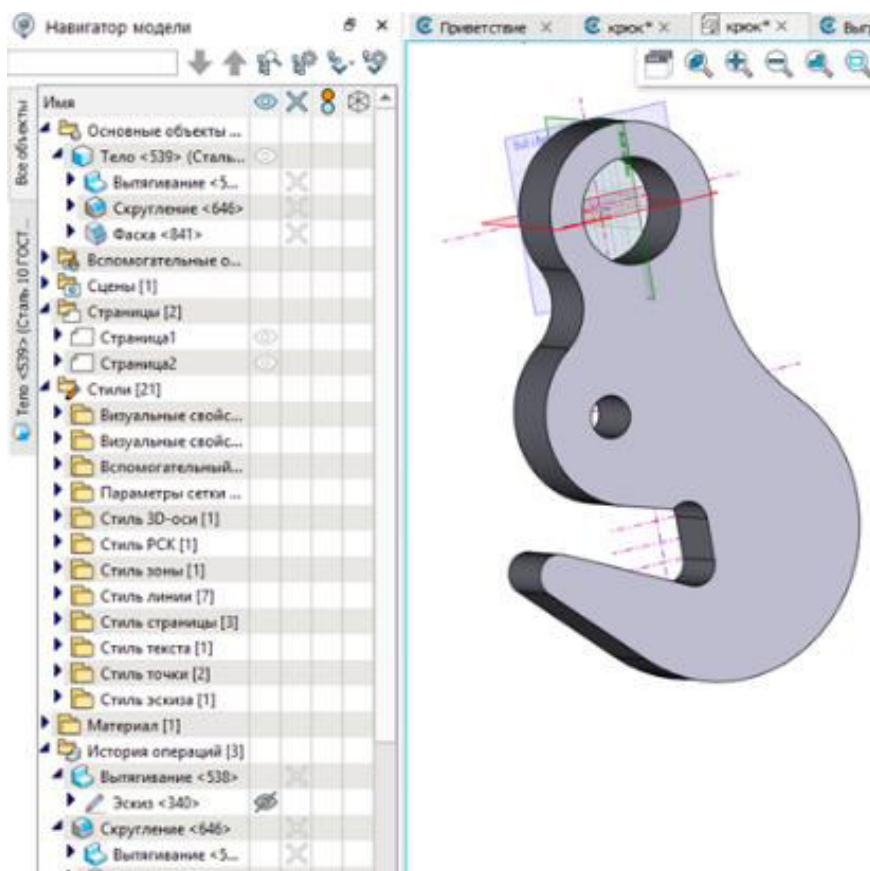


Рисунок 12. 3D-модель детали «Крюк чалочный» после выполнения операции «Выдавливания»

Для окончательного формирования 3D-модели необходимо выполнить операцию «Скругление». Ранее была проведена исследовательская работа в связи с проблемой построения скруглений в более ранних версиях «САРУС». Необходимо было проверить работоспособность ряда функций «Скругление» в СПЖЦ «САРУС», получившего сертификат САПР базового уровня. Был сделан вывод, что все алгоритмы, кроме управления опциями перекрытия скруглений – взаимной обрезки или сопряжения скруглений с наращиванием материала, оказались работоспособными [6], что было подтверждено выполнением операции при построении 3D-модели исходной детали.

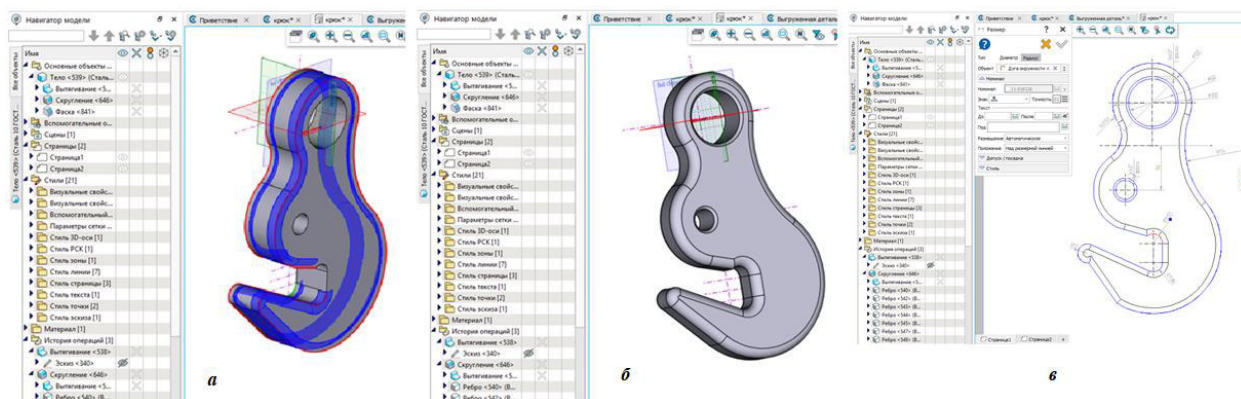


Рисунок 13. Окончание построения 3D-модели детали «Крюк чалочный»:

а – операция «Скругление»; б – окончательный вид модели; в – построение ассоциативного чертежа детали

В данном САД очень удобно, что все ребра выделяются сразу по контуру (за исключением небольшого участка в глубине крюка). Не нужно нажимать на каждую дугу, чтобы реализовать операцию (рис. 13, а). После добавления фасок и скруглений деталь принимает внешний вид, показанный на рисунке 13, б.

Заканчиваем работу построением ассоциативного чертежа с простановкой размеров (рис. 13, в).

## 2. Построения детали «Крюк чалочный» в ПО «КОМПАС-3D»

Для проведения сравнительного анализа, запланированного в задачах, создадим 3D-модель детали «Крюк чалочный» в ПО «КОМПАС-3D», аналогичный исходному.

Этапы проектирования эскиза в программном обеспечении «КОМПАС» представлены на рисунке 14.

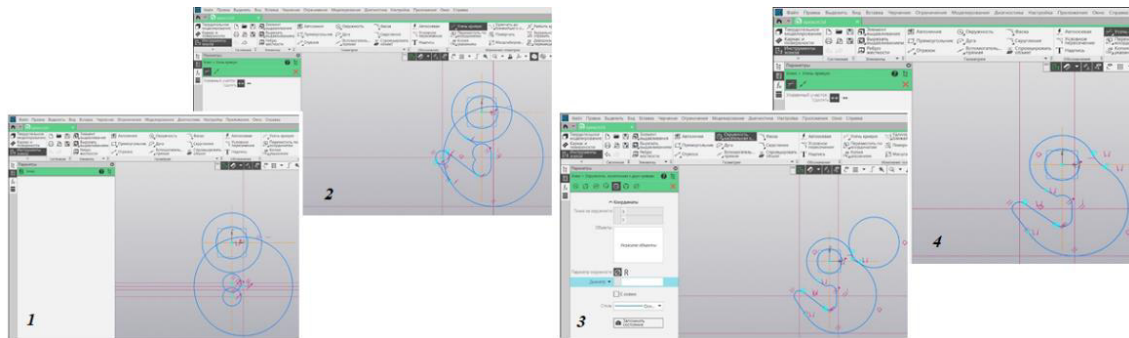


Рисунок 14. Этапы проектирования эскиза детали «Крюк чалочный» в ПО «КОМПАС-3D»

После построения эскиза в тот момент, когда необходимо было делать операцию «Выдавливание», в некоторых частях эскиза возникла ошибка, представленная на рисунке 15, а. Чтобы исправить эту ошибку пришлось уделить большое внимание просмотру всех стыков окружностей, всех соединений элементов. Это можно сделать вручную и потратить большое количество времени, а можно воспользоваться операцией «Проверка замкнутости» (рис. 15, б). При проверке были найдены незамкнутые контуры, представленные на рисунке 15, в. Стоит отметить, при проектировании в ПО «САРУС» таких проблем не возникло.

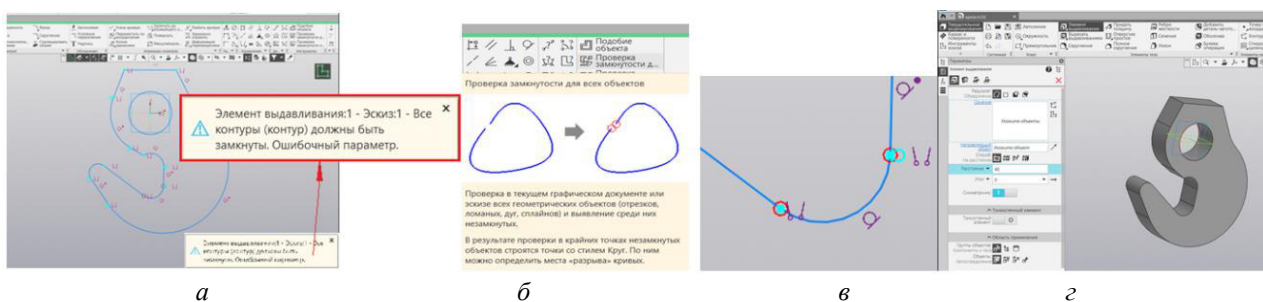


Рисунок 15. Выявление и исправление ошибки построения детали «Крюк чалочный» в ПО «КОМПАС-3D»: а – возникновение ошибки на этапе создания операции «Выдавливание»; б – выполнение операции «Проверка замкнутости»; в – выявление незамкнутых контуров; г – 3D-модель детали после операции «Выдавливание» и исправления ошибок

На заключительном этапе была применена операция «Скругления» (рис. 16, а). На этом этапе в ПО «КОМПАС» необходимо было выделить каждое ребро детали, что достаточно неудобно. В ПО «САРУС» такая операция применяется сразу для всех ребер. Окончательный вид детали «Крюк чалочный» показан на рисунке 16, б.

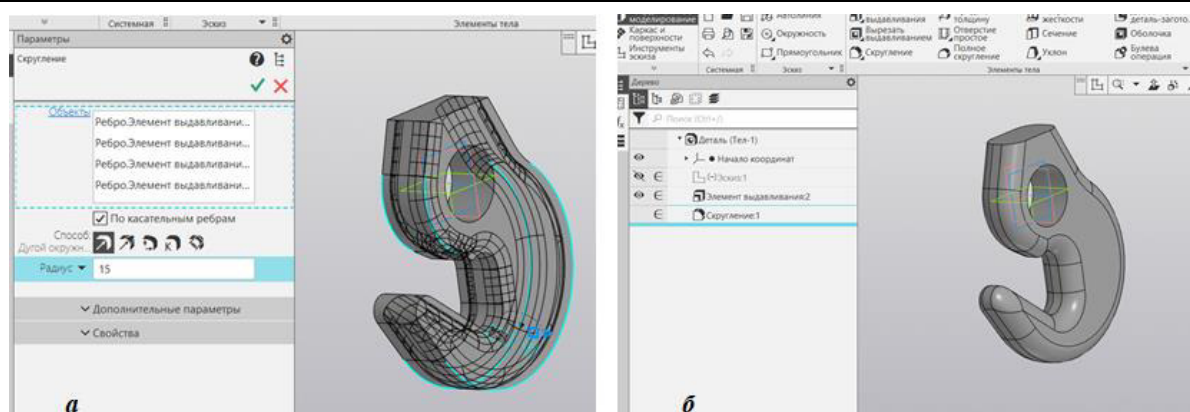


Рисунок 16. Окончание построения 3D-модели исходной детали:  
а – применение операции «Скругление»; б – окончательный вид детали «Крюк чалочный»

### Обсуждение результатов

При построении детали «Крюк чалочный» в КОМПАС-3D возникли некоторые сложности. В-первых, здесь оказался менее удобным функционал для нахождения незамкнутых контуров. Также при построении скруглений в ПО «КОМПАС» необходимо было выбрать каждое ребро с двух сторон, что не очень удобно при построении, в тот момент как в СПЖЦ «САРУС» можно выбрать одну сторону, и выделение будет сразу на всем участке.

Это привело к заключению, что 3D-модель детали «Крюк чалочный» в СПЖЦ «САРУС» построилась более корректно, чем в КОМПАС-3D.

### Заключение

Проведенное исследование показало, что функции группы «Ограничения» в САРУС эффективно срабатывают при выполнении задач по проектированию фасонных поверхностей. Основные результаты включают:

1. Эффективность реакции на изменения: при изменении значений параметров автоматически пересчитывались связанные с ними значения, что значительно упрощает процесс проектирования;
2. Поддержка сложных зависимостей: система позволяет учитывать как простые зависимости, так и более сложные логические условия, что повышает гибкость проектирования;
3. Выявление проблем: при тестировании были выявлены некоторые ограничения в реализации нескольких зависимостей одновременно, что требует доработки интерфейсов и алгоритмов.

Исследование работы функций группы «Ограничения» при построении фасонных поверхностей в системе проектирования жизненного цикла «САРУС» является важной задачей, поскольку эти функции позволяют контролировать процесс моделирования сложных геометрических форм.

Несмотря на существующие проблемы и вызовы, дальнейшее развитие и оптимизация этих функций могут привести к значительному улучшению процессов проектирования. Понимание и анализ срабатывания этих функций позволяет более эффективно управлять проектами и повышать качество конечных изделий. Будущие направления развития связаны с упрощением работы пользователей и оптимизацией ресурсов, что способно расширить применение ПО «САРУС» в промышленности.

### Источник финансирования

Договор от 25 марта 2024 г. № 96-2024/244 на выполнение научно-исследовательской работы по теме «Проведение научных исследований в интересах ФГУП «РФЯЦ-ВНИИЭФ» с использованием результатов для актуализации образовательных программ подготовки специалистов с высшим профессиональным образованием для ФГУП «РФЯЦ-ВНИИЭФ», Задача 10 «Исследование программно-аппаратных решений в области цифровых технологий, возможностей их отраслевого применения, разработка и версионное тестирование компонентов полностью защищенной системы полного жизненного цикла среднего класса СПЖЦ V3 «САРУС».

### Благодарность

Авторы благодарят кафедру цифровых технологий СарФТИ-НИЯУ МИФИ под руководством О.В. Кривошеева, к.т.н., зав. кафедрой, зам. директора ФГУП «РФЯЦ-ВНИИЭФ» по технологиям полного жизненного цикла – директора института цифровых технологий (ИТЦ) – главного конструктора СПЖЦ.



**Список литературы**

1. Отчёт о НИР «Исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖЦ V3» (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. номер НО/38-278.10-2023-О от 15.11.23. 2023. С. 64-88.
2. Отчет о научно-исследовательской работе исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖЦ V3 (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. № НО/37-278.10-2023-О. 2024. С. 8–27.
3. Отчет о научно-исследовательской работе исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖЦ V3 (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. № НО/37-278.10-2023-О. 2025. С. 10–38.
4. Словарь терминов по начертательной геометрии и инженерной графике / Новосиб. гос. аграр. ун-т, ИИ; сост. Т.В. Семенова, Г.А. Евдокимова, Е.В. Петрова. Новосибирск, 2010. 120 с. Режим доступа: <https://textarchive.ru/c-2996429-pall.html>
5. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие». Основная версия программного модуля. «Система конструкторского проектирования» (версия 2) / Руководство оператора 07623615.00423-06 34 01. 2021. 338 с.
6. Самарина Е.А., Денисова Н.А. Исследование операции «Скругление» в программном обеспечении «Система полного жизненного цикла «САРУС» // GraphiCon 2024 : материалы 34-й Междунар. конф. по компьютерной графике и машинному зрению (Россия, Омск, 17–19 сент. 2024 г.). Омск : Изд-во ОмГТУ, 2024. С. 766–775.

## **Жизненный цикл детали от проектирования сложных криволинейных поверхностей в ПМ САД до создания управляющей программы в ПМ САМ «САРУС»**

**И. И. Макаров, Н. А. Денисова**

Саровский физико-технический институт – филиал Национального исследовательского ядерного университета «МИФИ», Саров, Россия

**Аннотация.** Основные задачи, поставленные перед разработчиками отечественного программного обеспечения «Система полного жизненного цикла изделий «Цифровое предприятие» (актуальное наименование – СПЦЖ «САРУС») в РФЯЦ-ВНИИЭФ, – это следование концепции сквозного управления бизнес-процессами отраслевых предприятий на основе современных информационных технологий. Это позволяет конструировать трёхмерные объекты в защищенном формате, а также проектировать их изготовление в объеме жизненного цикла изделия. Данные задачи являются ключевыми в развитии отраслей машиностроения, авиастроения, судостроения, ракетостроения и многих других. В такой системе, как СПЦЖ «САРУС», возможна реализация всех стадий жизненного цикла проектирования.

В данной работе исследованы такие стадии, как создание управляющей параметризированной 3D-модели в системе САД «САРУС» как первичной; разработка управляющей программы обработки для станков с ЧПУ в системе САМ «САРУС» по первичной 3D-модели. Рассмотрено построение 3D-модели детали «Компрессор низкого давления турбореактивного двигателя». По модели выполнена управляющая программа для механической обработки.

**Ключевые слова:** программный модуль САД, программный модуль САМ, прямой метод взаимодействия; первичная 3D-модель; управляющая программа

## **The life cycle of a part starts from designing complex curved surfaces in PM CAD to creating a control program in PM CAM SARUS**

**I. I. Makarov, N. A. Denisova**

Sarov Institute of Physics and Technology – branch of the National Research Nuclear University "MEPhI", Sarov, Russia

**Abstract.** The main tasks assigned to the developers of the Russian software "Digital Enterprise Full Life Cycle System" (current name – FLCS SARUS) at RFNC-VNIIEPh are following the concept of end-to-end management of business processes of industry enterprises based on modern information technologies. It allows you to design three-dimensional objects in a secure format, as well as to design their manufacture in the scope of the product life cycle. These tasks are key to the development of the branches of mechanical engineering, aircraft construction, shipbuilding, rocket engineering and many others. In such a system as the SARUS FLCS, it is possible to implement all stages of the design lifecycle.

In this paper, we study such stages as the creation of a parameterized 3D control model in the SARUS CAD system as the primary one; the development of a control-processing program for CNC machines in the SARUS CAM system based on the primary 3D model. The construction of a 3D model of the "Low-pressure compressor of a turbojet engine" part is considered. A control program for mechanical processing is executing according to the model.

**Keywords:** CAD software module, CAM software module, direct interaction method, primary 3D model; control program.

### **Введение**

Основные задачи, поставленные перед разработчиками отечественного программного обеспечения «Система полного жизненного цикла изделий «Цифровое предприятие» (актуальное наименование – СПЦЖ «САРУС») в РФЯЦ-ВНИИЭФ, – это следование концепции сквозного управления бизнес-процессами отраслевых предприятий на основе современных информационных технологий. Это позволяет конструировать трёхмерные объекты в защищенном формате, а также проектировать их изготовление в объеме жизненного цикла изделия. Данные задачи являются ключевыми в развитии машиностроения, авиастроения, судостроения, ракетостроения и многих других отраслей.

Главной задачей современных высокотехнологических предприятий является оптимизация операций сквозного проектирования изделий в машиностроении. Сквозное проектирование является актуальной задачей, так как это в разы увеличивает производительность и эффективность в рамках

производства. Изменяя только параметры детали, не перестраивая модель полностью, можно достигнуть оптимизации процесса.

Одной из таких разработок является **Комплекс программ в защищенном исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»** – система автоматизированного проектирования (САПР), разработанной на базе интегрированной инженерной программной платформы (ИИПП) и программно-математического ядра трехмерного моделирования «РГЯ». САПР представляет собой автоматизированную организационно-техническую систему, реализующую информационную технологию выполнения функций проектирования и предназначенную для автоматизации процесса проектирования. СПЖЦ «Цифровое предприятие» функционально состоит из персонала и комплекса технических, программных и других средств автоматизации его деятельности [1].

В такой системе, как СПЖЦ «САРУС», возможна реализация всех стадий жизненного цикла проектирования. В данной работе исследованы следующие стадии:

- создание управляющей параметризированной 3D-модели в системе CAD «САРУС» как первичной;
- разработка управляющей программы обработки для станков с ЧПУ в системе CAM «САРУС» по первичной 3D-модели [2].

Как пример в работе рассмотрено построение 3D-модели детали «Компрессор низкого давления турбореактивного двигателя» (рис. 1). По модели выполнена управляющая программа для механической обработки. Ожидается, что результаты данной работы позволят в дальнейшем более эффективно создавать управляемые параметризованные модели сложных форм и управляемые программы для обработки их на станках с ЧПУ.

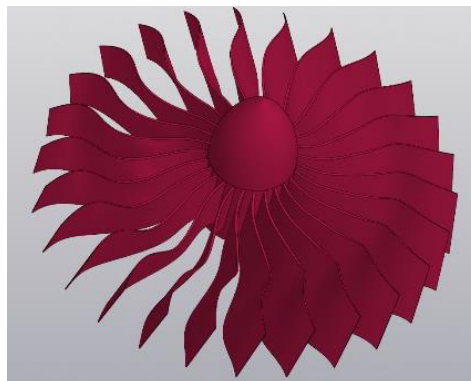


Рисунок 1. 3D-модель компрессора низкого давления турбореактивного двигателя

### Постановка задачи

Цель работы – провести практико-ориентированное исследование и аналитический анализ срабатывания программных модулей интегрированной инженерной программной платформы (ИИПП) «САРУС» путем создания первичной 3D-модели исходной детали и создания управляющей программы ее обработки в системе жизненного цикла изделия.

Задачи:

1. Создание управляющей параметризированной 3D-модели в программном модуле (ПМ) CAD системы «САРУС» как первичной;
2. Создание управляющей параметризированной 3D-модели в «КОМПАС-3D», сравнение трудоемкости построения;
3. Разработка управляющей программы обработки для станков с ЧПУ в программном модуле (ПМ) CAM системы «САРУС» по 3D-модели.

### Теория

В настоящее время для разработки производства какого-либо изделия применяются компьютерные средства моделирования на основе CAD-систем, что позволяет наглядно понимать конечный результат



и внешний вид изделия в режиме реального времени. Внедрение таких систем значительно повысило качество работы конструкторов и технологов, поскольку в них реализовано большое количество стандартных изделий в соответствии со стандартами ГОСТ и ISO. CAD (или САПР) представляет собой организационно-техническую систему, предназначенную для автоматизации процесса проектирования, состоящую из персонала и комплекса технических, программных и других средств автоматизации его деятельности [1].

Необходимо отметить, что внедрение разработанного изделия в производство требует применения современных компьютерных средств управления, обработки и хранения информации. Её эволюцию в виде нескольких «ступеней», включающих такие важные стадии, как концепция, разработка, производство, эксплуатация и окончательное выведение из эксплуатации обычно понимают как жизненный цикл технической (инженерной) системы. В последние десятилетия получили развитие технологии управления жизненным циклом изделия, которые включают в себя как CAD-системы, так и системы управления производственным процессом от создания технологии производства в соответствии с CAD-моделью до хранения архивной копии чертежей и технологии изготовления изделия.

В настоящее время политическая и экономическая обстановка диктует необходимость в использовании отечественных систем управления жизненным циклом изделия и актуальной становится цель – анализ доступных систем управления жизненным циклом изделия, построенных на импортозамещенных технологиях.

Управление данными в информационном пространстве, едином для различных автоматизированных систем, возлагается на систему управления жизненным циклом продукции – *PLM (Product Lifecycle Management)*. Технологии PLM объединяют методы и средства информационной поддержки изделий на всех этапах их жизненного цикла. При этом обеспечивается взаимодействие как средств автоматизации разных производителей, так и различных автоматизированных систем многих предприятий, то есть технологии PLM являются основой, интегрирующей информационное пространство, в котором функционируют САПР, ERP, PDM, SCM, CRM и другие автоматизированные системы разных предприятий [3].

Если вернуться к СПЖЦ «САРУС», то в Пояснительной записке Технического проекта [4; 20-23] отмечено, что программа «Система конструкторского проектирования» применяется в области разработки конструкторских документов, содержащих данные, необходимые на всех стадиях жизненного цикла изделий и их составных частей. Объектом автоматизации CAD-системы являются процессы конструкторско-технологической подготовки производства.

Система базируется на интегрированной инженерной программной платформе (далее – ТП) и состоит из программных модулей (ПМ), которые разрабатываются на единой и открытой программной архитектуре. Это обеспечивает возможность расширения состава задач за счёт разработки дополнительных программных модулей (приложений), решающих новые задачи на основе базовых платформенных инструментов.

При этом средства ТП должны предоставлять инструменты, обеспечивающие встраивание в конечную систему или комплекс систем дополнительных программных модулей, и средства, обеспечивающие их совместное (интегрированное) функционирование. Такие средства должны обеспечивать надёжную и своевременную процедуру обмена данным между CAD-системой и смежными программными модулями (рис. 2).

Интеграция между смежными программными модулями может выполняться по нескольким схемам, но на данном этапе тестирования нас интересует наиболее предпочтительный метод взаимодействия – прямое взаимодействие на основе единых программных функций и классов ТП, в связи с тем что большинство модулей конечной системы должны быть построены на основе классов и функций ТП. При таком методе взаимодействия обмен данными осуществляется прямым вызовом программного интерфейса смежного приложения или инструментов ТП [4; 24].

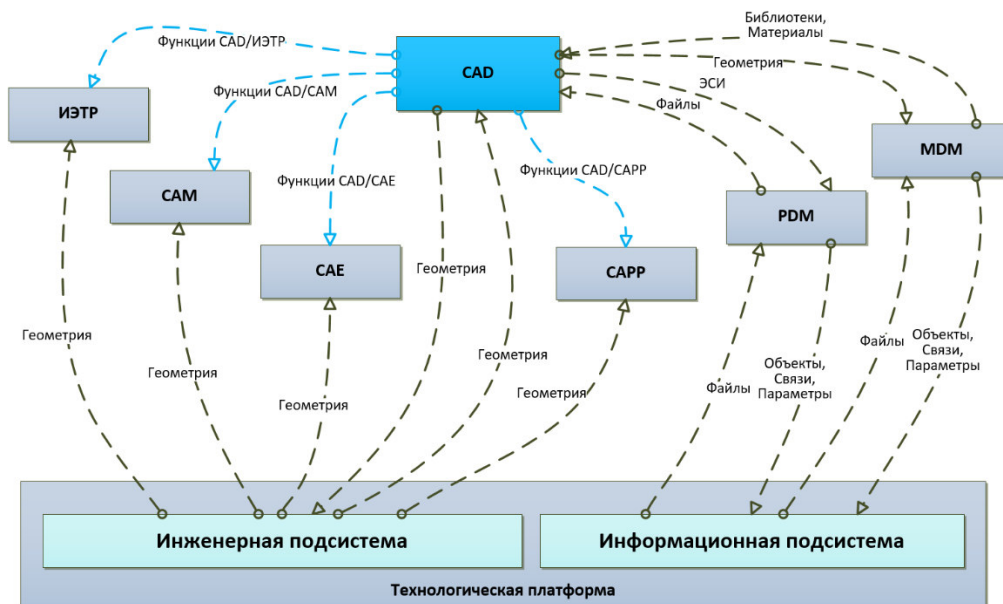


Рисунок 2. Схема взаимодействия модулей

Рассмотрим срабатывание метода и функций ПМ:

✓ CAD как средства автоматизированного проектирования, предназначенные для автоматизации двумерного и/или трехмерного геометрического проектирования, создания конструкторской и/или технологической документации, и САПР общего назначения;

✓ САМ как средства технологической подготовки производства изделий, обеспечивающие автоматизацию программирования и управления оборудования с ЧПУ или ГАПС (гибких автоматизированных производственных систем).

## Результаты экспериментов

### 1. Создание 3D-моделей детали в «САПУС» и «КОМПАС-3D»

#### 1.1. Создание управляющей параметризованной 3D-модели «Компрессор низкого давления для турбореактивного двигателя» в ПМ CAD-системы «САПУС»

Построение сложных кривых в плоскости можно получить различными способами:

- 1) сочетанием дуг (элементов окружности) и отрезков;
- 2) сочетанием нескольких сплайнов по заданным параметрам.

В решении данной задачи был выбран второй метод построения, а именно создание сплайнов по точкам. Этот выбор обусловлен тем, что параметризованные сплайны легче построить, чем сочетание дуг.

**Шаг 1.** Выбирается горизонтальная плоскость (вид сверху), это действие необходимо и пригодится для дальнейшей обработки модели в САМ-системе. Создается эскиз из двух сплайнов и дополнительных отрезков, задаются размеры.

Важной задачей на этом этапе является параметризация эскиза, т.е. ограничение геометрии будущей модели по всем степеням свободы, без допущения переопределенности.

Окрашивание эскиза в черный свет в данной программе указывает на то, что эскиз определен и построен правильно. Дальнейшее изменение параметров позволит правильно перестроить модель, не нарушая ее геометрии (рис. 3).

**Шаг 2.** Создается параллельная плоскость на расстоянии 70 мм. На ней строится второй профиль, повторяющий геометрию лопатки турбины (рис. 4).

Также необходимо задать все параметры и сделать эскиз определенным.

Стоит отметить, что толщина профиля лопатки уменьшается по мере удаления от оси вращения. Это можно регулировать длиной вспомогательных отрезков между сплайнами.

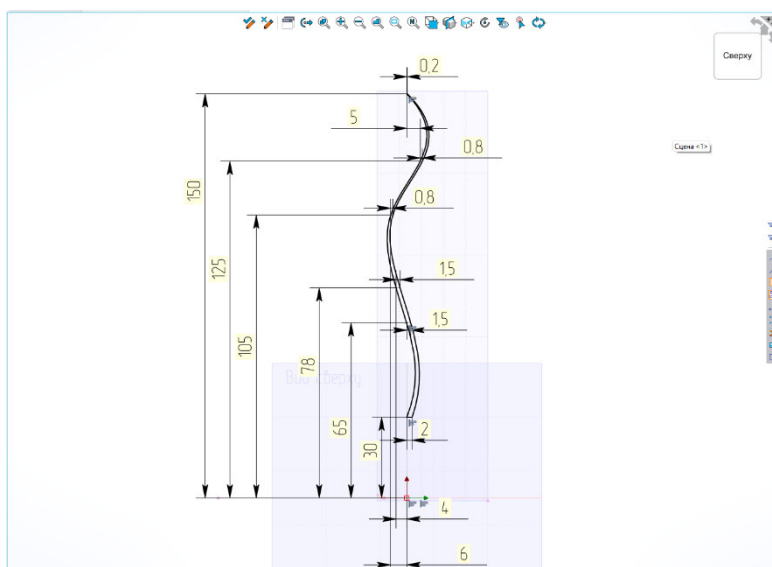


Рисунок 3. Шаг 1. Создание эскиза из двух сплайнов по заданным размерам

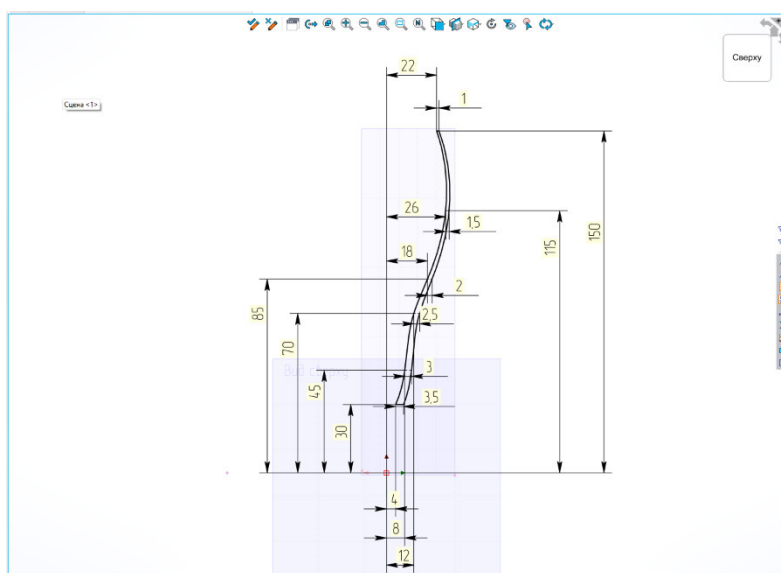


Рисунок 4. Шаг 2. Построение второго профиля на параллельной рабочей плоскости

**Шаг 3.** Создается криволинейная поверхность по сечениям. В мастере выполнения задач выбираются два сечения, и программа автоматически производит построение сложных поверхностей (рис. 5).

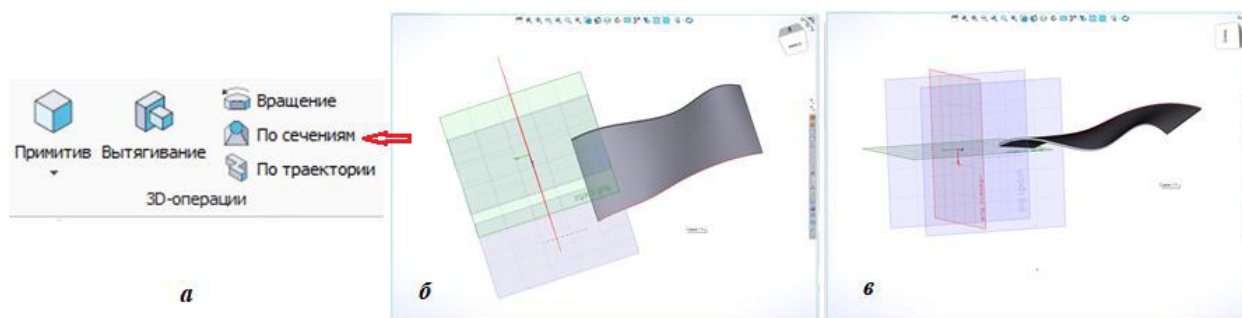


Рисунок 5. Шаг 3. Построение сложных криволинейных поверхностей:

а – выбор операции «По сечениям» из Ленты;  
б, в – созданные по эскизу криволинейные поверхности

Необходимо подчеркнуть, что «САРУС» с операцией «построение криволинейных поверхностей по сечениям» справляется лучше, чем «КОМПАС-3D».

Шаг 4. Создается эскиз конусного обтекателя для дальнейшего размещения на нем лопаток (рис. 6, а).

Шаг 5. Операцией вращения создается обтекатель (рис. 6, б).

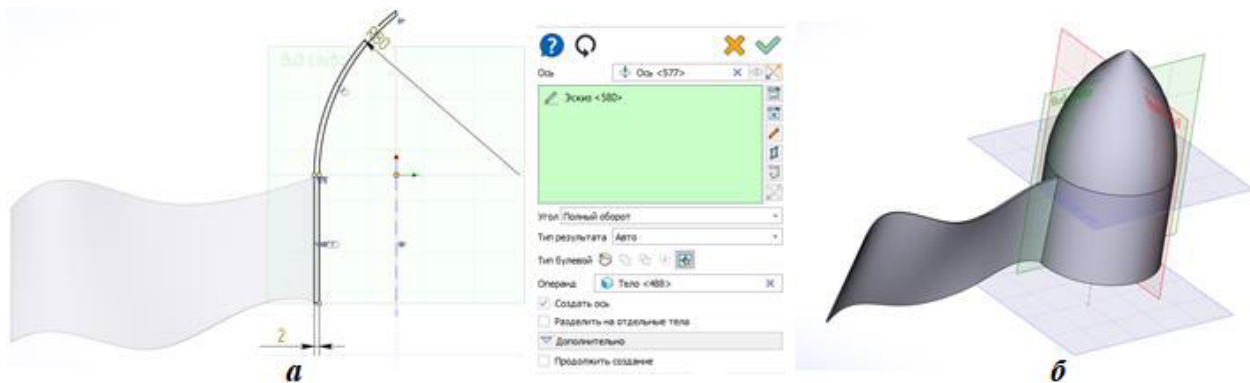


Рисунок 6. Создание конусного обтекателя:

а) шаг 4 – создание эскиза конусного обтекателя; б) шаг 5 – создание обтекателя операцией «Вращение»

Шаг 6. Операцией круговой массив создается 26 лопаток компрессора (рис. 7).

При построении кругового массива в данной программе предварительно необходимо строить ось вращения, это является недочетом.

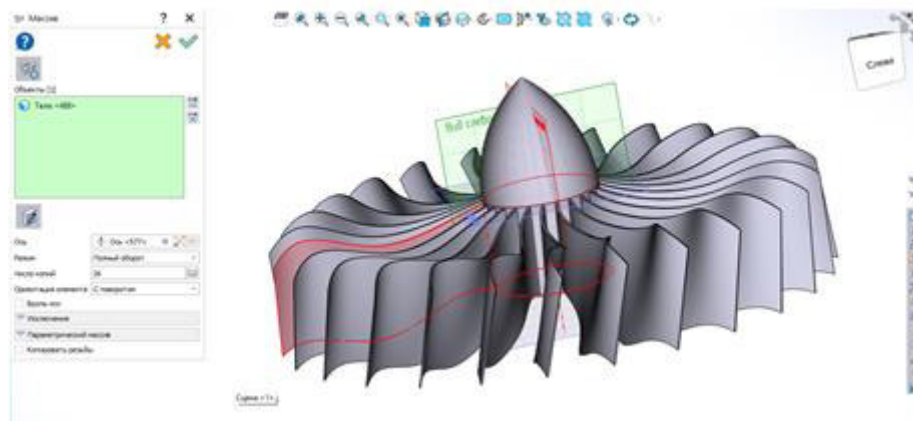


Рисунок 7. Создание лопаток компрессора операцией «Круговой массив»

### 1.2. Создание управляющей параметризированной модели в «КОМПАС-3D»

Процесс создания 3D-модели детали «Компрессор низкого давления для турбореактивного двигателя» в программе «КОМПАС-3D» аналогичен построению детали в СПЖЦ «САРУС». Алгоритм построения приведен на рисунках 8–11.

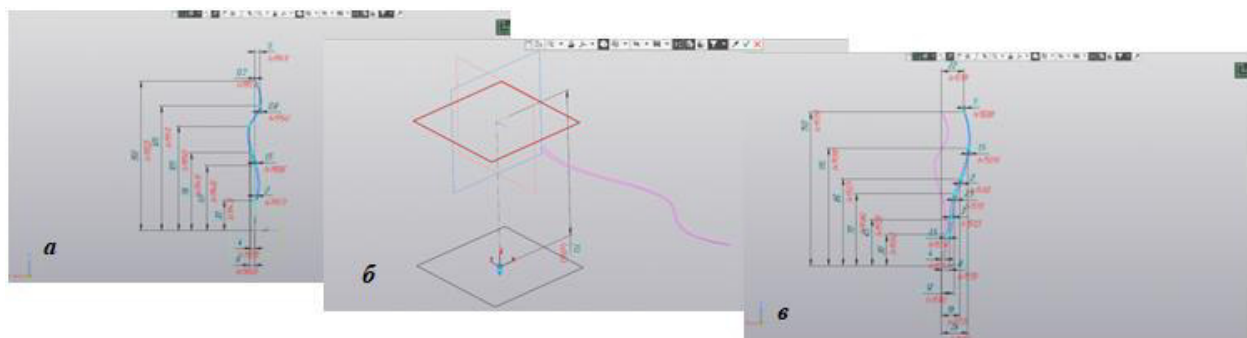


Рисунок 8. Построение эскизов лопастей: а) шаг 1 – создание эскиза из двух сплайнов по заданным размерам; б) шаг 2 – построение дополнительной рабочей плоскости; в) шаг 2 – построение второго профиля на параллельной рабочей плоскости

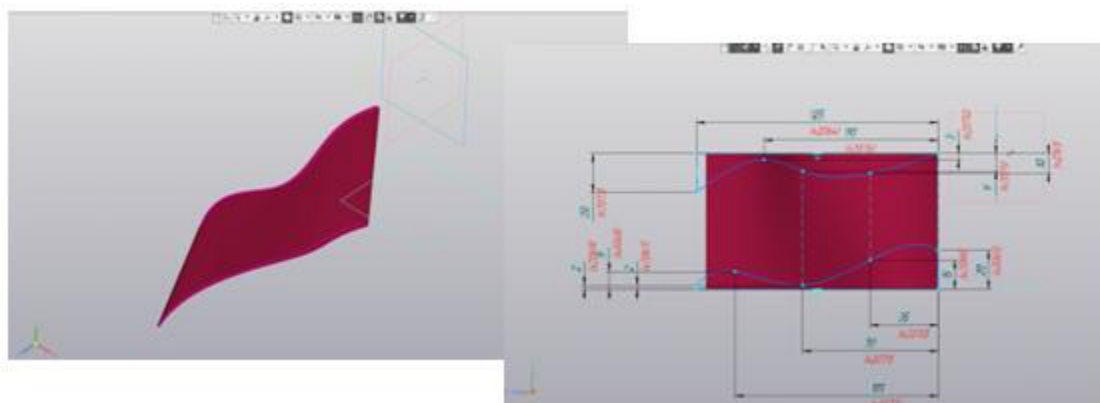


Рисунок 9. Шаг 3. Построение сложных криволинейных поверхностей «По сечениям»

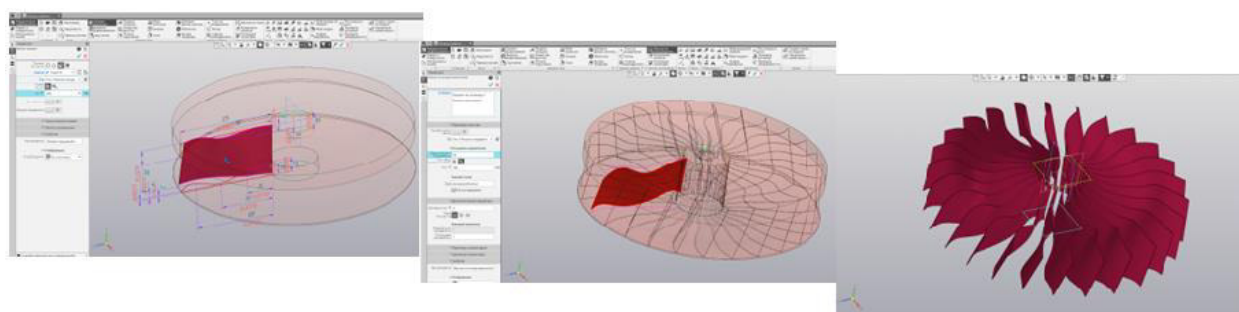


Рисунок 10. Шаг 4. Последовательность создания лопаток компрессора операций «Круговой массив»

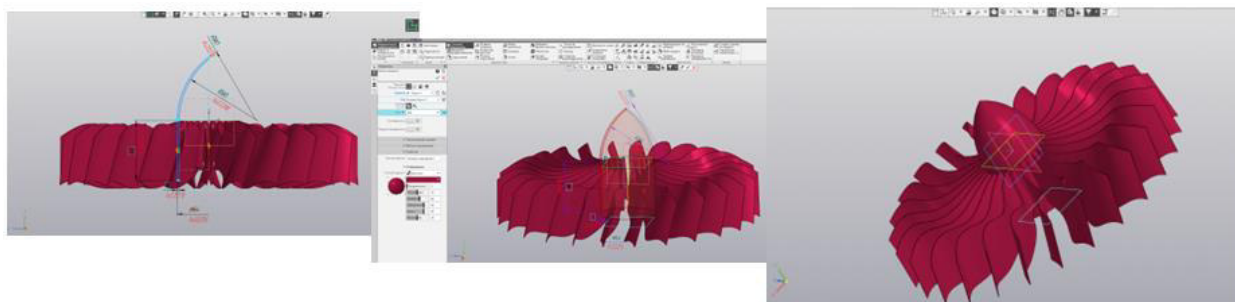


Рисунок 11. Шаг 5. Создание конусного обтекателя

По рисункам видно, что трудоемкость некоторых операций больше в «КОМПАС-3D», как уже отмечалось в предыдущем подпункте.

## **2. Разработка управляющей программы обработки для станков с ЧПУ в ПМ САМ-системы «САРУС» по созданной 3D-модели**

В данном разделе рассматриваются различные методы получения детали «Компрессор низкого давления для турбореактивного двигателя» в САМ-системе СПЖЦ «САРУС»

### **2.1. Попытка разработки управляющей программы для получения детали методом «Фрезерной обработки по двум направляющим 5D»**

**Шаг 1.** В разделе «ЧПУ» из верхней панели задач выбирается вкладка «Целевая геометрия» (рис. 12, а), на модели указываются все компоненты, требующие обработки (рис. 12, б).

**Шаг 2.** Следующая вкладка – «Заготовка» (рис. 13, а). На данном этапе необходимо определить поверхность, наиболее подходящую к исходной модели. Заготовку можно либо создать из предложенных объектов, либо отдельно построить сложную конфигурацию или применить геометрию другой обработки (рис. 13, б).



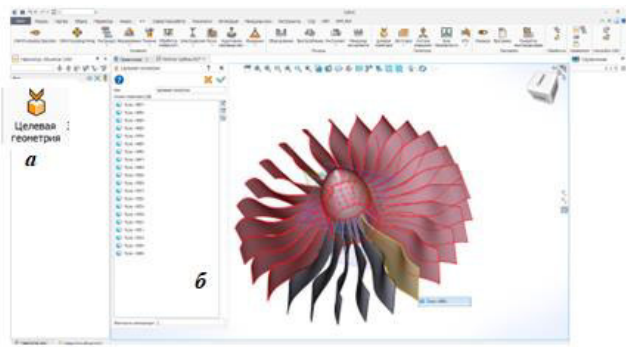


Рисунок 12. Шаг 1

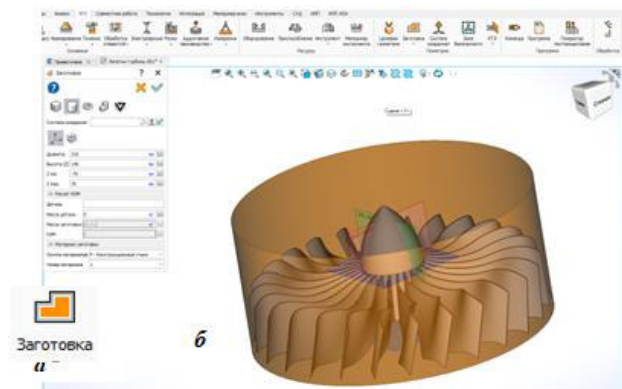


Рисунок 13. Шаг 2

**Шаг 3.** Определение системы координат (СК) (рис. 14, а). Важно учесть, что ось Z должна быть направлена вертикально, чтобы основание детали лежало в плоскости XOY (рис. 14, б). Положение оси Z определено параметрами большинства станков с ЧПУ.

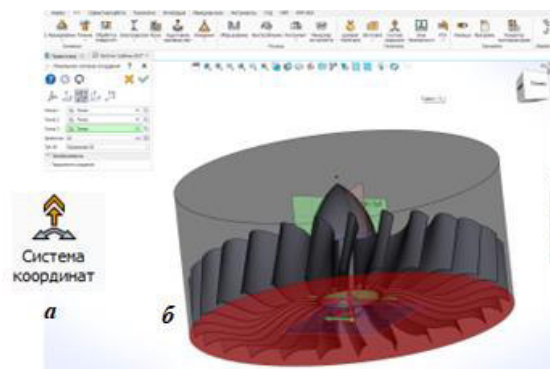


Рисунок 14. Шаг 3

**Шаг 4.** Выбор зоны безопасности (ЗБ). Данная область необходима для осуществления нерабочего перемещения инструмента (быстрые подачи) вне этой зоны (рис. 15).

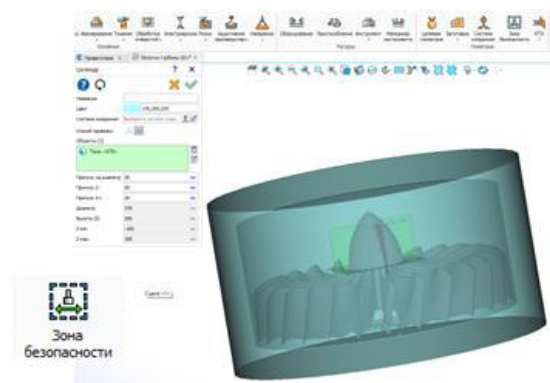


Рисунок 15. Шаг 4

**Шаг 5.** Выбор обрабатывающей программы «Фрезерование – по двум направляющим 5D» (рис. 16, а). Задаются основные параметры (рис. 16, б). Большинство отправных параметров программа задает автоматически. После заполнения данных для обработки (рис. 16, в) производится расчет (рис. 16, г).

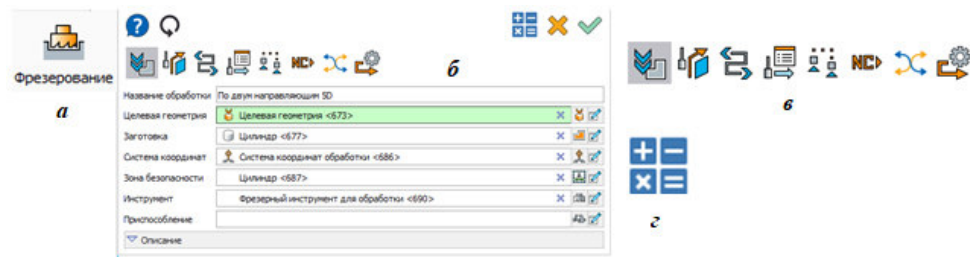


Рисунок 16. Шаг 5

**Шаг 6.** Реализация анимации обработки (рис. 17, а). Получение управляющих программ (G-код) для станков с ЧПУ – рис. 17, б).

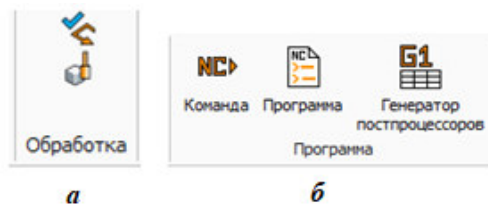


Рисунок 17. Шаг 6

**Вывод.** В данной работе не получилось произвести полноценную обработку с выводом программы на станок. Мы считаем, что это связано в первую очередь с тем, что такую сложную поверхность практически невозможно получить фрезерной обработкой. Во-вторых, технология требует более квалифицированного подхода.

**2.3. Попытка разработки управляющей программы для получения детали методом «Аддитивное производство (Послойная наплавка FFF)»**

**Шаг 1.** Создается: целевая геометрия, СК, ЗБ (рис. 18, а). Для упрощения расчета процесса обработки в качестве целевой геометрии выбирается одна лопасть и обтекатель компрессора. В разделе аддитивное производство выбирается «Послойная наплавка FFF» (рис. 18, б).

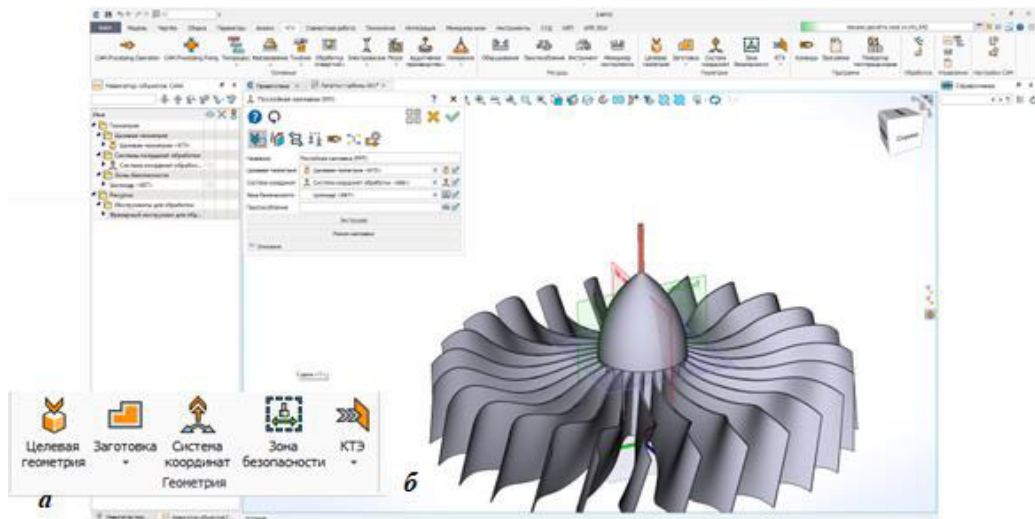


Рисунок 18. Шаг 1. Послойная наплавка FFF

**Шаг 2.** Для данной обработки задаются параметры, показанные на рисунках 19 и 20.



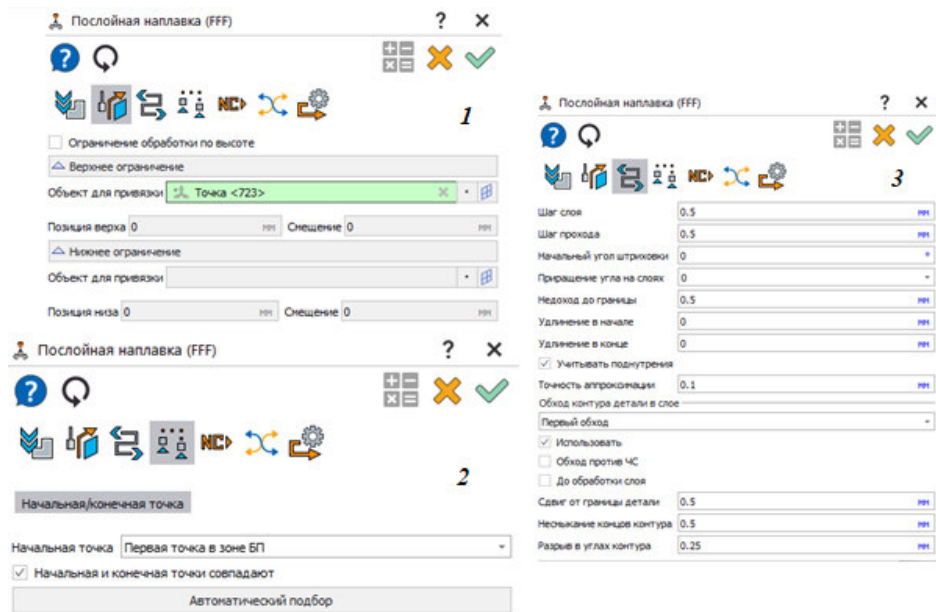


Рисунок 19. Шаг 2. Поэтапная настройка диалоговых окон на послойную наплавку

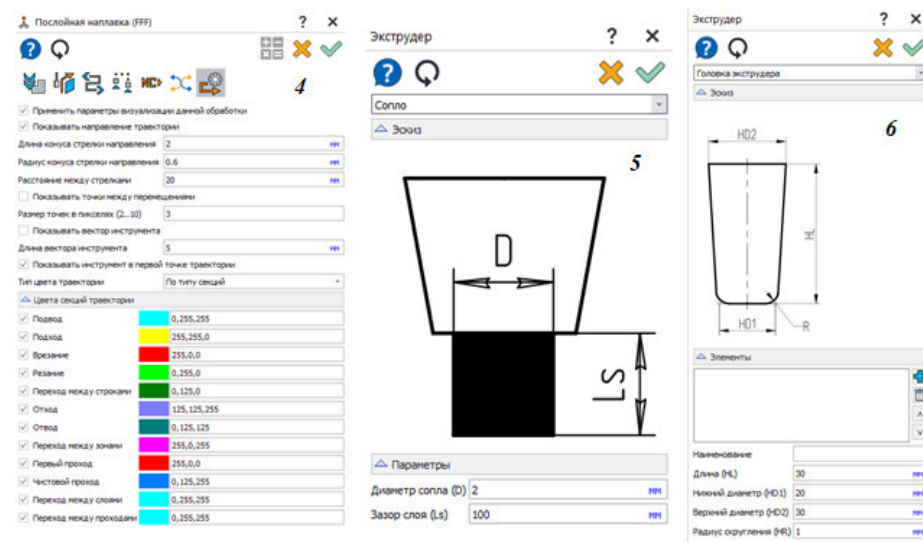


Рисунок 20. Шаг 2. Продолжение настройки

**Шаг 3.** Производится расчет программы обработки (рис. 21, а). Замечен первый положительный результат, а именно появление траектории обработки – розовый контур (рис. 21, б).

**Шаг 4.** Производится имитация обработки «Послойная наплавка FFF». Инструмент (в нашем случае сопло) движется по контуру, наплавляя материал; создается деталь (рис. 21, в).

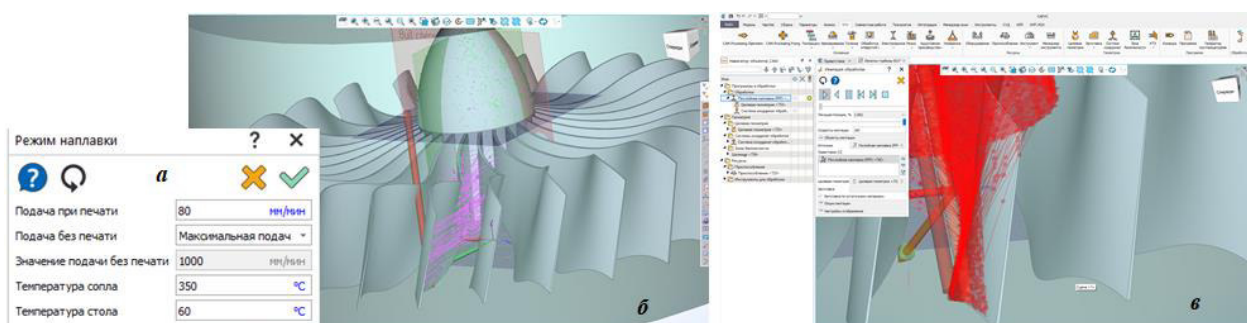


Рисунок 21. Расчет программы обработки

**Вывод.** Реализация «Послойной наплавки FFF» при изготовлении детали со сложными криволинейными поверхностями изделия «Компрессор низкого давления» прошла успешно. Были получены контур обработки, имитация обработки.

2.4. Попытка разработки управляющей программы для получения детали методом «Электроэрозионной двухконтурной обработки»

Шаг 1. Создаются целевая геометрия, заготовка, СК, ЗБ (рис. 22, а). В качестве целевой геометрии выбирается сборочная единица «Лопатка» основного изделия «Компрессор низкого давления». В разделе электроэрозионная обработка выбирается «Электроэрозионная двухконтурная обработка» (рис. 22, б)

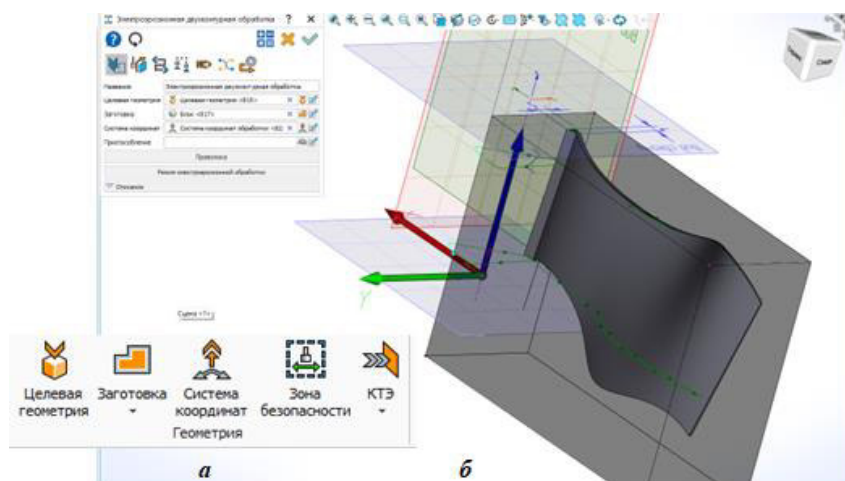


Рисунок 22. Настройка программы на электроэрозионную двухконтурную обработку

Шаг 2. Для данной обработки задаются следующие параметры (рис. 23, а, б, в): выбираются объекты верхнего и нижнего контуров по кривым целевой геометрии, находящимся в соответствующих плоскостях; задаются верхние точки привязки для входа и выхода инструмента–проволоки по верхнему контуру; нижние точки привязки соответственно для прохода инструмента по нижнему контуру.

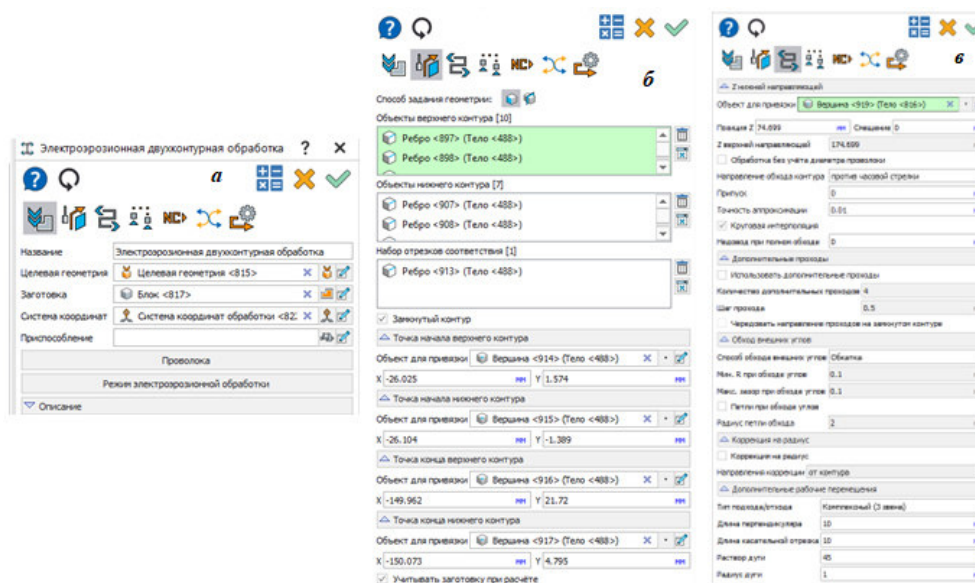


Рисунок 23. Выбор параметров для электроэрозионной двухконтурной обработки

Шаг 3. Осуществлена имитация обработки. В качестве инструмента используется проволока. Движение проволоки происходит по двум ранее заданным конурам одновременно (рис. 24).

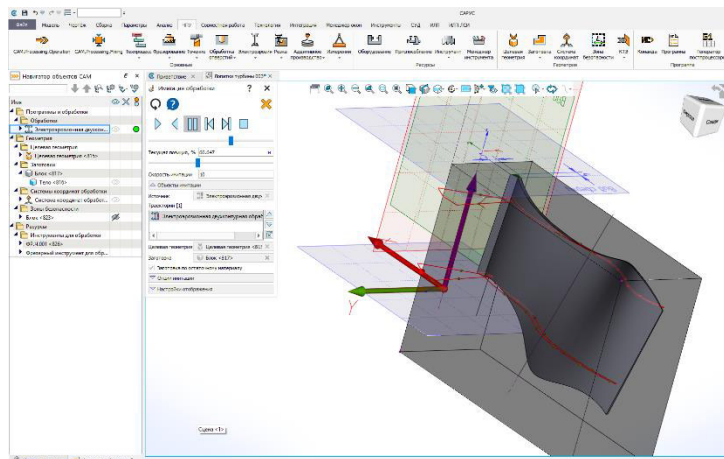


Рисунок 24. Результат настройки на электроэрозионную двухконтурную обработку

**Вывод.** Произведена двухконтурная электроэрозионная обработка сборочной единицы «Лопатка компрессора низкого давления». Наблюдается существенное отличие по скорости обработки, данная обработка по сравнению с предыдущими вариантами имеет большую производительность.

### Обсуждение результатов

#### 1. Сравнение построения 3D-модели в СПЖЦ «САРУС» с построением в ПО «КОМПАС-3D»

Преимущества CAD-системы «САРУС»:

✓ возможность построения сложных поверхностей при осуществлении операции выдавливания по трем и более плоскостям, в то время как в программе «КОМПАС-3D» корректно строятся сложные криволинейные поверхности только по двум плоскостям.

Недостатки CAD-системы «САРУС»:

- ✓ не имеет функцию отображать степени свободы, что немаловажно при построении параметризованного эскиза;
- ✓ не ловит оси начала координат при попытке создания кругового массива;
- ✓ не позволяет отменить выделенную линию путем повторного нажатия на нее;
- ✓ не ловит привязки к центру созданной ранее окружности;
- ✓ не позволяет вырезать в двух направлениях сразу, в каждом направлении на свою глубину, то есть не создает отдельные параметры для каждого направления;
- ✓ некорректно производит операцию «усечение кривой»;
- ✓ на данный момент не имеет рабочей функции «создать исполнение». Она есть, но только в замысле.

**Вывод.** СПЖЦ «САРУС» имеет ряд проблем, встречающихся при проектировании. Разработчики данной программы создают по результатам тестирования более новые версии.

#### 2. Анализ разработки управляющей программы обработки для станков с ЧПУ в ПМ САМ

В СПЖЦ «САРУС» проведены попытки разработки управляющих программ для разных видов механической обработки лопасти. Результативно проведена электроэрозионная двухконтурная обработка. Остальные виды обработки не проведены или проведены не полностью. Мы считаем, что это связано, во-первых, с тем, что такую сложную поверхность практически невозможно получить фрезерной обработкой. Во-вторых, технология требует более квалифицированного подхода.

В ПО «КОМПАС-3D» такой программный модуль не предусмотрен. Необходимо первичную 3D-модель экспортировать в другое ПО, например Gemma-3D.

### Выводы

Цели данного проекта достигнуты:

- изучены основные принципы построения сложных криволинейных поверхностей;
- попытки разобраться в алгоритмах создания управляющих программ для обработки на станках с ЧПУ привели к успеху.

Решены задачи:

- успешно разработаны 3D-модели в ПМ САД «САРУС» и «КОМПАС-3D»;
- проведено сравнение, выявлены преимущества и недостатки СПЖЦ «САРУС»;
- разработаны УП обработок в (ПМ) САМ-системы «САРУС» по 3D-модели:
  - фрезерная обработка по двум направляющим 5D – требует доработки;
  - послойная наплавка FFF – успешно;
  - электроэрозионная двухконтурная обработка – успешно.

Стоит отметить, что программный модуль САМ-системы «САРУС» находится на стадии тестирования и активной доработки.

#### **Источник финансирования**

Договор от 25 марта 2024 г. № 96-2024/244 на выполнение научно-исследовательской работы по теме «Проведение научных исследований в интересах ФГУП «РФЯЦ-ВНИИЭФ» с использованием результатов для актуализации образовательных программ подготовки специалистов с высшим профессиональным образованием для ФГУП «РФЯЦ-ВНИИЭФ», Задача 10 «Исследование программно-аппаратных решений в области цифровых технологий, возможностей их отраслевого применения, разработка и версионное тестирование компонентов полностью защищенной системы полного жизненного цикла среднего класса СПЖЦ V3 «САРУС».

#### **Благодарность**

Авторы благодарят кафедру цифровых технологий СарФТИ-НИЯУ «МИФИ» под руководством О.В. Кривошеева, к.т.н., заведующего кафедрой, зам. директора ФГУП «РФЯЦ-ВНИИЭФ» по технологиям полного жизненного цикла – директора института цифровых технологий (ИТЦ) – главного конструктора СПЖЦ.

#### **Список литературы**

1. ГОСТ 34.003-90. Информационная технология. Комплекс стандартов на автоматизированные системы. Термины и определения.
2. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»: Основная версия программного модуля. «Система конструкторского проектирования» (версия 2) / Руководство оператора 07623615.00423-06 34 01. 2021. 338 с.
3. Phiri Michael. Information Technology in Construction Design. London: Thomas Telford Publishing, 1999. P. 52. ISBN 0-7277-2673-0.
4. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»: Программный модуль «Система конструкторского проектирования». Пояснительная записка технического проекта 07623615.00423-12 81 01. Часть 1. 2021. 924 с.

УДК 087.2

DOI: 10.25686/978-5-8158-2474-4-2025-768-779

## Разработка пользовательского сценария создания 3D-модели детали «Вал» в программном модуле CAD СПЖЦ «САРУС»

О. А. Ключева, Н. А. Денисова

Саровский физико-технический институт – филиал Национального исследовательского ядерного университета «МИФИ», Саров, Россия

**Аннотация.** Одной из задач тестирования новой системы автоматизированного проектирования «Система полного жизненного цикла изделий «Цифровое предприятие» (актуальное наименование – СПЖЦ «САРУС») в связи с переходом РФЯЦ-ВНИИЭФ на использование нового комплекса является исследование оптимальных алгоритмов построения криволинейных поверхностей в программном модуле CAD, являющимся составной частью СПЖЦ «САРУС». Важно также подготовить методические материалы по обучению работы с ПМ CAD вышеназванной программы студентов и специалистов, имеющих опыт работы с рядом САПР как инструмента в направлении конструкторской подготовки производства.

В данной работе как пример описываются возможные методы построения трёхмерной модели, а также разработанный по заказу Отдела внедрения нового программного комплекса подробный и наиболее оптимальный алгоритм построения 3D-модели «Вал» для обучения студентов и специалистов предприятия – пользовательский сценарий.

**Ключевые слова:** программный модуль CAD, трёхмерные объекты, методы построения 3D-моделей, пользовательский сценарий.

## Development of a custom scenario for creating a 3D model of the detail shaft in the CAD software module of the SARUS FLCS

O. A. Klueva, N. A. Denisova

Sarov Institute of Physics and Technology – branch of the National Research Nuclear University "MEPhI", Sarov, Russia

**Abstract.** One of the tasks of testing the new computer-aided design system "Digital Enterprise Full Life Cycle System" (current name – FLCS SARUS) in connection with the transition of the RFNC-VNIIEPh to the use of the new complex is the study of optimal algorithms for constructing curved surfaces in the CAD software module, which is an integral part of the UIC SARUS. It is also important to prepare methodological materials for teaching students and specialists who have experience working with a number of CAD systems as a tool in the field of design preparation of production CAD as a tool in the direction of design preparation of production.

In this paper, as an example, possible methods of constructing a three-dimensional model are described, as well as a detailed and optimal algorithm for constructing a 3D model "Shaft" for training students and specialists of the enterprise, developed by order of the Department of Implementation of a new software package – a user scenario.

**Keywords:** CAD software module, three-dimensional objects, 3D model construction methods, user scenario

### Введение

Одной из задач тестирования системы автоматизированного проектирования *Комплекс программ в защищенном исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»* (далее – СПЖЦ «САРУС») в связи с переходом РФЯЦ ВНИИЭФ на использование нового комплекса является исследование оптимальных алгоритмов построения криволинейных поверхностей в программном модуле CAD (далее – ПМ CAD), являющимся составной частью СПЖЦ «САРУС», и подготовка методических материалов по обучению работы с ПМ CAD вышеназванной программы студентов и специалистов, имеющих опыт работы с рядом САПР как инструмента в направлении конструкторской подготовки производства [1].

На предыдущих этапах исследования [2, 3] были выявлены возможности применения операций, предлагаемых интерфейсом ПМ CAD по построению криволинейных поверхностей в 2D- и 3D-режимах простых и средней сложности деталей машиностроения. Выявлено, что в представленных версиях программного комплекса СПЖЦ «САРУС» 1.0.17.9 и 2023.1.0.1-RC16 при проверке работоспособности ряда функций в СПЖЦ «САРУС» установлено, что большинство алгоритмов оказались работоспособными.



Был проведен сравнительный анализ с известным САПР «КОМПАС-3D». По мнению тестировщиков, более производительным было создание 3D-моделей в «КОМПАС-3D». Для работы на производстве это имеет огромное значение. Но необходимо отметить, что и в той, и в другой САПР есть факты несрабатывания функций [4].

Однако «САРУС» удобен для работы технологом, так как в этой программе предусмотрены CAD-и САМ-системы и ещё много других модулей. При дальнейшей разработке и модернизации ПО СПЖЦ «САРУС» способно заменить широко известные CAD-программы [5].

Исходя из выводов по проделанной работе, в Техническом задании для реализации следующего этапа практико-ориентированного исследования запланирована работа с **деталью средней сложности** с подготовкой методических материалов по обучению работы с ПМ CAD «САРУС» студентов по направлению подготовки 15.03.04 – Конструкторско-технологическое обеспечение машиностроительных производств, а также специалистов, имеющих опыт работы с рядом САПР в направлении конструкторской подготовки [4].

В данной работе описываются возможные методы построения трёхмерной модели, а также подробный и наиболее оптимальный алгоритм построения 3D-модели «Вал» (рис. 1), описанный как «Пользовательский сценарий», разработанный по заказу Отдела внедрения нового программного комплекса для применения в обучении студентов и специалистов предприятия.

Разработка проведена в базовой версии, утвержденной для использования на цифровом предприятии в 2024 году, 2023.1.0.1-RC16.

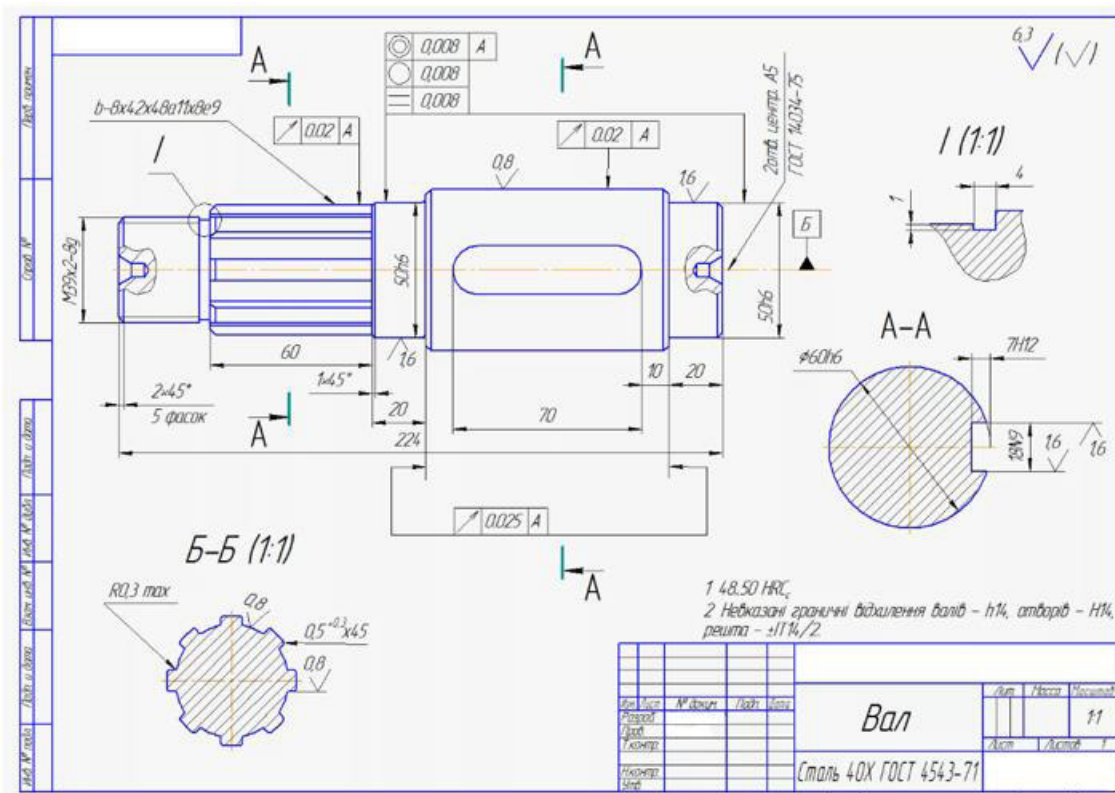


Рисунок 1. Чертеж детали «Вал». Выполнен в «КОМПАС-3D»

### Постановка задачи

Целью данной работы является исследование функционала и возможности применения команд ПМ CAD для построения криволинейных поверхностей и создания наиболее оптимального алгоритма построения 3D-модели «Вал», выявление достоинств и недостатков при использовании.

#### Основные задачи:

✓ тестирование новой системы автоматизированного проектирования в связи с переходом РФЯЦ ВНИИЭФ на использование нового комплекса;

- ✓ исследование методов построения 3D-моделей в ПМ CAD и выбор оптимального алгоритма для построения трёхмерной модели «Вал»;
- ✓ описание поэтапных действий для создания трёхмерной модели детали с применением функциональности ПМ CAD;
- ✓ сравнение полученных результатов замеров времени, затрачиваемого на процесс создания сборочных моделей в сравниваемых системах СПЖЦ «САРУС» и «КОМПАС-3D»;
- ✓ подготовка методических материалов по обучению работе с ПМ CAD студентов по направлению 15.03.04 – Конструкторско-технологическое обеспечение машиностроительных производств.

### Теория

Модуль ПМ CAD СПЖЦ «САРУС» позволяет создавать параметрические 2D-модели, параметрические твердотельные 3D-модели, ассоциативные чертежи и конструкторскую документацию [6]. ПМ CAD основан на командах трёхмерного моделирования и командах двухмерных построений. Очевидно, что ПМ является многонаправленным инструментом, поэтому в результате тестирования были выделены наиболее проблемные области: построение моделей и чертежей с замкнутым криволинейным контуром; совмещение двух методик построения по примитивам и по координатам; параметризация эскиза.

В программном модуле CAD СПЖЦ «САРУС» предусмотрено три варианта построения трёхмерных объектов.

1. *Метод построения по координатам (или по эскизу)*. Суть метода заключается в создании эскиза на рабочей плоскости с последующим его вытягиванием или же, наоборот, вырезанием до необходимой величины. В данном варианте предусмотрено проектирование не только базовых объектов, но и криволинейных поверхностей, трехмерных моделей с нестандартной конфигурацией, а также геометрически сложных деталей, т.е. является многофункциональным методом и имеет множество вариаций применения при построении конструкции.

Стоит отметить, данный метод является наиболее распространенным и более привычным для пользователей, ранее работавших в программе «Компас-3D», поскольку имеет схожий алгоритм действий. Время построения модели зависит от сложности её конфигурации. Вероятность появления ошибок в методе построения по координатам при работе в программе мала.

2. *Метод построения по примитивам*. Суть метода заключается в вытягивании или вырезании базовых объектов единой конфигурации (цилиндр, конус, параллелепипед и т.д.), есть возможность настроить их габаритные размеры, построение деталей иной формы невозможно. Данный метод имеет ограниченное применение, однако затраты времени на создание базовых объектов значительно меньше, чем при использовании способа по координатам (по эскизу). Производительность при создании моделей простой конфигурации будет выше при использовании примитивов.

Примитивы являются новой функцией для пользователей, ранее работавших в «Компас-3D», но использование их в ПМ CAD СПЖЦ «САРУС» значительно упрощает построение некоторых деталей. Вероятность возникновения ошибки мала.

3. *Комбинированный метод*, т.е. совмещение методов построения по координатам (по эскизу) или по примитивам. Суть метода заключается в создании трехмерных моделей, которые состоят из базовых объектов (примитивов) и сложных поверхностей, с целью сокращения затрат времени на конструирование.

Способ комбинирования имеет большие преимущества при сокращении времени, производительности и упрощении процесса проектирования. Однако в данном методе может возникать большое количество ошибок, например некорректное построение, возникновение нежелательных самопересечений контура и т.д. Их можно избежать, если следовать строго поставленному алгоритму. В тенденции развития метод имеет большие перспективы и в будущем будет иметь широкое применение [5].

В разработке пользовательского сценария создания 3D-модели детали «Вал» в программном модуле CAD СПЖЦ «САРУС» используется комбинированный метод построения, так как затрагиваются оба



метода, что позволяет пользователю освоить большое количество базовых команд сразу обоих способов, необходимых для освоения и понимания работы программы.

В CAD-системах метод создания моделей можно сравнить с обработкой заготовок деталей резанием, где постепенно удаляется припуск на обработку и при достижении заданных чертежом размеров получаются элементарные поверхности, которые при их сочетании формируют деталь. Есть ещё один метод – применение операции «Примитив», где модель создается постепенно добавлением простейших геометрических фигур – элементов детали, при котором целостная модель получается сложением, вычитанием или пересечением данных элементов друг с другом с помощью булевых операций.

Тогда классификация создания 3D-моделей может выглядеть следующим образом [6]:

1. Группа команд «3D-операции», объединяющая традиционные команды твердотельного моделирования.

1.1 Команда «Вытягивание» предназначена для создания твердого или поверхностного тела, которое представляет собой след от движения контура в заданном направлении на заданную длину.

1.2 Команда «Вращение» предназначена для создания тела как результата вращения контура относительно заданной оси на заданный угол.

1.3 Команда «По сечениям» предназначена для создания тел со сложной геометрией.

1.4 Команда «По траектории» предназначена для создания тела путем перемещения контура вдоль пространственной траектории.

2. Команда «Примитив» предназначена для создания базовых твердых тел.

3. Команда «Булева» предназначена для создания тела на основе двух или более ранее созданных тел.

4. Другие команды («Обрезка», «Скругление», «Фаска», «Оболочка», «Отверстие», пр.), способные окончательно сформировать форму детали согласно чертежу.

Деталь «Вал» представляет собой тело вращения – ступенчатый вал, усложненный шлицами на одной из ступеней. Такая форма поверхности требует специальных расчетов. Можно отнести деталь «Вал» к деталям средней конструктивной сложности, что соответствует техническому заданию.

### Результаты экспериментов

Для эффективного функционирования «пользовательского сценария» мы предлагаем представления его в форме таблицы с показом иллюстративного материала в качестве результатов проведенных операций. Мы считаем, что именно в таком варианте алгоритм создания 3D-модели будет наиболее обозреваем и понятен.

В зависимости от задач, которые нужно решить пользователю, выбирается необходимый прототип: модель, чертеж или спецификация. Это общая «приветственная» операция. Так как дан чертеж, по нему надо создать CAD-модель. Поэтому необходимо выбрать раздел «Модель» и нажать «ОК» (рис. 2, а). На вкладке «Модель» собраны команды, предназначенные для создания 3D-модели.

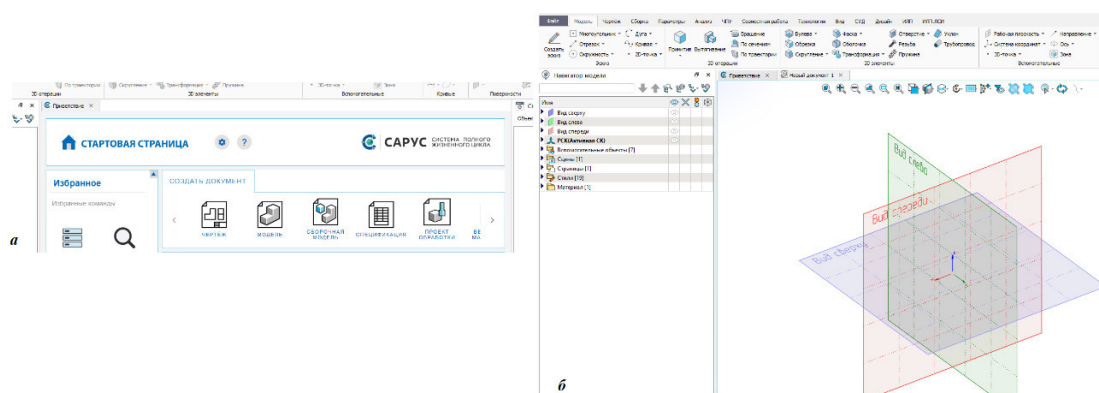


Рисунок 2. Выбор документа



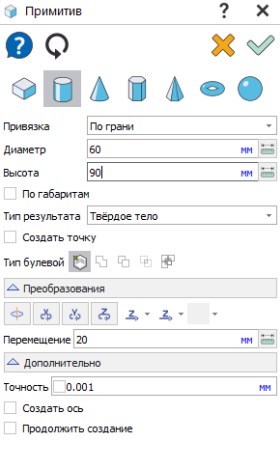
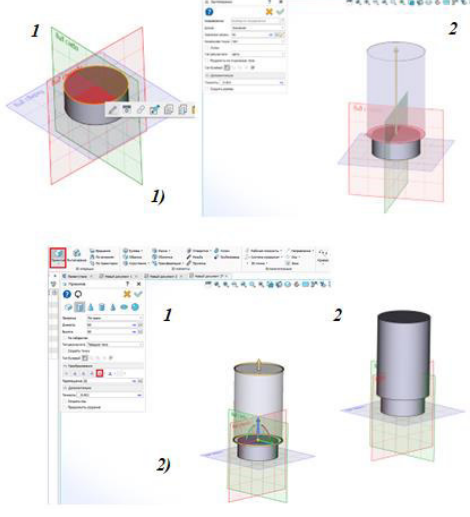




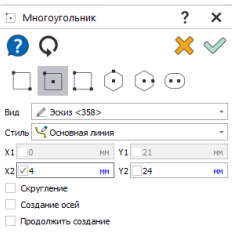
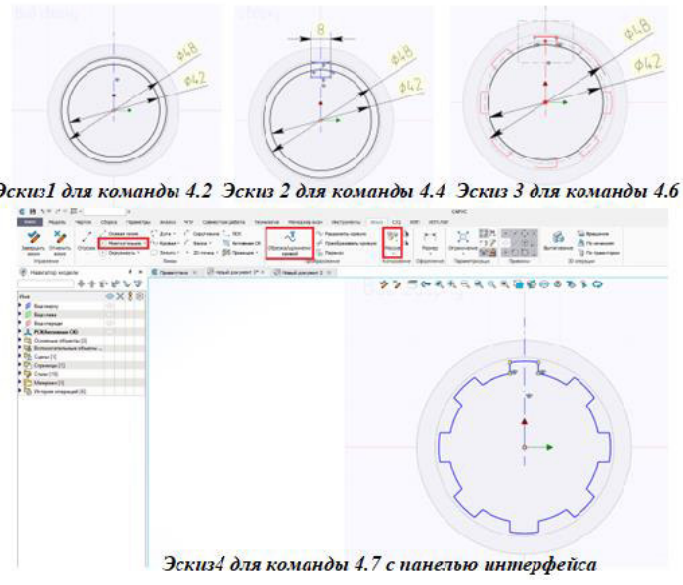
Появляется новое окно для создания CAD-модели (рис. 2, б). Сразу же сохраняем деталь на жестком диске, вызвав команду «Сохранить» на панели инструментов.

На следующем этапе переходим непосредственно к трехмерному моделированию детали. Данный этап является общим для создания любой модели, дальнейшие действия будут описаны в таблице. Таблица содержит столбцы, где описаны стандартные команды, показаны «иконки» команд, настройки в диалоговом окне, а также изображение на рабочем окне выполнения команд. Таблица в данной работе представлена в книжной ориентации, где стандартные команды описаны в объединенном столбце. Если таблицу представить в альбомной ориентации, раздел «Стандартные команды» можно выделить в отдельный столбец.

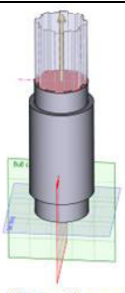


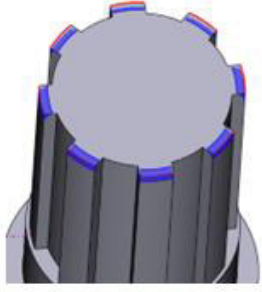

### Этапы выполнения работы

Команда	Настройки	Эскиз
<b>1. Создание эскиза.</b> Данную команду можно вызвать через: <ul style="list-style-type: none"> <li>– панель интерфейса. В появившемся окошке нужно выбрать рабочую плоскость;</li> <li>– рабочую плоскость. При нажатии на нужную рабочую плоскость высвечивается панель с командами. Выбираем команду <i>Создать эскиз</i></li> </ul>		
<b>1.1</b> 		<p>- Красный прямоугольник отмечает места, где можно выбрать иконки «Создать эскиз»</p> <p>- Не забываем нажимать на  , ее роль заключается в завершении и сохранении команды</p>
<b>2. Создание цилиндра.</b> Данную поверхность можно создать через: <ol style="list-style-type: none"> <li>1) эскиз. Выбрав команду <i>Окружность по центру и радиусу</i> (2.1.1), задаем в появившемся окне параметры (см. настройки). Затем, выбрав команду <i>Вытягивание</i> (2.1.2), задаем параметры (см. настройки);</li> <li>2) примитив. Выбираем команду <i>Примитив</i> (2.2.1) цилиндр. Задаем необходимые настройки и создаем эскиз</li> </ol>		
<b>2.1.1</b> 		<p>1)</p> <p>2)</p> <p>-Красным прямоугольником показано расположение команд и операций, которые должны быть выполнены.</p> <p>- Если операция выполнена правильно,  приобретает зеленый цвет</p> <p>- Не забываем нажимать на  для завершения и сохранения команды</p>
<b>2.1.2</b> 		
<b>2.2.1</b> 		

Продолжение таблицы

Команда	Настройки	Эскиз
<p><b>3. Построение ступеней вала.</b> Такие поверхности можно создать через:</p> <p>1) эскизы. Выбираем поверхность построения и нажимаем на команду <i>Окружность по центру и радиусу</i> (3.1.1), задаем в появившемся окне параметры (см. настройки). Затем, выбрав команду <i>Вытягивание</i> (3.1.2), задаем параметры (см. настройки);</p> <p>2) примитив. Выбираем команду <i>Примитив</i> (3.2.1) <i>цилиндр</i>. Задаем необходимые настройки и создаем эскиз. Выбираем перемещение по оси z и задаем необходимое значение</p>		
<p>3.1.1</p>  <p>3.1.2</p>  <p>Вытягивание</p> <p>3.2.1</p>  <p>Примитив</p>	<p>3.1.1 <math>\varnothing 60, h=90</math> 3.1.2 – рис. 1) 3.2.1 – рис. 2)</p> 	 <p>Примечание: Использование примитивов при построении значительно сокращает время работы. Но, если требуется построить нестандартную модель, т.е. нет нужного примитива, то необходимо использовать эскизное построение. Причем <u>возникает трудность</u> связать эти два метода построения модели. Решение данной проблемы в следующем пункте.</p> <p>- Аналогично строим следующую ступень вала <math>\varnothing 50, h=20</math></p>
<p><b>4. Построение зубьев вала</b></p> <p>1) Создаем эскиз на поверхности ступени вала.</p> <p>2) Используя команду <i>Окружность по центру и радиусу</i>, строим два круга (эскиз 1).</p> <p>3) Для лучшей ориентации строим <i>осевую линию</i>.</p> <p>4) Используем команду <i>Прямоугольник по центру и точке</i> (эскиз 2).</p> <p>5) Усекаем ненужные линии для образования зуба командой <i>Обрезка/удлинение</i>.</p> <p>6) Используем команду <i>Круговой массив</i>. Выделяем сначала один объект, <u>затем нажимаем на центр окружности</u> и только после этого выбираем оставшиеся объекты (эскиз 3). Необходимые параметры указаны в настройках. Создаем массив.</p> <p>7) Редактируем эскиз – убираем лишние линии командой <i>Обрезка/удлинение кривой</i>. Причем удаляем линии <u>начиная со второго зуба</u>. При удалении линий с первого зуба чертеж может построиться некорректно (эскиз 4).</p> <p>8) Используем команду <i>Вытягивание</i> и создаем модель (эскиз 5).</p> <p>9) Используя <i>булеву</i> операцию, <u>объединяем</u> все объекты (эскиз 6)</p>		
<p>4.4</p>  <p>4.5</p>  <p>Обрезка/удлинение кривой</p> <p>4.6</p>  <p>4.9</p> 	<p>4.4</p> 	 <p>Эскиз 1 для команды 4.2 Эскиз 2 для команды 4.4 Эскиз 3 для команды 4.6</p> <p>Эскиз 4 для команды 4.7 с панелью интерфейса</p>

Продолжение таблицы

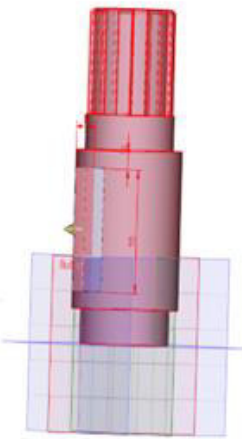
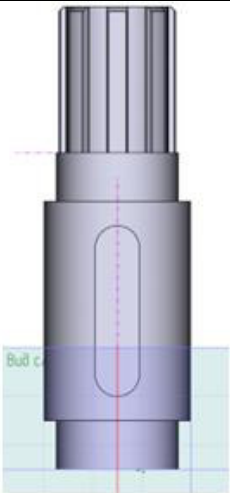

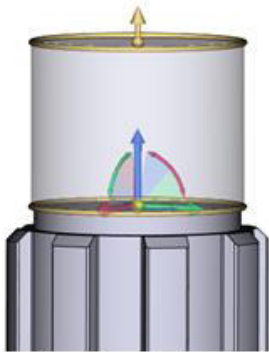
Команда	Настройки	Эскиз
	<p><b>4.5</b></p> <p>Массив ? X</p> <p>Объекты [3]</p> <ul style="list-style-type: none"> <li>Отрезок &lt;423&gt;</li> <li>Дуга окружности &lt;363&gt;</li> <li>Отрезок &lt;425&gt;</li> </ul> <p>Режим: Полный оборот</p> <p>Число копий: 1</p> <p>Шаг: 45 По точкам</p> <p>Угол: 360 По точкам</p> <p>Исходная точка</p> <p>X: -4 Y: 22.14</p> <p>Объект: Выберите объект</p> <p>Центр</p> <p>X: 0 Y: 0</p> <p>Объект: Начальная точка (Отрезок &lt;365&gt;)</p> <p>ΔX: -0.0000000000000001 ΔY: -22.308</p> <p>Длина: 22.498 Угол: 280.241</p> <p>Ассоциативно: Нет</p> <p><input checked="" type="checkbox"/> Создать группу</p> <p><input checked="" type="checkbox"/> Непрерывный ввод</p> <p><input type="checkbox"/> Продолжить создание</p>	 <p>Эскиз 5 для команды 4.8</p>  <p>Эскиз 6 для команды 4.9</p>
<p><b>5. Фаска.</b> Используем команду <i>Фаска по углу</i>. Задаем параметры, указанные в окне настроек. Выделяем зубья вала и создаем фаску (эскиз 1)</p>		
<p><b>5.1</b></p> 	<p><b>5.1</b></p> <p>Фаска ? X</p> <p>Общее сечение: 2</p> <p>Общий угол: 45</p> <p>Объекты для фаски [8]</p> <ul style="list-style-type: none"> <li>Ребро (Тело &lt;329&gt;)</li> <li>Ребро (Тело &lt;329&gt;)</li> </ul> <p>Сечение: 2</p> <p>Угол: 45</p> <p><input checked="" type="checkbox"/> Обратное направление</p> <p><input type="checkbox"/> Создать размер</p> <p><input type="checkbox"/> Продолжить создание</p>	 <p>Эскиз 1 для команды "Фаска"</p>
<p><b>6. Скругления</b></p> <p>1) Используем команду <i>Скругление с постоянным радиусом</i>. Выделяем грани внешнего диаметра зубьев (эскиз 1). Необходимые параметры указаны в настройках. Создаем скругления.</p> <p>2) Таким же методом создаем скругления для внутреннего диаметра (эскиз 2) и между ступенями вала (эскиз 3)</p>		
<p><b>6.1</b></p> 	<p><b>6.1</b></p> <p>Скругление ? X</p> <p>Управление набором</p> <p>Радиус: Постоянный</p> <p>Значение: 0.5</p> <p><input type="checkbox"/> Изменить направление ребра</p> <p>Сферическая вершина</p> <p>Параметры округления</p> <p>Дополнительно</p> <p><input type="checkbox"/> Продолжить создание</p>	

Продолжение таблицы

Команда	Настройки	Эскиз
	<div><div>6.2</div><div><div>Скругление</div><div><div><div><div><div><div>?</div><div>×</div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div></div><div><div>Управление набором</div><div>Радиус</div><div>Постоянный</div><div>Значение</div><div>0,3</div><div>mm</div><div><div><div><div><div>Изменить направление ребра</div><div>Сферическая вершина</div><div>Параметры скругления</div><div>Дополнительно</div><div>Продолжить создание</div></div></div></div></div></div></div><div><div>6.3</div><div><div>Скругление</div><div><div><div><div><div><div>?</div><div>×</div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div></div><div><div>Управление набором</div><div>Радиус</div><div>Постоянный</div><div>Значение</div><div>2</div><div>mm</div><div><div><div><div><div>Изменить направление ребра</div><div>Сферическая вершина</div><div>Параметры скругления</div><div>Дополнительно</div><div>Продолжить создание</div></div></div></div></div></div></div></div><div><div>Эскизы для команды "Скругление"</div><div><div><div>Эскиз 1</div><div>Эскиз 2</div></div><div><div>Эскиз 3</div></div></div><div><div>Примечание.</div><div>При нажатии на команду <i>Скругление</i> с постоянным радиусом через панель интерфейса переносит на <i>скругление с переменным радиусом</i>. Это некорректная работа ссылок. В данной ситуации создаем скругления по окну настроек, как указано в сценарии</div></div></div></div></div></div></div></div></div></div>	
<div><div>7. Построение шпоночного отверстия</div><div><div>1. Используем команду <i>Создание параллельной рабочей плоскости</i>. Выбираем любую плоскость, перпендикулярную основанию вала, затем переносим на расстояние, заданное в окне настроек, т.е. радиус цилиндра, на котором располагается шпоночное отверстие (эскиз 1).</div><div>2. Создаем эскиз на полученной плоскости.</div><div>3. Для удобства строим осевую линию. Так будет проще поймать нужные точки и образмерить чертеж шпонки.</div><div>4. Создаем прямоугольник, используя команду <i>Прямоугольник по центру и точке</i>. Центр выбираем приблизительный и пока не задаем размеры.</div><div>5. Используем команду <i>Линейный размер</i>, от верхней точки задаем нужный размер, а затем и длину шпоночного отверстия. Создаем эскиз (эскиз 2).</div><div>6. Используем команду <i>Вытягивание</i>. В окне настроек указываем в значении длины отрицательное значение, так как у нас <b>вычитание</b>, или меняем направление. Указываем тело, из которого вырезаем шпонку (т.е. выделяем вал), и в типе булевой операции выбираем команду <i>Вычитание</i> (эскиз 3).</div><div>7. Используем команду <i>Скругление</i> для того, чтобы убрать углы (эскиз 4).</div></div></div>		
<div><div>7.1</div><div><div><div>7.5</div><div><div><div><div><div><div>?</div><div>×</div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div><div><div><div>?</div><div>↺</div></div><div><div>✗</div><div>✓</div></div></div></div><div><div>Базовый объект</div><div>Вид спереди</div><div>Объект</div><div>Выберите объект</div><div>Расстояние</div><div>30</div><div>mm</div><div><div><div><div><div>Другой вариант</div><div>Визуальные границы плоскости</div><div>Левая граница</div><div>-50</div><div>mm</div><div>Правая граница</div><div>50</div><div>mm</div><div>Нижняя граница</div><div>-50</div><div>mm</div><div>Верхняя граница</div><div>50</div><div>mm</div><div>Маски</div><div>Тип</div><div>Нет</div></div></div></div></div></div></div></div><div><div>Эскиз 1 для команды 7.1</div><div>Эскиз 2 для команды 7.3 - 7.5</div></div></div></div></div></div>		



Продолжение таблицы

Команда	Настройки	Эскиз	
	<div><div>7.5</div><div><div>Размер</div><div><div>?</div><div>✕</div></div><div><div>?</div><div>↺</div><div>✕</div><div>✓</div></div><div><div>↔</div><div>↕</div><div>↻</div><div>↔</div><div>↕</div></div><div><div>Постановка</div><div>Между объектами</div><div>По отрезку</div><div>»</div></div><div><div>Тип размера</div><div>↔</div><div>↕</div><div>↻</div></div><div><div>Первый объект</div><div>Конечная точка (Отрезок &lt; ✕</div></div><div><div>Второй объект</div><div>Середина линии (Отрезок &lt; ✕</div></div><div><div>✓</div><div>Показывать в 3D</div></div><div><div>Номинал</div><div>Номинал</div><div>10</div><div>У</div></div><div><div>Знач</div><div>Точность</div><div>0.01</div></div><div><div>Текст</div><div>До</div><div>После</div><div>45°</div></div><div><div>Под</div></div><div><div>Разнесение</div><div>Автоматическое</div></div><div><div>Положение</div><div>Над размерной линией</div></div><div><div>Допуск / посадка</div></div><div><div>Стиль</div></div><div><div>Управление цепочкой</div></div><div><div>Продолжить создание</div></div></div></div>	<div><div></div><div></div></div> <div><div>Эскиз 3 для команды 7.6</div><div>Эскиз 4 для команды 7.7</div></div>	
	<div><div>7.6</div><div><div>Вытягивание</div><div><div>?</div><div>✕</div></div><div><div>?</div><div>↺</div><div>✕</div><div>✓</div></div><div><div>Направление</div><div>Выберите направление</div></div><div><div>Длина</div><div>Значение</div></div><div><div>Значение длины</div><div>-7</div><div>MM</div></div><div><div>Начальная точка</div><div>Нет</div></div><div><div>Уклон</div></div><div><div>Тип результата</div><div>Авто</div></div><div><div>Тип булевой</div><div>✕</div><div>✕</div><div>✕</div><div>✕</div></div><div><div>Операнд</div><div>Тело &lt; Вычитание</div><div>✕</div></div><div><div>Дополнительно</div></div><div><div>Точность</div><div>0.001</div><div>MM</div></div><div><div>Создать размер</div></div><div><div>✓</div><div>Продолжить создание</div></div></div></div>		
	<div><div>7.7.</div><div><div>Скругление</div><div><div>?</div><div>✕</div></div><div><div>?</div><div>↺</div><div>✕</div><div>✓</div></div><div><div>Управление наборами</div></div><div><div>Радиус</div><div>Постоянный</div></div><div><div>Значение</div><div>0</div><div>MM</div></div><div><div>Изменить направление ребра</div></div><div><div>Сферическая вершина</div></div><div><div>Параметры скругления</div></div><div><div>Дополнительно</div></div><div><div>Продолжить создание</div></div></div></div>		
<div>8. Построение канавки</div> <div>Выбираем поверхность построения, затем используем команду <i>Примитив - цилиндр</i>. Задаем необходимые настройки (8.1) и создаем эскиз (эскиз 1). Выбираем перемещение по оси z и задаем значение.</div> <div>9. Построение ступени</div> <div>Принцип аналогичный (эскиз 2). Необходимые настройки указаны в столбце «настройки» (9.1).</div>			
<div>8.1</div> <div><div>Примитив</div><div>Примитив</div></div> <div>9.1</div> <div><div>Примитив</div><div>Примитив</div></div>	<div><div>8.1</div><div><div>Примитив</div><div><div>?</div><div>✕</div></div><div><div>?</div><div>✕</div><div>✓</div></div><div><div>Привязка</div><div>По грани</div></div><div><div>Диаметр</div><div>38</div><div>MM</div></div><div><div>Высота</div><div>4</div><div>MM</div></div><div><div>По габаритам</div></div><div><div>Тип результата</div><div>Твёрдое тело</div></div><div><div>Тип булевой</div><div>✕</div><div>✕</div><div>✕</div><div>✕</div></div><div><div>Операнд</div><div>Тело &lt;329&gt;</div><div>✕</div></div><div><div>Преобразования</div></div><div><div>↔</div><div>↕</div><div>↻</div><div>↔</div><div>↕</div></div><div><div>Перенесение</div><div>190</div><div>MM</div></div><div><div>Дополнительно</div></div><div><div>Точность</div><div>0.001</div><div>MM</div></div></div></div>	<div><div></div><div></div></div> <div><div>Эскиз 1 для команды "Построение канавки"</div><div>Эскиз 2 для команды "Построение ступени"</div></div>	





Окончание таблицы

Команда	Настройки	Эскиз
	<div><div>11.2</div><div><div>Фаска</div><div><div><div><div></div><div></div><div></div></div><div><div></div><div></div><div></div></div></div><div>Общее сечение 1</div><div>Объекты для фаски [1]</div><div>Ребро (Тело &lt;329&gt;)</div><div>Сечение 1</div><div><div><div>Создать размер</div><div>Продолжить создание</div></div></div></div></div></div> <div><p>Примечание: при построении фаски на эскизе 2, могут возникнуть трудности из-за выбора неверных граней, здесь важно курсором мыши выделить грань, предназначенную для фаски, не задев зубья шлицев</p></div>	
<div>12. Создание центровых отверстий</div> <div>1. Создаем эскиз на верхней поверхности вала. Затем на панели интерфейса выбираем команду 2D-точка (12.1). В появившемся окне выбираем команду Точка по координатам (12.2) и строим точку в центре (эскиз 1). Это необходимо для построения отверстия. Далее сохраняем эскиз.</div> <div>2. На панели интерфейса выбираем команду Отверстие (12.3). В появившемся окне выбираем команду Глухое отверстие (12.4). Затем задаем параметры, указанные в столбце настройки (12.2). Исходные тела – выделяем поверхность, на которой будет будущее отверстие. Положение – указываем точку, которую мы построили ранее.</div> <div>3. Аналогичный порядок действий выполняется для построения центрального отверстия на нижней поверхности вала</div>		
<div><div>12.1</div><div>2D-точка</div><div>12.2</div><div>12.3</div><div>Отверстие</div><div>12.4</div></div>	<div><div>12.2</div><div><div>Отверстие</div><div><div><div><div></div><div></div><div></div></div><div><div></div><div></div><div></div></div></div><div>Форма отверстия</div><div>Простая</div><div>Начало</div><div>Размер фаски 2</div><div>Угол фаски 45</div><div>Диаметр 5</div><div><div><div>Создать управляющий размер</div><div>Угол конуса 120</div></div></div><div>Глубина 5</div><div>Общие параметры</div><div>Грань Грань &lt;686&gt; (Скругление &lt;685&gt;)</div><div>Исходные тела [1]</div><div>Тело &lt;329&gt;</div><div><div><div>Направление</div><div>Положение</div></div><div>Точка</div><div><div><div>Создать ось</div><div>Продолжить создание</div></div></div></div></div></div></div>	<div><div><div>Эскиз 1</div><div>Эскиз 2</div></div></div>

Обсуждение результатов

При разработке сценария на примере детали «Вал» было проведено сравнение с построением этой же 3D-модели в ПО «КОМПАС-3D» и сделаны следующие выводы:

1. Затраты времени в сравнении с программой «Компас-3D». «САРУС» – 29 мин., «КОМПАС-3D» – 24 мин. Полученные результаты показывают незначительные различия скорости создания модели. Разница в 5 минут связана с более долгой загрузкой (срабатывания команд) в программе «САРУС» (при создании фасок или скруглений), параметризация эскизов требует больше времени, чем в «КОМПАС-3D», а также возникновение некоторых программных ошибок, которые быстро решаются, но затрачивается время на их исправление.

2. *Влияние владения командами CAD СПЖ «САРУС».* Несколько раз была построена 3D-модель детали «Вал» различными способами. По мере повторения одного и того же алгоритма время, затрачиваемое на построение модели, становилось с каждым разом все меньше, поскольку выработался навык и были освоены базовые команды.

3. *Тенденции развития.* При исключении программных ошибок, недочетов и более быстрой загрузке срабатывания команд по мере их исправления и введения обновлений CAD СПЖ «САРУС» будет иметь колоссальное преимущество во времени, функционале и применяемости.

### Выводы

Проведенное исследование выявило основные сложности, за счёт которых увеличивается затрачиваемое время на построение 3D-модели. Разработанный «Пользовательский сценарий» включает алгоритм построения, который является наиболее оптимальным и подходит для освоения базовых команд для пользователя.

При создании 3D-модели используются основные команды ПМ CAD, которые являются ключевыми для понимания и дальнейшего освоения данной программы. Овладев базовыми командами, пользователь получает знания и навыки для проектирования более сложных конструкций.

### Источник финансирования

Договор от 25 марта 2024 г. № 96-2024/244 на выполнение научно-исследовательской работы по теме «Проведение научных исследований в интересах ФГУП «РФЯЦ-ВНИИЭФ» с использованием результатов для актуализации образовательных программ подготовки специалистов с высшим профессиональным образованием для ФГУП «РФЯЦ-ВНИИЭФ», Задача 10 «Исследование программно-аппаратных решений в области цифровых технологий, возможностей их отраслевого применения, разработка и версионное тестирование компонентов полностью защищенной системы полного жизненного цикла среднего класса СПЖ V3 «САРУС».

### Благодарности

Авторы благодарят кафедру цифровых технологий СарФТИ-НИЯУ МИФИ под руководством О.В. Кривошеева, к.т.н., зав. кафедрой, зам. директора ФГУП «РФЯЦ-ВНИИЭФ» по технологиям полного жизненного цикла – директора Института цифровых технологий (ИТЦ) – главного конструктора СПЖ;

- научно-исследовательский отдел поддержки и внедрения Института цифровых технологий РФЯЦ-ВНИИЭФ (руководитель Г.С. Лазарев).

### Список литературы

1. Денисова Н.А., Федоренко Г.А. Применение в учебном процессе вуза тестирования программного модуля CAD вновь создаваемого САПР «САРУС» // GraphiCon 2023: труды 33-й Междунар. конф. по компьютерной графике и машинному зрению (Москва, 19–21 сент. 2023 г.). М.: Институт прикладной математики им. М.В. Келдыша РАН, 2023. С. 848-859.

2. Отчёт о НИР «Исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖ V3» (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. номер НО/38-278.10-2023-О от 15.11.23. 2023. С. 64–88.

3. Отчет о научно-исследовательской работе «Исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖ V3». (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. № НО/37-278.10-2023-О. 2024. С. 8–27.

4. Отчет о научно-исследовательской работе «Исследование цифровых технологий и возможностей их отраслевого применения, разработка и тестирование компонентов СПЖ V3». (Договор на выполнение НИР от 2 марта 2023 г. № 96-2023/203, Задача 10). Инв. № НО/37-278.10-2023-О. 2025. С. 14–46.

5. Прыткова Ю.Б., Денисова Н.А. Роль системы координат в построении 3D-модели в СПЖ «САРУС» // GraphiCon 2024 : материалы 34-й Междунар. конф. по компьютерной графике и машинному зрению (Россия, Омск, 17–19 сент. 2024 г.). Омск : Изд-во ОмГТУ, 2024. С. 795-804.

6. Комплекс программ в защищённом исполнении «Система полного жизненного цикла изделий «Цифровое предприятие»: Основная версия программного модуля. «Система конструкторского проектирования» (версия 2) / Руководство оператора 07623615.00423-06 34 01. 2021. 338 с.



Секция 5

# ***ОБРАБОТКА И АНАЛИЗ БИОМЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ***

---

PROCESSING AND ANALYSIS  
OF BIOMEDICAL IMAGES



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия

УДК 004.93'1

DOI: 10.25686/978-5-8158-2474-4-2025-781-789

## Комплексная система анализа фундус-изображений с применением методов глубокого обучения

**К. Д. Аксенов<sup>1,2</sup>, Л. Е. Аксенова<sup>1,2</sup>**<sup>1</sup>ООО «Пространство интеллектуальных решений», Новороссийск, Краснодарский край, Россия<sup>2</sup>Новороссийский политехнический институт (филиал) Кубанского государственного технологического университета, Новороссийск, Краснодарский край, Россия

**Аннотация.** Целью работы является создание вычислительного конвейера для автоматизированного анализа фундус-изображений, включающего классификацию патологий, сегментацию биомаркеров и сосудистой сети, а также интеграцию алгоритмов в специализированную медицинскую платформу. Для обучения моделей использованы открытые наборы данных. Для классификации применена мультимодальная трансформерная модель с архитектурой Vision Transformer и BERT, которая показала значение AUC 0.9155 на тестовой выборке. Для сегментации сосудов использовали архитектуры семейства U-Net (U-Net++, Attention U-Net), а для биомаркеров — U-Net с энкодером ResNeSt-50 и Mask R-CNN, что обеспечило высокие значения метрик Dice и IoU на стандартных датасетах. Все алгоритмы интегрированы в микросервисную платформу, включающую сервер Django, DICOM-хранилище Orthanc, систему управления данными (PostgreSQL, Redis) и модуль AI-ассистента формирования отчётов. Нагрузочные испытания подтвердили устойчивость системы. Полученные результаты демонстрируют возможность создания комплексных решений, которые способны автоматизировать несколько этапов анализа изображений глазного дна. В то же время требуется и дополнительная валидация предложенного подхода на клинических данных для оценки его диагностической ценности.

**Ключевые слова:** искусственный интеллект, офтальмология, фундус-изображения, сегментация, классификация

## A comprehensive system for fundus image analysis using deep learning methods

**K. D. Aksenov<sup>1,2</sup>, L. E. Aksenova<sup>1,2</sup>**<sup>1</sup>LLC Predict Space, Novorossiysk, Krasnodar Region, Russia<sup>2</sup>Novorossiysk Polytechnic Institute (branch), Kuban State Technological University, Novorossiysk, Krasnodar Region, Russia

**Abstract.** The aim of this study is the development of a computational pipeline for automated fundus image analysis, including pathology classification, biomarker and vessel segmentation, as well as the integration of algorithms into a specialized medical platform. Open datasets were used for model training. For classification, a multimodal transformer model combining Vision Transformer and BERT architectures was employed, achieving an AUC of 0.9155 on the test set. Vessel segmentation was performed using U-Net family architectures (U-Net++, Attention U-Net), while biomarker segmentation was carried out with a U-Net model using a ResNeSt-50 encoder and Mask R-CNN, yielding high Dice and IoU metrics on standard datasets. All algorithms were integrated into a microservice-based platform comprising a Django server, an Orthanc DICOM storage, a data management system (PostgreSQL, Redis), and an AI assistant module for report generation. Load testing confirmed the robustness of the system. The results demonstrate the feasibility of developing comprehensive solutions capable of automating multiple stages of fundus image analysis. At the same time, additional validation on clinical data is required to assess the diagnostic value of the proposed approach.

**Keywords:** artificial intelligence, ophthalmology, fundus images, segmentation, classification

### Введение

Скрининг заболеваний сетчатки по изображениям глазного дна (фундус-изображениям) остаётся важной и одновременно сложной задачей современной офтальмологии. Согласно данным литературы, отмечается устойчивое развитие программ скрининга заболеваний сетчатки с использованием фундус-камер и технологий искусственного интеллекта (ИИ) в мире. К примеру, сеть MinuteClinic (CVS Health) предлагает телемедицинские обследования сетчатки в более чем 1200 филиалах с применением портативного RetinaVue 700 Imager и удалённой оценки офтальмологами, что позволяет повысить охват пациентов и доступность скрининга [1]. Кроме того, совсем недавно FDA одобрило портативную систему Optomed Aurora с модулем AI-анализа, которая может проводить автономный скрининг заболеваний сетчатки менее чем за минуту, что важно в условиях нехватки специалистов и высоких временных затрат стандартной диагностики [2]. Несмотря на то, что фундус-съёмка является неинвазивной и относительно недорогой процедурой, её широкое применение ограничено такими факторами, как нехватка квалифицированных специалистов-офтальмологов, высокая нагрузка на

медицинские учреждения, а также вариации качества изображений, обусловленные разными типами камер, освещением, степенью прозрачности сред глаза и др. [3, 4]. В реальных клинических условиях диагностика часто требует тщательного осмотра, времени и ресурсов, что может приводить к задержкам выявления заболеваний, особенно в регионах с ограниченным доступом к офтальмологам.

Развитие машинного обучения (и в особенности глубокого обучения) за последние годы дало мощный толчок в решении этих проблем. Модели на основе сверточных нейронных сетей, трансформеров и их гибридов доказали свою способность автоматически распознавать проявления диабетической ретинопатии, возрастной макулярной дегенерации, окклюзий сосудов, — не только классифицировать состояния, но и сегментировать отдельные биомаркеры (кровоизлияния, микроаневризмы, экссудаты и др.). Например, в работе Duan с соавт. был разработан гибридный алгоритм на основе методов глубокого обучения, который использует DenseNet121 и XGBoost, для классификации 16 заболеваний по широкоугольным фундус-снимкам [5]. Другие исследования подчеркивают рост числа разработок, направленных на раннее выявление и автоматизацию анализа фундус-изображений при диабетической ретинопатии, глаукоме и возрастной макулярной дегенерации [6]. Среди конкретных работ по сегментации сосудов и биомаркеров особенно выделяются модели, которые объединяют визуальную сегментацию сосудистой сети и очагов поражения. Например, в работе Kim и др. представлена модель, сочетающая механизмы внимания, многомасштабное слияние признаков (JPU) и блоки трансформеров [7].

Таким образом, уже существует достаточное число публикаций, подтверждающих, что проблемы, связанные с нехваткой специалистов, вариабельностью данных, необходимостью быстрых и автоматизированных решений, могут быть решены средствами машинного и глубокого обучения [8]. Однако многие модели ориентированы на одну задачу — либо классификацию заболеваний, либо сегментацию структур, либо обнаружение конкретного биомаркера. В то же время лишь незначительное количество работ посвящено комплексным системам, которые одновременно автоматизируют несколько этапов оценки изображений врачом — от дифференциальной диагностики и выделения биомаркеров до создания отчета [9].

### Постановка задачи

Целью исследования являлось создание вычислительного конвейера обработки изображений глазного дна и его интеграции в медицинскую информационную систему. Задачами исследования являлись разработка моделей классификации и сегментации изображений глазного дна, разработка архитектуры платформы, а также интеграция моделей в платформу.

### Теория

#### 1. Классификация патологий

Для обучения и валидации модели классификации использовали открытый набор данных RFMiD [6]. Примеры изображений приведены на рисунке 1.

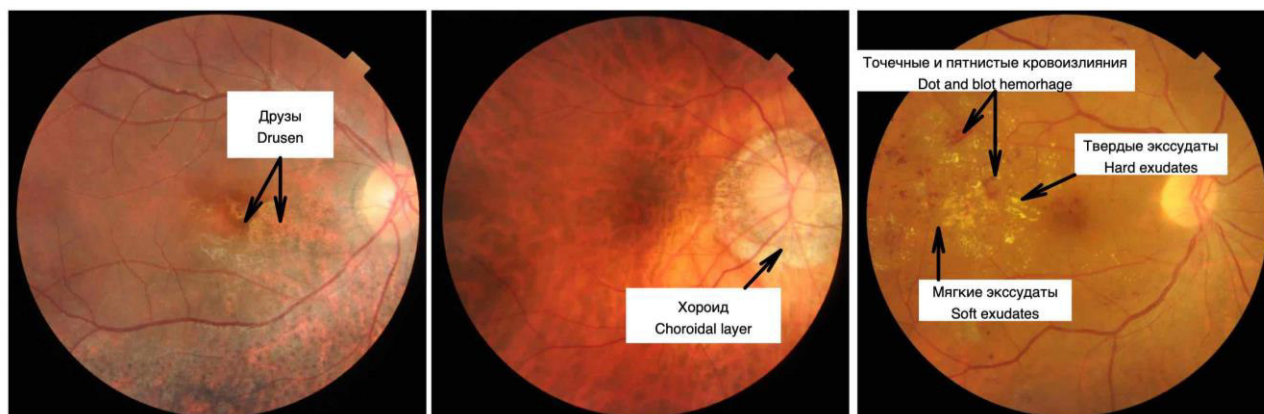


Рисунок 1. Примеры изображений глазного дна из набора данных RFMiD для классов: а) друзы (Drusens, DN); б) миопия (Myopia, MYA); в) окклюзия ветви центральной вены сетчатки (Branch Retinal Vein Occlusion, BRVO) и определяющие данные классы характерные визуальные признаки

Он состоит из 3200 снимков дна, полученных с помощью трех различных фундус-камер и размечен относительно 46 классов двумя экспертами. В настоящем исследовании использована вторая версия набора данных RFMID, в которой все классы с числом изображений менее 10 были объединены в одну обобщённую категорию «Другие». Таким образом, итоговое количество классов для обучения и оценки составило 29. Общий набор был разделен на обучающую (1920), валидационную (640) и тестовую (640) выборки. Количество данных приведено в таблице 1.

Таблица 1. Распределение изображений по классам заболеваний в наборе данных

Класс	Все	Тренинг	Валидация	Тест	Класс	Все	Тренинг	Валидация	Тест
Норма (NL)	669	401	134	134	Хориоретинит (CRS)	54	32	11	11
Диабетическая ретинопатия (DR)	632	379	126	127	Окклюзия центральной вены сетчатки (CRVO)	45	27	9	9
Помутнение оптических сред (MH)	523	314	105	104	Клетки пигментного эпителия сетчатки (RPEC)	32	19	6	7
Ковшевидное углубление диска зрительного нерва (ODC)	445	267	89	89	Макулярный рубец (MS)	27	16	5	6
Тесселяция (TSLN)	304	182	61	61	Эпиретинальная мембрана (ERM)	26	16	5	5
Друзы (DN)	230	138	46	46	Передняя ишемическая нейропатия зрительного нерва (AION)	26	16	5	5
Возрастная макулярная дегенерация (ARMD)	169	101	34	34	Астероидный гиалоз (AH)	25	15	5	5
Миопия (MYA)	167	100	33	34	Тракция сетчатки (RT)	25	15	5	5
Окклюзия ветви центральной вены сетчатки (BRVO)	119	71	24	24	Экссудация (EDN)	24	14	5	5
Побледнение диска зрительного нерва (ODP)	115	69	23	23	Парафовеолярная телеангиэктазия (PT)	17	10	3	4
Отек диска зрительного нерва (ODE)	96	58	19	19	Макулярное отверстие (MHL)	17	10	3	4
Лазерные рубцы (LS)	79	47	16	16	Оптоцилиарный шунт (ST)	11	7	2	2
Ретинит (RS)	71	43	14	14	Пигментный ретинит (RP)	10	6	2	2
Центральная серозная ретинопатия (CSR)	61	37	12	12	Извитые сосуды (TV)	10	6	2	2
					Другие	67	40	13	14

Для решения задачи многоклассовой классификации заболеваний по изображениям глазного дна в настоящем исследовании использовалась трансформерная мультимодальная модель, сочетающая визуальный и текстовый энкодеры. Визуальный модуль представляет собой Vision Transformer (ViT), извлекающий глобальные пространственные признаки из цветных фундус-фотографий (CFP), а текстовый модуль основан на архитектуре BERT и обрабатывает клинические отчёты, получая семантические текстовые представления. На этапе предварительного обучения модель обучалась на триплетах «левый глаз – правый глаз – текст» в парадигме контрастного обучения, где формировались сопоставления между изображениями и текстами на трёх уровнях: левый глаз, правый глаз и уровень пациента в режиме самоконтролируемого обучения на 1,6 млн неразмеченных фундус-изображений [10]. Таким образом, модель может одновременно учитывать как локальные особенности отдельных глаз, так и информацию о пациенте в целом. Веса визуальных энкодеров для обоих глаз были общими (shared weights). В рамках настоящего исследования предварительно обученная модель была дообучена на наборе данных RFMID, содержащем изображения глазного дна с разметкой по 29 классам. Дообучение выполнялось в течение 75 эпох с использованием оптимизатора AdamW, фиксированной скоростью обучения  $3 \times 10^{-5}$  и размером батча 256.



## 2. Сегментация сосудистой сети

Для задачи сегментации сосудистой сети на фундус-изображениях использовались архитектуры семейства U-Net (рис. 2). В качестве базовой модели применялись U-Net и её модификации, включая Attention U-Net и U-Net++, специально адаптированные для выделения сосудистых структур различного калибра. Энкодер выполнял последовательное извлечение признаков с понижением пространственного разрешения, что обеспечивало захват глобальных характеристик изображения, тогда как декодер восстанавливал пространственную структуру посредством апсемплинга и skip-соединений. Такая архитектура позволяла сохранять мелкие сосудистые ветви, важные для диагностики микроангиопатий. Для повышения точности использовались механизмы внимания (Attention Gates), позволяющие модели фокусироваться на области сосудистой сети и подавлять фоновые шумы. В качестве функции потерь применялась комбинированная метрика Dice Loss и Binary Cross-Entropy, что позволило учитывать как глобальную форму сосудистой сети (перекрывание предсказанной и истинной маски), так и точность локального распознавания мелких сосудов [11].

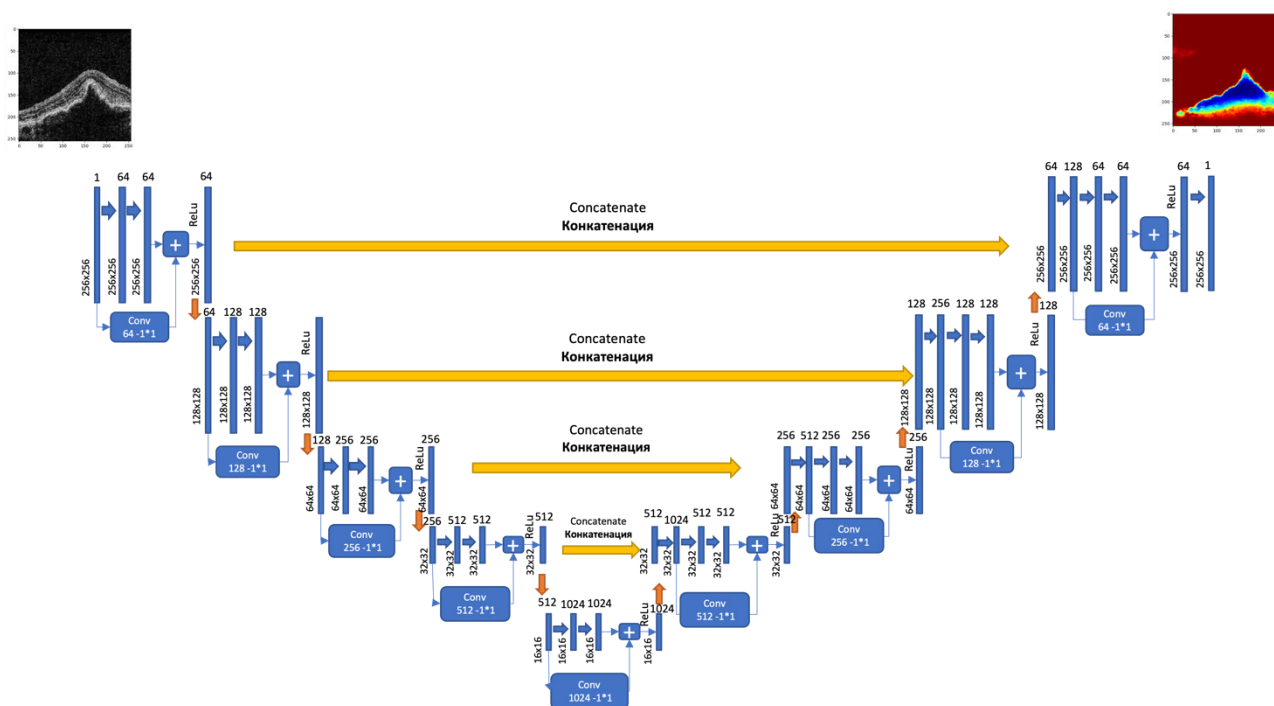


Рисунок 2. Архитектура U-Net

## 3. Сегментация биомаркеров

Для решения задачи автоматизированного выделения патологических структур на изображениях глазного дна были использованы два подхода: пиксельная сегментация и объектная детекция. Их реализация основана на архитектурах глубокого обучения, интегрированных в библиотеку fundus-lesions-toolkit [12]. Для семантической сегментации патологических очагов применялась архитектура U-Net с энкодером ResNeSt-50, обеспечивающим извлечение многоуровневых признаков из входного изображения. В качестве функции потерь использовались комбинированные Dice Loss и Focal Loss. Выход модели представляет собой многоканальную карту вероятностей, соответствующую четырём типам патологических очагов: мягкие экссудаты (cotton wool spots), твёрдые экссудаты, кровоизлияния и микроаневризмы.

Для задачи локализации и подсчёта очагов поражения использовалась архитектура Mask R-CNN, включающая базовую архитектуру ResNet с пирамидальной сетью признаков (FPN). На первом этапе Region Proposal Network (RPN) формирует кандидаты областей интереса, для которых затем выполняется уточнение с использованием операции RoIAlign. Для каждого выделенного региона параллельно решаются три подзадачи: классификация по типу поражения, регрессия координат ограничивающего прямоугольника и сегментация области. Такой подход позволяет не только



определять принадлежность очагов к конкретной категории, но и выделять их границы на изображении.

Для обеспечения воспроизводимости и сокращения времени обучения использовались предобученные модели, доступные через платформу HuggingFace Hub. Механизм загрузки весов позволял выбирать варианты моделей, обученные на различных открытых наборах данных (IDRiD, MESSIDOR, DDR, FGADR, RETINAL-LESIONS), а также использовать ансамблевые решения на основе объединённых датасетов. Это обеспечило высокую обобщающую способность алгоритмов.

#### 4. Цифровая платформа

В процессе проектирования и реализации цифровой платформы применялись современные методы архитектурного и программного проектирования. Архитектура системы была спроектирована с использованием принципов микросервисной модели, что обеспечило модульность, масштабируемость и гибкость интеграции компонентов. В качестве основного фреймворка для серверной части использован Django, реализующий бизнес-логику и REST API, а для клиентской части – React, обеспечивающий динамичный пользовательский интерфейс. Хранение структурированных данных организовано на базе реляционной СУБД PostgreSQL, в то время как кэширование реализовано с помощью Redis. Диагностические изображения обрабатываются и хранятся на DICOM-сервере Orthanc, обеспечивающем интеграцию с офтальмологическим оборудованием. Для задач автоматизированного анализа изображений применялись модели глубокого обучения, разработанные с использованием PyTorch и TensorFlow, с выносом тяжёлых вычислений в асинхронные задачи через Celery. Инфраструктура развернута в изолированной среде с использованием Docker, что гарантирует переносимость и удобство развертывания. Проектирование системы было основано на предварительном анализе клинического рабочего процесса врача-офтальмолога, что позволило структурировать этапы сбора, хранения и анализа данных. Были использованы следующие типы данных: демографическая информация пациентов, структурированные записи осмотров, данные медицинской визуализации, а также результаты аннотированных выборок для обучения ИИ-моделей. Для обеспечения совместимости и безопасности применены стандарты HL7, DICOM, механизмы авторизации, шифрования и логирования доступа к медицинским данным. Особое внимание уделено стандартизации интерфейсов взаимодействия модулей и унификации идентификации пациента.

### Результаты

В рамках работы были разработаны и протестированы алгоритмы анализа изображений глазного дна, включающие многоклассовую классификацию патологий, сегментацию сосудистой сети и выявление биомаркеров.

#### 1. Классификация патологий

Для задачи диагностики офтальмологических и системных заболеваний по фундус-изображениям использовалась мультимодальная трансформерная модель, объединяющая визуальные и текстовые признаки. Модель обучена на датасете RFMiD (3200 изображений, 29 классов) и показала высокие результаты:  $AUC = 0.9155$  на тестовой выборке без признаков переобучения (рис. 3).

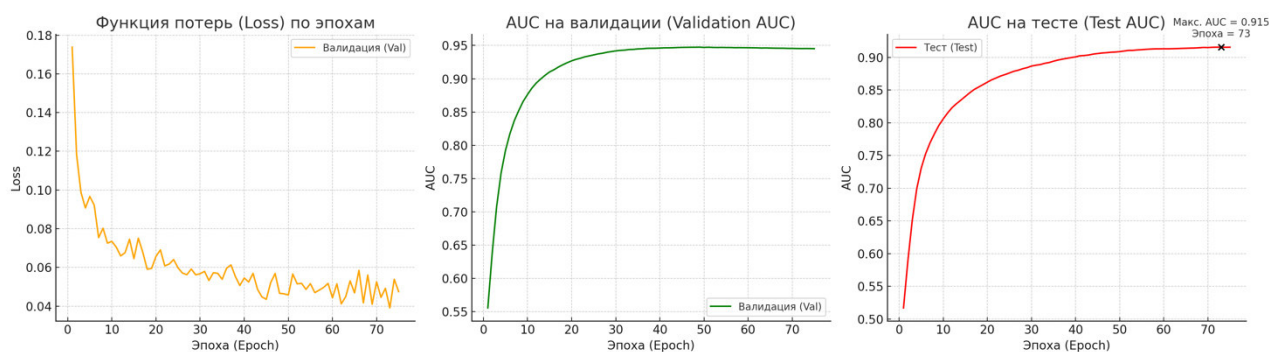


Рисунок 3. Динамика обучения модели: функция потерь (Loss), AUC на валидационном и AUC на тестовом наборах по эпохам

## 2. Сегментация биомаркеров и сосудов

Для выделения ключевых структур использованы специализированные модели:

- fundus-lesions-toolkit — сегментация очагов поражения (экссудаты, кровоизлияния, микроаневризмы);
- fundus-vessels-toolkit — сегментация сосудистой сети.

Обе модели адаптированы и интегрированы в единую вычислительную среду (рис. 4).

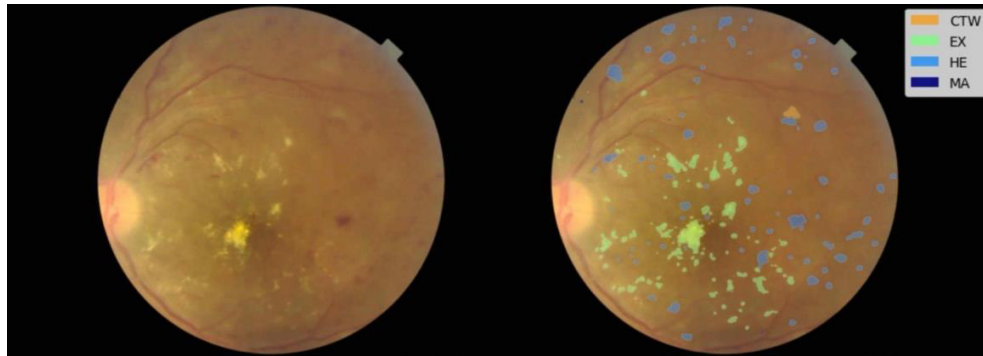


Рисунок 4. Примеры сегментации биомаркеров на реальных изображениях глазного дна: очаги мягких экссудатов (cotton wool spots - CTW), твёрдые экссудаты (exudates - EX), кровоизлияния (hemorrhages - HE) и микроаневризмы (microaneurysms - MA)

Результаты, полученные с помощью моделей классификации и сегментации, передаются на анализ большой языковой моделью, которая формирует предварительное заключение: предполагаемый диагноз, список биомаркеров, а также рекомендации по дополнительному обследованию. Такой подход обеспечивает интерпретацию данных в удобном для врача формате (рис. 5).

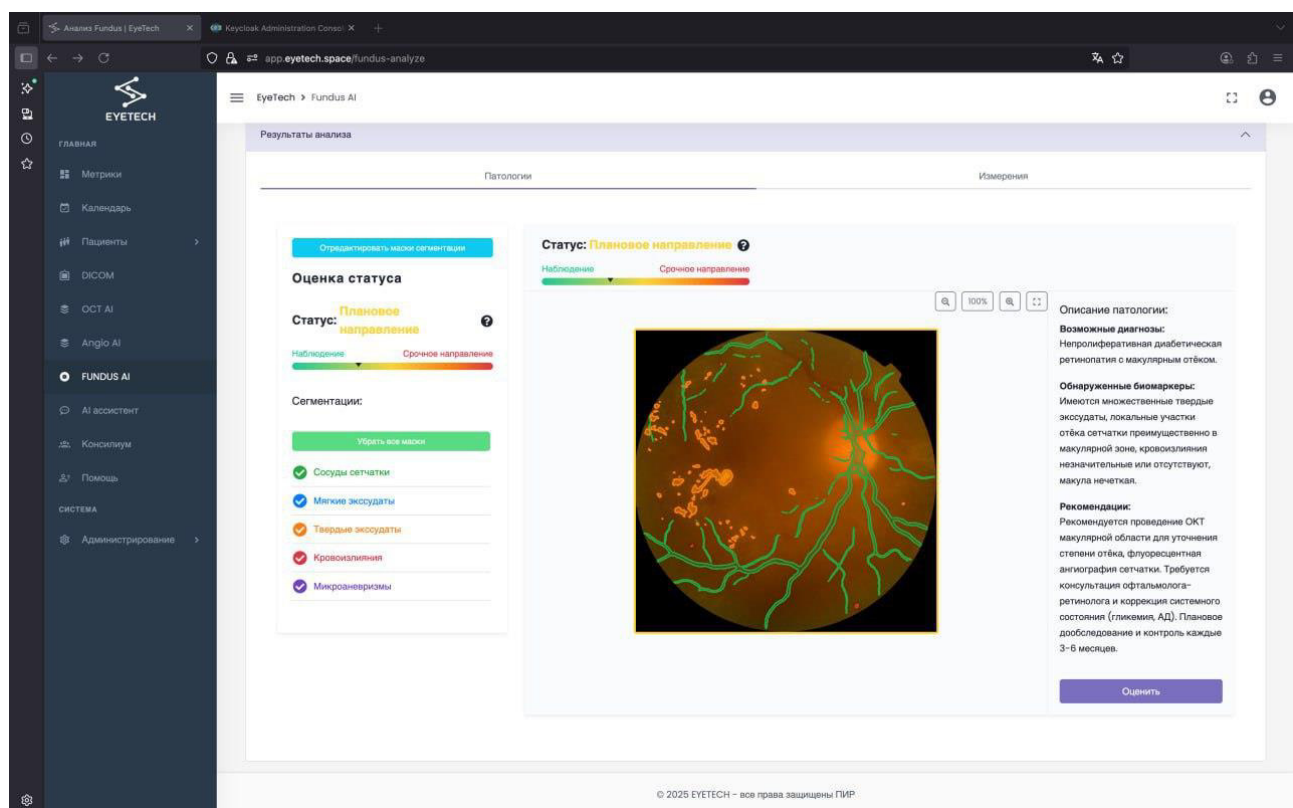


Рисунок 5. Графический интерфейс платформы с интегрированными моделями глубокого обучения и большой языковой моделью

### 3. Архитектура и инфраструктура платформы.

Для интеграции алгоритмов создана специализированная офтальмологическая платформа на основе микросервисной архитектуры (рис. 6).

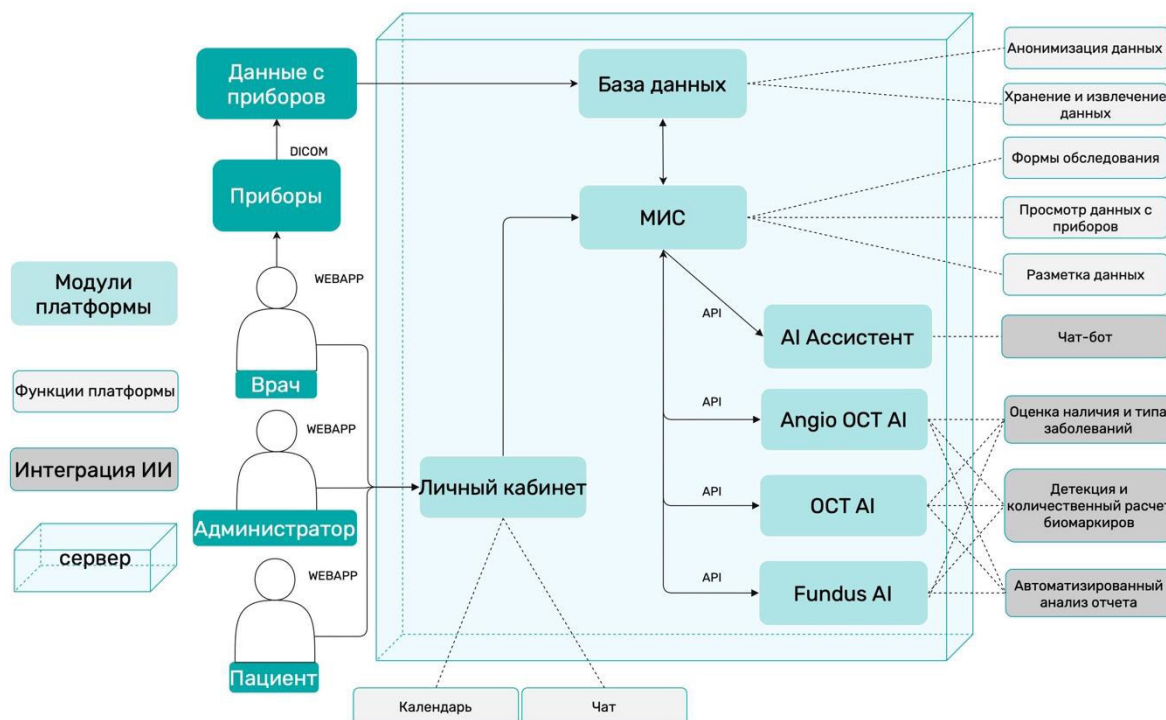


Рисунок 6. Высокоуровневая архитектура платформы

На рисунке 7 представлена микросервисная архитектура платформы, которая включает брокер задач и распределённую систему обработки данных. Такая архитектура обеспечивает масштабируемость и устойчивость системы за счёт декомпозиции на отдельные функциональные компоненты, каждый из которых отвечает за специфический набор задач: веб-интерфейс, обработку запросов, работу с базой данных, асинхронное выполнение задач и анализ изображений средствами машинного обучения. Центральное взаимодействие осуществляется через Django-бэкенд, который координирует обмен данными между пользователем, хранилищем медицинских изображений (ORTHANC), PostgreSQL-базой данных, брокером Redis и сервисами ИИ, реализованными на FastAPI.

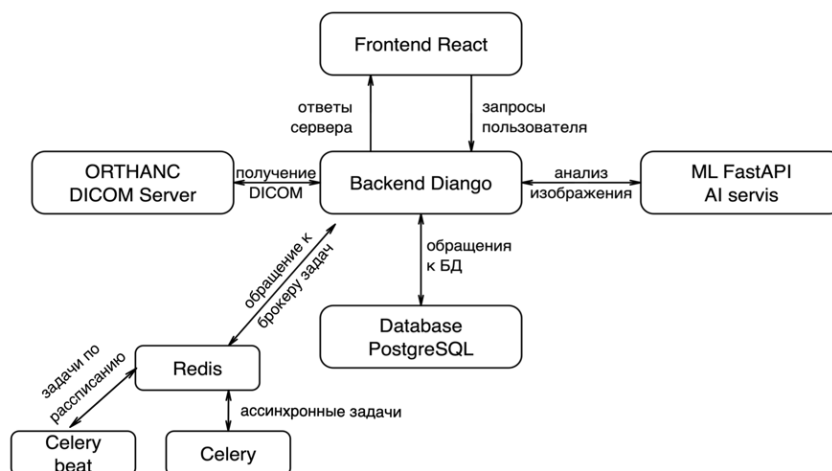


Рисунок 7. Микросервисная архитектура платформы

Нагрузочные тесты подтвердили масштабируемость системы: стабильная работа сохраняется при до 800 одновременных пользователей и 960 запросах/с.

### Обсуждение результатов

В ходе работы была разработана и протестирована многоуровневая система для вычислительного конвейера обработки изображений глазного дна и его интеграции в медицинскую информационную систему. Данная система представляет собой цифровую платформу, которая включает модули многоклассовой классификации патологий, сегментации сосудистой сети и выявления биомаркеров, DICOM-сервер Orthanc, систему управления данными (PostgreSQL, Redis, Django backend) и AI-ассистента. Полученные результаты показывают, что предложенные методы демонстрируют высокую точность и стабильное обучение на тестовых выборках из открытых наборов данных. В частности, мультимодальная трансформерная модель классификации заболеваний достигла значения AUC 0.9155 без признаков переобучения. Таким образом, результаты классификации можно рассматривать как основу для более широкого внедрения автоматизированных систем поддержки принятия решений в офтальмологической практике. Для задач сегментации сосудов и биомаркеров интеграция специализированных моделей (U-Net, Attention U-Net, Mask R-CNN) позволила достичь высоких значений метрик Dice и IoU, обеспечив корректное выделение как сосудистой сети, так и локальных патологических очагов. Однако стоит отметить, что точность для отдельных категорий поражений (например, микроаневризм) может быть ограничена качеством исходной разметки и дисбалансом классов. На следующем этапе анализа необходимо провести детальную оценку точности по каждому классу относительно различных метрик (чувствительность, специфичность, F1-score) с учётом клинической значимости каждого биомаркера.

В ходе исследования было показано, что офтальмологическая платформа является масштабируемой и устойчивой. Микросервисная архитектура обеспечивает возможность интеграции новых алгоритмов и модулей без модификации всей системы, что делает решение перспективным для тиражирования в другие области медицинской диагностики. Включение модуля анализа на основе большой языковой модели позволило автоматически формировать клинические заключения за счет объединения результатов классификации и сегментации в интерпретируемый для врача формат. Такой подход существенно повышает удобство использования системы и может способствовать сокращению времени постановки диагноза.

Несмотря на полученные положительные результаты, исследование имеет ряд ограничений. Во-первых, обучение и тестирование моделей проводилось на открытых наборах данных (RFMiD, IDRiD, MESSIDOR и др.), которые могут не полностью отражать клиническое разнообразие реальных случаев. Во-вторых, качество разметки, выполненной экспертами, не исключает вариабельности, что особенно важно для мелких структур, таких как микроаневризмы. В-третьих, в исследовании не проводилась внешняя валидация на клинических данных локальных медицинских учреждений, что ограничивает обобщаемость результатов. Следовательно, дальнейшие исследования должны быть направлены на адаптацию алгоритмов к реальным клиническим данным, оценку их диагностической ценности в условиях практического применения, а также на сравнение с результатами, получаемыми врачами-офтальмологами.

### Заключение

Разработанная система демонстрирует высокую степень автоматизации и поддержки принятия врачебных решений, а также отвечает требованиям надёжности, масштабируемости и совместимости с медицинской инфраструктурой. Полученные результаты подтверждают перспективность интеграции современных информационных решений в офтальмологическую практику и указывают на возможность их дальнейшего применения для других направлений медицинской диагностики.

### Источник финансирования

Исследование выполнено при финансовой поддержке Кубанского научного фонда, ООО ПИР в рамках проекта № НТИП-24.1/1 «Портативный оптический прибор для визуализации глазного дна со встроенными технологиями искусственного интеллекта».

**Список литературы**

1. Gustaitis J. CVS now offering retinopathy exams. *Diabetes Self-Management*. March 4, 2022. <https://www.diabetesselfmanagement.com/news-research/2022/03/04/cvs-now-offering-retinopathy-exams/> Accessed August 30, 2023.
2. Lupidi M., Danieli L., Fruttini D., Nicolai M., Lassandro N., Chhablani J., Mariotti C. Artificial intelligence in diabetic retinopathy screening: clinical assessment using handheld fundus camera in a real-life setting. *Acta Diabetol*. 2023 Aug;60(8):1083-1088. doi: 10.1007/s00592-023-02104-0. Epub 2023 May 8. PMID: 37154944; PMCID: PMC10166040
3. Cleveland, S. D., Baker, M. J., Erdman, A. G., Nazari, H. Current and future directions for the use of handheld fundus cameras in telehealth. *Expert Review of Medical Devices*. 2025, 22(7): 657–665. <https://doi.org/10.1080/17434440.2025.2508877>
4. Chan T.Y., Wang J.H., Chen N., Chiu C.J. The Assessment of Retinal Image Quality Using a Non-Mydriatic Fundus Camera in a Teleophthalmologic Platform. *Diagnostics (Basel)*. 2024 (Jul): 17;14(14):1543. doi: 10.3390/diagnostics14141543. PMID: 39061681; PMCID: PMC11275639
5. Duan M-M., Tu X. Deep learning-based classification of multiple fundus diseases using ultra-widefield images. *Front. Cell Dev. Biol*. 2025; 13:1630667. doi: 10.3389/fcell.2025.1630667
6. Hassan, B., Raja, H., Hassan, T. et al. A comprehensive review of artificial intelligence models for screening major retinal diseases. *Artif Intell Rev*. 2024, 57, 111. <https://doi.org/10.1007/s10462-024-10736-z>
7. Kim, Hyung-Joo, Hassan Eesaar, Kil To Chong. Transformer-Enhanced Retinal Vessel Segmentation for Diabetic Retinopathy Detection Using Attention Mechanisms and Multi-Scale Fusion. *Applied Sciences*. 2024. 14, no. 22: 10658. <https://doi.org/10.3390/app142210658>
8. Liang X., Bao Y., Du Y., Kong N. AI-Assisted Screening for Diabetic Retinopathy and Fundus Abnormalities in a Large-Scale Physical Examination Population. *Clin Ophthalmol*. 2025 Aug 22;19:2889-2900. doi: 10.2147/OPTH.S538020. PMID: 40873655; PMCID: PMC12380004
9. Zedadra, Amina, Mahmoud Yassine Salah-Salah, Ouarda Zedadra, and Antonio Guerrieri. Multi-Modal AI for Multi-Label Retinal Disease Prediction Using OCT and Fundus Images: A Hybrid Approach. *Sensors*. 2025, 25(14): 4492. <https://doi.org/10.3390/s25144492>
10. Zhou, Y., Chia, M.A., Wagner, S.K. et al. A foundation model for generalizable disease detection from retinal images. *Nature*. 2023, 622: 156–163. <https://doi.org/10.1038/s41586-023-06555-x>
11. Roy, S. K., Ahmed, M., Rangayyan, R. M. Deep learning-based retinal vessel segmentation: A comprehensive review and benchmark. *IEEE Trans. Med. Imaging*, Dec. 2023; 42(12):3567–3583, doi: 10.1109/TMI.2023.10230840].
12. Lesions Segmentation Toolkit for Fundus Images. Режим доступа: <https://github.com/ClementPla/fundus-lesions-toolkit> (дата обращения: 02.08.2025).

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-790-810

## Сети KAN и MLP в био- и медицинской диагностике на примере стресса растений

А. И. Ларионов, В. Е. Турлапов

Центр искусственного интеллекта, Университет им. Н.И. Лобачевского, Нижний Новгород, Россия

**Аннотация.** Потери урожая, связанные с болезнями растений, остаются актуальной проблемой современного сельского хозяйства. Раннее детектирование – ключевой фактор в сокращении ежегодных потерь урожая, которые по различным позициям могут достигать 40 %. Выполнено сравнительное исследование эффективности раннего обнаружения стресса растений как на различных открытых датасетах, так и с помощью недавно появившихся нейронных сетей Колмогорова–Арнольда на фоне многослойного перцептрона (MLP) с аналогичным числом обучаемых параметров. Также проведено сравнение обычной сверточной нейронной сети с сетью, сверточные слои которой используют обучаемые функции активации. KAN-сеть показала хорошие результаты по отношению к MLP как на задаче регрессии математической функции, так и на задачах бинарной и мультиклассовой классификаций болезней растений, относительно количества обучаемых параметров. Исследованию предшествует обзор современных модификаций KAN и их применения в медицине и биологии.

**Ключевые слова:** искусственный интеллект, сеть Колмогорова–Арнольда, стресс растений, обучаемые функции активации, объяснимый ИИ, машинное обучение, глубокое обучение, компьютерное зрение

## KAN and MLP networks in bio- and medicine diagnostics on the example of plant stress

Larionov Aleksei Igorevich, Turlapov Vadim Evgenievich

Center for Artificial Intelligence, N.I. Lobachevsky University, Nizhny Novgorod, Russia

**Abstract.** Crop losses due to plant diseases remain a pressing challenge in modern agriculture. Early detection is a key factor in reducing annual yield losses, which in some cases can reach up to 40%. A comparative study of the efficiency of early detection of plant stress was performed, both on various open datasets and using recently emerged Kolmogorov–Arnold neural networks against the background of a multilayer perceptron (MLP) with a similar number of trainable parameters. A comparison was also made between a conventional convolutional neural network and a network whose convolutional layers use trainable activation functions. The KAN network showed good results in relation to MLP, both on the problem of regression of a mathematical function and on the problems of binary and multiclass classification of plant diseases, relative to the number of trainable parameters. The study is preceded by a review of modern modifications of KAN and their application in medicine and biology.

**Keywords:** artificial intelligence, Kolmogorov–Arnold network, plant stress, trainable activation functions, explainable AI, machine learning, deep learning, computer vision

### Введение

В работе исследуется использование сетей Колмогорова–Арнольда (KANs) как перспективной альтернативы традиционным многослойным перцептронам (MLPs), которые в настоящее время широко применяются для решения различных задач. Сети Колмогорова–Арнольда предлагают новую архитектуру, основанную на теореме представления Колмогорова–Арнольда, которая позволяет обучать не веса, а сами функции активации. В этой работе представлены результаты применения различных вариантов KAN к теоретическим и прикладным задачам, сравнение эффективности неглубокой KAN, использующей кубические сплайны для обучаемых функций активации и классической однослойной нейронной сети с обучаемыми весами и фиксированной функцией активации ReLU, исследование объяснимости результатов обучения KAN.

Рассмотрена возможность применения различных сетей Колмогорова–Арнольда к прикладной задаче раннего детектирования болезней растений. Болезни представляют существенную угрозу мировому производству продуктов питания. Глобально от болезней растений и вредителей мировое сельское хозяйство ежегодно теряет от 16 до 40 % урожая. Несмотря на разработку новых методов защиты растений, потери все равно остаются ощутимыми – раннее обнаружение заболевания или стресса растения является ключевым фактором минимизации потерь урожая. Основными методами обнаружения болезней растений являются методы на основе анализа ДНК, технологии на основе



анализа измерений летучих выделений растений, визуальный анализ, в том числе с применением нейронных сетей.

### Обзор публикаций

С момента публикации [1] в 2024 году, реализующей KAN, научное сообщество проявило явный интерес к новому полю для исследований и его возможностям. С тех пор вышло как минимум 60 публикаций. Среди них есть Kolmogorov-Arnold Networks Meet Science [2], в которой авторы делают упор на научное применение сетей KAN. Они рассматривают KAN как мост между современными принципами Artificial Intelligence (AI), основанными на коннекционизме, и символической наукой. В данной публикации был расширен функционал оригинальной библиотеки PyKAN такими функциями, как например: *kanpiler* – создает KAN из математической формулы, *tree converter* – преобразует модель в дерево, *MultKAN* – модель KAN с умножением вместо сложения на узлах. Также было внесено множество улучшений производительности по сравнению с первой публикацией библиотеки, множество функций, требующих больших затрат производительности или недоступных для параллелизации, стали опциональными, проведена оптимизация расчёта сплайнов – для сети глубины  $L$ , ширины слоя  $N$  с количеством интервалов  $G$  расход памяти был уменьшен с  $O(LN^2G)$  до  $O(LNG)$ .

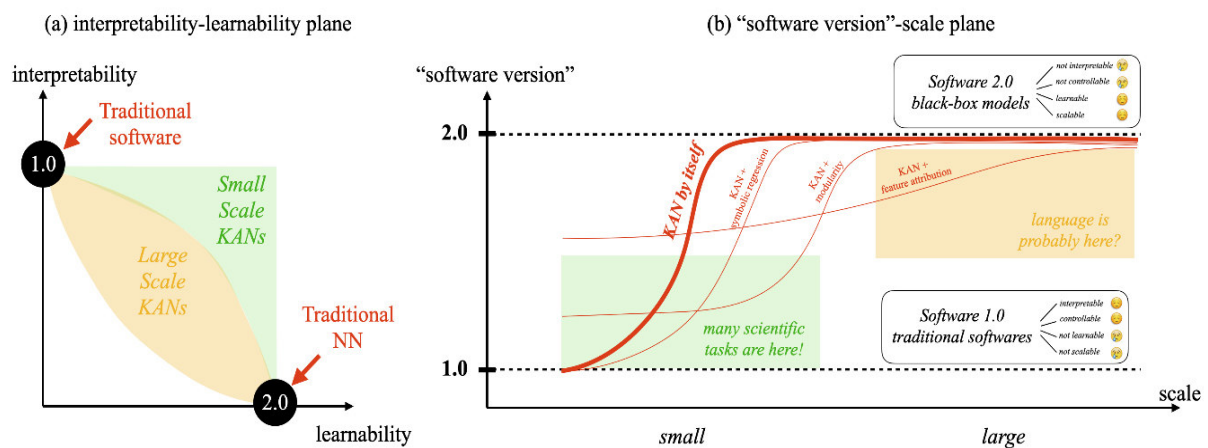


Рисунок 1 [2]

На рисунке 1 показано представление авторов о месте KAN на текущем ландшафте научных и производственных задач. KAN стремится занять промежуточную позицию между классическим программным обеспечением (software 1.0), ориентированным на интерпретируемость, и современным программным обеспечением (software 2.0), ориентированным на обучаемость, обеспечивая баланс между интерпретируемостью и способностью к обучению, что иллюстрируется их положением на фронте Парето в пространстве «интерпретируемость-масштабируемость». Количество интерпретаций, которое можно получить с помощью KAN, зависит от масштаба задачи и выбранных методов интерпретации. Также они предлагают KAN как универсальные базовые модели для физических исследований и демонстрируют их сильные стороны на исследовании с помощью KAN таких физических задач и концепций, как поиск сохраняющихся величин, поиск Лагранжианов на примере задачи маятника и релятивистской массы в однородном поле, поиск скрытой симметрии, на примере черной дыры Шварцшильда.

В публикации DropKAN [11] исследуется задача предотвращения ко-адаптации функций активации. Исследование показало, что разрежение сети KAN с помощью механизма dropout между слоями сети KAN (отключение узлов) может приводить к тому, что функции активации все равно будут испытывать эффект от обратного распространения ошибки. Метод DropKAN, основанный на случайном маскировании выходов либо после активаций (DropKANpa), либо только сплайновых компонентов (DropKANps) внутри слоя KAN, показал лучшие результаты по сравнению со стандартными dropout между слоями (табл. 1).



Таблица 1. Классификация на KAN с различными реализациями DROPOUT [11]

Dataset	KANs Accuracy				
	No-Drop	Dropout <sub>w/o scale</sub>	Dropout <sub>w/scale</sub>	DropKAN <sub>ps</sub>	DropKAN <sub>pa</sub>
dermatology	73.78%±5.94	89.46%±5.60	86.49%±1.66	92.43%±1.21	<b>92.70%±1.54</b>
german	66.70%±1.35	74.30%±2.02	73.40%±3.49	70.70%±3.03	<b>76.6%±1.98</b>
semeion	89.66%±0.0	94.55%±4.50	97.43%±0.52	<b>99.81%±0.42</b>	97.62%±0.42
car	<b>91.56%±1.18</b>	86.07%±1.60	77.69%±6.96	89.65%±0.63	85.66%±2.95
abalone	24.98%±0.95	23.71%±1.96	23.35%±1.85	27.68%±0.87	<b>27.85%±0.49</b>
adult	84.94%±0.17	84.88%±0.15	84.66%±0.53	<b>85.27%±0.12</b>	85.08%±0.19
bank-full	90.13%±0.37	90.06%±0.26	90.12%±0.20	90.28%±0.20	<b>90.33%±0.13</b>
connect-4	67.26%±4.31	70.53%±0.39	63.85%±5.77	<b>72.14%±0.31</b>	68.67%±1.72
diabetic	52.56%±4.44	56.06%±1.58	51.52%±2.80	<b>57.95%±0.27</b>	56.55%±0.76
census-income	94.78%±0.21	<b>94.90%±0.08</b>	94.11%±0.23	94.58%±0.11	94.69%±0.03

Kolmogorov-Arnold Transformer [12] рассматривает использование KAN слоев в архитектуре трансформеров вместо MLP, для глубокого обучения. Авторы публикации ускорили сети KAN путем использования рациональных функций, более оптимальных для расчёта на GPU, группировкой граней для оптимизации вычисления параметров и улучшили инициализацию весов для обеспечения стабильной дисперсии. Они сравнили производительность трех вариантов архитектуры – классический трансформер, трансформер, в котором MLP заменен на стандартную реализацию KAN, и трансформер, в котором MLP заменен на KAN с перечисленными выше улучшениями Group-Rational KAN (GR-KAN).

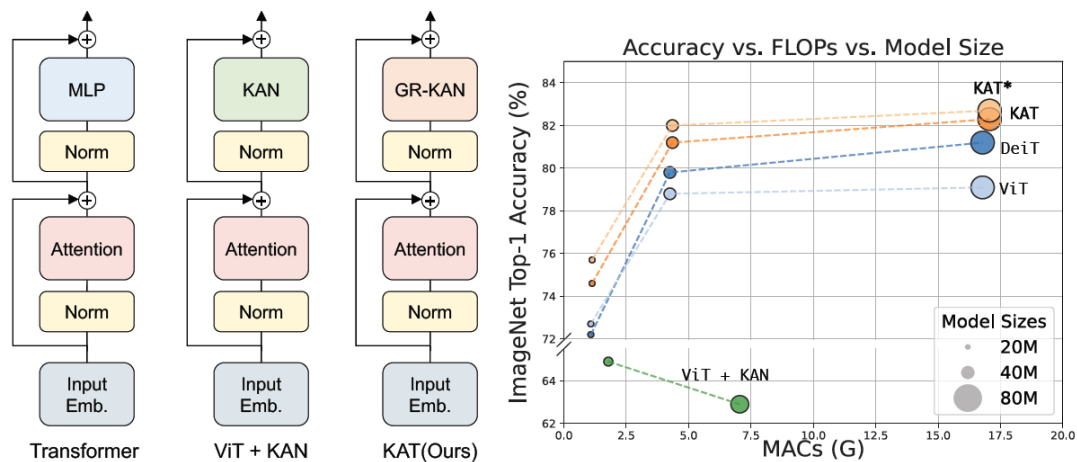


Рисунок 2. Сравнение результатов архитектур, предложенных на датасете ImageNet [12]

Версия трансформера, использующая GR-KAN, показала лучшие результаты в тестировании на нескольких задачах.

Kolmogorov-Arnold Convolutions [13] исследует применение методов KAN к задачам Computer Vision (CV), рассматривает такие замены для сплайнов, как вейвлеты и множество многочленов Грэма. Также были предложены реализации механизмов самовнимания и фокальной модуляции.

Как показано в таблице 2, в работе достигнуты передовые результаты в задачах сегментации с использованием архитектур, аналогичных U-Net, с KAN-свёртками на датасетах BUSI, GlaS и CVC.

Kolmogorov-Arnold Fourier Networks [14] представляет собой дальнейшее развитие идей интерпретируемых сетей KAN для решения задач с высоким числом признаков.

Таблица 2. Результаты эксперимента сегментации медицинских изображений [13]

Methods	BUSI [34]		GlaS [35]		CVC [36]	
	IoU↑	F1↑	IoU↑	F1↑	IoU↑	F1↑
U-Net [6]	57.22	71.91	86.66	92.79	83.79	91.06
Att-Unet [37]	55.18	70.22	86.84	92.89	84.52	91.46
U-Net++ [10]	57.41	72.11	87.07	92.96	84.61	91.53
U-NeXt [38]	59.06	73.08	84.51	91.55	74.83	85.36
Rolling-UNet [39]	61.00	74.67	86.42	92.63	82.87	90.48
U-Mamba [40]	61.81	75.55	87.01	93.0	84.79	91.63
U-KAN [13]	63.38	76.40	87.64	93.37	85.05	91.88
U <sup>2</sup> -KAGNet, Small (Ours)	59.44	74.56	<b>89.13</b>	<b>94.25</b>	85.62	92.26
U <sup>2</sup> -KAGNet, BottleNeck, Small (Ours)	62.95	77.26	85.99	92.47	87.61	93.39
U <sup>2</sup> -KAGNet, BottleNeck (Ours)	55.91	71.72	87.73	93.47	<b>88.86</b>	<b>94.1</b>
U <sup>2</sup> -KAGNet (Ours)	58.13	73.52	88.14	93.7	87.07	93.09
UKAGNet (Ours)	<b>63.45</b>	<b>77.64</b>	87.31	93.23	76.85	86.91

Авторы отмечают, что при всех теоретических преимуществах KAN возникает проблема взрывного роста числа параметров и сложности в захвате высокочастотных признаков в многомерных задачах. В работе предлагается интеграция обучаемых случайных Фурье-признаков (Random Fourier Features, RFF) и новой гибридной функции активации GELU-Fourier, что позволяет добиться баланса между компактностью параметров и спектральной выразительностью модели.

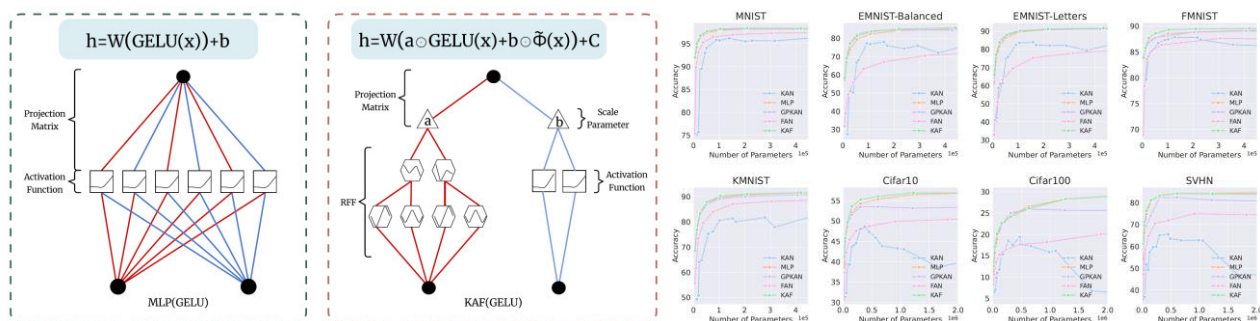


Рисунок 3 [14]

На рисунке 3 слева – стандартный MLP с активацией GELU и KAF с активацией GELU. MLP включает в себя проекционную матрицу, за которой следует GELU, в то время как KAF добавляет случайные признаки Фурье (RFF) и параметры масштабирования, предлагая большую гибкость в преобразованиях признаков.

Справа на рисунке 3 показано сравнение производительности различных моделей (KAN, MLP, GPKAN, FAN, KAF) в нескольких наборах данных (MNIST, EMNIST, FMNIST, KMNIST, Cifar10, Cifar100, SVHN). Результаты показывают, что KAF обычно достигает более высокой точности с меньшим количеством параметров.

Wav-KAN [15] – реализация KAN сетей с функциями активаций на основе вейвлетов. Итоги тестирования авторами показали, что реализация функций активации с помощью вейвлета Difference of Gaussians (DOG) превосходит базовую реализацию KAN по точности. Также вейвлеты позволяют оптимизировать количество обучаемых параметров.

СохKAN [16] – это специализированная модификация сети KAN, предложенная для задач анализа выживаемости с акцентом на достижение баланса между производительностью и

интерпретируемостью. В отличие от классической модели пропорциональных рисков Кокса, CoxKAN способен автоматически выявлять сложные взаимодействия признаков и формулировать аналитические (символические) выражения для функции риска, что повышает прозрачность результатов. Авторы показывают, что CoxKAN успешно восстанавливает символические формулы и осуществляет автоматический отбор признаков на синтетических данных, а также демонстрирует лучшие или сопоставимые результаты с глубокими нейросетями и классическими моделями на реальных медицинских и геномных датасетах, что отражено в таблице 3.

Таблица 3. Результаты на клинических датасетах [16]

Dataset	CoxPH	DeepSurv	CoxKAN Trained	CoxKAN Pruned	CoxKAN Symbolic
SUPPORT	0.583074 (0.581, 0.585)	0.618308* (0.616, 0.620)	0.624482 (0.622, 0.625)	<b>0.624485</b> <b>(0.622, 0.625)</b>	0.623755 (0.623, 0.626)
GBSG	0.656291 (0.655, 0.662)	0.668402* (0.665, 0.671)	0.678294 (0.676, 0.682)	0.679219 (0.675, 0.681)	<b>0.682796</b> <b>(0.678, 0.684)</b>
METABRIC	0.632363 (0.628, 0.637)	0.643375* (0.639, 0.647)	0.647177 (0.644, 0.652)	0.648004 (0.646, 0.654)	<b>0.649618</b> <b>(0.644, 0.651)</b>
FLCHAIN	<b>0.797854</b> <b>(0.797, 0.802)</b>	0.794520 (0.793, 0.798)	0.797064 (0.796, 0.801)	0.795911 (0.792, 0.797)	0.796281 (0.795, 0.800)
NWTCO	0.698347 (0.693, 0.703)	0.698300 (0.692, 0.703)	0.719843 (0.714, 0.725)	0.720721 (0.708, 0.718)	<b>0.722225</b> <b>(0.715, 0.725)</b>

Residual Kolmogorov-Arnold Network (RKAN) [17] – это архитектурное расширение для сверточных нейронных сетей, направленное на преодоление ограничений стандартных глубоких моделей, связанных с их линейной природой и фиксированными функциями активации. В RKAN используются иерархические KAN- слои на основе остаточных связей, что способствует увеличению эффективности обучения, выразительности моделей и стабильности, снижая риски затухающих и взрывающихся градиентов, особенно на небольших датасетах. Особое внимание уделено использованию многочленов Чебышёва вместо B-сплайнов, что позволяет повысить вычислительную эффективность и масштабируемость. В экспериментах на стандартных наборах данных (CIFAR-100, Food-101, ImageNet) RKAN стабильно улучшал точность топ-1 как минимум на 1 % относительно базовых моделей, при этом на больших архитектурах прирост был более заметным, что показано в таблице 4.

Таблица 4. Сравнение ТОП-1 точности (%) базовых моделей и моделей, дополненных RKAN на датасетах CIFAR-100, FOOD-101 [17]

	CIFAR-100			Food-101		
	res.	Rkan	base	res.	Rkan	base
ResNeXt-101	128	<b>86.15</b>	85.28	224	<b>90.82</b>	89.87
ResNeXt-50	128	<b>85.08</b>	84.40	224	<b>90.00</b>	89.20
ResNet-152	128	<b>85.40</b>	84.63	224	<b>90.36</b>	89.70
ResNet-101	128	<b>85.12</b>	84.00	224	<b>90.09</b>	89.29
ResNet-50	128	<b>84.56</b>	84.12	224	<b>89.48</b>	88.84
RegNetY-32GF	128	<b>87.03</b>	85.44	224	<b>91.62</b>	90.72
RegNetY-8GF	128	<b>86.11</b>	84.77	224	<b>91.17</b>	90.43
RegNetY-3.2GF	128	<b>85.46</b>	84.68	224	<b>90.09</b>	89.54
RegNetY-800MF	128	<b>83.19</b>	82.74	224	<b>89.00</b>	88.39
DenseNet-201	128	<b>85.35</b>	84.28	224	<b>89.58</b>	88.83
DenseNet-169	128	<b>84.84</b>	84.00	224	<b>89.74</b>	89.17
DenseNet-121	128	<b>84.73</b>	84.09	224	<b>89.43</b>	88.98

Модуль также демонстрирует лучшую стабильность (меньше коэффициент вариации), эффективное использование параметров и памяти, а также повышенную скорость прямого и обратного прохода по сравнению с другими подходами.

Construction of the Kolmogorov-Arnold representation using the Newton-Kaczmarz method [29] предлагает свою реализацию KAN на языке C++, ускоряющую обучение данного типа сетей. Одна из рассмотренных в публикации проблем – нахождение определителя квадратной матрицы размерности 5, она включала 10 миллионов записей и 25 параметров. В этой задаче обучение KAN заняло до 10 минут, тогда как обучение аналогичной модели многослойного перцептрона занимало от 4 до 8 часов, обучение останавливалось при достижении коэффициента Пирсона  $>0.9$  на валидирующем датасете из 2 миллионов записей.

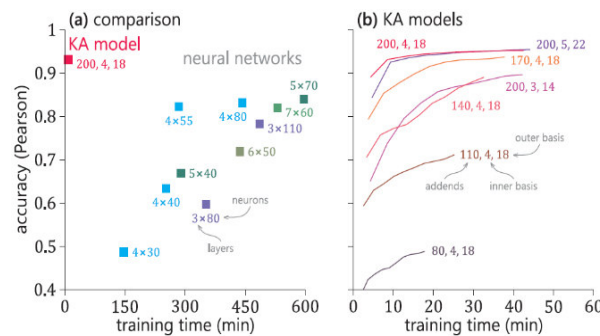


Рисунок 4 [29]

На рисунке 4а показано сравнение производительности сети Колмогорова-Арнольда (КА-model) и встроенных моделей MATLAB на задаче регрессии – поиска определителя для квадратной матрицы размерности 5. На рисунке 4б – эволюция точности различных КА-моделей. Такое ускорение достигается за счет использования метода Ньютона-Качмарца и его модификации для параллельного исполнения. Дальнейшее ускорение происходит за счет того, что при обучении не фиксируются базисные функции, большая часть проходов может использовать кусочно-линейные функции, а более сложные и выразительные базисные функции могут быть «включены» на поздних этапах обучения.

## Материалы и методы

### 1. Теорема Колмогорова-Арнольда

Теорема Колмогорова-Арнольда о представлении непрерывных функций нескольких переменных в виде суперпозиции функций меньшего числа переменных математически формулируется как

$$f(x_1, \dots, x_d) = \sum_{q=0}^{2d} \Phi_q \left( \sum_{p=1}^d \Phi_{q,p}(x_p) \right), \quad (1)$$

где  $f: [0,1]^d \rightarrow R$  – непрерывная функция  $d$  переменных;  $d$  – размерность входного пространства;  $\Phi_q$  – непрерывные функции одной переменной;  $\Phi_{q,p}$  – непрерывные функции одной переменной.

На основе данной теоремы в 2024 году была представлена работа Z. Liu с соавторами [1], в которой описан подход к построению нейронных сетей, использующих обучаемые функции активации на ребрах и операцию сложения на узлах. В этой работе такая архитектура сети предлагается как альтернатива классическим нейронным сетям, которые опираются на универсальную теорему аппроксимации.

Универсальная теорема аппроксимации (теорема Цыбенко):

$$G(x, w, \alpha, \theta) = \sum_{i=1}^N \alpha_i \varphi(w_i^T x + \theta_i), \quad (2)$$

где  $w_i$  – веса между входными нейронами и нейронами скрытого слоя,

$\alpha$  – веса между связями от нейронов скрытого слоя и выходным нейроном,

$\theta$  – смещения для нейронов входного слоя.

Ниже представлены схематические изображения структуры однослойного перцептрона и однослойной сети Колмогорова-Арнольда. Для простоты и KAN, и перцептрон, с одним скрытым слоем, будем называть однослойными.

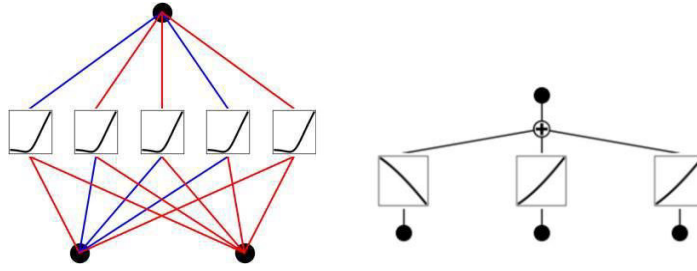


Рисунок 5. Однослойный перцептрон (слева) и однослойная сеть Колмогорова-Арнольда (справа)

Для KAN количество обучаемых параметров будет считаться

$$p = (d_{in} \times d_{out}) \times (G + K + 3) + d_{out}, \quad (3)$$

где  $G$  – интервалы,  $K$  – порядок сплайна, константа 3 – веса В-сплайнов, shortcut-веса и смещение.

Для MLP количество параметров определяется следующим образом:

$$p = (d_{in} \times d_{out}) + d_{out}. \quad (4)$$

## 2. Реализация сети Колмогорова-Арнольда

По теореме Колмогорова-Арнольда непрерывная функция многих переменных на интервале может быть выражена через конечную композицию непрерывных функций одной переменной и операции сложения. Определим KAN-слой как матрицу из одномерных функций:

$$\Phi = \{\Phi_{q,p}\}, \quad p = 1, 2, \dots, n_{in}, \quad q = 1, 2, \dots, n_{out}. \quad (5)$$

По образу MLP сетей мы можем создавать KAN-слои необходимой ширины и последовательно подключать их для достижения глубины.

В матричной форме один слой представляется как

$$X_{l+1} = \Phi_l X_l, \quad (6)$$

где  $X_l$  – тензор входных параметров,  $X_{l+1}$  – тензор выходных параметров,  $\Phi_l$  – матрица функций  $l$ -го слоя KAN.

$$\Phi_l = \begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \dots & \phi_{1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \dots & \phi_{2,n_l}(\cdot) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \dots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix} \quad (7)$$

$$KAN(x) = (\Phi_{L-1} \boxtimes \Phi_{L-2} \boxtimes \dots \boxtimes \Phi_1 \boxtimes \Phi_0)x. \quad (8)$$

Так как все операции дифференцируемы, может быть использован алгоритм обратного распространения ошибки. Функция активации в KAN определяется как

$$\phi(x) = w_b b(x) + w_s spline(x), \quad b(x) = silu(x) = x/(1 + e^{-x}), \quad (9)$$

где  $spline(x)$  – линейная комбинация В-сплайнов:

$$spline(x) = \sum_i c_i B_i(x). \quad (10)$$

Здесь обучаемыми параметрами являются  $w_b$ ,  $w_s$  и  $c_i$ . Функции активации инициализируются  $w_s = 1$  и  $spline(x) \approx 0$  по принципу инициализации Ксавье, используемому для инициализации линейных слоев в MLP. Так как сплайны определены на ограниченных отрезках, а признаки могут выйти за пределы этих отрезков – интервалы для сплайнов обновляются «на лету».

Реализация функций активации остается обширным полем для исследований и экспериментов, коэффициенты  $w_b$ ,  $w_s$  – обучаемы, но не обязательны, они могут быть выражены самими  $b(x)$  и  $spline(x)$ . Функция  $b(x)$  может быть отброшена, а вместо  $spline(x)$  можно использовать другие непрерывные параметризуемые функции, как это уже сделано в работах [1,2].

### 3. Организация среды для исследования KAN

Для работы использовалась реализация KAN из библиотеки PyKAN. Она основана на torch, и в ней реализованы такие инструменты, как создание датасета из математической функции, генерация изображения сети на текущем этапе обучения, отображение графика функций активации на ребрах сети. С помощью данной библиотеки можно создавать как однослойные KAN, так и многослойные, возможно настроить количество сегментов, диапазон и порядок для сплайнов – функций активации. Также была использована библиотека convKAN для тестирования сверточной сети. Для MLP-сетей, обработки датасетов при обучении и валидации была использована библиотека torch 2.7.0 для версии cu128. Для подготовки датасетов к обучению, разделению на обучающие и тестовые наборы, расчета весов классов для обучения – pandas, sklearn, numpy. Для визуализации отношений признаков к классам для анализа использовалась библиотека seaborn. Для визуализации графиков, диаграмм изображений, отчетов – matplotlib. Для обработки изображений – rembg, cv2, PIL. Для контроля версий и экспериментов была использована библиотека clearml. Исследование проводилось в локальном Jupyter Notebook в IDE Visual Studio Code, часть кода была исполнена в Py файлах, так как у Jupyter Notebook бывают проблемы с работой мультипоточности, например для torch.utils.data.DataLoader для использования num\_workers > 0.

Работа выполнена на Python 3.12.10. Оборудование, на котором проводилось исследование:

- GPU: Nvidia GeForce RTX 3060 12Gb, Cuda Toolkit v12.8, CUDNN v9.10;
- CPU: Intel core i-5 12400f, RAM: DDR4 – 48Gb.

### 4. Механизмы и маркеры стресса растений

Одним из ранних признаков стресса для растений является повышение температуры листьев [8–10] на 0.2°. Такой признак непросто детектировать, так как это повышение совсем незначительное и происходит еще до потери воды растением и видимых изменений на листьях и связано с изменением режима дыхания для сбережения воды. Обнаружение стресса до начала потери воды считается «ранним» обнаружением. Другим признаком является изменение в спектральной отражательной способности, которое появляется из-за того, что болезни и стресс вызывают изменения в пигментном составе и структуре тканей листа, что влияет на отражающие характеристики листа в различных участках спектра. В детектировании таких признаков помогают гиперспектральные камеры. В силу высокого спектрального разрешения (3-10 нм) они позволяют вычислять различные вегетационные индексы, которые чувствительны к состоянию зеленой массы растения [8, 9].

Существуют и более высокоуровневые признаки, позволяющие довольно точно определять болезни растений на ранних стадиях. Исследования показывают, что энтропия, рассчитанная на основе гистограммы вегетационных индексов пикселей изображения растения, может служить универсальным и объяснимым высокоуровневым признаком для ранней диагностики стресса. Кроме энтропии, другие статистические характеристики гистограммы (минимум, максимум, среднее, стандартное отклонение и max-min) также несут информацию о состоянии растения.

### 5. Вегетационные индексы

Для раннего детектирования болезней растений часто используют вегетационные индексы, которые можно получить из гиперспектральных и тепловизионных изображений растений. В данной работе использованы индексы:

$NDVI = \frac{NIR-R}{NIR+R}$  – отношение разности коэффициентов отражения, ближнего инфракрасного (NIR) и красного (R) диапазонов спектра и их суммы, указывает на активность фотосинтетической биомассы;

$VARI = \frac{G-R}{G+R-B}$  – охватывает весь видимый сегмент спектра, позволяет минимизировать влияние атмосферных эффектов и разницы в освещении;

$NDBlue = \frac{G-B}{\max(G-B)-\min(G-B)}$  – полезен для отделения растительности от фона.

Для извлечения статистических характеристик использовались стандартные R, G, и B каналы изображения и индексы NDVI, VARI. Для расчёта NDVI канал NIR был заменен на канал G:  $NDVI = \frac{G-R}{G+R}$ , VARI изначально предназначен для RGB-изображений.



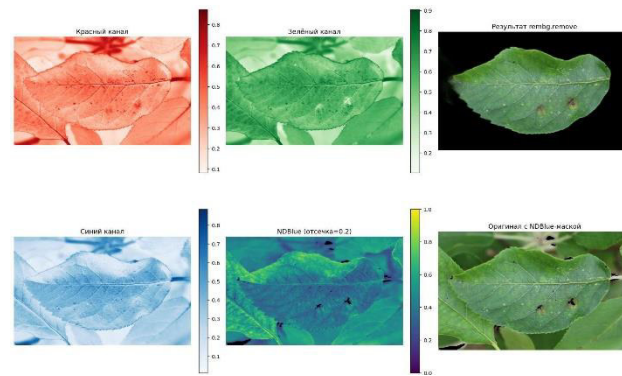


Рисунок 6

На рисунке 6 показан пример обработки изображения, полученного в полевых условиях (неоднородный фон, состоящий из других листьев). Четыре изображения со шкалами – каналы изображения Красный, Синий, Зеленый и индексный NDVI-канал. Изображения справа сверху – сегментация с помощью библиотеки rembg (U-Net) и снизу – сегментация с помощью маски значений NDBIue (threshold = 0.2).

#### 6. Используемые датасеты

В ходе исследования использованы и обработаны различные датасеты, включая Plant Village, BRACOL, Plant Pathology, RoCoLe, Citrus и Rice Leaf Diseases, ссылки на которые перечислены в таблице 5. Каждый из этих датасетов включает изображения растений (как здоровых, так и поражённых различными заболеваниями), что обеспечивает широкий спектр тестовых данных для валидации моделей. Техническое описание датасетов дано в таблице 5.

Таблица 5. Обзор публичных датасетов, использованных в исследовании [19]

Dataset	DiaMOSPlant [5]	BRACOL [12]	RoCoLe [11]	Plant Pathology [14]	Rice Leaf Diseases[13]	Citrus [15]	APDA [16]	Plant-Village [7]
Plant/Crop	Pear	Coffee	Coffee	Apple	Rice	Citrus	Rose	Multiple
Dataset size	3505(3006 leaf images+499 fruit images)	4407	1560	3651	120	759 (609 leaf images + 150 fruit images)	40	54.309
Symptoms No	4	4	2	3	3	5	2	26
Acquisition device	Smartphone DSRL	Smartphone	Smartphone	DSLR Camera, Smartphone	DSLR camera	DSLR camera	Smartphone	Smartphone
Color	RGB	RGB	RGB	RGB	RGB	RGB	RGB	RGB
Image resolution	Multiple	2048 × 1024	Multiple	2048 × 1365	2848 × 4288	256 × 256	N.d.	Multiple
Annotation	Polygon, Label	Polygon, Label	Polygon, Label	Label	Label	Label	Label	Label
Annotation format	csv, YOLO	csv	COCO, JSON, Pascal, VOC, csv	csv	Folder structure	Folder structure	N.d.	Folder structure
Data sharing platform	Zenodo	GitHub	Mendeley Data	Kaggle	UCI Machine Learning Repository	Mendeley Data	MathWorks	Github
Acquisition place	Field	Laboratory	Field	Field	Laboratory	Laboratory	Laboratory	Laboratory
Side of the leaf	Adaxial	Abaxial	Adaxial, Abaxial	Adaxial	Adaxial	Adaxial	Adaxial	Adaxial
Object of interest	Fruit, leaf	Leaf	Leaf	Leaf	Leaf	Fruit, leaf	Leaf	Leaf



Для обеспечения единообразия и повышения качества обобщающей способности моделей на каждом датасете были переопределены и стандартизированы классы, что позволило более точно оценить эффективность классификации как внутри каждого датасета, так и в сравнении с эталонным датасетом Plant Village. Для каждого датасета было подготовлено две версии: одна содержит исходные изображения, а другая – изображения, преобразованные в нормализованные статистические характеристики. Такой подход позволил провести комплексное тестирование и выявить сильные и слабые стороны моделей при работе как с визуальными, так и с абстрактными признаками. В частности, исходный датасет Plant Village включает 54309 изображений размером  $256 \times 256$  пикселей в формате RGB и содержит 38 несбалансированных классов растений и заболеваний. Анализ, проведенный в работе [18], выявил значительные закономерности в данных, влияющие на обобщающую способность сверточных нейронных сетей. Например, многие изображения одного класса сняты в одинаковых условиях освещенности, на одном фоне, с характерными тенями и высокой контрастностью, тогда как другой класс мог отличаться по этим параметрам, но оставаться однородным внутри себя. Это приводило к высокой точности классификации (приближающейся к 100 %) на тестовой выборке внутри датасета, но к резкому падению точности (до 30 %) при попытке классифицировать изображения из других датасетов [23-25, 27, 28]. Для устранения этих закономерностей и повышения устойчивости моделей к различным условиям съёмки, фон изображений был удалён с помощью индекса NDBI и библиотеки Rembg, основанной на сети U-Net. Кроме того, при подготовке данных для обучения использовалась конвертация изображений в нормализованные статистические характеристики, что должно позволить сфокусироваться на признаках, связанных непосредственно с заболеванием растений, исключая пространственные и визуальные артефакты. Такой подход, несмотря на возможные ограничения, обеспечивает более объективную оценку классификаторов и закладывает основу для последующего анализа.

## Результаты экспериментов и обсуждение

### 7. Аппроксимация математической функции

В первую очередь рассмотрим задачу аппроксимации математической функции на примере функции

$$f(x, y) = e^{(\sin(\pi \cdot x) + y^2)}.$$

Для сети KAN будем использовать 5 скрытых нейронов, 3 интервала для функций активации, сплайны третьего порядка:  $p = l_1 + l_2 = ((2 * 5) * (3 + 3 + 3) + 5) + ((5 * 1) * (3 + 3 + 3) + 1) = 141$ .

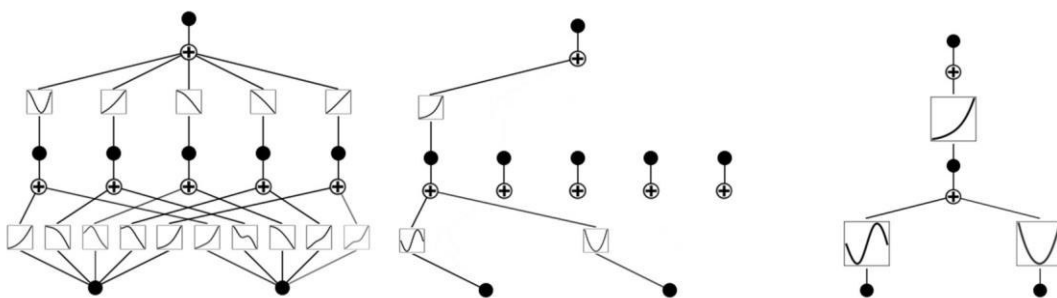


Рисунок 7

На рисунке 7 слева направо представлены модель при инициализации, модель после 50 эпох обучения, модель после операции `prune()` и дообучения. После 50 эпох большая часть нейронов «отключается» и остаются только самые активные. «Отключившиеся» нейроны удаляются с помощью функции `prune()` для оптимизации процесса обучения. После этапа обучения модель можно автоматически привести к символьной формуле, для исходного выражения  $f(x, y) = e^{(\sin(\pi \cdot x) + y^2)}$  была получена формула  $e^{y^2 + \sin(3.1416 \cdot x)}$ .



Рисунок 8. График потерь KAN, 200 эпох

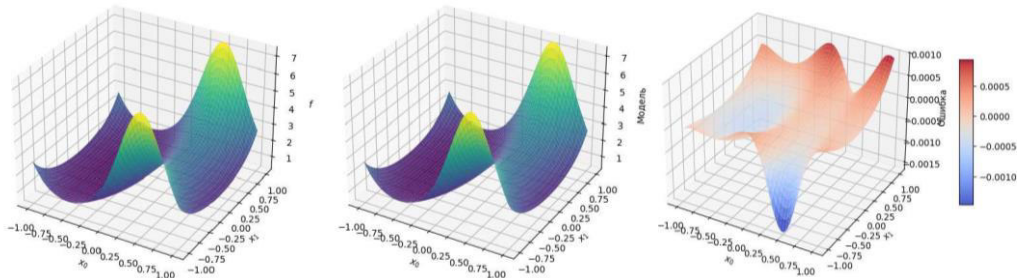


Рис. 9.

На рисунке 9 слева направо приведены графики: истинной функции, предсказание модели, отклонение модели от истинной функции. Обучение, после 200 эпох, снижает отклонение модели до величин не более  $4e-5$ . Длительность обучения составила примерно 30 секунд. Значение MSE составило  $6.4e-18$ . Эксперимент был выполнен на GPU.

Для однослойного перцептрона было использовано 35 скрытых нейронов. Так,  $p = l_1 + l_2 = ((2 * 35) + 35) + ((35 * 1) + 1) = 141$ .

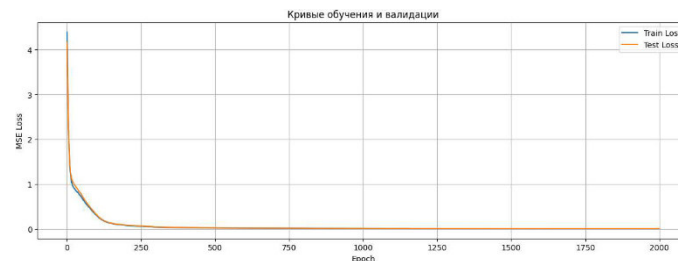


Рисунок 10. Функции потерь для SLP, 2000 эпох

Кривая потерь прекращает существенно уменьшаться после 500 эпох.

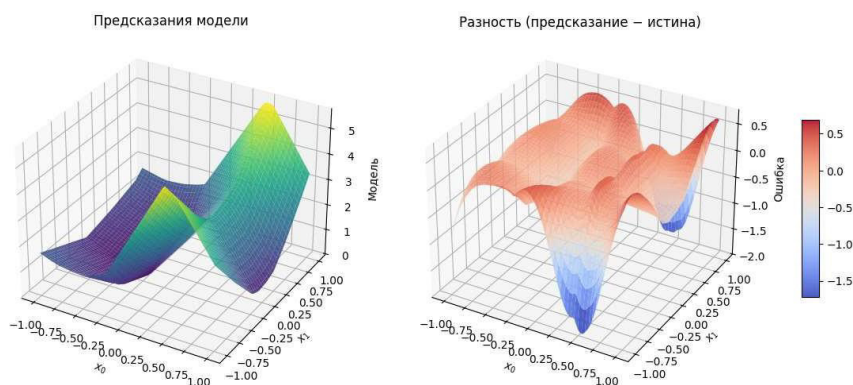


Рисунок 11. SLP после 200 эпох

После 200 эпох MSE составил  $5,12e-1$ . После 2000 эпох этот показатель составил  $9,8e-3$ . Данный эксперимент выполнен на GPU.

#### 8. Детектирование болезней растений

##### 8.1. Эксперимент 1: Бинарная классификация Plant Village

Ниже дано описание модели KAN в виде таблицы 6. Нейросеть принимает 30 значений, нормированных до интервала  $[0, 1]$ .

Таблица 6. Характеристики сети, тестируемой в эксперименте 1

Модель	Форма	Интервалы	Степень сплайна	Датасет обучения	Классификация
KAN	[30,32,2]	10	3	Plant Village	Бинарная

В этом и дальнейших тестах использованы:

- функция потерь: CrossEntropyLoss с весами классов для борьбы с несбалансированностью обучающей выборки;
- оптимизатор: AdamW с начальным `learning_rate = 5.2e-4` и `weight_decay = 1e-3`;
- планировщик: ReduceLROnPlateau с параметрами `patience = 20`, `factor = 0.8`, `mode = 'min'`, `min_lr = 1e-6`;
- размер батча: 256.

Датасеты предразбиты на стратифицированные обучающую, валидирующую и тестовую выборки в пропорции 7:1:1 для воспроизводимости экспериментов.

Обучали на протяжении 20 эпох при использовании GPU. Обучение продлилось 19 секунд, модель достигла 90,80 % точности на тестовой выборке из датасета Plant Village. Дальнейшее обучение практически не имеет смысла. Для такой сети обучение на 200 эпохах заняло 3 минуты 20 секунд и достигло точности 92.61 % на валидирующей выборке.

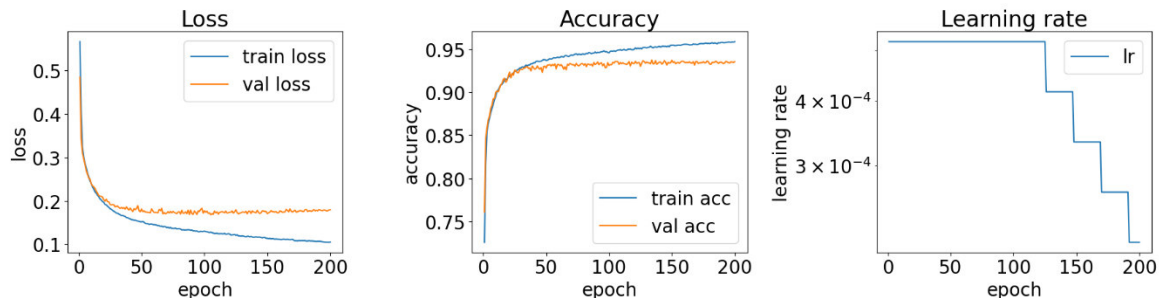


Рисунок 12. График потерь для обучающей и валидирующей выборок, график точности и график learning rate

Минимальное значение функции потерь на валидирующей выборке было достигнуто на 105 эпоху и составило 0.1679. Обучение до чекпоинта, на котором было достигнуто минимальное значение потерь на валидации, заняло 137 секунд, полное обучение до 200 эпох заняло 259 секунд. На тестовой выборке была достигнута точность 93,2 %. На основании этих данных можно сделать вывод, что модель близка к лимиту своих обобщающих возможностей и для улучшения результата потребуется более глубокая модель или широкие слои.

По диаграмме видно, что наиболее значимыми оказались R\_std, B\_std, G\_mean, G\_std, что расходится с результатами исследований [8, 9], это можно объяснить плохим качеством разметки датасета, о чем говорится в работе [18]. Попробуем проверить это, верифицировав результаты обучения на других датасетах перечисленных ранее, а также попробуем обучить сеть с аналогичными параметрами на других датасетах. Для датасетов, содержащих преимущественно больные листья, показатели F1, точности и полноты остаются высокими, что говорит о хорошей способности модели фиксировать патологию. Однако по мере увеличения доли здоровых растений качество классификации заметно снижается: модель чаще выдаёт ложные положительные ответы, но и доля ложных отрицаний остаётся ощутимой. На сбалансированных по классам наборах, таких как RaCoLe, точность падает до уровня

около 41 % – значительно ниже как простого случайного выбора, так и взвешенного случайного угадывания. Существенную роль в этом может играть и разнородность условий съёмки: часть наборов получена не в лабораторной среде, а в полевых условиях, что дополнительно осложняет задачу. Аналогичные выводы о падении качества вне контролируемых условий подтверждаются и в работе [18], критикующей датасет Plant Village.

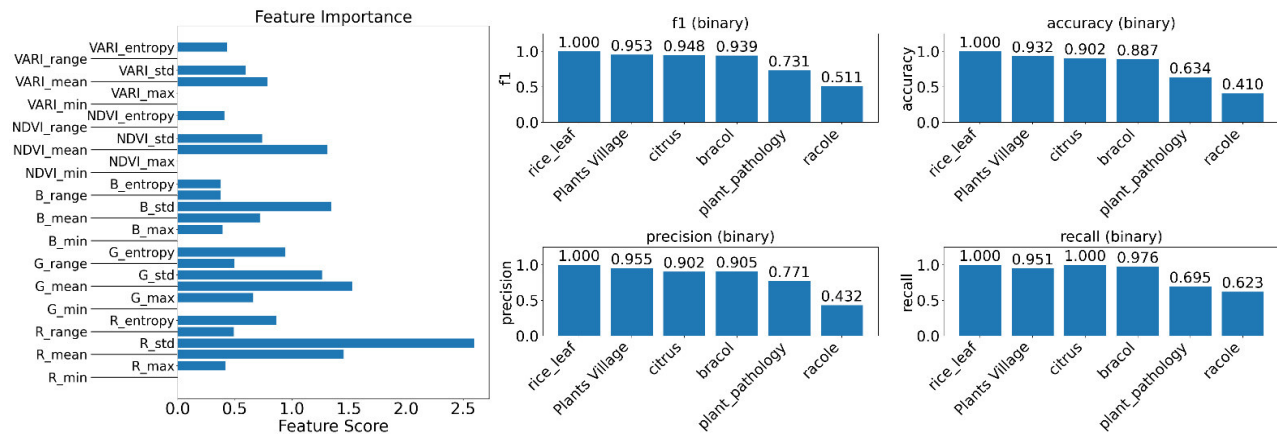


Рисунок 13. График важности признаков и метрик (F1, accuracy, precision, recall) на тестовых выборках

Попробуем обучить аналогичную сеть на датасете RaCoLe и валидировать ее на оставшихся.

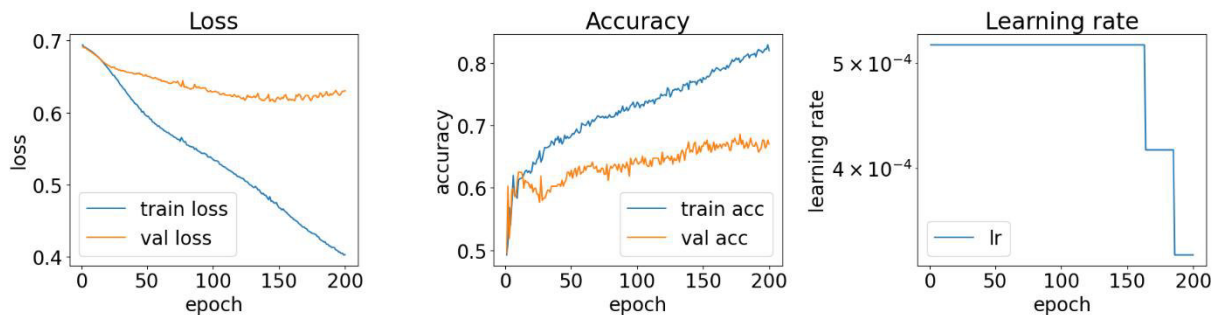


Рисунок 14. Функции потерь, точности и learning rate

При обучении аналогичной сети на датасете RaCoLe видим похожую картину – модель сильно переобучается. Минимальное достигнутое значение потерь на валидации 0.615 на 147 эпохе обучения. Обучение 200 эпох заняло 11 секунд. Заметен тренд на переобучение для KAN-моделей, здесь это также связано с тем, что данный датасет значительно меньше: 1560 записей против 54309.

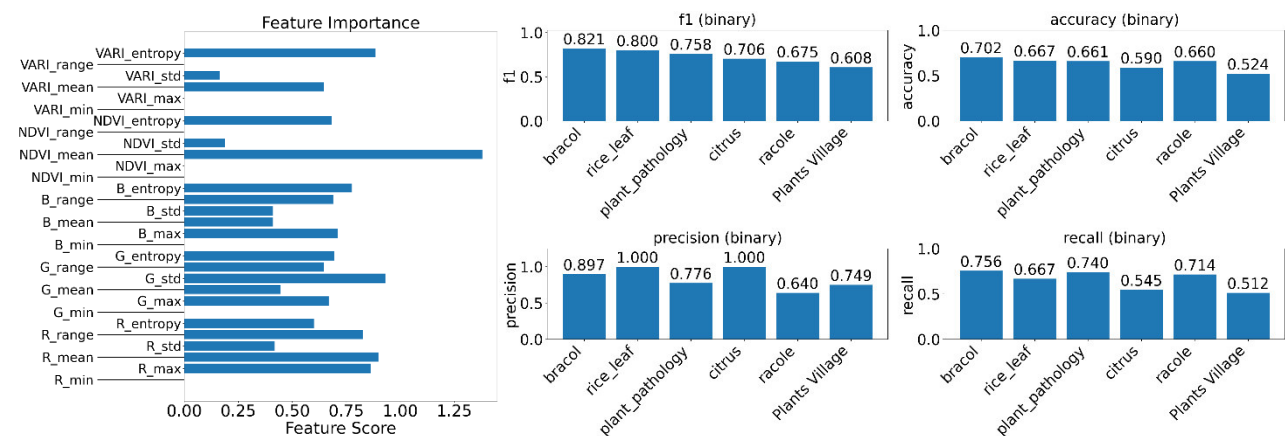


Рисунок 15. График важности признаков и метрик (F1, accuracy, precision, recall) на тестовых выборках

Диаграмма важности признаков показывает более равномерное распределение значимости: выраженных доминирующих параметров, как в предыдущем эксперименте, не наблюдается. Наибольший вклад в классификацию дает показатель NDVI\_mean, однако его превосходство над остальными не столь велико. Итоговая проверка на внешних датасетах выявила снижение точности: модель лучше адаптировалась к собственному тренировочному набору, но при этом её способность к обобщению оказалась хуже. В отличие от обучения на Plant Village, где наблюдался более высокий перенос, здесь рост качества на «своём» датасете сопровождается ухудшением результатов на остальных.

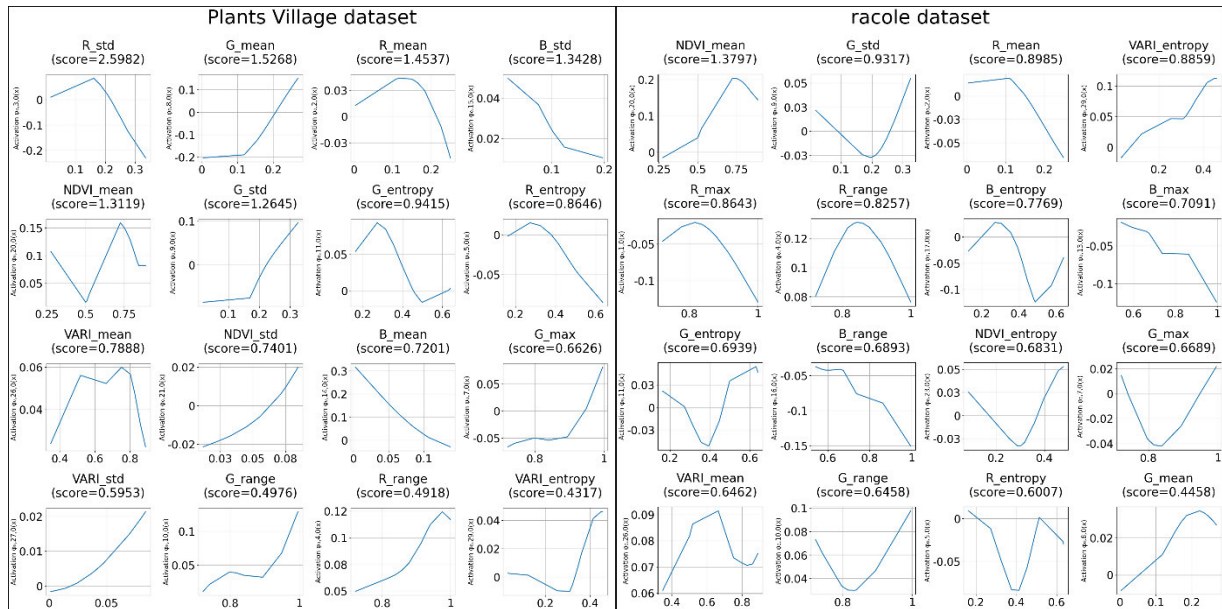


Рисунок 16. Графики функций активации модели, обученной на датасете Plant Village (слева) и датасете RaCoLe (справа)

На визуализации некоторых функций активации входного слоя мы видим, что некоторые функции активации, полученные после обучения на Plant Village и отсортированные по их score, имеют практически линейный вид, в то время как функции активации, обученные на датасете RaCoLe, вносят больше нелинейности. Стоит отметить, что на этом и последующем графиках функций активации оценивалась только общая форма функций, интервалы на которых определены после обучения, эти функции, конечно, играют важную роль, но пока неясно, как анализировать эти данные, так что они были оставлены малоразличимыми.

Заключительный эксперимент для бинарной классификации – обучение на комбинированном датасете, составленном из всех использованных, и последующая валидация.

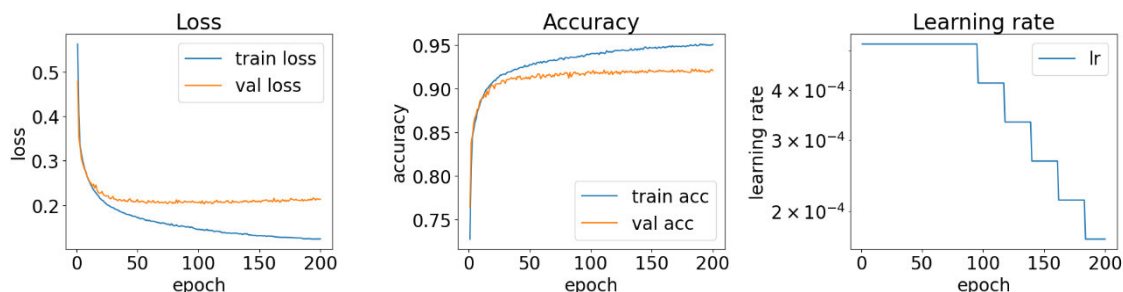


Рисунок 17. Графики функции потерь, точности и learning rate

На комбинированном датасете минимальное значение потерь на валидирующей выборке составило 0.2036 на 75-й эпохе за 108 секунд, обучение до 200-й эпохи заняло 284 секунды. Валидация, как и в



предыдущих экспериментах, выходит на плато, в то время как потери на тренирующей выборке продолжают падать.

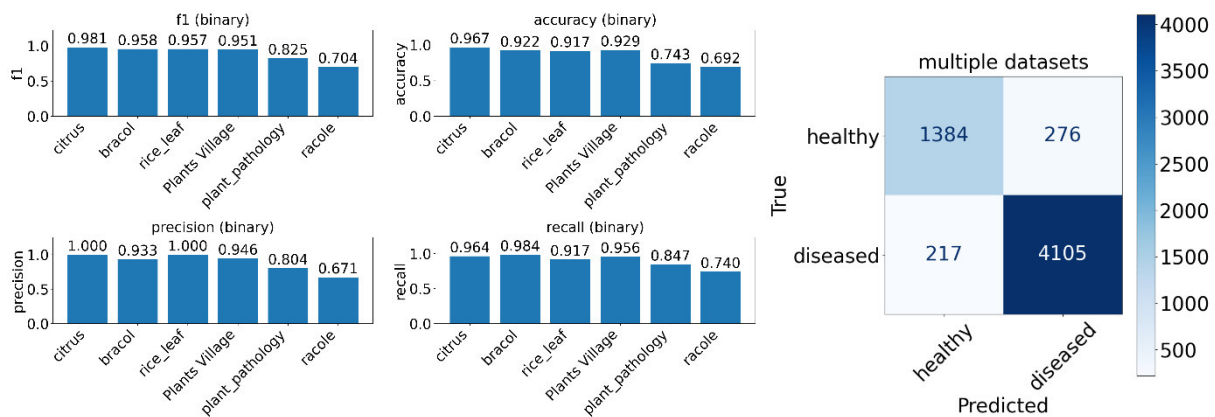


Рисунок 18. График метрик (F1, ассурасу, precision, recall) на тестовых выборках и общая матрица несоответствий

По итогу обучения есть видимый переко в сторону False negatives. Точность при валидации на всех датасетах повысилась, теперь наименьшая точность 69,2 %. На датасете Citrus точность бинарной классификации достигла 96,7 %.

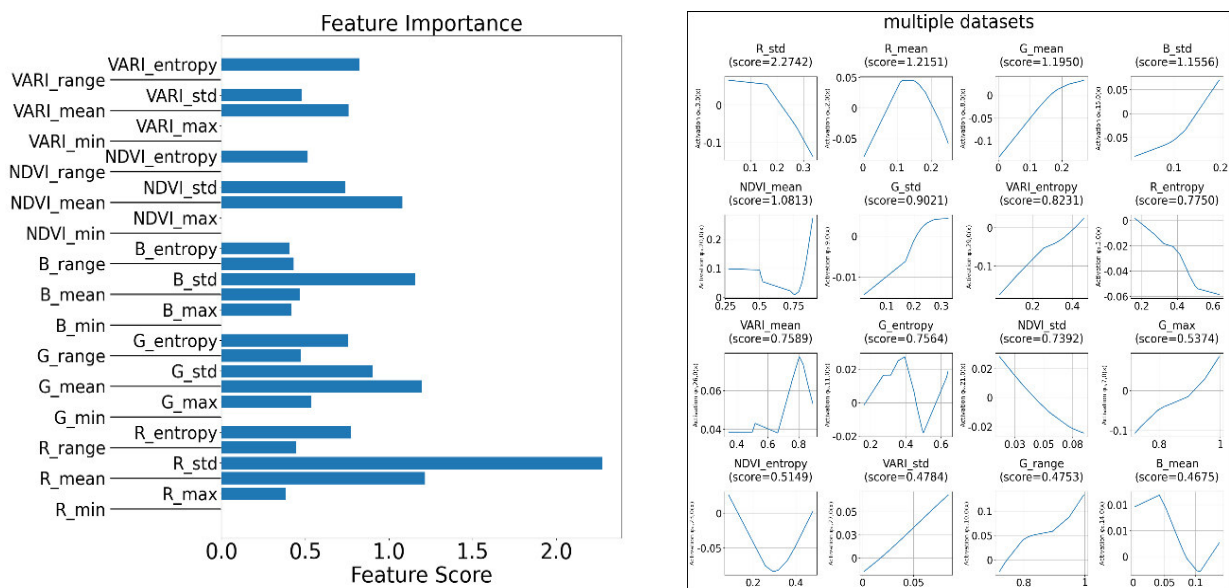


Рисунок 19. Диаграмма важности признаков и графики функций активации обученной модели

Итоговый вклад входных признаков очень близок к тому, что мы видели при обучении просто на Plant Village, это предсказуемо, так как даже комбинация всех остальных датасетов составляет небольшую долю от основного датасета. Аналогично для функций активации на входном слое.

Сравним полученную сеть с MLP с аналогичным количеством параметров, что по формулам (3) и (4) соответствует примерно 497 скрытым нейронам. Чтобы не перегружать публикацию, приводим результаты обучения MLP сразу на комбинированном датасете, результаты экспериментов, аналогичных двум предыдущим, будут указаны в итоговой таблице.

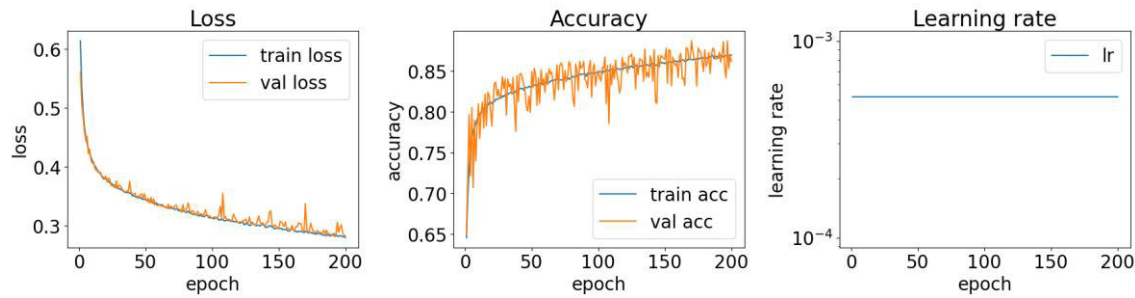


Рисунок 20. График функции потерь и точности MLP, обученной на 200 эпох

Обучение минимальное, значение потерь на валидации составило 0.2827 на 200 эпохе после 76 секунд обучения.

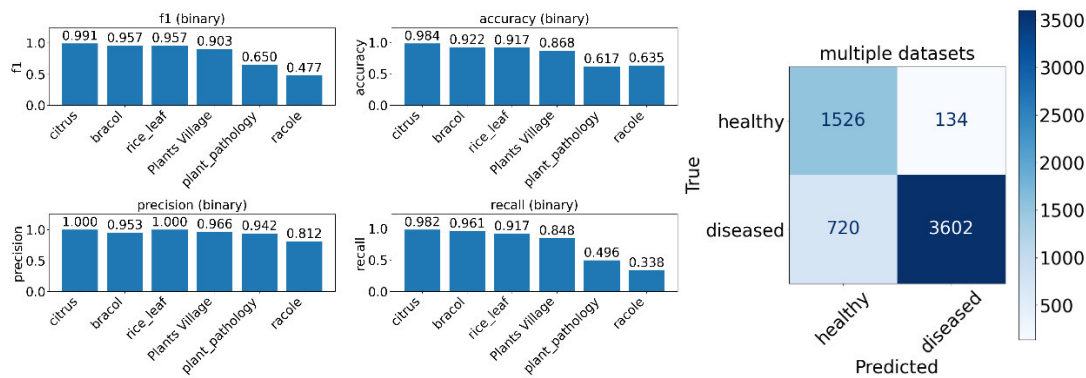


Рисунок 21. График точности на датасетах и матрица несоответствий обученной модели

На тестовой выборке перекося в сторону False negatives более сильный, чем у KAN, обученной на аналогичном датасете. По итогу модель MLP хуже обобщает, с минимальным результатом 61 % точности на одном из проверочных датасетов и дает худший Recall на общей тестовой выборке.

8.2. Эксперимент 2: мультиклассовая классификация растений

Таблица 7. Характеристики сети в эксперименте 2

Модель	Форма	Интервалы	Порядок сплайна	Датасет обучения	Классификация
KAN	[30,61,5]	20	3	Plant Village	Мульти класс

Проведем эксперимент, подобный эксперименту 1, но в этот раз будем делить растения по типам болезней из таблицы 1. Классы в обучающей выборке распределены неравномерно, всего 5 классов по типам болезней.

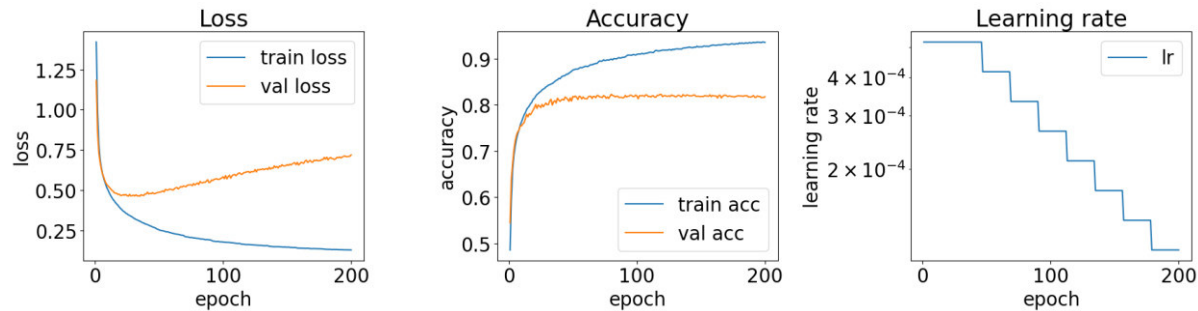


Рисунок 22. График функции потерь, точности и learning rate KAN при обучении задачи классификации из 5 классов на протяжении 200 эпох



Минимальное значение функции потерь на валидации составило 0.4638 и было достигнуто к 26-й эпохе через 34 секунды обучения, после чего модель начала переобучаться. Обучение 200 эпох заняло 259 секунд.

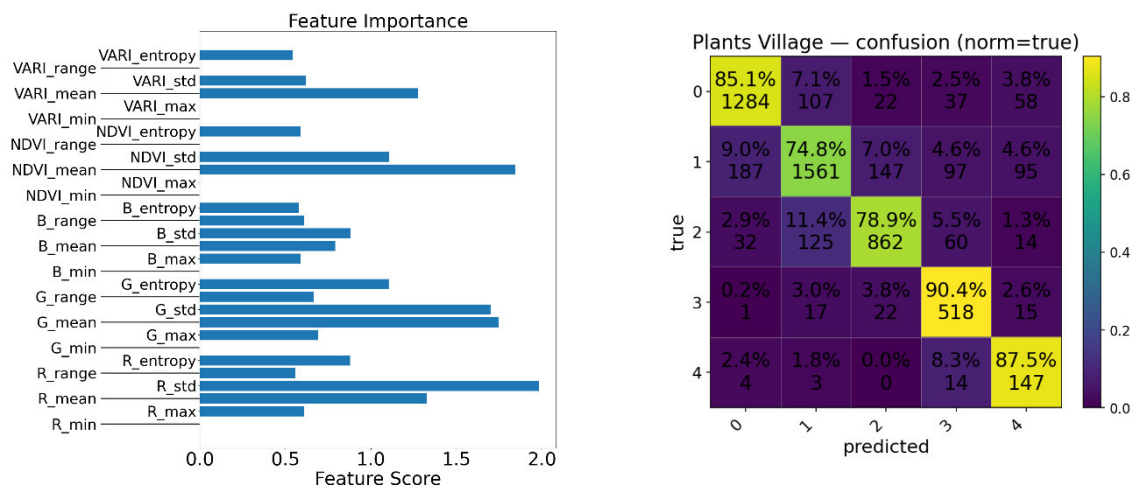


Рисунок 23. Диаграмма важности параметров и матрица несоответствий

Распределение оценок признаков отличается от полученного при бинарной классификации, более заметно разделение признаков по «важности», одними из важных признаков оказались B\_std, B\_mean, G\_std, G\_mean, R\_std, G\_entropy.

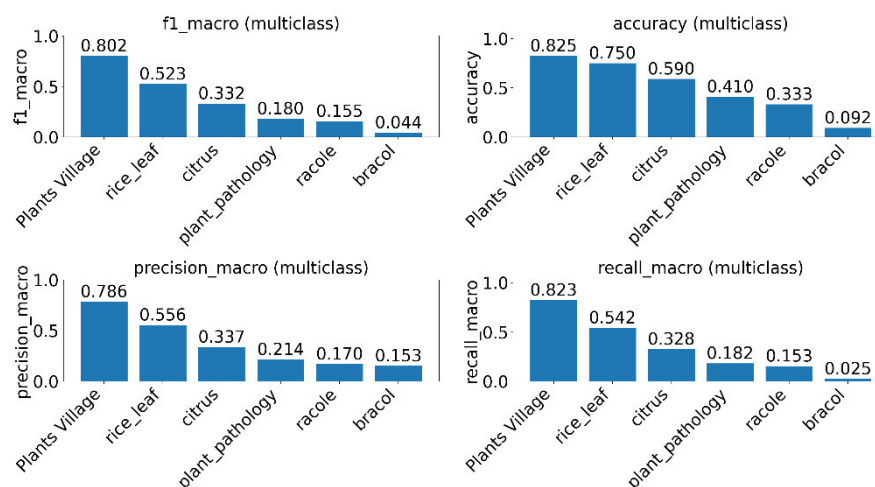


Рисунок 24. Диаграммы валидации на дополнительных датасетах

Попробуем обучить сеть сразу на комбинированном датасете. Минимальное значение потерь на валидации 0.5165 достигнуто на 20-й эпохе после 29 секунд обучения. Обучение до 200-й эпохи заняло 289 секунд. Как и в предыдущем эксперименте, обучение на комбинированном датасете выявило больше «важных» признаков на входном слое, но основные признаки остались как при тестировании на Plant Village.

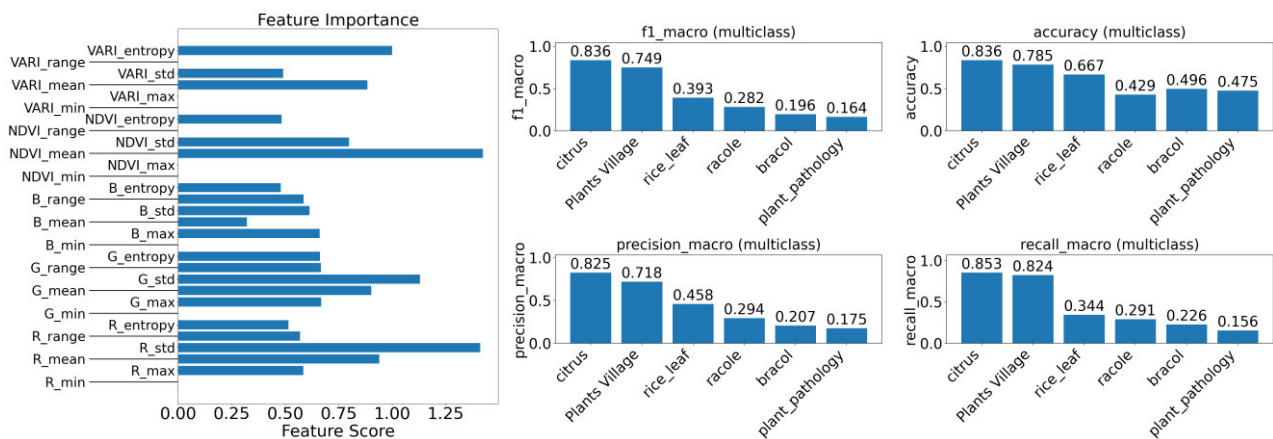


Рисунок 25. Диаграммы важности признаков и валидации на дополнительных датасетах

Обучение на комбинированном датасете позволило достичь высокой точности на всех датасетах. Минимальная точность составила 42.9 %, модель показала неплохую обобщающую способность.

Для сравнения была обучена модель MLP с 1540 нейронов на скрытом слое, что дает практически равное количество обучаемых параметров с нашей KAN-моделью в соответствии с формулами (3) и (4).

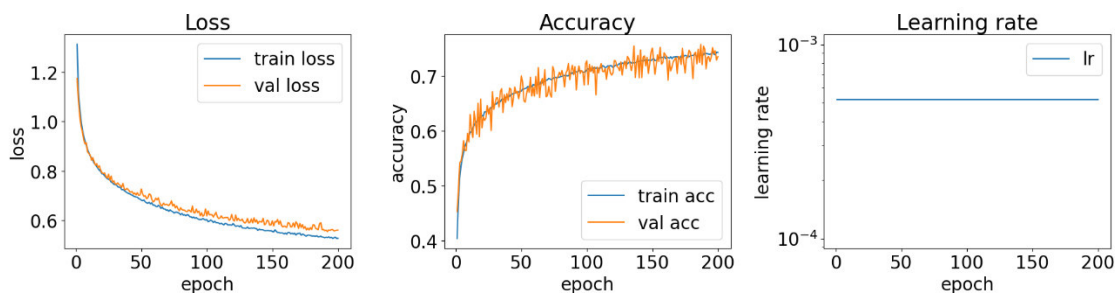


Рисунок 26. График функции потерь и точности MLP при обучении задачи классификации из 5 классов на протяжении 200 эпох

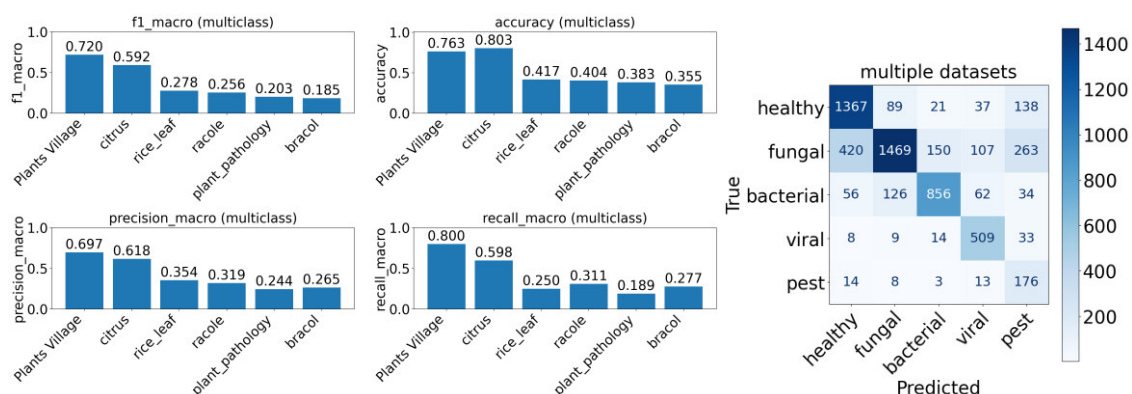


Рисунок 27. Диаграммы валидации на дополнительных датасетах и распределения предсказаний по классам

Классы на верифицирующей выборке определяются менее точно, точность на других датасетах в среднем тоже ниже, хотя обучение проходило на комбинированной версии датасета. На валидирующих датасетах модель показала неплохие результаты в мульти классовой задаче с лучшим показателем в 80.3 % точности.

### 2.3. Эксперимент 3: мультиклассовая классификация растений, сверточная KAN

В эксперименте рассмотрена сверточная KAN для классификации изображений 38 классов. Изображения отмасштабированы до размера 224\*224 пикселей и содержали в себе 3 канала: RGB. На

основе архитектуры Convolutional Neural Network (CNN), использованной в работе [34], собрана аналогичная модель Torch со сверточными слоями KAN, где ядро свертки состоит из обучаемых функций активации. Результаты обучения сверточных вариантов нейросети будут приведены в итоговой таблице. На данный момент сверточные KAN-слои требуют слишком больших вычислительных мощностей. При обучении на видеокарте A100 40GB VRAM пришлось прибегнуть к понижению порядка сплайнов до 1 и снижению количества интервалов до 4. Сверточная сеть на основе KAN прошла 10 эпох обучения примерно за 11 часов, в то время как оригинальная CNN прошла 10 эпох обучения примерно за час. Для дальнейших исследований стоит опробовать менее глубокую архитектуру сверточной KAN и проверить, дадут ли обучаемые функции активации возможность достичь результатов, сравнимых с более глубокими CNN. Для будущих исследований подготовлено программное обеспечение, визуализирующее результат применения фильтров сверточных слоев (рис. 30).

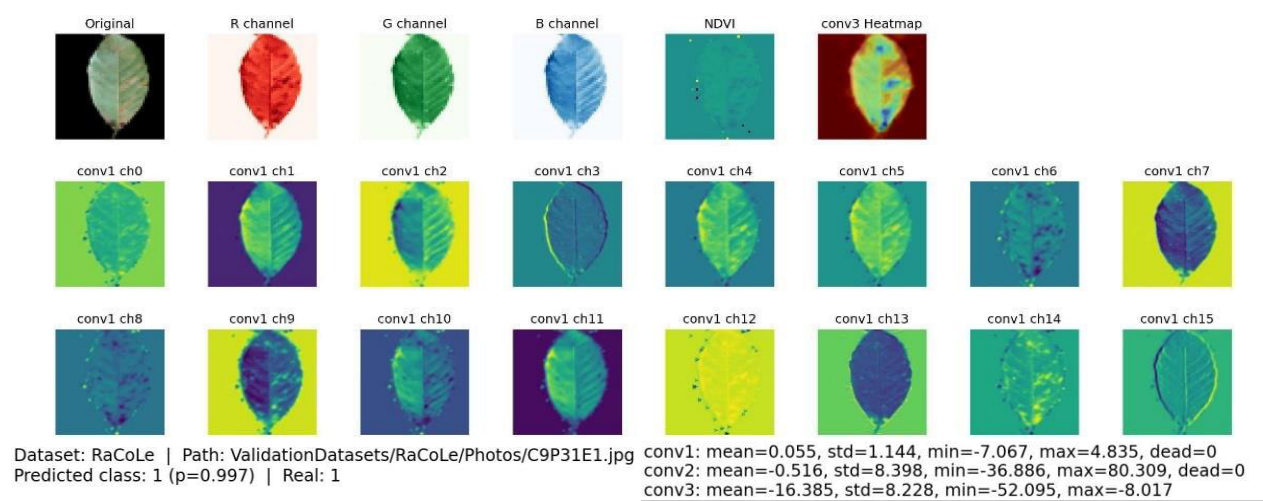


Рисунок 30. Пример отчета для дальнейшего исследования сверточных сетей на предмет объяснимости

Таблица 8. Общий результат тестирования

Описание эксперимента			Валидация					
Классификация	Модель	Обучающая выборка	BRACOL	Plant Pathology	RoCoLe	Citrus	Rice Leaf Diseases	Plant Village
Бинарная	KAN(2)	Plant Village	88,70%	63,40%	41,00%	90,20%	<b>100,00%</b>	93,20%
		RoCoLe	70,20%	66,10%	66,00%	59,00%	66,70%	52,40%
		Комбинированный датасет	<b>92,20%</b>	<b>74,30%</b>	<b>69,20%</b>	96,70%	97,70%	<b>92,90%</b>
	MLP(2)	Plant Village	90,10%	53,60%	50,60%	88,50%	<b>100,00%</b>	87,00%
		RoCoLe	<b>92,20%</b>	71,60%	<b>69,20%</b>	52,50%	83,30%	59,20%
		Комбинированный датасет	<b>92,20%</b>	61,70%	63,50%	<b>98,40%</b>	91,70%	86,80%
Мультиклассовая	KAN(5)	Plant Village	9,20%	41,00%	33,30%	59,00%	<b>75,00%</b>	<b>82,50%</b>
		Комбинированный датасет	<b>49,60%</b>	47,50%	42,90%	<b>83,60%</b>	66,70%	78,50%
	MLP(5)	Plant Village	0,01%	1,37%	1,03%	59,00%	50,00%	75,80%
		Комбинированный датасет	33,00%	<b>51,00%</b>	<b>48,00%</b>	76,00%	70,00%	78,00%
	convKAN (38)	Plant Village	59,33%	-	-	-	-	-
	CNN(38)	Plant Village	83,10%	-	-	-	-	-

## Заключение

KAN-сеть показала хорошие результаты по отношению к MLP как на задаче регрессии математической функции, так и на задачах бинарной и мультиклассовой классификаций болезней растений относительно количества обучаемых параметров. Стоит отметить, что KAN-сети обучаются дольше из-за того, что расчёт B-сплайнов не распараллеливается на GPU, и из-за того, и что для их обучения требуется производить большее количество вычислений. Однако этот недостаток уже активно исправляется путем использования параллелизуемых базисных функций.

В плане дальнейшей работы изучить различия KAN и MLP по времени обучения; реализацию KAN на основе RBF – параллелизуемых на GPU; другие подходы к использованию KAN; сверточные KAN-слои; KAN трансформеры. Особый интерес вызывает дополнение стандартных архитектур нейронных сетей KAN-слоями. Подобная концепция уже рассматривалась в работе [29] почти одновременно с выходом первой публикации по KAN.

## Источник финансирования

Работа выполнена при поддержке Министерства экономического развития Российской Федерации (соглашение о предоставлении гранта № 139-15-2025-004 от 17 апреля 2025 г., ИГК 000000Ц313925P3X0002).

## Список литературы

- [1] KAN: Kolmogorov-Arnold Networks / Z. Liu, Y. Gao, Z. Zhang, et al. // arXiv preprint arXiv:2404.19756. 2024. Режим доступа: <https://doi.org/10.48550/arXiv.2404.19756>
- [2] KAN 2.0: Kolmogorov-Arnold Networks meet science / Z. Liu, Y. Gao, Z. Zhang et al. // arXiv preprint arXiv:2408.10205. 2024. Режим доступа: <https://doi.org/10.48550/arXiv.2408.10205>
- [3] KAN or MLP: A Fairer Comparison / R. Yu, W. Yu, X. Wang // arXiv preprint arXiv:2407.16674. 2024. Режим доступа: <https://doi.org/10.48550/arXiv.2407.16674>
- [4] Schmidt-Hieber J. The Kolmogorov–Arnold representation theorem revisited // Neural networks. 2021. Vol. 137. P. 119-126. Режим доступа: <https://doi.org/10.48550/arXiv.2007.15884>
- [5] Chebyshev Polynomial-Based Kolmogorov-Arnold Networks: An Efficient Architecture for Nonlinear Function Approximation / S.S. Sidharth, A.R. Keerthana, R. Gokul, K.P. Anas // arXiv preprint arXiv:2405.07200. 2024. Режим доступа: <https://arxiv.org/html/2405.07200>
- [6] Single-Layer Learnable Activation for Implicit Neural Representation (SL<sup>2</sup>A-INR) / M. Heidari, J. Wei, Z. Wang et al. // arXiv preprint arXiv:2409.10836. 2024. Режим доступа: <https://doi.org/10.48550/arXiv.2409.10836>
- [7] Lai M.J., Shen Z. The Kolmogorov superposition theorem can break the curse of dimensionality when approximating high dimensional functions // arXiv preprint arXiv:2112.09963. 2021. Режим доступа: <https://doi.org/10.48550/arXiv.2112.09963>
- [8] Entropy as a High-Level Feature for XAI-Based Early Plant Stress Detection / M. Lysov, I. Maximova, E. Vasiliev, A. Getmanskaya, V. Turlapov // Entropy. 2022. Vol. 24, no. 11. P. 1597. Режим доступа: <https://doi.org/10.3390/e24111597>
- [9] Ensuring Explainability and Dimensionality Reduction in a Multidimensional HSI World for Early XAI-Diagnostics of Plant Stress / M. Lysov, K. Pukhkiy, E. Vasiliev, A. Getmanskaya, V. Turlapov // Entropy. 2022. Vol. 25, no. 5. P. 801. Режим доступа: <https://doi.org/10.3390/e25050801>
- [10] Barbedo J.G.A. Plant disease identification from individual lesions and spots using deep learning // Biosystems Engineering. 2019. Vol. 180. P. 96-107. Режим доступа: <https://doi.org/10.1016/j.biosystemseng.2019.02.002>
- [11] DropKAN: Regularizing KANs by masking post-activations / Z. Liu, T. Zhang, W. Wang, et al. // arXiv preprint arXiv:2406.09516. 2024. Режим доступа: <https://arxiv.org/pdf/2407.13044>
- [12] Kolmogorov-Arnold Transformer / Y. Zhang, Z. Liu, T. Zhang, et al. // arXiv preprint arXiv:2409.10594. 2024. Режим доступа: <https://arxiv.org/abs/2409.10594>
- [13] Kolmogorov-Arnold Convolutions: Design Principles and Empirical Studies / J. Jiang, Z. Liu, T. Zhang, et al. // arXiv preprint arXiv:2407.01092. 2024. Режим доступа: <https://arxiv.org/abs/2407.01092>
- [14] Kolmogorov-Arnold Fourier Networks / T. Chen, Z. Liu, T. Zhang, et al. // arXiv preprint arXiv:2502.06018. 2024. Режим доступа: <https://arxiv.org/abs/2502.06018>
- [15] Wav-KAN: Wavelet Kolmogorov-Arnold Networks / J. Xu, Z. Liu, T. Zhang, et al. // arXiv preprint arXiv:2405.12832. 2024. Режим доступа: <https://arxiv.org/abs/2405.12832>
- [16] CoxKAN: Kolmogorov-Arnold Networks for Interpretable, High-Performance Survival Analysis / H. Zhang, Z. Liu, T. Zhang, et al. // arXiv preprint arXiv:2409.04290. 2024. Режим доступа: <https://arxiv.org/abs/2409.04290>
- [17] RKAN: Residual Kolmogorov-Arnold Network / Y. Li, Z. Liu, T. Zhang, et al. // arXiv preprint arXiv:2410.05500. 2024. Режим доступа: <https://arxiv.org/abs/2410.05500>
- [18] Noyan M. A. Uncovering bias in the PlantVillage dataset // arXiv preprint arXiv:2206.04374. 2022. Режим доступа: <https://arxiv.org/abs/2206.04374>

- 
- [19] Fenu G., Mallocci F.M. DiaMOS Plant: A Dataset for Diagnosis and Monitoring Plant Disease // *Agronomy*. 2021. Vol. 11, no. 11. P. 2107. Режим доступа: <https://doi.org/10.3390/agronomy11112107>
- [20] A Novel Machine Learning Method for Estimating Biomass of Grass Swards Using a Photogrammetric Canopy Height Model, Images and Vegetation Indices Captured by a Drone / N. Viljanen, E. Honkavaara, R. Näsi, et al. // *Agriculture*. 2018. Vol. 8, no. 5. P. 70. Режим доступа: <https://doi.org/10.3390/agriculture8050070>
- [21] The global burden of pathogens and pests on major food crops / S. Savary, L. Willocquet, S.J. Pethybridge, et al. // *Nature Ecology & Evolution*. 2019. Vol. 3, no. 3. P. 430–439. DOI: 10.1038/s41559-018-0793-y
- [22] Oerke E.-C. Crop losses to pests // *The Journal of Agricultural Science*. 2006. Vol. 144, no. 1. P. 31–43. DOI: 10.1017/S0021859605005708
- [23] Hughes D., Salathé M. An open access repository of images on plant health to enable the development of mobile disease diagnostics // *arXiv preprint arXiv:1511.08060*. 2015. Режим доступа: <https://arxiv.org/abs/1511.08060>
- [24] RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition / J. Parraga-Alava, K. Cusme, A. Loor, E. Santander // *Data in Brief*. 2019. Vol. 25. P. 104414. DOI: 10.1016/j.dib.2019.104414
- [25] Krohling R., Esgario J., Ventura J.A. BRACOL-A Brazilian Arabica Coffee Leaf images dataset to identification and quantification of coffee diseases and pests // *Mendeley Data*. 2019. V1. DOI: 10.17632/yy2k5y8mxg.1
- [26] Prajapati H.B., Shah J.P., Dabhi V.K. Detection and classification of rice plant diseases // *Intelligent Decision Technologies*. 2017. Vol. 11. P. 357–373. DOI: 10.3233/IDT-170301
- [27] The Plant Pathology Challenge 2020 data set to classify foliar disease of apples / R. Thapa, K. Zhang, N. Snively, et al. // *Applications in Plant Sciences*. 2020. Vol. 8, no. 6. e11390. DOI: 10.1002/aps3.11390
- [28] A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning / H.T. Rauf, B.A. Saleem, M.I.U. Lali, et al. // *Data in Brief*. 2019. Vol. 26. P. 104340. DOI: 10.1016/j.dib.2019.104340. Heidari M. et al.
- [29] Poluektov M., Polar A. Construction of the Kolmogorov-Arnold representation using the Newton-Kaczmarz method // *arXiv preprint arXiv:2305.08194*. 2023.

## Адверсативная аугментация физически информированных нейронных операторов для задачи одномерного уравнения диффузии

Н. Д. Локшин, А. С. Крылов

Московский государственный университет имени М. В. Ломоносова, Москва, Россия

**Аннотация.** Известно, что нейронные сети уязвимы к adversarial-атакам – данным с тщательно подобранными возмущениями, которые незаметны человеческому глазу. В задачах медицинской визуализации это может представлять серьёзную угрозу при построении предсказаний на основе глубоких нейронных сетей. В данной статье мы изучаем влияние adversarial-атак на MLP и Fourier Neural Operator (FNO) при решении уравнения диффузии, которое иногда используется в задачах подавления шума в медицинских изображениях. Мы показываем, что при использовании adversarial augmentation ошибка  $L_2$  на adversarial-решениях для возмущённых начальных условий немного улучшается, в то время как ошибка на решениях для чистых начальных условий уменьшается, что указывает на более высокую устойчивость по сравнению с обучением без adversarial augmentation. Дополнительно мы вводим простую, но новую adversarial-атаку, которая эффективна против Fourier Neural Operator – Low Frequency Attack.

**Ключевые слова:** PINO, Fourier Neural Operator, Adversarial Attacks, Deep Learning, Diffusion Equation, Medical Imaging

## Adversarial augmentation in the task of PINO for 1D diffusion equation

N. D. Lockshin, A. S. Krylov

Lomonosov Moscow State University, Moscow, Russia

**Abstract.** Neural networks are known to be vulnerable against adversarial attacks – data with carefully crafted adversarial perturbations that are imperceptible to the human eye. In medical imaging tasks this can be a major threat for making predictions based on deep neural network solutions. In this paper we study the effect of adversarial attacks on MLP and Fourier Neural Operator (FNO) for solving the diffusion equation, sometimes used in medical image denoising tasks. We show that with adversarial augmentation the  $L_2$  error on the adversarial solutions for perturbed initial conditions marginally improves, whereas the error on the solutions for clean initial conditions decreases, indicating higher robustness than without adversarial augmentation. Additionally, we introduce a simple yet novel adversarial attack that is shown to be effective against Fourier Neural Operators – the Low Frequency Attack.

**Keywords:** PINO, Fourier Neural Operator, Adversarial Attacks, Deep Learning, Diffusion Equation, Medical Imaging

### Введение

Многие задачи науки и техники связаны с многократным решением сложных систем уравнений в частных производных (PDE) для различных значений некоторых параметров. Примеры возникают в молекулярной динамике, микромеханике и турбулентных течениях. Часто такие системы требуют тонкой дискретизации, чтобы уловить моделируемое явление. В результате традиционные численные решатели оказываются медленными и иногда неэффективными. Например, при проектировании материалов, таких как аэродинамические профили, необходимо решать обратную задачу, где требуется провести тысячи прогонов прямой модели. Быстрый метод делает такие задачи осуществимыми.

Традиционные решатели PDE, такие как методы конечных элементов (FEM) и конечных разностей (FDM), решают уравнение путём дискретизации пространства. Таким образом, возникает компромисс между скоростью и точностью: грубые сетки быстры, но менее точны; мелкие сетки точны, но медленны. Сложные PDE-системы обычно требуют очень мелкой дискретизации и поэтому крайне трудоёмки для традиционных методов. С другой стороны, методы, основанные на данных, могут напрямую обучаться по данным, что делает их на порядки быстрее традиционных решателей.

Physics-Informed Neural Networks (PINNs) [1] произвели революцию в решении PDE, включая физические законы в процесс обучения. PINNs могут интегрировать экспериментальные данные разной точности и модальности с различными формулировками уравнений Навье–Стокса для несжимаемых потоков [2, 3], а также для сжимаемых потоков [4] и биомедицинских течений [5]. Однако PINNs ограничены в своей способности устойчиво решать широкий класс PDE, так как любое



изменение начальных или граничных условий требует полного переобучения нейросети. Для решения этой проблемы были предложены PhysicsInformed Neural Operators (PINOs) [6], которые расширяют PINNs. PINO преобразовали решение PDE, обучая отображения операторов, которые обобщаются на разные входные функции, предлагая более высокую эффективность для задач вроде одномерного уравнения диффузии. PINO интегрируют физические законы в операторные нейросетевые структуры, такие как DeepONet [7] или Fourier Neural Operator (FNO) [8], что позволяет моделировать сложные системы с различными начальными и граничными условиями.

Тем не менее, глубокие нейронные сети уязвимы к adversarial-атакам — малым возмущениям на входе, которые приводят к значительным ошибкам предсказания, — что представляет собой вызов для надёжных научных вычислений. Adversarial-атаки и методы защиты широко изучены в компьютерном зрении [9, 10, 11, 12, 13]. Однако лишь немногие работы посвящены adversarial-атакам в задачах PINN и PINO. Недавние исследования, такие как [14], предлагают методы вроде AT-PINNs для повышения устойчивости, хотя их применение к уравнению теплопроводности пока мало изучено. Другие работы, включая исследования PINNs для задач теплопередачи [15], подчёркивают необходимость устойчивых моделей, но не рассматривают adversarial-атаки. Как и PINNs, PINOs уязвимы к adversarial-атакам, что ставит под угрозу их надёжность [16]. Кроме того, методы компьютерного зрения, основанные на решении PDE, например гибридные подходы [17], использующие метод Перона–Малика для устранения шума [18], также могут быть уязвимы.

В этой статье мы исследуем влияние таких атак на PINO для одномерного уравнения теплопроводности и эффективность adversarial augmentation в повышении устойчивости моделей, представляя метрики для оценки устойчивости и выделяя направления для будущих исследований, чтобы обеспечить надёжные предсказания в критических приложениях, таких как медицинская визуализация и теплотехника.

## Постановка задачи

### 1. Одномерное уравнение теплопроводности

Мы рассматриваем одномерное уравнение теплопроводности

$$\frac{\partial t}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}, \quad x \in (0,1), t \in (0,1], \quad (1)$$

где  $u(x, t)$  — функция распределения температуры,  $t$  — время,  $x$  — пространственная координата, а  $\alpha > 0$  — коэффициент теплопроводности.

Для решения этого уравнения нам необходимо задать начальное и граничные условия. Пусть начальное распределение температуры задано:

$$u(x, 0) = f(x), \forall x \in (0,1), \quad (2)$$

где  $f(x)$  — произвольная функция начального распределения температуры вдоль стержня длиной 1. Рассмотрим однородные граничные условия Дирихле

$$u(0, t) = 0, \forall t > 0, \quad (3)$$

$$u(1, t) = 0, \forall t > 0. \quad (4)$$

С помощью синус-преобразования решение можно записать как

$$u(x, t) = \sum_{n=1}^{\infty} b_n \sin(n\pi x) e^{-\alpha(n\pi)^2 t}, \quad (5)$$

где коэффициенты  $b_n$  определяются начальными условиями

$$b_n = 2 \int_0^1 f(x) \sin(n\pi x) dx. \quad (6)$$

Подставив эти коэффициенты, получаем полное решение  $u(x, t)$ , которое используется в наших экспериментах как ground truth.

### 2. Adversarial-атаки

Adversarial-атаки — это малые возмущения, добавляемые к исходному образцу, которые могут приводить к ошибкам классификации или регрессии нейросети. В контексте PDE атаки могут быть

направлены на начальные или граничные условия, а также на входную сетку. В данной работе мы сосредотачиваемся на атаках на начальные условия.

Формулировка задачи:

$$\max_{\xi} L(r(v + \xi; \theta), r(v + \xi)), \quad (7)$$

$$\text{s. t. } \|\xi\|_{\infty} \leq \epsilon, \quad (8)$$

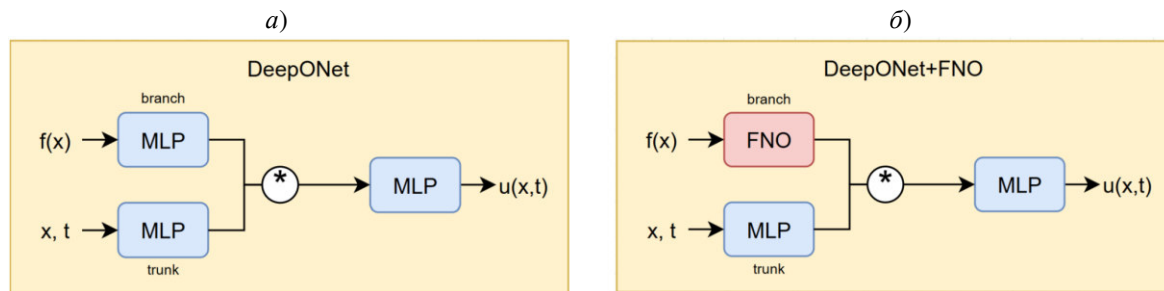
где  $L(\cdot)$  — функция расстояния между  $(r(v + \xi; \theta)$  и  $r(v + \xi)$ , а  $\epsilon$  — максимально допустимая величина возмущения. В нашей работе используется  $L_2$  — норма.

### 3. DeepONet

DeepONet — архитектура нейронной сети для обучения операторов [7]. Она вдохновлена теоремой об универсальной аппроксимации операторов и состоит из двух подсетей: branch network и trunk network. Branch network кодирует входные функции, а trunk network кодирует точки, в которых вычисляется оператор. В наших экспериментах мы используем три варианта DeepONet:

- оригинальный DeepONet с MLP в качестве branch и trunk;
- DeepONet с FNO в качестве branch;
- Stacked DeepONet с FNO в branch и разными Nmodes.

Также мы добавляем дополнительный слой MLP после произведения branch и trunk сетей, как показано на рисунке.



Варианты архитектуры DeepONet, используемые в данной статье: а) DeepONet с MLP в качестве branch и trunk сетей, с дополнительным слоем MLP в конце; б) DeepONet с нейронным оператором Фурье в качестве branch сети и MLP в качестве trunk, с дополнительным слоем MLP в конце

### Теория

Для обучения нейросети на каждой итерации мы сэмплируем  $x \sim \text{Uniform}(0, 1)$ ,  $t \sim \text{Exponential}(\lambda = 10.0)$ . Экспоненциальное распределение времени выбрано, чтобы избежать переобучения на состояниях с температурами, близкими к нулю. Формула плотности этого распределения имеет вид

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (9)$$

Начальные условия формируются как сумма синусоид с случайными коэффициентами

$$f(x) = \sum_{i=1}^{Nc} a_i \sin(ipx), \quad (10)$$

где  $Nc = 100$ .

Для генерации adversarial-атак на начальные условия мы используем:

1. Low Frequency Attack (LFA) — добавление низкочастотной компоненты

$$f_{adv}(x) = f(x) + \epsilon \sin(kpx), \quad (11)$$

где  $k = 1$ ,  $\epsilon = 0.1$ .

2. Fast Gradient Sign Method (FGSM) [9]:

$$f_{adv}(x) = f(x) + \epsilon \text{sign} \nabla_x L(u(x, t; \theta), u(x, t)). \quad (12)$$

При adversarial training на каждой итерации с вероятностью  $p = 0.5$  мы используем adversarial начальное условие, сгенерированное с помощью предварительно обученной модели. Решение для такого условия берётся как ground truth для обучения.

### Результаты экспериментов

Мы оцениваем модели, вычисляя ошибку  $L_2$  между предсказаниями и двумя решениями: чистым (без возмущений) и adversarial (с возмущением в начальных условиях). Формула

$$L_2 \approx \frac{1}{N_x N_t} \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_t-1} (u(\frac{i}{N_x}, \frac{j}{N_t}; \theta) - u(\frac{i}{N_x}, \frac{j}{N_t}))^2, \quad (13)$$

при  $N_x = N_t = 1024$ . Результаты приведены в таблице. Первый столбец – названия моделей. Надстрочный индекс adv в названии модели указывает, что модель обучалась с использованием adversarial-усиления. Далее названия столбцов обозначают тип adversarial-атаки на начальное условие; надстрочный индекс adv обозначает ошибку в норме  $L_2$  относительно adversarial-решения, а отсутствие индекса – на ошибку относительно чистого решения.

**Численные результаты, показывающие эффективность ADVERSARIAL-усиления**

Модель \ Атака:	$L_2$	$L_2$ LFA(k=1, $\epsilon=0.1$ )	$L_2^{\text{adv}}$ LFA(k=1, $\epsilon=0.1$ )	$L_2$ FGSM( $\epsilon=0.1$ )	$L_2^{\text{adv}}$ FGSM( $\epsilon=0.1$ )
MLP	0.0009	0.0009	0.006	0.001	0.001
MLP <sup>adv</sup>	0.0005	0.002	0.001	0.001	0.0008
FNO <sub>16</sub>	0.0009	0.001	0.005	0.016	0.015
FNO <sub>16</sub> <sup>adv</sup>	0.0003	0.003	0.0006	0.005	0.004
FNO <sub>32</sub>	0.0006	0.001	0.003	0.009	0.005
FNO <sub>32</sub> <sup>adv</sup>	0.0006	0.002	0.002	0.004	0.001
FNO <sub>64</sub>	0.04	0.04	0.05	0.05	0.05
FNO <sub>64</sub> <sup>adv</sup>	0.04	0.04	0.05	0.05	0.05
StackFNO <sub>8,16,32,64</sub>	0.001	0.001	0.006	0.01	0.004
StackFNO <sub>8,16,32,64</sub> <sup>adv</sup>	0.0003	0.003	0.0009	0.005	0.001

Можно видеть, что практически в любом случае, при любом выборе  $N_{\text{modes}}$  для FNO, модели с состязательной аугментацией превосходят свои аналоги без состязательной аугментации на состязательных решениях, что свидетельствует о повышенной устойчивости. Кроме того, состязательная аугментация может повысить производительность на чистых данных, как показано во втором столбце.

### Заключение

В этой статье предложен метод adversarial augmentation для обучения PINO. Показано, что он повышает устойчивость моделей на чистых и adversarial-данных. Необходимы дальнейшие исследования влияния adversarial augmentation и выбора атак для других PDE (поток Дарси, уравнение Бюргерса, Навье–Стокса), а также методов медицинской визуализации, основанных на 2D-фильтре Перона–Малика.

### Список литературы

1. M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational physics* 378 (2019) 686–707.
2. X. Jin, S. Cai, H. Li, G. E. Karniadakis, Nsfnets (navier-stokes ow nets): Physics-informed neural networks for the incompressible navier-stokes equations, *Journal of Computational Physics* 426 (2021) 109951.
3. M. Raissi, A. Yazdani, G. E. Karniadakis, Hidden uid mechanics: Learning velocity and pressure fields from flow visualizations, *Science* 367 (2020) 1026–1030.
4. Z. Mao, A. D. Jagtap, G. E. Karniadakis, Physics-informed neural networks for high-speed flows, *Computer Methods in Applied Mechanics and Engineering* 360 (2020) 112789.

5. M. Yin, X. Zheng, J. D. Humphrey, G. E. Karniadakis, Non-invasive inference of thrombus material properties with physics-informed neural networks, *Computer Methods in Applied Mechanics and Engineering* 375 (2021) 113603.
6. N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar, Neural operator: Learning maps between function spaces with applications to PDEs, *Journal of Machine Learning Research* 24 (2023) 1–97.
7. L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis, Learning nonlinear operators via deeponet based on the universal approximation theorem of operators, *Nature machine intelligence* 3 (2021) 218–229.
8. Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations, *arXiv preprint arXiv:2010.08895* (2020).
9. I. J. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples, *arXiv preprint arXiv:1412.6572* (2014).
10. X. Chen, X. Gao, J. Zhao, K. Ye, C.-Z. Xu, Advdiuser: Natural adversarial example synthesis with diffusion models, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4562–4572.
11. C. Xiao, B. Li, J.-Y. Zhu, W. He, M. Liu, D. Song, Generating adversarial examples with adversarial networks, *arXiv preprint arXiv:1801.02610* (2018).
12. P. Zhu, G. Osada, H. Kataoka, T. Takahashi, Frequency-aware gan for adversarial manipulation generation, in: *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4315–4324.
13. W. Ma, Y. Li, X. Jia, W. Xu, Transferable adversarial attack for both vision transformers and convolutional networks via momentum integrated gradients, in: *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4630–4639.
14. Y. Li, S. Shi, Z. Guo, B. Wu, Adversarial training for physics-informed neural networks, 2023. URL: <https://arxiv.org/abs/2310.11789>. *arXiv:2310.11789*.
15. S. Cai, Z. Wang, S. Wang, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks for heat transfer problems, *Journal of Heat Transfer* 143 (2021) 060801.
16. A. D. Adesoji, P.-Y. Chen, Evaluating the adversarial robustness for fourier neural operators, *arXiv preprint arXiv:2204.04259* (2022).
17. N. Wang, Y. Shang, Y. Chen, M. Yang, Q. Zhang, Y. Liu, Z. Gui, A hybrid model for image denoising.
18. P. Pietro, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. PAMI* 12 (1990).

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-816-827

## Обзор наборов данных электронной микроскопии мозга для задач сегментации

М. Э. Куракин, Н. А. Соколов, А. А. Серебрякова

Исследовательский центр в области искусственного интеллекта, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

**Аннотация.** В данной работе представлен систематический обзор публично доступных размеченных наборов данных для электронной микроскопии (ЭМ). Приведен детальный анализ ключевых характеристик существующих датасетов, включая разрешение изображений, плотность аннотаций, баланс классов и согласованность разметки. Результаты исследования демонстрируют существенную неоднородность качества и полноты существующих разметок, что создает значительные сложности для обучения универсальных моделей сегментации. Работа представляет ценность для исследователей в области компьютерного зрения и нейробиологии, предлагая методическую основу для оценки качества ЭМ-данных и выявления наиболее перспективных направлений для создания новых специализированных датасетов.

**Ключевые слова:** электронная микроскопия, сегментация изображений, обзор датасетов, гистограммный анализ, нейробиология, компьютерное зрение, мозг мыши

## Brain electron microscopy dataset review

M. E. Kurakin, N. A. Sokolov, A. A. Serebryakova

Research Center for Artificial Intelligence, Lobachevsky State University, Nizhny Novgorod, Russia

**Abstract.** This paper presents a systematic review of publicly available annotated datasets for electron microscopy (EM). We conduct a detailed analysis of key characteristics of existing datasets, including image resolution, annotation density, class balance, and labeling consistency. The study reveals significant heterogeneity in the quality and completeness of existing annotations, which poses major challenges for training universal segmentation models. The work provides valuable insights for researchers in computer vision and neurobiology by offering a methodological framework for assessing EM data quality and identifying promising directions for developing new specialized datasets.

**Keywords:** electron microscopy, image segmentation, datasets overview, histogram analysis, neurobiology, computer vision, mouse brain

## Введение

Современные методы исследования биологических тканей зачастую содержат алгоритмы машинного обучения, в том числе и глубокие нейронные сети. Однако первые нейросетевые модели, появившиеся в середине XX века, такие как перцептрон Розенблатта, предложенный в 1957 году, не получили широкого распространения как из-за нехватки вычислительных мощностей, так и данных для обучения. С появлением специализированных графических адаптеров в начале XXI века нейросетевые технологии стали активно развиваться, но только с появлением большого количества размеченных данных их начинают активно использовать. Так, с появлением AlexNet в 2012 году сверточные нейронные сети начали активно использоваться для обработки изображений, а появившаяся архитектура U-Net [1] в 2015 году стала основой для сегментации медицинских изображений.

Среди биологических тканей особый интерес представляют наборы данных изображений клеток головного мозга мыши. Для такой задачи, как исследование структуры строения и работы мозга, одним из нескольких основных методов получения данных являются электронные микроскопы (ЭМ), а для обработки этих данных часто применяются методы сегментации при помощи сверточных нейронных сетей (СНС). К сожалению, успешность применения методов глубокого обучения снижается нехваткой данных для обучения, поэтому в этой статье будет проведен обзор и анализ доступных датасетов мозга мыши.

## Описание датасетов

Датасеты естественным образом встречаются в работах по архитектурам нейросетей, для аккумуляции датасетов создаются облачные хранилища и сайты с СУБД, а также исследователи собирают в свои таблички и публикуют в статьях, как, например, в статье авторов Aswath A. и др. [2].

**Набор данных EPFL**, или Lucchi, или Lucchi++, или ITMM, (рис. 1) стал одним из первых общедоступных датасетов для 3D-сегментации нейронных структур в электронной микроскопии. Он

появился в 2012 году и до сих пор служит эталоном для проверки алгоритмов автоматического анализа митохондрий. Данные, пример которых представлен на рисунке 1а, были получены с помощью серийной электронной микроскопии (SEM) и включали аннотированные срезы, взятые из области гиппокампа СА1 мозга мыши, что позволило исследователям тестировать методы машинного обучения на реальных биологических структурах.

Lucchi++ остаётся важным историческим бенчмарком и до сих пор используется в образовательных целях и для отладки новых моделей [3-7].

Размер датасета составляет 1065x2048x1536 вокселей, что при размере вокселя (разрешении) 5x5x5 нм составляет 5325x10240x7680 нм.

Размеченная часть датасета состоит из двух непересекающихся областей, названных как test и train с разрешением по 1024x768x165 вокселей. Для EPFL датасета существуют три версии разметки: первоначальная EPFL [8], улучшенная Lucchi++ [9] (пример на рисунке 1б) и расширенная ITMM [10]. Первые две реализации разметки содержат только разметку митохондрий, расширенная версия содержит меньшее количество слоев, но разметку на 6 классов: митохондрии, мембраны клеток, постсинаптические уплотнения (PSD), везикулы, аксоны, границы митохондрий. Шестиклассовая разметка всё ещё находится в процессе работы и на момент публикации содержит 60 слоев из набора данных train и 10 из test с разрешением по 1024x768.

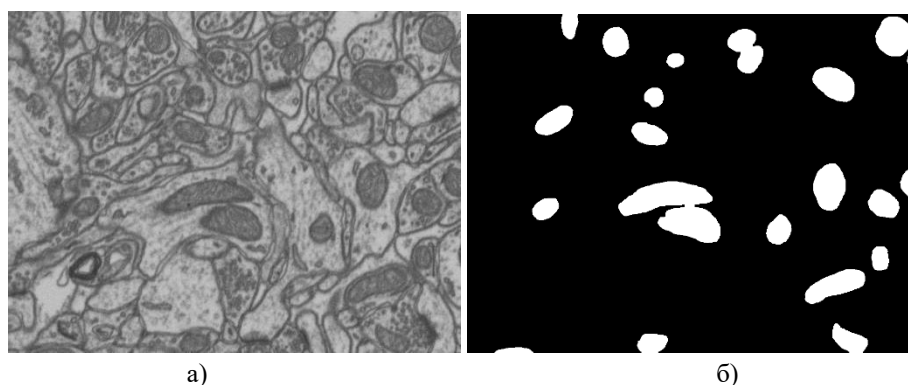


Рисунок 1. Пример данных Lucchi++: а – слой данных; б – разметка митохондрий

**Набор данных Kasthuri** (рис. 2) содержит аннотации митохондрий 3-цилиндрового объема коры мыши [11]. Ткань представляет собой плотный нейропил млекопитающих из слоев 4 и 5 первичной соматосенсорной коры S1, полученный с помощью серийной секционной электронной микроскопии (ssEM).

Датасет состоит из двух стеков размерами 1463x1613x85 и 1334x1553x75 вокселей. Размер вокселя 3x3x30 нм. Данный датасет соответствует ткани 4389x4839x2550 и 4002x4659x2250 нм соответственно.

Было обнаружено две версии разметки, содержащие маски для митохондрий: первоначальная и улучшенная (см. рис. 2б).

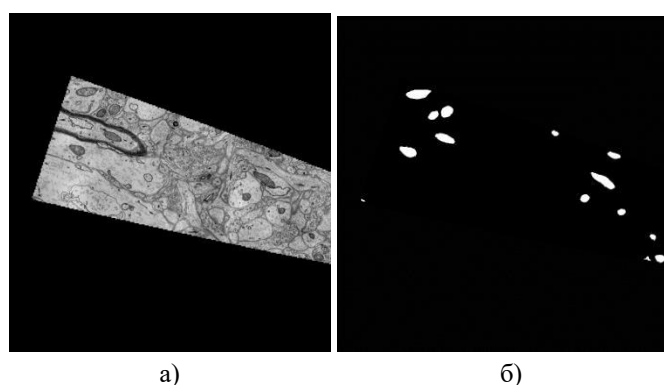


Рисунок 2. Пример данных Kasthuri++: а – слой данных; б – разметка митохондрий



**Набор данных AC4** (рис. 3) был использован в ISBI 2013 EM Challenge [12] в качестве набора для обучения и тестирования. Состоит из стеков, полученных при помощи ЭМ из коры головного мозга мыши (см. рис. 3а). Набор данных состоит из двух стеков, обозначенных как train и test.

Размер стеков составляет  $1024 \times 1024 \times 100$  вокселей, что при размере вокселя (разрешении)  $6 \times 6 \times 30$  нм соответствует двум экземплярам ткани по  $6144 \times 6144 \times 3000$  нм.

Разметка содержит экземплярную (instance) маркировку нейритов (рис. 3б).

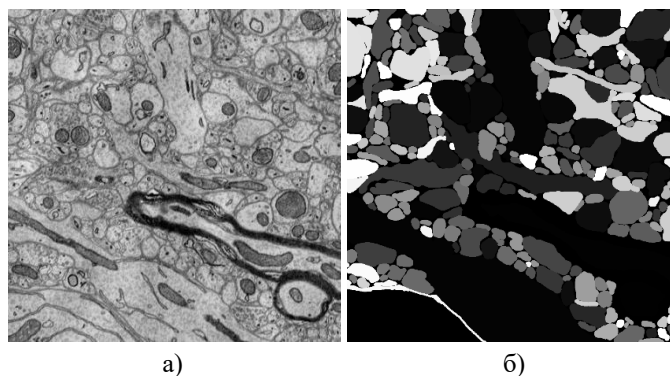


Рисунок 3. Пример данных AC4: а – слой данных; б – разметка клеток после трансформации в полный диапазон 0-255 для визуального восприятия

**Набор данных MitoEM** (рис. 4) представлен в статье [13].

Размер датасета составляет  $512 \times 512 \times 128$  вокселей, что при размере вокселя (разрешении)  $8 \times 8 \times 30$  нм составляет  $4096 \times 4096 \times 3840$  нм.

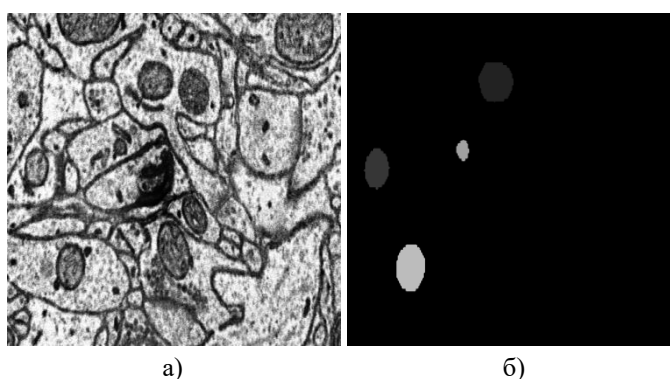


Рисунок 4. Пример данных MitoEM: а – слой данных; б – разметка митохондрий

**Набор данных AxonEM** (рис. 5), представленный в статье [14], состоит из 4 наборов по  $1536 \times 1536 \times 90$  вокселей, что при размере вокселя в  $30 \times 30 \times 30$  нм составляет  $46080 \times 46080 \times 2700$  нм.

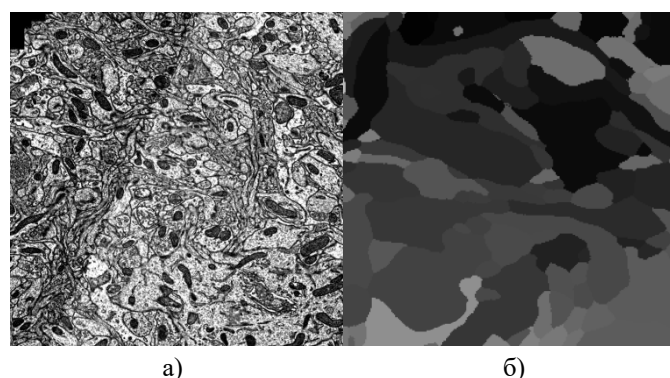


Рисунок 5. Пример данных AxonEM: а – слой данных; б – разметка клеток

**Набор данных CDeep3M-S** (рис. 6) представлен в статье [15].

Размер датасета составляет 1024x1024x21 воксель, что при размере вокселя (разрешении) 2,4x2,4x2,4 нм составляет 2458x2458x50 нм.

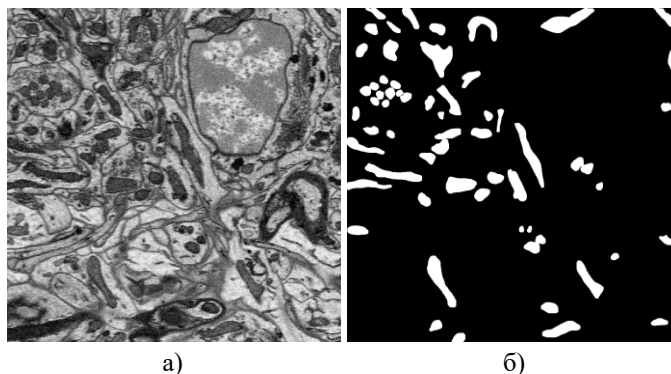


Рисунок 6. Пример данных CDeep3M-S: а – слой данных; б – разметка митохондрий

**Набор данных NucMM** (рис. 7), представленный в статье [16], состоит из 8 размеченных наборов по 192x192x192 вокселя, что при разрешении вокселя 4x4x30 нм составляет 768x768x5760 нм.

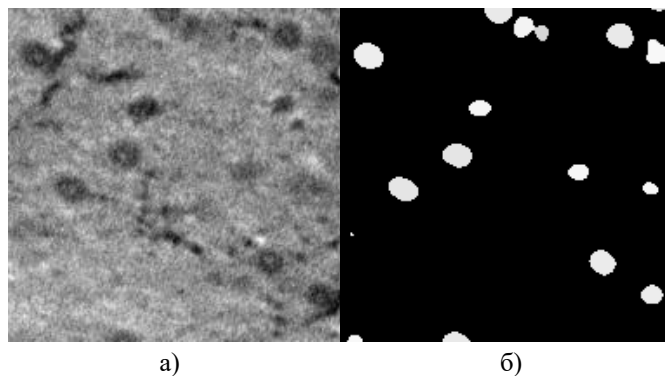


Рисунок 7. Пример данных NucMM: а – слой данных; б – разметка ядер

После фильтрации датасетов с сайта <https://openorganelle.janelia.org/datasets> по снимкам мозга мыши было найдено 16 датасетов: 1-4. Mouse nucleus accumbens; 5-7. Cultured mouse cerebellar granule neurons; 8-9. Mouse dorsal striatum; 10-11. Mouse deep nuclei; 12. Mouse cerebral cortex; 13-15. Mouse hippocampus; 16. Mouse choroid plexus.

**Mouse nucleus accumbens** состоит из ЭМ-снимков ткани прилегающего ядра мыши. Состоит из 4 датасетов.

#### **Датасет Mouse nucleus accumbens jrc\_mus-nacc-1**

Размер датасета составляет 2361x2442x563 вокселя, что при размере вокселя (разрешении) 4x4x4 нм составляет 9444x9768x2252 нм (рис. 8).

Размеченная часть датасета представляет собой фрагмент исходного слоя объемом 300x300x300 вокселей. Разметка состоит из следующих 20 классов, которые можно разбить на три группы: 1) основные клеточные структуры (тело клетки, плазматическая мембрана, внеклеточное пространство); 2) органеллы (митохондрия целиком, внутренний матрикс митохондрии, мембраны митохондрии, рибосомы, ассоциированные с митохондриями, эндоплазматический ретикулум целиком, внутреннее содержимое эндоплазматического ретикулума, мембраны эндоплазматического ретикулума); 3) цитоплазму и включения (цитоплазма, везикулы, содержимое везикул, мембраны везикул, микротрубочки, внутренние участки микротрубочек, внешние участки микротрубочек) (рис. 8б).

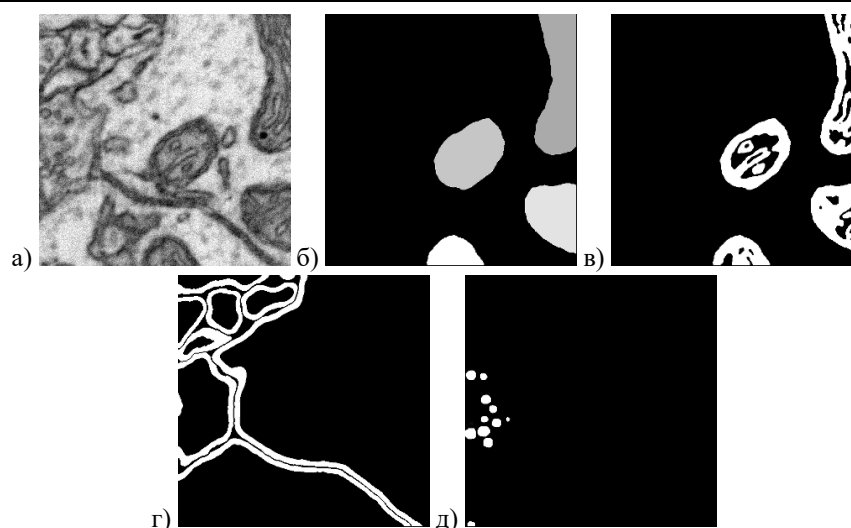


Рисунок 8. Пример данных Mouse nucleus accumbens jrc\_mus-nacc-1: а – слой данных; б – разметки митохондрий; в – разметка границ митохондрий; г – разметка мембран; д – разметка везикул

Исходный датасет имеет объем  $2361 \times 2442 \times 563$  вокселя, в то время как размер предоставленной разметки составляет  $600 \times 600 \times 600$  вокселей. Первоначальное несоответствие в размерах, особенно тот факт, что глубина разметки превышает глубину исходных данных, объясняется разным разрешением вокселей. Это означает, что размеченный кубический участок не соответствует вокселям оригинала по масштабу. Для корректного совмещения требовалось провести пространственное согласование: уменьшить размер разметки в 2 раза по всем осям, преобразовав её до  $300 \times 300 \times 300$  вокселей. Поскольку в данных не было координат размеченной области, её пришлось искать вручную. При анализе первого слоя была обнаружена область, визуально похожая на разметку, но соответствующая размеру  $300 \times 300 \times 300$  вокселей, что и подтвердило необходимость масштабирования и позволило сопоставить разметку с исходным изображением.

#### Датасет Mouse nucleus accumbens jrc\_mus-nacc-2

Размер датасета составляет  $2596 \times 2520 \times 564$  вокселя, что при размере вокселя (разрешении)  $4 \times 4 \times 2,96$  нм составляет  $10384 \times 10080 \times 1669$  нм (рис. 9).

Маркировка содержит набор автоматической разметки (с помощью алгоритма лапласиана Гаусса (LoG)) ядер [17] объёмом  $324 \times 315 \times 70$  вокселей (см. рис. 9 (справа)). Такая же автоматическая разметка применена к датасетам с разметкой ядер (см. ниже).

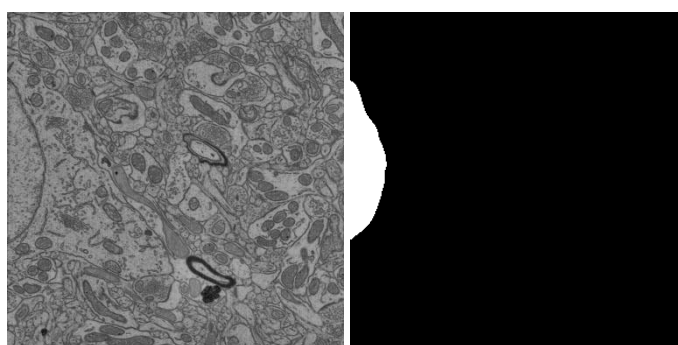


Рисунок 9. Пример данных Mouse nucleus accumbens jrc\_mus-nacc-2 слоя (слева) и разметки ядра (справа)

#### Датасет Mouse nucleus accumbens jrc\_mus-nacc-3

Размер датасета составляет  $5108 \times 5250 \times 732$  вокселя, что при размере вокселя (разрешении)  $4 \times 4 \times 3,83$  нм составляет  $20432 \times 21000 \times 2804$  нм (рис. 10).

Маркировка содержит набор автоматической разметки ядер [17] объёмом 638х656х91 вокселей.

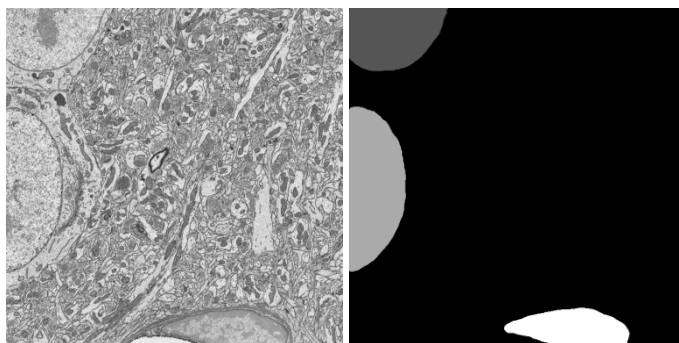


Рисунок 10. Пример данных Mouse nucleus accumbens jrc\_mus-nacc-3 слоя (слева) и разметки ядра (справа)

#### Датасет Mouse nucleus accumbens jrc\_mus-nacc-4

Размер датасета составляет 5083х5123х763 вокселя, что при размере вокселя (разрешении) 4х4х3,99 нм составляет 20332х20492х3044 нм (рис. 11).

Маркировка содержит набор автоматической разметки ядер [17] объёмом 635х640х95 вокселей.

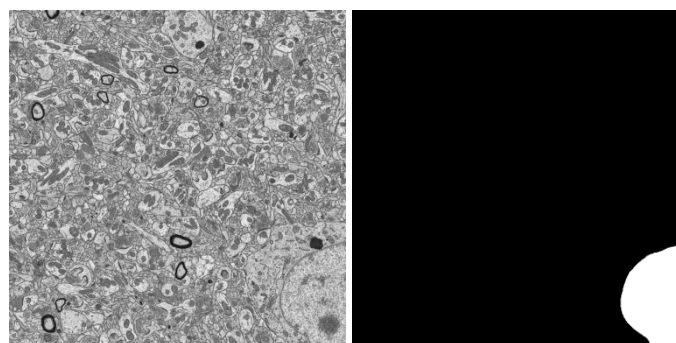


Рисунок 11. Пример данных Mouse nucleus accumbens jrc\_mus-nacc-4 слоя (слева) и разметки ядра (справа)

**Cultured mouse cerebellar granule neurons** состоит из ЭМ-снимков ткани гранулярных нейронов мозжечка мыши, включает 3 датасета.

#### Cultured mouse cerebellar granule neurons jrc\_mus-granule-neurons-1

Размер датасета составляет 13000х1360х4350 вокселей, что при размере вокселя (разрешении) 4х4х3,76 нм составляет 52000х5440х16356 нм (рис. 12).

Маркировка содержит набор автоматической разметки ядер [17] объёмом 812х85х271 воксель.

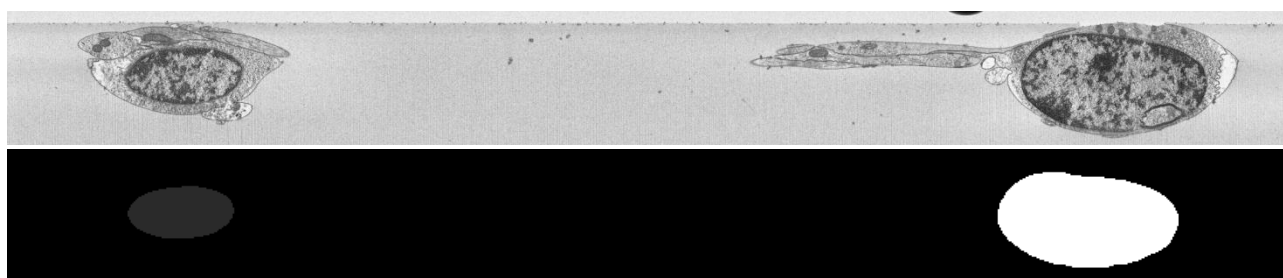


Рисунок 12. Пример данных Cultured mouse cerebellar granule neurons jrc\_mus-granule-neurons-1 слоя (вверху) и разметки ядра (внизу)

### Cultured mouse cerebellar granule neurons jrc\_mus-granule-neurons-2

Размер датасета составляет 2262x537x3946 вокселей, что при размере вокселя (разрешении) 16x16x15,72 нм составляет 36192x8592x62031 нм (рис. 13).

Маркировка содержит набор автоматической разметки ядер [17] объёмом 565x134x986 вокселей.

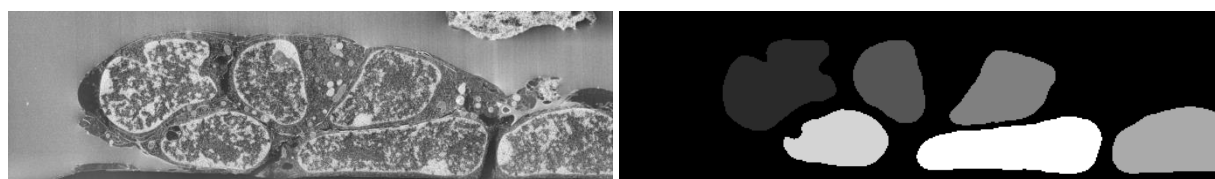


Рисунок 13. Пример данных Cultured mouse cerebellar granule neurons jrc\_mus-granule-neurons-2 слоя (слева) и разметки ядра (справа)

### Cultured mouse cerebellar granule neurons jrc\_mus-granule-neurons-3

Размер датасета составляет 7800x3000x3759 вокселей, что при размере вокселя (разрешении) 4x4x5,37 нм составляет 31200x12000x20186 нм (рис. 14).

Маркировка содержит набор автоматической разметки ядер [17] объёмом 487x187x234 вокселя.

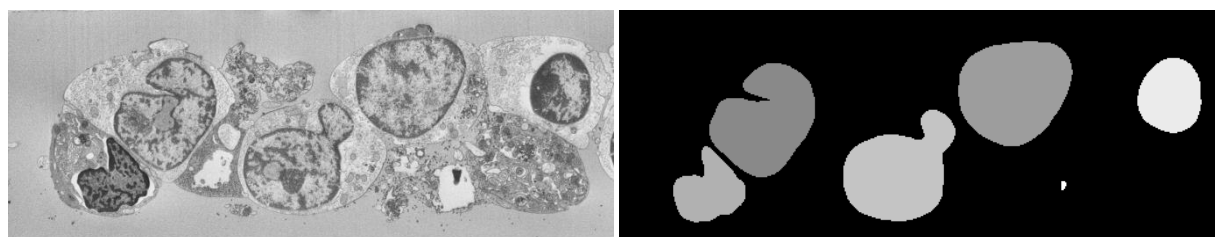


Рисунок 14. Пример данных Cultured mouse cerebellar granule neurons jrc\_mus-granule-neurons-3 слоя (слева) и разметки ядра (справа)

**Mouse dorsal striatum** состоит из снимков ЭМ ткани полосатого тела мыши, включает два датасета.

### Mouse dorsal striatum jrc\_mus-dorsal-striatum

Размер датасета составляет 2554x2564x1353 вокселя, что при размере вокселя (разрешении) 4x4x4,44 нм составляет 10216x10256x6007 нм (рис. 15а). Не содержит разметку.

### Mouse dorsal striatum jrc\_mus-dorsal-striatum-2

Размер датасета составляет 1793x1597x1591 вокселя, что при размере вокселя (разрешении) 4x4x4 нм составляет 7172x6388x6363 нм (рис. 15б). Не содержит разметку.

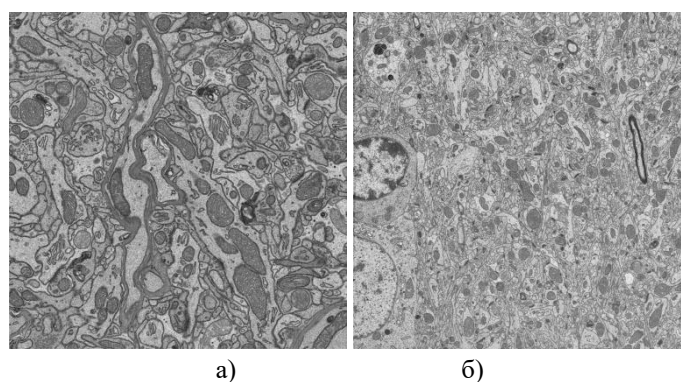


Рисунок 15. Пример данных Mouse dorsal striatum: а – jrc\_mus-dorsal-striatum; б – striatum jrc\_mus-dorsal-striatum-2

**Mouse deep nuclei** состоит из снимков ЭМ-ткани глубоких ядер мозжечка мыши, включает два датасета.

**Mouse deep nuclei jrc\_mus-cerebellum-4**

Размер датасета составляет 6326x6291x1088 вокселей, что при размере вокселя (разрешении) 4x4x4 нм составляет 25304x25164x4352 нм (рис. 16a). Не содержит разметку.

**Mouse deep nuclei jrc\_mus-cerebellum-5**

Размер датасета составляет 1793x1597x1591 воксель, что при размере вокселя (разрешении) 4x4x4 нм составляет 7172x6388x6363 нм (рис. 16б). Не содержит разметку.

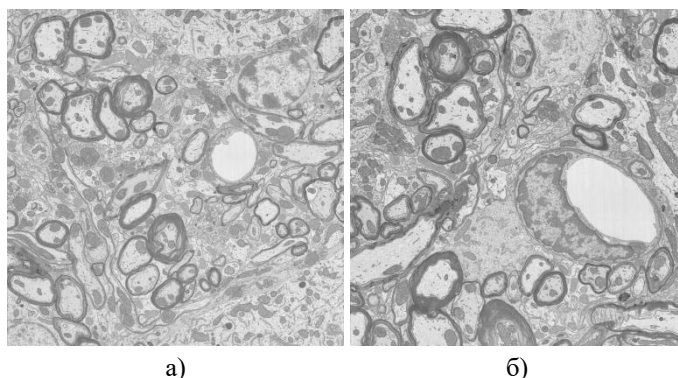


Рисунок 16. Пример данных Mouse deep nuclei: а – jrc\_mus-cerebellum-4; б – jrc\_mus-cerebellum-5

**Mouse cerebral cortex** состоит из снимков ЭМ тканей коры больших полушарий мыши.

Размер датасета составляет 6351x6395x5910 вокселей, что при размере вокселя (разрешении) 4x4x4 нм составляет 25404x25580x23640 нм (рис. 17). Не содержит разметку.

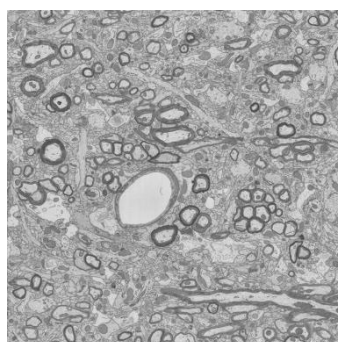


Рисунок 17. Пример данных Mouse cerebral cortex

**Mouse hippocampus** состоит из снимков ЭМ-ткани гиппокампа мыши, включает три датасета.

**Mouse hippocampus jrc\_mus-hippocampus-1**

Размер датасета составляет 17672x17829x24548 вокселей, что при размере вокселя (разрешении) 8x8x8 нм составляет 196384x142632x141376 нм (рис. 18). Содержит автоматическую разметку ядер [17] (рис. 18, справа).

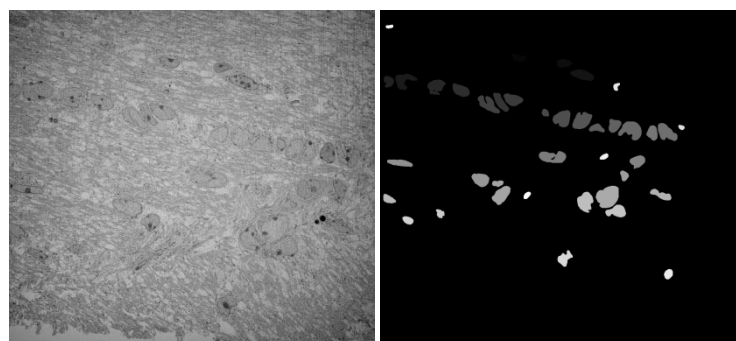


Рисунок 18. Пример данных Mouse hippocampus jrc\_mus-hippocampus-1 (слева) и разметки ядра (справа)

**Mouse hippocampus jrc\_mus-hippocampus-2**

Размер датасета составляет 5622x5612x6882 вокселя, что при размере вокселя (разрешении) 8x8x8 нм составляет 44976x44896x55056 нм (рис. 19a). Не содержит разметку.

**Mouse hippocampus jrc\_mus-hippocampus-3**

Размер датасета составляет 7613x10160x5634 вокселя, что при размере вокселя (разрешении) 8x8x8 нм составляет 60904x81280x45072 нм (рис. 19б). Не содержит разметку.

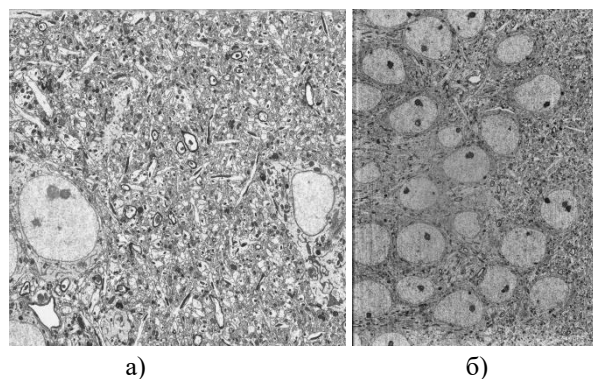


Рисунок 19. Пример данных: а – Mouse hippocampus jrc\_mus-hippocampus-2;  
б – Mouse hippocampus jrc\_mus-hippocampus-3

**Mouse choroid plexus** состоит из снимков ЭМ ткани сосудистого сплетения мыши.

Размер датасета составляет 7500x7500x7560 вокселей, что при размере вокселя (разрешении) 8x8x8 нм составляет 60000x60000x60480 нм (рис. 20). Не содержит разметку.

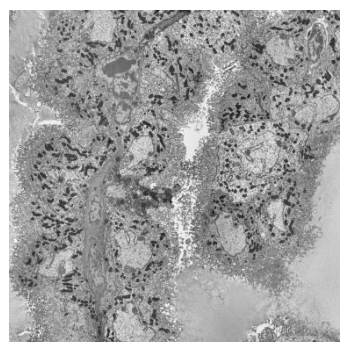


Рисунок 20. Пример данных Mouse choroid plexus

Среди рассмотренных с сайта 16 датасетов изображений тканей мозга мыши ручная разметка доступна только в Mouse nucleus accumbens, охватывающая множество классов, но аннотации покрывают не весь слой, а только определенную область. Остальные датасеты содержат либо автоматическую разметку ядер, либо не имеют разметки. Это ограничивает их применение для задач сегментации клеточных структур.

**Анализ датасетов**

В рамках данного исследования проведен анализ датасетов, включающий оценку их объема и временной динамики формирования. Построен график роста числа размеченных датасетов и размеченных пикселей по годам (рис. 21). За 12 лет развития области в публичный доступ было выложено 9 размеченных наборов данных, менее одного набора данных за год. Рост количества публично выложенной разметки сильно меньше, чем количество выложенных неразмеченных датасетов и количество предложенных архитектур. Присутствует положительная динамика появления новых датасетов, особенно в последние годы, что свидетельствует об актуальности темы исследования мозга при помощи ЭМ (см. рис. 21). Заметим, что на графике присутствуют только те датасеты, которые смогли успешно скачать. Их несколько меньше, чем в обзоре [2], так как часть ссылок в 2025 году не открылись, поэтому исследователям в каждый момент времени доступны не все описанные датасеты. Несмотря на



появляющиеся время от времени инструменты полуавтоматической разметки, получение качественной разметки так и остается сложной задачей.

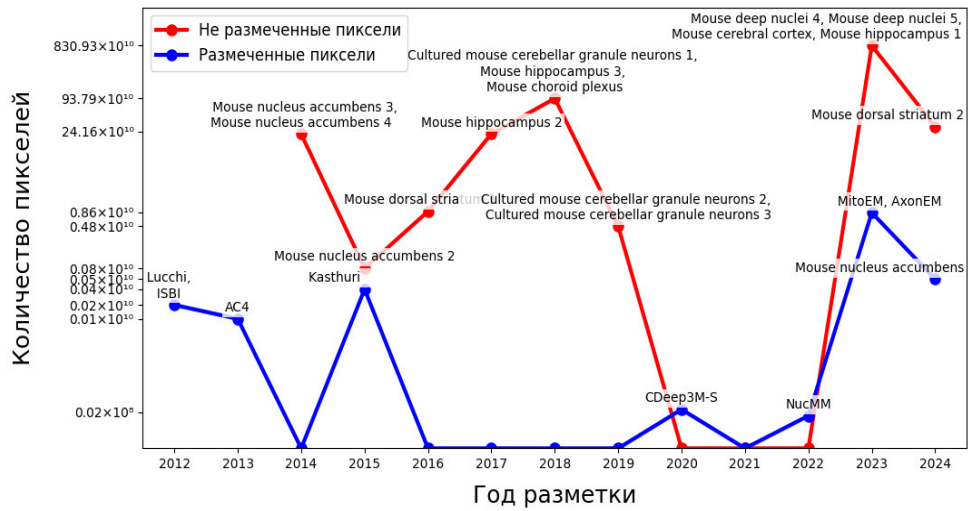


Рисунок 21. График появления размеченных датасетов

В ходе анализа ключевых публичных датасетов для сегментации компартментов мозга на изображениях, полученных с электронного микроскопа, были построены диаграммы распределения пикселей органелл (рис. 22).

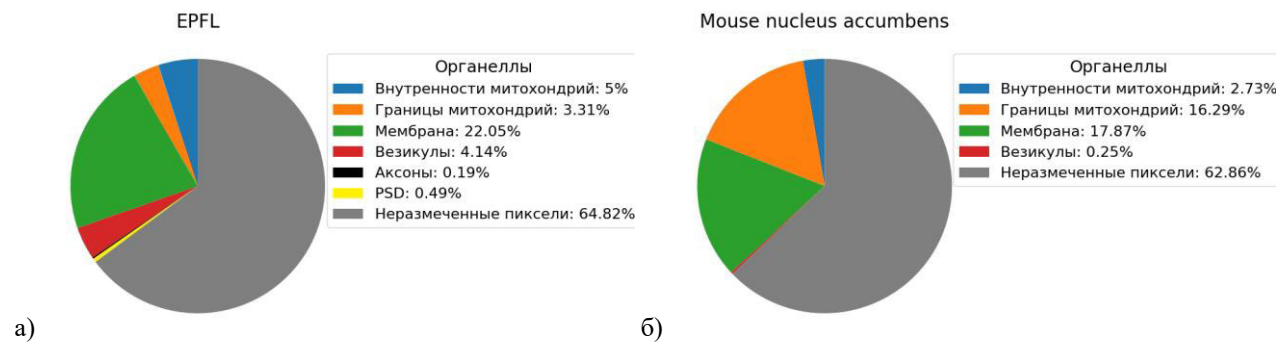


Рисунок 22. Диаграммы, построенные по наборам данных: а – EPFL; б – Mouse nucleus accumbens

Важным аспектом анализа данных является оценка баланса классов. Как видно из представленных диаграмм (рис. 22), для всех проанализированных датасетов наблюдается значительный дисбаланс в распределении пикселей по классам. Например, класс митохондрии на несколько порядков превышает по объему такие классы, как аксоны или PSD. Анализ общего количества размеченных пикселей (табл. 1) также показывает значительную несбалансированность классов.

Таблица 1. Общее количество размеченных пикселей для классов

Класс	Количество размеченных пикселей	Датасеты
Митохондрии	186 932 838	Lucchi, Kasthuri, mnac, CDeep3M-S
Границы митохондрий	5 958 553	ITMM, mnac
Мембраны	15 227 809	ITMM, mnac
Везикулы	2 021 510	ITMM, mnac
Аксоны	88 467	ITMM
PSD	229 141	ITMM
Нейтриты	84 322 554	AC4
Клетки	1 917 180	NucMM

Кроме баланса, ещё одной из ключевых характеристик класса в наборе данных является его цвет (или значения его интенсивности на изображении). Для анализа этой характеристики построены гистограммы датасетов, содержащих несколько классов (рис. 23).

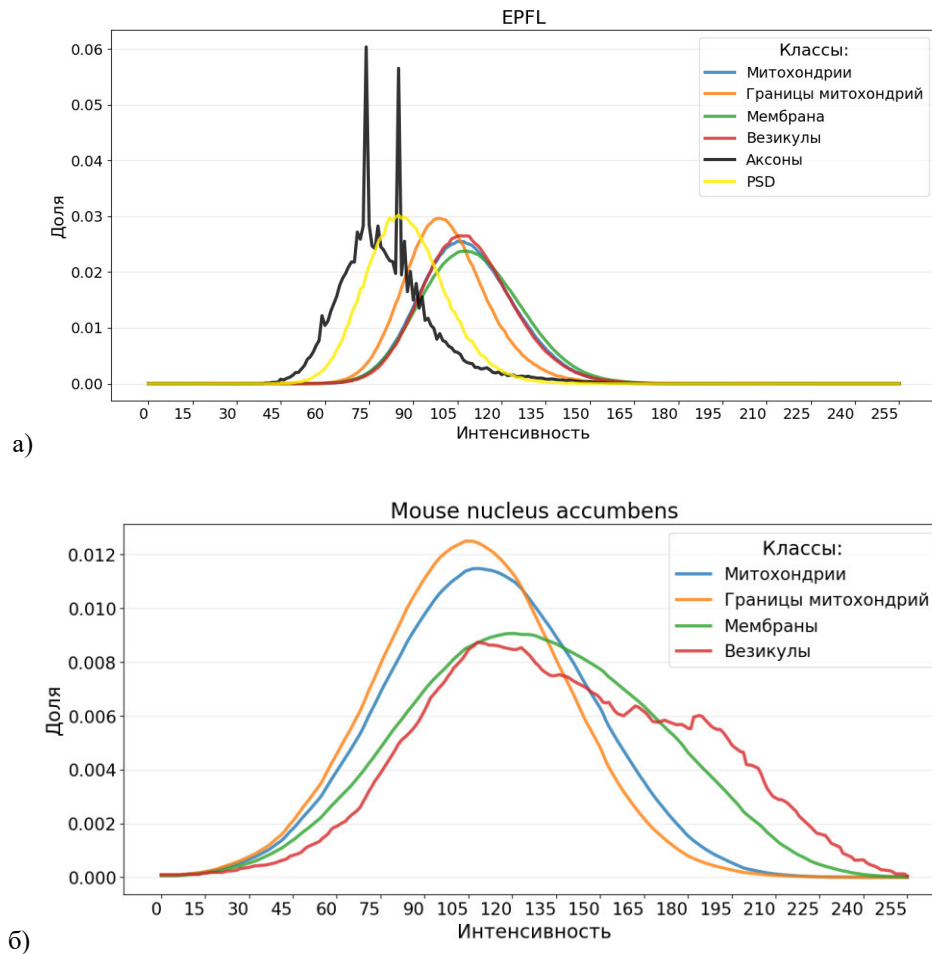


Рисунок 23. Гистограммы, построенные по наборам данных: а – EPFL; б – Mouse nucleus accumbens

Полученные графики имеют визуальное сходство между различными структурами, которое сохраняется между различными датасетами, что не позволяет классифицировать данные по значениям интенсивности. Это указывает на то, что такое распределение является общей характеристикой ЭМ-изображений рассматриваемых биологических тканей и не является существенной характеристикой. Пилообразность некоторых графиков связана с малой представленностью класса в наборе данных.

Схожесть гистограмм одного класса на разных наборах данных мозга мыши показывает наличие единого подхода для получения данных при помощи ЭМ, что дает возможность создать универсальный алгоритм как для сегментации, так и для генерации данных.

### Заключение

Данные – это фундамент, на котором строится глубокое обучение, это стратегический актив, определяющий возможности и ограничения искусственного интеллекта. Ценность данных определяется их объемом, репрезентативностью, чистотой и балансом.

Проведенный анализ публичных наборов данных по электронной микроскопии однозначно подтверждает проблему данной работы: основным препятствием для эффективного и широкого применения нейронных сетей в автоматической сегментации электронно-микроскопических изображений мозга является острая нехватка качественных, объемных и разнообразных размеченных датасетов. Несмотря на значительный прогресс в разработке специализированных архитектур, методов снижения зависимости от разметки и инструментов для работы с данными, все эти решения в конечном

итоге упираются в проблему данных. Трудоемкость и медлительность ручной экспертной разметки, особенно для многоклассовой сегментации, дисбаланс классов, сложность создания реалистичных синтетических данных и отсутствие детально размеченных датасетов для многих структур мозга остаются ключевыми барьерами. Преодоление этой зависимости от данных является критически важной задачей для дальнейшего развития надежных, точных и универсально применимых методов автоматической сегментации в нейробиологии.

#### Источник финансирования

Работа выполнена при поддержке Министерства экономического развития Российской Федерации (соглашение о предоставлении гранта № 139-15-2025-004 от 17 апреля 2025 г., ИГК 000000Ц313925P3X0002).

#### Список литературы

1. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*. 2015: 234-241.
2. Aswath A., et al. Segmentation in large-scale cellular electron microscopy with deep learning: A literature survey. *Medical image analysis*, 2023, 89, 102920.
3. Marcén A. C., Iglesias A., Lapeña R., Pérez F., Cetina C. A Systematic Literature Review of Model-Driven Engineering Using Machine Learning, in *IEEE Transactions on Software Engineering*, Sept. 2024, 50(9), 2269-2293, doi: 10.1109/TSE.2024.3430514.
4. Kazimi B., Ruzaeva K., Sandfeld S. Self-Supervised Learning with Generative Adversarial Networks for Electron Microscopy, 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA, 2024, pp. 71-81, doi: 10.1109/CVPRW63382.2024.00012.
5. Ali M., Benfante V., Basirinia G., Alongi P., Sperandeo A., Quattrocchi A., Giannone A.G., Cabibi D., Yezzi A., Di Raimondo D., et al. Applications of Artificial Intelligence, Deep Learning, and Machine Learning to Support the Analysis of Microscopic Images of Cells and Tissues. *J. Imaging*, 2025, 11, 59. <https://doi.org/10.3390/jimaging11020059>
6. Lun Che, Zhongping He, Kaiyuan Zheng, Tianyu Si, Meiling Ge, Hong Cheng, Lingrong Zeng. Deep learning in alloy material microstructures: Application and prospects, *Materials Today Communications*, 2023, 37, 107531, ISSN 2352-4928, <https://doi.org/10.1016/j.mtcomm.2023.107531>.
7. Azhagu Madhavan Sivalingam, Darshitha D. Sureshkumar, Vijayalakshmi Pandurangan. Cerebellar pathology in forensic and clinical neuroscience, *Ageing Research Reviews*, 2025, 106, 102697, ISSN 1568-1637, <https://doi.org/10.1016/j.arr.2025.102697>.
8. Lucchi A., Smith K., Achanta R., Knott G., Fua P. Electron microscopy dataset. 2023. <https://www.epfl.ch/labs/cvlab/data/data-em/>.
9. Casser V., Kang K., Pfister H., Haehn D. Fast mitochondria detection for connectomics, *Proceedings of Machine Learning Research*, 2020, 1, 10.
10. Getmanskaya A.A., Sokolov N.A. Turlapov V.E. Multiclass U-Net Segmentation of Brain Electron Microscopy Data Using Original and Semi-Synthetic Training Datasets. *Program Comput Soft*, 2022, 48, 164–171.
11. Kasthuri N., Hayworth K. J., Berger D. R., et al. Saturated reconstruction of a volume of neocortex, *Cell*, 2015, 162(3), 648–661.
12. Arganda-Carreras I., Turaga S. C., Berger D. R., et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in Neuroanatomy*, 2015, 9, 142.
13. Wei Donglai, Lin Zudi, Franco-Barranco Daniel, et al. MitoEM Dataset: Large-Scale 3D Mitochondria Instance Segmentation from EM Images. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020: 66-76.
14. Wei D. et al. AxonEM Dataset: 3D Axon Instance Segmentation of Brain Cortical Regions. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. MICCAI 2021. *Lecture Notes in Computer Science*, 12901. Springer, Cham, 2021. [https://doi.org/10.1007/978-3-030-87193-2\\_17](https://doi.org/10.1007/978-3-030-87193-2_17)
15. Haberl M.G., Churas C., Tindall L., et al. CDeep3M—Plug-and-Play cloud-based deep learning for image segmentation. *Nature Methods*, 2018, 15 (9), 677-680.
16. Lin Z., Wei D., Petkova M.D., et al. NucMM dataset: 3d neuronal nuclei instance segmentation at sub-cubic millimeter scale. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, 164-174.
17. Xu C.S., Pang S., Shtengel G. et al. An open-access volume electron microscopy atlas of whole cells and tissues. *Nature*, 2021, 599, 147–151. <https://doi.org/10.1038/s41586-021-03992-4>

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-828-834

## Влияние объема синтетически сгенерированных изображений в наборе данных для задачи сегментации аксонов в электронной микроскопии мозга

А. И. Миронов, Н. А. Соколов, А. А. Серебрякова

Исследовательский центр в области искусственного интеллекта, Нижегородский государственный университет им. Н. И. Лобачевского, Нижний Новгород, Россия

**Аннотация.** В данной работе исследуется влияние объема синтетических данных на качество сегментации изображений электронной микроскопии мозга. В качестве исходного набора данных использованы широко применяемый набор EPFL и разметка ITMM на шесть классов. Для генерации синтетических данных использовались геометрические алгоритмы, а для решения задачи сегментации применена модификация модели U-Net. Основное внимание уделено анализу зависимости точности сегментации аксонов и постсинаптических уплотнений от соотношения реальных и синтетических данных. Применение синтетических данных продемонстрировало значительное улучшение качества сегментации при ограниченном объеме реальных данных.

**Ключевые слова:** синтетические данные, медицинские данные, сегментация, электронная микроскопия

## The effect of the volume of synthetically generated images in the dataset for the task of axon segmentation in brain electron microscopy

A. I. Mironov, N. A. Sokolov, A. A. Serebryakova

Research Center for Artificial Intelligence, Lobachevsky State University of Nizhny Novgorod, Russia

**Abstract.** This paper examines the impact of synthetic data volume on the quality of brain electron microscopy image segmentation. The widely used EPFL dataset and the six-class ITMM labeling were used as the initial dataset. Geometric algorithms were used to generate the synthetic data, and a modification of the U-Net model was applied to solve the segmentation problem. The primary focus is on analyzing the dependence of axon and postsynaptic densification accuracy on the ratio of real to synthetic data. The use of synthetic data demonstrated a significant improvement in segmentation quality with a limited real data volume.

**Keywords:** Synthetic data, Medical data, Segmentation, Electron microscopy, U-Net, Axon

### Введение

Современные методы компьютерного зрения, особенно в области сегментации изображений, значительно продвинулись благодаря использованию глубоких нейронных сетей. Семантическая сегментация — это задача компьютерного зрения, в которой каждому пикселю изображения присваивается метка, соответствующая классу объекта, что позволяет выделить и идентифицировать различные объекты и области на изображении. Однако успешное применение моделей часто ограничивается доступностью размеченных данных, особенно в специализированных областях, таких как медицина.

В последние годы синтетические данные стали важным инструментом для преодоления этих ограничений. Их генерация позволяет создавать разнообразные и реалистичные примеры, которые могут использоваться для обучения нейронных сетей [1, 3]. Однако влияние объема синтетических данных на качество сегментации остается недостаточно изученным. Недостаточное количество данных может привести к низкой точности сегментации, в то время как избыток некачественной синтетики приведет к снижению качества сегментации на реальных данных. В контексте сложных медицинских изображений это особенно актуально, так как нет генераторов синтетики, точно воспроизводящей настоящие данные. Также в медицинских изображениях, как правило, существует сильный дисбаланс классов [4], что может приводить к смещению прогнозируемых вероятностей в сторону доминирующих классов и низкой точности распознавания редких объектов. Улучшение распознавания редких объектов может быть достигнуто при помощи добавления синтетических изображений в обучающий набор данных.

Данная работа фокусируется на применении синтетических данных для сегментации изображений электронной микроскопии мозга. Особое внимание уделяется анализу зависимости качества

сегментации аксонов и постсинаптических уплотнений от соотношения реальных и синтетических данных. Аксоны в датасете EPFL 5 являются самым редким классом, из-за чего их классификация становится невозможной без синтетических данных.

Мы ставим эксперименты, чтобы определить оптимальный баланс, обеспечивающий наибольшую точность сегментации при добавлении минимальных объемов синтетических данных.

## Постановка задачи

### 1. Датасет

В исследовании используется Electron Microscopy Dataset, предоставленный исследовательским университетом Швейцарии EPFL 5 вместе с разметкой ITMM 6. Датасет представляет участок размером  $5 \times 5 \times 5$  микрометров, взятый из области гиппокампа CA1 мозга, что соответствует объёму  $1065 \times 2048 \times 1536$ . Все три доступные разметки датасета работают с двумя выделенными частями разрешением  $1024 \times 768 \times 165$ . Мы используем разметку ITMM 6, в которой, несмотря на малое количество размеченных слоев (70 из 330), они размечены на шесть классов, в отличие от двух других доступных разметок для одного класса митохондрий.

Для обучения модели было выбрано 10 слоёв, разбитых на тайлы размером  $256 \times 256$ . Каждый слой разбивался при помощи окна  $256 \times 256$  и смещения 128 пикселей. На изображениях с аннотациями определенного класса наличие объекта отмечается белым цветом, а отсутствие – черным. Пример тайла представлен на рисунке 1.

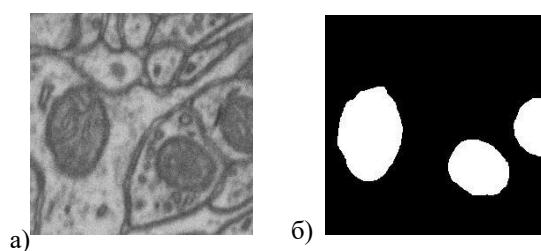


Рисунок 1. Пример тайла данных из датасета "EPFL Electron Microscopy Dataset": а – входной тайл  $256 \times 256$ ; б – размеченная маска митохондрий

Для тестирования использовались 5 слоёв, которые в процессе тестирования также разбиваются на тайлы со смещением, а потом предсказание собирается в бинарную маску размера исходного слоя.

### 2. Синтетические данные

Для генерации синтетических данных электронной микроскопии использован параметризуемый алгоритм, основанный на геометрических правилах. Формы органелл моделируются с использованием случайной выборки параметров, таких как размер, положение и ориентация, что позволяет эффективно создавать разнообразные и в то же время контролируемые изображения.

Процесс генерации включает четыре этапа:

- 1) синтез органелл (аксоны, митохондрии, PSD (postsynaptic density, постсинаптическое уплотнение), везикулы);
- 2) размещение объектов на изображении с учётом отсутствия пересечений;
- 3) генерация мембран;
- 4) добавления размытия и шума.

Синтетические данные генерировались с конфигурациями, включающими следующее количество объектов: митохондрии — 0, а количество аксонов, PSD и областей везикул может варьироваться до 3 на одном изображении. Эти объекты были случайным образом распределены на изображениях размером  $256 \times 256$ . Более подробное описание алгоритма описано в статье [7]. Генератор изображений доступен в репозитории GitHub 8. Преимуществом такого подхода является наличие точной разметки сразу для всех классов в виде масок, а также эффективность генерации с минимальными вычислительными затратами, по сравнению, например, с диффузионными нейросетями. Генерация одного геометрического тайла занимает примерно 3 секунды на центральном процессоре i7-13700k

3400MHz, тогда как генерация тайла с использованием диффузионной модели занимает 15 секунд с использованием NVIDIA A100. На рисунке 2 приведен пример сгенерированного изображения.

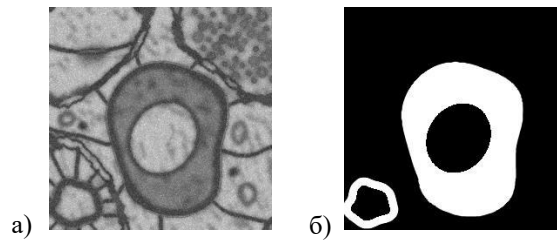


Рисунок 2. Пример синтетических данных: а – сгенерированное изображение; б – сгенерированная маска

### 3. Tiny U-Net

U-Net 9 — это архитектура нейронной сети, изначально предложенная для задач сегментации медицинских изображений, в частности полученных с помощью микроскопии. Разработанная в 2015 году для сегментации биомедицинских изображений архитектура зарекомендовала себя [10] как одна из самых эффективных для обработки изображений с небольшим количеством размеченных данных.

В работе используется модификация классической архитектуры — Tiny U-Net [11]. Модификация имеет меньшее количество параметров, что уменьшает вычислительные затраты как на этапе обучения, так и на этапе инференса. Кроме того, Tiny U-Net лучше подходит для задач, где количество обучающих данных сильно ограничено. Используемая реализация нейронной сети находится в репозитории GitHub [12].

### 4. Оценка качества

Для оценки качества сегментации использовался коэффициент Dice-Score (DSC, Dice), широко применяемый в медицинской сегментации изображений. Значения метрики варьируются от 0 до 1. Пусть TP — количество пикселей, правильно классифицированных как принадлежащие целевому классу (True Positive); TN — правильно классифицированные фоны (True Negative); FP — ошибочно отнесённые к целевому классу (False Positive); FN — ошибочно отнесённые к фону (False Negative). Тогда метрика Dice определяется следующим образом:

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (1)$$

Так как в работе рассматривается многоклассовая сегментация, применяется векторная форма метрики Dice — отдельное значение  $DSC_i$  для каждого класса. Чтобы использовать её как функцию потерь при обучении нейронной сети, необходимо агрегировать её в скалярную величину. Для этого применяется линейная комбинация

$$Loss = \sum_{i=1}^N \alpha_i (1 - DSC_i), \alpha_i \geq 0, \sum_{i=1}^N \alpha_i = 1, \quad (2)$$

где  $Loss$  — итоговая скалярная функция потерь;  $N$  — количество классов;  $\alpha_i$  — весовой коэффициент для  $i$ -го класса, выбранный равным  $\frac{1}{N}$ .

## Эксперименты

Для определения оптимального количества синтетических изображений, необходимого для достижения высокого качества сегментации, были выполнены следующие шаги.

1. *Подготовка данных.* Для обучения модели было выбрано 10 слоёв. Каждый слой разбили на 1170 тайлов размерностями 256x256 пикселей. Разбиение проводилось для каждого слоя с шагом 128 пикселей, в результате итоговые изображения перекрываются. Входные данные содержат вручную размеченные маски для каждого класса.

2. *Генерация синтетических данных.* С использованием геометрического генератора создано 300 синтетических изображений размером 256x256.

3. *Формирование обучающих данных.* Обучающие наборы формировались с числом настоящих данных 0, 585 и 1170 тайлов (соответствует 0, 5 и 10 изображениям). В каждый настоящий набор добавлялись синтетические изображения в объёме от 0 до 300 тайлов с шагом 30.

4. *Подсчёт органелл.* Для каждого тайла определялись количество органелл и их суммарная площадь как в синтетических, так и в реальных данных. Количество пикселей конкретного объекта считалось по маскам.

5. *Обучение модели.* Модель Tiny U-Net обучалась на каждом из наборов 5 раз с различным значением генератора случайных чисел (seed) для устойчивости оценки.

6. *Тестирование модели.* После обучения каждая модель тестировалась на выделенной тестовой выборке, качество сегментации оценивалось с помощью Dice-Score. Метрики усреднялись.

Проведено два эксперимента. Они различаются в способе генерации синтетических данных. В первом эксперименте синтетические данные содержали только аксоны, во втором – аксоны, постсинаптические уплотнения и везикулы. Выбор именно таких классов для создания геометрической синтетики обусловлен стартовыми условиями исходного набора данных EPFL и особенностями расположения компартментов относительно друг друга. В тренировочном наборе EPFL аксон представлен на 36 слоях и совершенно не похож на аксон в тестовом наборе данных. В этом случае мы ограничены в использовании других методов расширения датасета, таких как аугментация и обучаемые нейросетевые модели, в том числе диффузионные [13]. В этом же наборе данных интенсивность PSD пересекается с интенсивностью миелиновой оболочки аксона, поэтому нейросети сложно различать эти два класса. Класс везикулы появляется в синтетических данных благодаря тому, что в реальных биологических структурах скопления везикул располагаются рядом с PSD.

### Результаты экспериментов

Результаты первого эксперимента на рисунках 3 и 4 показывают положительную зависимость качества сегментации от количества добавленных синтетических изображений. Красный цвет отображает эксперимент, в котором использовалось 5 слоёв оригинальных данных, зеленый – 10 слоёв соответственно. На графиках важно выделить начальные точки, где синтетические данные ещё не использовались. Отмечается существенный рост качества сегментации при добавлении 90 синтетических тайлов, что составляет около 15 % от общего числа данных в наборе, содержащем 5 реальных слоёв, и 7,5 % – в наборе, содержащем 10 реальных слоёв (рис. 3). Эти точки соответствуют количеству аксонов 177 и 193 (165 синтетических аксонов) соответственно (рис. 4). Наилучшее качество наблюдается при увеличении количества синтетических изображений до 210 (рис. 3). Это указывает на то, что увеличение количества данных улучшает способность модели различать аксоны на изображениях. После добавления более 210 синтетических изображений не наблюдается повышение качества.

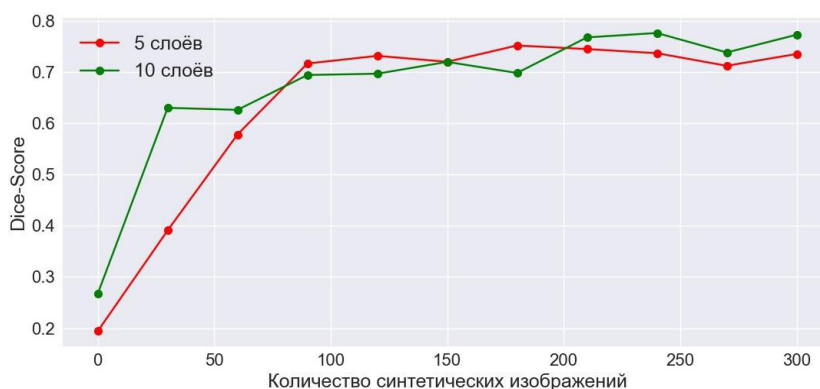


Рисунок 3. Зависимость качества сегментации от количества синтетических изображений



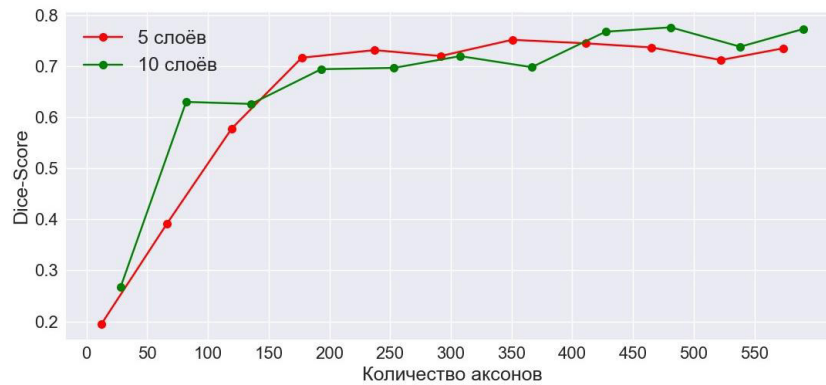


Рисунок 4. Зависимость качества сегментации от количества аксонов

Во время проведения эксперимента оказалось, что модель начинает путать постсинаптические уплотнения и аксоны, что заметно на рисунке 6. Это может происходить из-за отсутствия PSD в синтетических изображениях, поэтому во втором эксперименте они были добавлены при генерации.

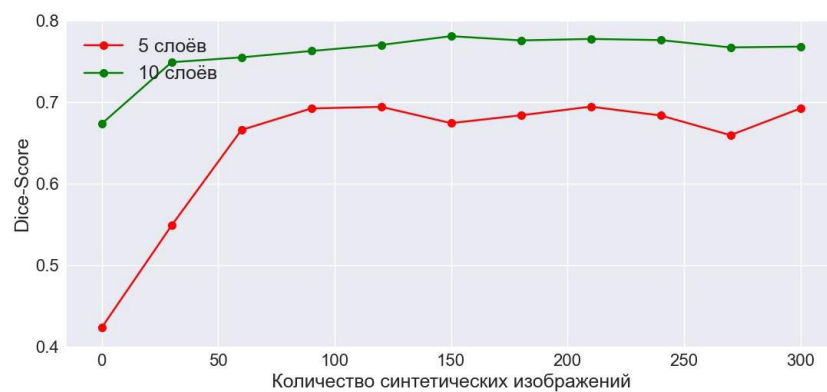


Рисунок 5. Зависимость качества сегментации PSD от количества добавленных данных с синтетическими аксонами

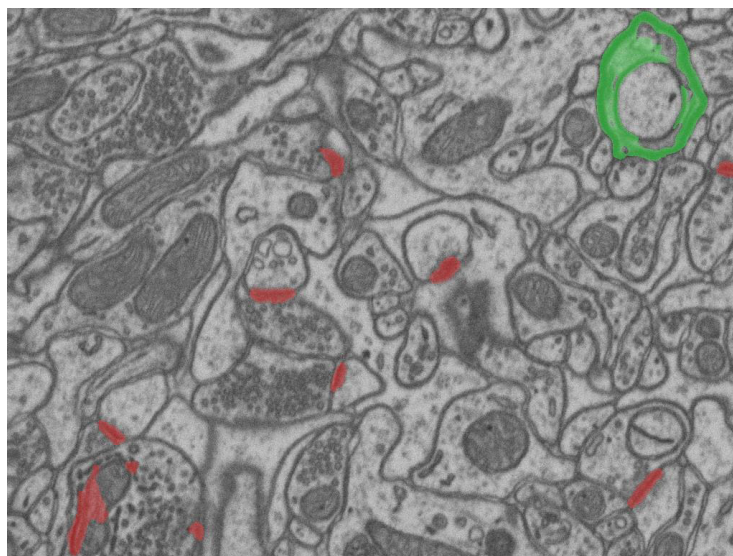


Рисунок 6. Маска сегментации аксонов. Зеленое наложение совпадает с разметкой, красное – ложная разметка

Второй эксперимент, изображенный на рисунках 7 и 8, также показывает положительную зависимость качества сегментации от количества добавленных синтетических изображений. Синий цвет отображает эксперимент, в котором использовались только синтетические данные. Красный цвет

отображает эксперимент, в котором использовались 5 слоёв оригинальных данных, остальные – синтетические, зеленый цвет – 10 слоёв оригинальных данных соответственно. На графиках важно выделить начальные точки, где синтетические данные ещё не использовались. Результаты довольно схожи с первым экспериментом: при добавлении 180 синтетических аксонов и больше качество сегментации Dice становится больше 0,74, но в отличие от первого эксперимента рост качества не прекращается на 0,78, а достигает 0,795 для красного графика при количестве аксонов 300 (рис. 8), что вдвое меньше, чем для первого эксперимента.



Рисунок 7. Зависимость качества сегментации от количества синтетических данных

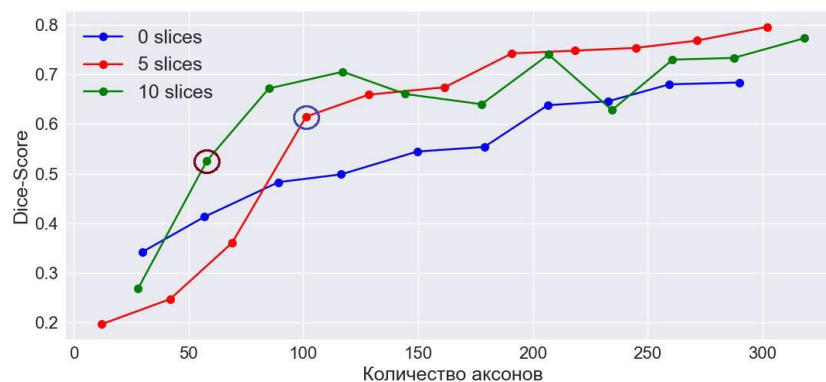


Рисунок 8. Зависимость качества сегментации от количества аксонов

На рисунке 9 видно, что качество сегментации PSD также растет и довольно быстро достигает уровня в 75 % в эксперименте с 10 оригинальными изображениями и 72 % с 5 оригинальными изображениями. В первом эксперименте с 5 оригинальными изображениями качество сегментации PSD меньше 70 %, однако с 10 изображениями качество сегментации достигает 77 %. Добавление PSD в синтетические данные не несет изменений в качестве их сегментации, однако аксоны определяются увереннее.

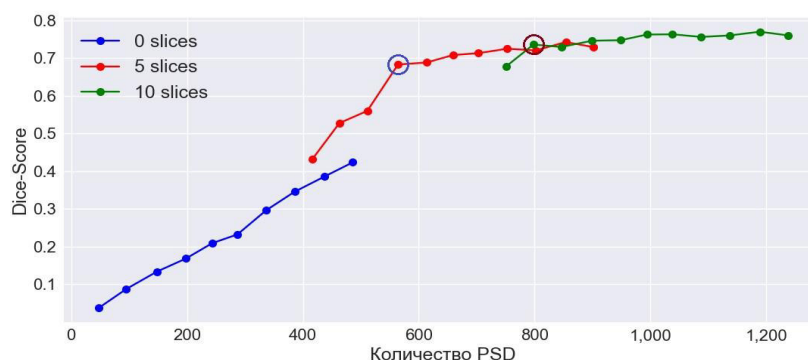


Рисунок 9. Зависимость качества сегментации PSD от их количества

По результатам экспериментов заметно, что для обучения нейронной сети с синтетическими данными очень важно добавлять настоящие данные или генерировать синтетику, неотличимую от настоящих данных. Это позволяет улучшить качество сегментации. Синтетические данные, в свою очередь, могут быть полезны для увеличения объема обучающего набора и создания разнообразия, что особенно важно в случаях, когда реальные данные ограничены или труднодоступны.

### Заключение

В рамках исследования на малой части датасета электронной микроскопии было проанализировано влияние добавленного объема синтетических данных на качество сегментации, редко встречаемого в исходном наборе данных класса аксон. Полученные результаты показали, что использование синтетических данных может существенно повысить качество сегментации. При сравнении с исследованием [13], где использовались 42 слоя из того же датасета и 1000 синтетических тайлов, сгенерированных аналогичным способом и включающих все 5–6 классов, различия в качестве сегментации аксонов оказались незначительными. Это связано с сопоставимым количеством аксонов в итоговых выборках, на 1000 тайлов генерации приходится примерно 475 аксонов.

Несбалансированные наборы данных создают сложности в предсказаниях как для редких классов, так и для классов, схожих с ними, что приводит к неопределенности в предсказаниях. Мы продемонстрировали, что добавление простых синтетических изображений для редких классов существенно улучшает качество сегментации. Это особенно выражено при небольшом дополнении как для основного редкого класса аксон, так и для схожего по интенсивности класса PSD, что позволяет снизить путаницу между ними.

### Источник финансирования

Работа выполнена при поддержке Министерства экономического развития Российской Федерации (соглашение о предоставлении гранта № 139-15-2025-004 от 17 апреля 2025 г., ИГК 000000Ц313925P3X0002).

### Список литературы

1. Sergey I. Nikolenko. Synthetic Data for Deep Learning (2019). ArXiv preprint arXiv:1909.11512
2. Jonas Rabensteiner, Cynthia I. Ugwu, Oswald Lanz. Improving Semantic Segmentation Models through Synthetic Data Generation via Diffusion Models (2024). ICLR 2024. Workshop DMLR.
3. Daniel Saragih, Atsuhiko Hibi MSc, Pascal N. Tyrrell PhD (2024). Using Diffusion Models to Generate Synthetic Labeled Data for Medical Image Segmentation. arXiv preprint arXiv:2310.16794
4. V. P. Agostina J. Larrazabal, Nicol'as Nieto. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. PNAS, 2020.
5. EPFL Computer Vision Lab. Electron Microscopy Dataset.
6. ITMM, 6-Class Labels for EPFL EM Dataset (2023). GitHub repository. <https://github.com/GraphLabEMproj>.
7. N.A. Sokolov, E.P. Vasiliev, A.A. Getmanskaya. Generation and study of the synthetic brain electron microscopy dataset for segmentation purpose. Computer Optics, 2023.
8. GraphLab EM Project (2021). Synthetics: Synthetic Data Generator for EM Segmentation, GitHub repository <https://github.com/GraphLabEMproj/Synthetics>
9. Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation (2015). ArXiv preprint arXiv:1505.04597.
10. Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval. Medical Image Segmentation Review: The Success of U-Net. arXiv preprint arXiv:2211.14830 (2022).
11. A. A. Getmanskaya, Nikolai A. Sokolov, V. E. Turlapov. Multiclass U-Net Segmentation of Brain Electron Microscopy Data Using Original and Semi-Synthetic Training Datasets. Programming and Computer Software 48 (2022) 164–171.
12. Nikolai Sokolov. UnetClass: GitHub repository (2024). <https://github.com/NikolaySokolov152/UnetClass>
13. Nikolay Sokolov, Alexandra Getmanskaya, Vadim Turlapov. AI Diffusion Model-Based Technology for Automating the Multi-Class Labeling of Electron Microscopy Datasets of Brain Cell Organelles for Their Augmentation and Synthetic Generation. Technologies, 2025.

## Изучение нескольких ResNet-подобных моделей для классификации атипичных митотических фигур

Александр Трыкин<sup>1,2</sup>, Евгений Васильев<sup>1</sup>, Вадим Турлапов<sup>1</sup>

<sup>1</sup>Исследовательский центр искусственного интеллекта, Институт информационных технологий, математики и механики, Нижний Новгород, Россия

<sup>2</sup>Лаборатория системной медицины старения, Научно-исследовательский институт биологии старения, Нижний Новгород, Россия

**Аннотация.** Атипичные изменения в клеточном делении могут указывать на опасные патологические состояния, связанные с высокоагрессивными формами злокачественных новообразований. Идентификация митотических фигур — субъективная и трудоёмкая задача в гистопатологии. В данной работе представлены ResNet-подобные модели для решения задачи классификации нормальных/атипичных митотических фигур. Были взяты два общедоступных набора данных AMi-Br и MIDOG25, посвящённых этой относительно новой проблеме, и объединены, чтобы максимально использовать имеющиеся данные и охватить различные условия. Рассмотрено использование нескольких функций потерь, таких как Focal Loss (FL), Recall Loss (RL) и Focal Recall Loss (FRL). Наилучшие значения сбалансированной точности 84.93 % и AUROC 92.12 % продемонстрировала модель ResNetV2-34D с Recall Loss, F1 76,81 % показала ResNeSt-26D также с Recall Loss.

**Ключевые слова:** атипичный митоз, гистопатология, глубокое обучение, несбалансированное обучение

## Exploring several ResNet-like models for atypical mitotic figure classification

Alexander Trykin<sup>1,2</sup>, Evgenii Vasiliev<sup>1</sup> and Vadim Turlapov<sup>1</sup>

<sup>1</sup>Artificial Intelligence Research Center, Institute of Information Technologies, Mathematics and Mechanics, Lobachevsky State University, Nizhny Novgorod, Russia

<sup>2</sup>Laboratory of Systems Medicine of Ageing, Research Institute for Biology of Aging, Lobachevsky State University, Nizhny Novgorod, Russia

**Abstract.** Atypical changes in cell division may indicate dangerous pathological conditions associated with highly aggressive forms of malignant neoplasms. Identification of mitotic figures is a subjective and labour-intensive task in histopathology. In this paper, we present ResNet-like models for solving the problem of normal/atypical mitotic figure classification. We took two publicly available datasets devoted to this relatively new problem: AMi-Br and MIDOG25, and combined them to use as much available data as possible and cover different conditions. We have looked at the use of several loss functions, such as Focal Loss (FL), Recall Loss (RL), and Focal Recall Loss (FRL). We were able to achieve the best values of balanced accuracy of 84.93% and AUROC of 92.12% was shown by the ResNetV2-34D model with Recall Loss, and F1 of 76.81% was ResNeSt-26D also with Recall Loss.

**Keywords:** atypical mitosis, histopathology, deep learning, imbalanced learning

### Introduction

Cell division is a fundamental biological process, ensuring the growth, development, and maintenance of multicellular organisms. A critical aspect of this process is mitosis, where genetic material is equally distributed between daughter cells. Under normal conditions, mitotic figures exhibit a standardized morphology, reflecting the highly regulated orchestration of chromosomal and cellular events. However, in certain pathological states, such as cancer, mitotic figures can deviate from their typical appearance, giving rise to what are commonly referred to as atypical mitotic figures (AMF's). These anomalies often signal chromosomal instability, a hallmark of tumorigenesis, and are associated with highly aggressive malignancies, poor prognoses, and resistance to therapeutic interventions [1, 2, 3].

While traditional histopathology has played a significant role in identifying AMF's, the process is often limited by subjectivity and requires significant expertise. Recent advancements in imaging, artificial intelligence (AI), and computational pathology have opened new pathways for automated, precise, and reproducible classification of atypical mitotic figures [4, 5, 6].

The recent Mitosis Domain Generalization Challenge 2025 (MIDOG25) [7] provides a standardized benchmark for creating accurate and reliable solutions, calling on researchers to contribute to the development of AI cancer assessment tools to assist pathologists. The authors of the benchmark note several key issues: (1)

severe class imbalance, (2) high morphological variability within each class, (3) little difference between normal and atypical figures, and (4) a wide range of different tumours, types, staining methods, and scan areas.

Although the challenge is relatively new, work on this topic is just beginning to emerge. The authors of the original Ami-Br dataset presented results for their dataset [8]. They tried to deal with imbalance using various techniques such as weighted cross-entropy, focal loss, and weighted sampling. They tried two deep learning models: DenseNet-121 and EfficientNet V2-S. They report achieving an average balanced accuracy of around 80.6%. In [9] authors presented two new datasets: AtNorM-Br, a mitotic figure dataset from the TCGA breast cancer cohort, and AtNorM-MD, a multi-domain mitotic figure dataset from a subset of the MIDOG++ training set. They tried approaches including end-to-end trained deep learning models, foundation models with linear probing, and foundation models fine-tuned with low-rank adaptation (LoRA). They report achieving 81.35%, 77.88%, and 77.23% average balance accuracy on the in-domain AMi-Br and the out-of-domain AtNorm-Br and AtNorM-MD datasets, respectively

In this paper, we consider several ResNet-like models for the atypical mitotic figure classification task. We use two public datasets, AMi-Br and MIDOG25, to make the most of the available data and thus cover different tumours, tissues, scanners, i.e., very different conditions. We use on-the-fly data augmentation and weighted random sampling to combat class imbalance. Moreover, we also test several loss functions, such as Focal Loss (FL), Recall Loss (RL), and Focal Recall Loss (FRL).

## Materials

### 1. Used Datasets

We utilized two publicly available datasets, AMi-Br [8] and MIDOG25 [10], which were specifically designed to support the development and evaluation of mitotic figure classification methods. These datasets consist of annotated histopathological images of tissue samples from patients with tumours originating from a variety of organ sites. More detailed information about each dataset is given in Tab. 1 and below:

Table 1. **Datasets description**

Property	AMi-Br	MIDOG25
Image Size	128 x 128	128 x 128
Total MFs	3720	11939
Atypical	832	1748
Normal	2888	10191
AMF Rate (%)	22.4%	14.6%
Annotation Type	3-expert vote	3-expert vote
Expert Agreement	78.2%	70%
Source Datasets	TUPAC16, MIDOG21	MIDOG++ (MIDOG21, MIDOG22)
Species	Human	Human + Canine

- **AMi-Br:** This dataset consists of two previously largest and most diverse human breast cancer challenges, namely the TUPAC [11] using an improved alternative version of annotations and the MIDOG 2021 [12]. The samples were processed in several laboratories: UMC Utrecht, Symbiant Pathology Expert Center (Alkmaar & Zaandam), and using several scanners: Hamamatsu NanoZoomer XR (C1200022), Hamamatsu NanoZoomer S360, Aperio Scanscope CS2, Aperio ScanScope XT, Leica Aperio GT 450, Leica SCN400. All histological sections were stained with standard H&E stain and scanned using a 40× objective, giving a scanning resolution of 0.23, 0.25 or 0.26  $\mu\text{m}/\text{px}$  depending on the scanner. Three expert pathologists were recruited to create the annotations, independently of each other.

- **MIDOG25:** The MIDOG25 dataset is part of the Mitosis Domain Generalization Challenge 2025 (task 2), created to benchmark AI models for mitotic figure classification under varying imaging conditions and biological domains. Cases of three tumour types were obtained from human patients (breast carcinoma, pancreatic and gastrointestinal neuroendocrine tumours, and cutaneous melanoma) and four tumour types were obtained from dog patients (pulmonary carcinoma, lymphosarcoma mostly in lymph nodes, cutaneous mast cell tumours, and (sub)cutaneous soft tissue sarcoma). The samples were processed in several laboratories: UMC Utrecht, VMU Vienna, FU Berlin, AMC New York, and using several scanners: Hamamatsu XR

(C12000-22), Hamamatsu S360 (0.5 N/A), Leica ScanScope CS2, 3DHitech Panoramic Scan II, Aperio ScanScope CS2. All histological sections were stained with standard H&E stain and scanned using a 40× objective, giving a scanning resolution of 0.23 or 0.25  $\mu\text{m}/\text{px}$  depending on the scanner. Three expert pathologists were recruited to create the annotations, who had extensive experience in annotating mitotic figures.

Importantly, the datasets were curated to represent diverse imaging conditions, emulating real-world variability and enabling the development of robust, domain-generalizable classification approaches.

## 2. Short Data Overview

Mitosis is the division of one cell nucleus, resulting in the formation of two daughter nuclei with identical genetic information [1]. Mitotic figures (MF) are structures that can be identified using light microscopy. MFs are defined as nuclear aggregates containing short rods or spikes of chromosomes. Depending on the phase of mitosis, the morphology of MFs varies. The phases of normal mitosis include prophase, prometaphase, metaphase, anaphase, and telophase. It is important to note that mitosis is a continuous process, and cells observed in histological preparations may pause during the transition between phases. The duration of each phase may vary depending on the animal species and cell type. Examples of normal mitotic figures are shown in fig. 1.

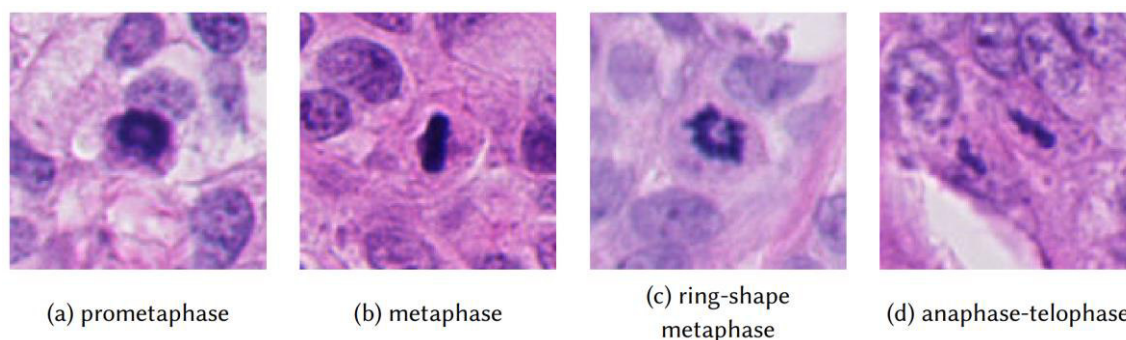


Figure 1. Examples of normal mitotic figures (NMF) taken from the presented datasets: (a)–(d) represent different phases of mitosis in which mitotic figures are normal

The process of cell division can be prone to errors in chromosome segregation. When these errors occur in neoplastic (cancerous) cells, they result in genetic abnormalities that are morphologically detectable as AMF. These abnormalities can be broadly classified into two groups: (1) issues related to mitotic or polar asymmetry, and (2) improper segregation of chromosomes. More intricate genetic aberrations can develop from the interplay of these fundamental categories. Examples of atypical mitotic figures are shown in fig. 2.

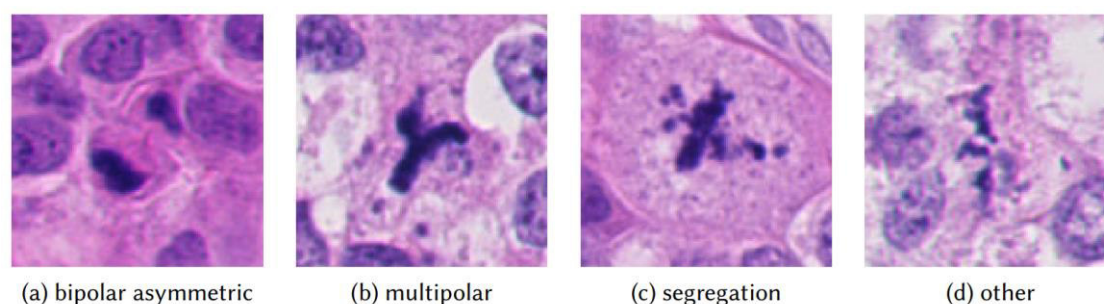


Figure 2. Examples of atypical mitotic figures (AMF) taken from the presented datasets: (a)–(d) represent various anomalies of mitosis in which mitotic figures are atypical

## Methods

### 1. Evaluation Metrics

Since the problem is unbalanced in terms of data (only about 20 % are atypical figures), it is important to use appropriate metrics for proper evaluation. We chose balanced accuracy as our main metric. We also consider the F1 score and area under ROC curve for more information on testing.



Balanced Accuracy (BA): is a metric designed to address imbalanced datasets by accounting for class-specific performance. It calculates the average of recall values across all classes, ensuring equal importance is given to each class regardless of its prevalence. It can be evaluated in following way:

$$BA = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{FN}{FN + FP} \right). \quad (1)$$

F1 Score (F1): is a harmonic mean of precision and recall, offering a single metric that reflects a model's ability to balance false positives and false negatives. It calculated as:

$$BA = \frac{2TP}{2TP + FP + FN}. \quad (2)$$

Area Under ROC Curve (AUROC): evaluates a classification models ability to distinguish between classes at various decision thresholds. It plots the true positive rate (sensitivity) against the false positive rate and computes the area under this curve. AUC-ROC values range from 0 to 1, where a value closer to 1 signifies excellent discriminative ability. This metric is particularly effective when comparing model performance across different threshold settings or when dealing with imbalanced datasets, as it evaluates performance independent of a specific decision boundary

## 2. Convolutional Neural Networks

We decided to focus on several ResNet-like models, which are deep convolutional neural networks. ResNet still maintains several advantages and may well be the optimal choice, particularly in certain scenarios. Architectures like ResNet are computationally efficient compared to transformers such as Vision Transformers (ViT) and MLP-Mixers, especially for small and medium datasets or in resource-constrained situations [13]. Transformers are also often more sensitive to hyperparameter tuning. We would also like to focus on simpler and more well-known models first.

A detailed description of each model's architecture is provided below.

### 2.1. ResNet & ResNetV2

It is known that if you simply increase the number of layers in the network, it begins to learn worse. This result is counterintuitive, as the intuitive assumption is that increased network depth should linearly enhance representational capacity. The fact is that one of the big problems of deep networks is the problem of vanishing gradients [14].

The organization of the network in the form of residual blocks, invented by a team from the Asian division of Microsoft Research, made it possible to significantly increase the depth of trained networks in 2015 [15, 16]. Their idea is to add a shortcut (skip connection) that skips one or more layers.

At the same time, short connections have a positive effect on the landscape of the error function [17, 18]. This does not make the function convex, which would significantly simplify the learning process, but allows the surface to become smoother, which means that gradient descent and its modifications will work better.

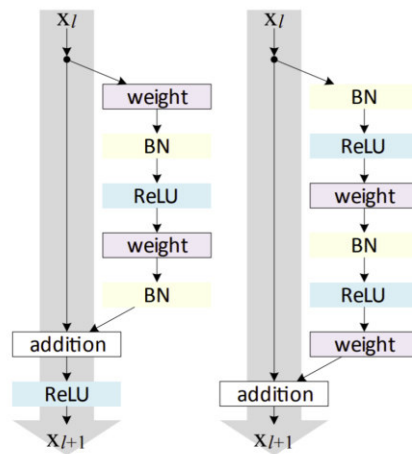


Figure 3. Differences between ResNet (left) and ResNetV2 (right)

While the original ResNet paper did not draw direct inspiration from biological systems, subsequent research has explored parallels between ResNet's architecture and certain biological learning mechanisms. Notably, advancements in neuroscience have revealed similar structures in the brains of organisms such as



fruit fly larvae through uncovering the entire connectome. A 2023 study [19] identified "multilayer shortcuts" within the insect brain that function similarly to the skip connections utilized in ResNets and other artificial neural networks. This finding suggests intriguing links between artificial intelligence architectures and the neural designs observed in biological systems, providing a fascinating intersection of technology and biology.

### 2.2. DenseNet

The success of ResNet suggested that short connections allow training deeper networks and making the model more accurate. The authors analyzed this fact and presented DenseNet (Densely Connected Convolutional Network) in 2017 [20]. They reworked the residual block and presented a dense block. The main difference from the residual block is that now the feature maps are not summed up, but concatenated, and each layer receives information from the previous ones.

The developers claim that in addition to improving efficiency, this network organization improves the flow of information and gradients throughout the network, which facilitates training, thereby helping deep networks. In addition, dense blocks have a regularizing effect that reduces overfitting in problems with small training sets.

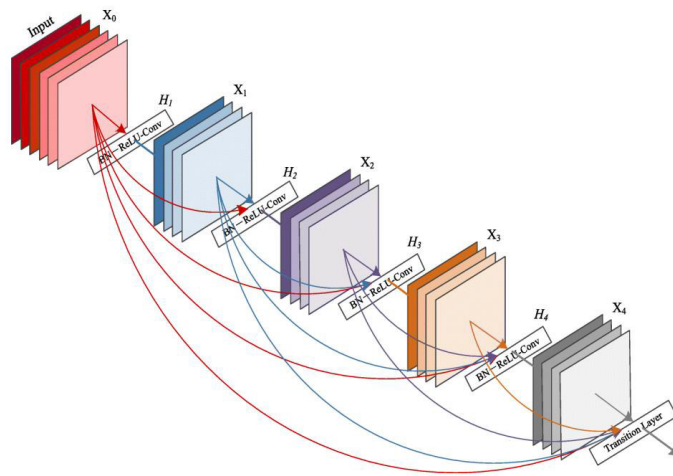


Figure 4. One of the variants for implementing the DenseNet

### 2.3. ResNeSt

The ResNeSt network extends the classic ResNet framework by integrating a nested skip connection strategy [21]. This approach allows for multiple pathways through which features can flow, thus enabling the model to learn from various levels of abstraction within the input data. Specifically, the architecture incorporates a series of "split-attention" blocks that enable the network to aggregate information based on the importance of different features.

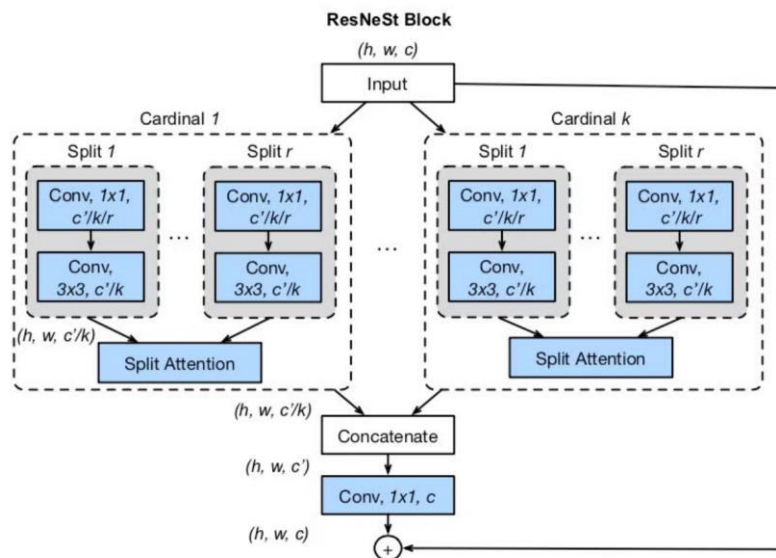


Figure 5. ResNeSt block [21]

Each block consists of two stages: the first stage applies convolutional operations to generate feature maps, while the second stage uses attention mechanisms to weigh the output of these feature maps (fig. 5). This dual-stage processing allows ResNeSt to capture both local and global contextual information, leading to enhanced representational power.

The model's ability to efficiently capture intricate patterns in images translates into significant gains in various tasks, including image classification, object detection, and semantic segmentation. It boosts the performance of downstream models such as Mask R-CNN, Cascade R-CNN and DeepLabV3.

### 3. Implementation Details

We used models implemented in timm (PyTorch Image Models) with Apache-2.0 licence [22]. timm is a deep-learning framework that offers a comprehensive suite of state-of-the-art computer vision models, encompassing layers, utilities, optimizers, schedulers, data handling tools, augmentation techniques, and training/validation scripts. Notably, it empowers users to replicate ImageNet training outcomes.

We have chosen a non-standard optimizer, Adan [23], which is probably not as well-known as SGD or Adam. Adan (ADaptive Nesterov momentum algorithm) introduces a Nesterov momentum estimation approach to achieve stable and precise estimations of the first and second moments of gradients in adaptive gradient algorithms, enhancing acceleration. Compared to previous adaptive optimizers like Adam, Adan demonstrates provably faster convergence [24]. Empirical results reveal that Adan delivers significantly improved generalization performance, outperforming state-of-the-art deep optimizers on both CNNs and transformers. We set the learning rate to  $10^{-3}$  and weight decay to  $10^{-4}$ .

We used two datasets and split each into training, validation, and test parts in a 70 : 15 : 15 ratio. Then we combined each part of the first dataset and the second one respectively (training with training, and so on). During training, we monitored the validation balanced accuracy and saved the best model for it, which was then used on the test data. All models were trained for 25 epochs. The input data shape for each model is  $3 \times 128 \times 128$  (CHW order). The batch size is set to 64 for all models. All the algorithms were trained using an NVIDIA RTX 3080 Ti (12 GB) with Python 3.11, PyTorch 2.7 and CUDA 11.8.

### 4. On-The-Fly Data Augmentation

Data augmentation plays a vital role in enhancing the performance of deep learning models, particularly in domains with limited data [25]. On-the-fly data augmentation (also online augmentation) refers to the process of generating augmented data in real-time during the training process, rather than pre-computing and storing augmented samples before training begins. This method has gained popularity due to its efficiency and effectiveness in reducing overfitting, improving model generalization, and increasing the diversity of the training dataset. We have chosen the well-proven Albumentations [26] to create augmentations. Training augmentations consisted of affine transform (scale, translate and rotate), H&E staining, Gaussian noise, blurring (motion, median or box blur), random changes of brightness/contrast, image compression (JPEG) and grid dropout. We perform these augmentations with some probabilities on-the-fly.

### 5. Weighted Random Sampling

Weighted random sampling (WRS) is a method for selecting items from a collection where each item has a different probability of being chosen, based on its assigned weight [27]. Higher weights indicate a greater chance of an item being selected.

This is in contrast to standard random sampling where each item has an equal probability of being chosen. This technique can be applied to solve the problems of imbalanced dataset. We use it during training, where the sample weight is inversely proportional to the class frequency (from the training set). This is one of the simplest techniques, which is also natively implemented in PyTorch.

### 6. Loss Functions

In our study, we explored several loss functions to optimize the performance of our mitotic figure classification models, specifically focusing on Focal Loss, Recall Loss, and Focal Recall Loss. Each of these loss functions provides unique advantages tailored to the challenges presented by imbalanced datasets.

#### 6.1. Focal Loss

Focal Loss is designed to address class imbalance by applying a modulating factor to the standard cross-entropy loss [28]. This loss function is particularly effective in scenarios where the number of negative samples

vastly exceeds the number of positive samples, as is common in mitotic figure identification tasks. The formulation of Focal Loss is given by:

$$L_F = -\alpha_t(1 - p_t)^\gamma \log p_t, \quad (3)$$

where  $p_t$  is the predicted probability of the true class,  $\alpha_t$  is a balancing factor for each class, and  $\gamma$  is a focusing parameter that adjusts the rate at which easy examples are down-weighted

### 6.2. Recall Loss

Recall Loss focuses specifically on the model's ability to correctly identify positive instances in the dataset [29]. Unlike traditional loss functions that may prioritize overall accuracy, Recall Loss emphasizes the true positive rate, guiding the model toward minimizing false negatives. The formulation for Recall Loss can be defined as:

$$L_R = 1 - \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i}, \quad (4)$$

where TP and FN can be calculated as contiguous way (let  $y$  be the one-hot encoded target vector and  $\hat{y}$  be the predicted probabilities for each class):

$$TP = \sum_{i=1}^N (y_i \cdot \hat{y}_i), \quad FN = \sum_{i=1}^N (y_i \cdot (1 - \hat{y}_i)). \quad (5)$$

The objective of Recall Loss is to ensure high sensitivity, which is critical in pathological assessments where missing atypical mitotic figures could lead to severe diagnostic errors.

### 6.3. Focal Recall Loss

Focal Recall Loss combines the advantages of Focal Loss and Recall Loss, targeting both the challenge of class imbalance and the importance of high sensitivity [30, 31]. This loss function modifies the standard Recall Loss by incorporating a focusing parameter that allows for downweighting contributions from well-classified examples. The formulation for Focal Recall Loss is expressed as:

$$L_{FR} = -\frac{1}{N} \sum_{i=1}^N \left(1 - \frac{TP_i}{TP_i + FN_i}\right)^\beta \cdot \alpha_t(1 - p_t)^\gamma \log p_t. \quad (4)$$

In this equation,  $p_t$  represents the predicted probability of the true class, while the additional terms  $\alpha_t$ ,  $\gamma$  and  $\beta$  serve to modulate the loss based on classification performance. Focal Recall Loss aims to enhance the model's focus on correctly identifying atypical mitotic figures in a challenging dataset, effectively balancing the need for precision in the context of clinical relevance. For all experiments, we set the values of  $\alpha_t$ ,  $\gamma$  and  $\beta$  to 1, 2 and 2, respectively.

## Results

In this section, we present the performance metrics of various models evaluated on the combined dataset (AMI-Br + MIDOG25) using three different loss functions: Focal Loss (FL), Recall Loss (RL) and Focal Recall Loss (FRL). The models assessed include ResNet and DenseNet architectures of various depths, specifically ResNet-34D, ResNetV2-34D, DenseNet-121, and ResNeSt-26D. Table 2 summarizes the test results, showcasing model performance.

Table 2. Test results on the combined dataset (ami-br + midog25)

Model	BA (%)	F <sub>1</sub> (%)	AUROC (%)
ResNet-34D (FL)	79.21	73.78	88.92
ResNet-34D (RL)	79.49	74.50	88.89
ResNet-34D (FRL)	79.55	74.40	88.09
ResNetV2-34D (FL)	80.72	72.83	89.89
ResNetV2-34D (RL)	<b>84.93</b>	76.73	<b>92.12</b>
ResNetV2-34D (FRL)	82.42	73.26	90.85
DenseNet-121 (FL)	81.70	73.38	89.28
DenseNet-121 (RL)	82.81	76.67	90.39
DenseNet-121 (FRL)	81.99	75.96	89.85
ResNeSt-26D (FL)	82.71	74.66	90.56
ResNeSt-26D (RL)	82.48	<b>76.81</b>	90.27
ResNeSt-26D (FRL)	81.22	70.52	89.32

Quite good quality metrics were obtained for all models. The best quality in terms of BA (84.93) and AUROC (92.12) was shown by the ResNetV2-34D model with Recall Loss. The best model in terms of F1 (76.81) was ResNeSt-26D also with Recall Loss.

We can note that in many cases, using Recall Loss or Focal Recall Loss also has a positive effect on the results. Although here, extensive experiments with more careful tuning of the optimizer, network hyperparameters or loss function may be required. It is obvious that there is a gap for improving the results, and also that the existing labelling may be inaccurate, which introduces an error.

## Conclusions

In this work, we presented several ResNet-like models for normal/atypical mitotic figure classification, a relatively new task in computational histology. We tested several well-known models, such as ResNet-34D, ResNetV2-34D, DenseNet-121, and ResNeSt-26D. We used two public datasets, AMi-Br and MIDOG25, to best capture the existing variable real-world histological imaging conditions. We also considered several error functions: Focal Loss (FL), Recall Loss (RL), and Focal Recall Loss (FRL).

Good results were demonstrated, which obviously can be further improved. Several potential directions of work can be noted: (1) using other neural network architectures, such as Transformers or others, (2) additional use of other datasets, (3) conducting experiments to test generalizability to different tissues, scanners, etc.

## References

1. T. A. Donovan, F. M. Moore, C. A. Bertram, R. Luong, P. Bolfa, R. Klopffleisch, H. Tvedten, E. N. Salas, D. B. Whitley, M. Aubreville, D. J. Meuten, Mitotic figures normal, atypical, and imposters: A guide to identification, *Veterinary Pathology* 58 (2020) 243257. <http://dx.doi.org/10.1177/0300985820980049>
2. A. Lashen, M. S. Toss, M. Alsaleem, A. R. Green, N. P. Mongan, E. Rakha, The characteristics and clinical significance of atypical mitosis in breast cancer, *Modern Pathology* 35 (2022) 13411348. <http://dx.doi.org/10.1038/s41379-022-01080-0>
3. C. A. Bertram, A. Bartel, T. A. Donovan, M. Kiupel, Atypical mitotic figures are prognostically meaningful for canine cutaneous mast cell tumors, *Veterinary Sciences* 11 (2023) 5. <http://dx.doi.org/10.3390/vetsci11010005>
4. M. Cui, D. Y. Zhang, Artificial intelligence and computational pathology, *Laboratory Investigation* 101 (2021) 412422. <http://dx.doi.org/10.1038/s41374-020-00514-0>
5. A. N. N. Wong, Z. He, K. L. Leung, C. C. K. To, C. Y. Wong, S. C. C. Wong, J. S. Yoo, C. K. R. Chan, A. Z. Chan, M. D. Lacambra, M. H. Y. Yeung, Current developments of artificial intelligence in digital pathology and its future clinical applications in gastrointestinal cancers, *Cancers* 14 (2022) 3780. <http://dx.doi.org/10.3390/cancers14153780>
6. M. Aubreville, J. Ganz, J. Ammeling, T. A. Donovan, R. H. Fick, K. Breininger, C. A. Bertram, Deep Learning-based Subtyping of Atypical and Normal Mitoses using a Hierarchical Anchor-free Object Detector, *Springer Fachmedien Wiesbaden*, 2023, p. 189195. [http://dx.doi.org/10.1007/978-3-658-41657-7\\_40](http://dx.doi.org/10.1007/978-3-658-41657-7_40)
7. J. Ammeling, M. Aubreville, S. Banerjee, C. A. Bertram, K. Breininger, D. Hirling, P. Horvath, N. Stathonikos, M. Veta, Mitosis domain generalization challenge 2025, 2025. <https://zenodo.org/doi/10.5281/zenodo.15077361>
8. C. A. Bertram, V. Weiss, T. A. Donovan, S. Banerjee, T. Conrad, J. Ammeling, R. Klopffleisch, C. Kaltenecker, M. Aubreville, Histologic Dataset of Normal and Atypical Mitotic Figures on Human Breast Cancer (AMi-Br), *Springer Fachmedien Wiesbaden*, 2025, p. 113118. [http://dx.doi.org/10.1007/978-3-658-47422-5\\_25](http://dx.doi.org/10.1007/978-3-658-47422-5_25)
9. S. Banerjee, V. Weiss, T. A. Donovan, R. H. J. Fick, T. Conrad, J. Ammeling, N. Porsche, R. Klopffleisch, C. Kaltenecker, K. Breininger, M. Aubreville, C. A. Bertram, Benchmarking deep learning and vision foundation models for atypical vs. normal mitosis classification with cross-dataset evaluation, 2025. <https://arxiv.org/abs/2506.21444>. doi:10.48550/ARXIV.2506.21444.
10. J. Ammeling, M. Aubreville, S. Banerjee, C. A. Bertram, K. Breininger, D. Hirling, P. Horvath, N. Stathonikos, M. Veta, Mitosis domain generalization challenge 2025, 2025. <https://zenodo.org/doi/10.5281/zenodo.15077360>
11. M. Veta, Y. J. Heng, N. Stathonikos, B. E. Bejnordi, F. Beca, T. Wollmann, K. Rohr, M. A. Shah, D. Wang, M. Rousson, M. Hedlund, D. Tellez, F. Ciompi, E. Zerhouni, D. Lanyi, M. Viana, V. Kovalev, V. Liauchuk, H. A. Phoulady, T. Qaiser, S. Graham, N. Rajpoot, E. Sjöblom, J. Molin, K. Paeng, S. Hwang, S. Park, Z. Jia, E. I.-C. Chang, Y. Xu, A. H. Beck, P. J. van Diest, J. P. Pluim, Predicting breast tumor proliferation from wholeslide images: The tupac16 challenge, *Medical Image Analysis* 54 (2019) 111121. <http://dx.doi.org/10.1016/j.media.2019.02.012>
12. M. Aubreville, N. Stathonikos, C. A. Bertram, R. Klopffleisch, N. ter Hoeve, F. Ciompi, F. Wilm, C. Marzahl, T. A. Donovan, A. Maier, J. Breen, N. Ravikumar, Y. Chung, J. Park, R. Nateghi, F. Pourakpour, R. H. Fick, S. Ben Hadj, M. Jahanifar, A. Shephard, J. Dextl, T. Wittenberg, S. Kondo, M. W. Lafarge, V. H. Koelzer, J. Liang, Y. Wang, X. Long, J. Liu, S. Razavi, A. Khademi, S. Yang, X. Wang, R. Erber, A. Klang, K. Lipnik, P. Bolfa, M. J. Dark, G. Wasinger,

- M. Veta, K. Breininger, Mitosis domain generalization in histopathology images the midog challenge, *Medical Image Analysis* 84 (2023) 102699. <http://dx.doi.org/10.1016/j.media.2022.102699>
13. J. Mauricio, I. Domingues, J. Bernardino, Comparing vision transformers and convolutional neural networks for image classification: A literature review, *Applied Sciences* 13 (2023) 5521. <http://dx.doi.org/10.3390/app13095521>
  14. S. Basodi, C. Ji, H. Zhang, Y. Pan, Gradient amplification: An efficient way to train deep neural networks, *Big Data Mining and Analytics* 3 (2020) 196207. <http://dx.doi.org/10.26599/BDMA.2020.9020004>
  15. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, p. 770778. <http://dx.doi.org/10.1109/CVPR.2016.90>
  16. K. He, X. Zhang, S. Ren, J. Sun, Identity Mappings in Deep Residual Networks, Springer International Publishing, 2016, p. 630645. [http://dx.doi.org/10.1007/978-3-319-46493-0\\_38](http://dx.doi.org/10.1007/978-3-319-46493-0_38)
  17. H. Li, Z. Xu, G. Taylor, C. Studer, T. Goldstein, Visualizing the loss landscape of neural nets, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, Curran Associates Inc., Red Hook, NY, USA, 2018, p. 63916401.
  18. L. Wang, B. Shen, N. Zhao, Z. Zhang, Is the skip connection provable to reform the neural network loss landscape?, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI'20, 2021.
  19. M. Winding, B. D. Pedigo, C. L. Barnes, H. G. Patsolic, Y. Park, T. Kazimiers, A. Fushiki, I. V. Andrade, A. Khandelwal, J. Valdes-Aleman, F. Li, N. Randel, E. Barsotti, A. Correia, R. D. Fetter, V. Hartenstein, C. E. Priebe, J. T. Vogelstein, A. Cardona, M. Zlatić, The connectome of an insect brain, *Science* 379 (2023). <http://dx.doi.org/10.1126/science. add9330>
  20. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017. <http://dx.doi.org/10.1109/CVPR.2017.243>
  21. H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, A. Smola, Resnest: Split-attention networks, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2022, p. 27352745. <http://dx.doi.org/10.1109/CVPRW56347.2022.00309>
  22. R. Wightman, Pytorch image models, <https://github.com/rwightman/pytorch-image-models>, 2019. doi:10.5281/zenodo.4414861.
  23. X. Xie, P. Zhou, H. Li, Z. Lin, S. Yan, Adan: Adaptive nesterov momentum algorithm for faster optimizing deep models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46 (2024) 95089520. <http://dx.doi.org/10.1109/TPAMI.2024.3423382>
  24. A. R. Ismail, M. Z. R. Azhary, N. A. Hitam, Evaluating Adan vs. Adam: An Analysis of Optimizer Performance in Deep Learning, Springer Nature Switzerland, 2025, p. 251263. [http://dx.doi.org/10.1007/978-3-031-82931-4\\_19](http://dx.doi.org/10.1007/978-3-031-82931-4_19)
  25. J. Ma, C. Hu, P. Zhou, F. Jin, X. Wang, H. Huang, Review of image augmentation used in deep learning-based material microscopic image segmentation, *Applied Sciences* 13 (2023) 6478. <http://dx.doi.org/10.3390/app13116478>
  26. A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A. A. Kalinin, Albumentations: Fast and flexible image augmentations, *Information* 11 (2020) 125. <http://dx.doi.org/10.3390/info11020125>
  27. P. Efraimidis, P. Spirakis, Weighted Random Sampling: 2005; Efraimidis, Spirakis, Springer US, 2008, p. 10241027. [http://dx.doi.org/10.1007/978-0-387-30162-4\\_478](http://dx.doi.org/10.1007/978-0-387-30162-4_478)
  28. T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (2020) 318327. <http://dx.doi.org/10.1109/TPAMI.2018.2858826>
  29. J. Tian, N. C. Mithun, Z. Seymour, H.-P. Chiu, Z. Kira, Striking the right balance: Recall loss for semantic segmentation, in: 2022 International Conference on Robotics and Automation (ICRA), IEEE, 2022, p. 50635069. <http://dx.doi.org/10.1109/ICRA46639.2022.9811702>
  30. Z. Chen, Z. Wang, J. Qiu, The alzheimers disease segmentation model based on focal recall loss, *Computational Science and Applications* 12 (2022) 95–107. doi:10.12677/CSA.2022. 121011.
  31. X. Liu, L. Wang, L. Ma, C. Wang, Drfl: Dynamic-recall focal loss for image classification and segmentation, *Applied Artificial Intelligence* 38 (2024). <http://dx.doi.org/10.1080/ 08839514.2024.2411845>.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-844-851

## Локализация клеток с использованием моделей на основе YOLO в анализе жизнеспособности живых-мертвых клеток

Александр Трыкин<sup>1,2</sup>, Михаил Кривоносов<sup>1,2</sup>, Елизавета Калинина<sup>3</sup>

<sup>1</sup> Исследовательский центр искусственного интеллекта, Институт информационных технологий, математики и механики, ННГУ, Нижний Новгород, Россия

<sup>2</sup> Лаборатория системной медицины старения, НИИ биологии старения, ННГУ, Нижний Новгород, Россия

<sup>3</sup> Институт биологии и биомедицины, ННГУ, Нижний Новгород, Россия

**Аннотация.** Анализ жизнеспособности клеток критически важен для оценки клеточного здоровья особенно в таких областях, как скрининг лекарственных препаратов, исследования рака и тканевая инженерия. В данной работе мы рассматриваем задачу анализа жизнеспособности живых и мертвых клеток через призму обнаружения объектов, используя модели на основе YOLO для точной локализации и классификации живых и мертвых клеток на микроскопических изображениях. Наш вклад заключается в следующем: (1) представляем полностью аннотированный набор данных для анализа жизнеспособности живых и мертвых клеток в качестве задачи обнаружения объектов; (2) проводим эксперименты с относительно небольшой обучающей выборкой (2-10 % из 252 изображений) и демонстрируем способность нейронной модели к обучению и обобщению, достигая при этом довольно высоких показателей качества; (3) обучаем новейшие модели YOLO (с версии 8 по версию 12) и достигаем высоких и сопоставимых показателей качества для каждой модели; лучшие из представленных моделей достигают точности до 97 % для синих клеток и 96.5 % для красных клеток с достаточным количеством полноты.

**Ключевые слова:** обработка изображений, флуоресцентная микроскопия, обнаружение объектов, глубокое обучение.

## Cell Localization with YOLO-based Models in Live-Dead Cell Viability Assay

Alexander Trykin<sup>1,2</sup>, Mikhail Krivonosov<sup>1,2</sup>, Elizaveta Kalinina<sup>3</sup>

<sup>1</sup> Artificial Intelligence Research Center, Institute of Information Technologies, Mathematics and Mechanics, Lobachevsky State University, Nizhny Novgorod, Russia

<sup>2</sup> Laboratory of Systems Medicine of Ageing, Research Institute for Biology of Aging, Lobachevsky State University, Nizhny Novgorod, Russia

<sup>3</sup> Institute of Biology and Biomedicine, Lobachevsky State University, Nizhny Novgorod, Russia

**Abstract.** Cell viability assays are critical for assessing cellular health, particularly in applications such as drug screening, cancer research, and tissue engineering. In this work, we address the challenge of live-dead cell viability analysis through the lens of object detection, utilizing YOLO-based models to accurately localize and classify live and dead cells in microscopy images. Our contributions are as follows: (1) we present a fully annotated dataset for live-dead cell viability analysis as an object detection task; (2) we conduct experiments with a relatively small training sample size (2-10% of 252 images) and demonstrate the ability of the neural model to learn and generalize, while achieving fairly high quality metrics; (3) we train the latest YOLO models (from v8 to v12) and achieve high and comparable quality scores for each model, the best of presented models achieve up to 97% precision for blue cells and 96.5% for red cells with sufficient recall numbers.

**Keywords:** image processing, fluorescence microscopy, object detection, deep learning

### Introduction

Live-dead cell viability assays are laboratory methods used to determine the proportion of live and dead cells in a population. These assays are commonly used in cell biology, drug discovery, and biomedical research to assess cell viability, cytotoxicity, and overall cell health.

The principle behind live-dead cell viability assays is to distinguish between live cells with intact membranes (indicating cell viability) and dead cells with compromised membranes (indicating cell death). Various techniques and dyes are utilized to differentiate between these two populations on the basis of their membrane integrity.

In recent years, convolutional neural networks (CNNs) have revolutionized the field of computer vision, and their application in biomedical image analysis has shown immense promise.

The paper [1] proposes a modification of YOLOv5 with FPN (Feature Pyramid Network) to handle cells of different sizes in the images. Their dataset consisted of 283 images, which were reasoned to 600 using



techniques such as rotation, scale and flip. They were able to achieve an AP (Average Precision) of 0.8 and a processing time of 43.9 ms per image ( $\sim 144$  FPS) on a public dataset.

The work [2] proposes the use of the E-U-Net segmentation network in the task of "live-dead" analysis on unlabelled cells using phase imaging with computational specificity (PICS). They developed a U-Net-like neural network with an EfficientNet encoder using transfer learning on live adherent HeLa and CHO cell cultures. They obtained classification of the viable state of a single cell with an accuracy of 95%.

The paper [3] proposes a YOLOv5-based approach for automatic cell recognition and counting in a case study of laboratory cell detection using images from a CytoSMART Exact FL microscope. The authors used a dataset of 21 expert-labelled images, as well as an additional Sperm DetectionV dataset of 1024 images for transfer learning. The proposed method showed significantly better performance compared to the U-Net model. YOLOv5 had accuracy, precision, recall, and F1 scores of 92%, 84%, 91%, and 87%, respectively.

In paper [4], a novel two-stage DL-based detection method (cGAN-YOLO) is developed to further improve the cell counting efficiency by combining a DL-based transmitted-light images and a DL-based cell detection model. The dataset of transmitted-light images and the fluorescent images consisted of 4000  $512 \times 512$  images of human umbilical vein endothelial cells (HUVEC) and MadinDarby canine kidney cells (MDCK). The recognition accuracy (RA) results were obtained as follows: 96.67% for HUVEC and 92.58% for MDCK.

This study explores the application of YOLO-based models for cell localization in live-dead viability assays. By leveraging the robustness of YOLO algorithms, we aim to provide a method that is not only fast and accurate, but also scalable across different assay conditions and imaging platforms.

## Materials

### 1. Proposed Dataset

The data transmitted to us were collected by the laboratory of the Institute of Biology and Biomedicine of Lobachevsky University. There are a total of 180 pairs of images in PNG format and with a resolution of  $2560 \times 1920$ : 180 with blue fluorescent emission and 180 with red fluorescent emission. These pairs belong to different experiments, thereby forming series (sequences). An example of data is shown in fig. 1.

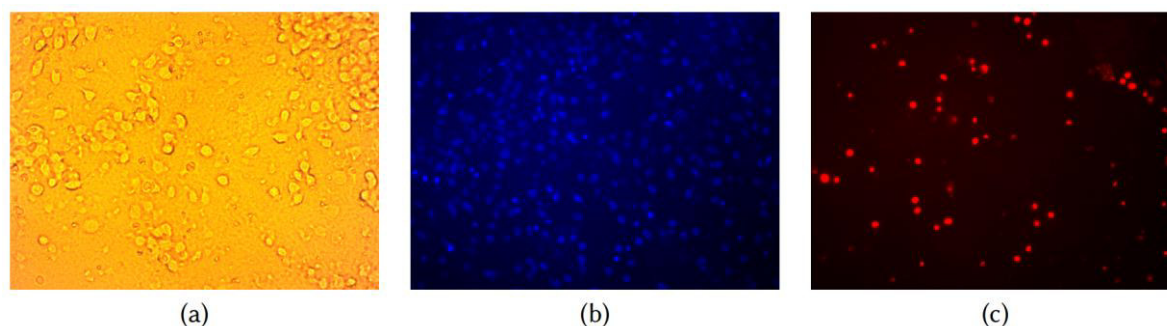


Figure 1. Example images from collected dataset. (a) brightfield/phase-contrast: cell morphology and density are visible without fluorescence, (b) blue fluorescence: dye penetrates nuclei, (c) red fluorescence: indicator of dead cells, cannot cross intact plasma membranes, so it only penetrates cells with compromised membranes

At the beginning of the work, we had only a labelled part of the entire dataset: 34 with a blue emission and 34 with a red emission. We used the Roboflow [5] to perform collaborative management of the dataset, and its labelling. We have made our dataset public: <https://universe.roboflow.com/unnn/cytotoxicity-hypoxia-neuroadapt-od> (during the work there were several versions of the dataset, v8 is the current one, on which the experiments were conducted).

We prepared the dataset by annotating cells with a bounding box in YOLOv8 Annotation Format. This involved converting the bounding box coordinates to  $(x, y, w, h)$  format, where  $(x, y)$  represents the coordinates of the centre of the bounding box and  $(w, h)$  represents the width and height of the bounding box. The coordinates and dimensions of the box were specified in a normalized form, depending on the height and width of the image. The class label for each bounding box was also encoded as an integer. Quality checks were performed to ensure the accuracy and consistency of the annotations.



After annotation, we randomly split the dataset into training (70%), validation (20%) and test (10%) sets. In any case, researchers are free to use any other split according to their needs.

It is important to note that cell labelling turned out to be quite a complex task, since in some cases even a person cannot always determine with certainty whether an area is a cell or a background. Such situations arise due to the complex structure of cells, low contrast or high noise in the images, which complicates the verification process. As a result, errors in the dataset labelling are possible due to subjective interpretation of complex areas of the image. Despite this, the creation of such a dataset remains an important step for training and evaluating models working with cell image analysis.

## 2. Preliminary Data Analysis

The cells in the presented images are round and differ in size. Cells are usually discrete but may "clump" together with varying degrees of overlap, which can be a problem in cell detection. The cells themselves have different luminosities, from brightly fluorescent with a fairly clear border to weakly fluorescent with a vague border that almost merges with the background.

In different images, in the same image, but in different parts of the frame, background heterogeneity may be observed. This is clearly visible in example 2 (a), where the background inside the red rectangle is taller (lighter) than the background inside the yellow rectangle (darker).

And in examples 2 (b, c) one can see the possible variation of the background in different images from almost imperceptible to the level of strong noise. This may be due to the peculiarities of the scanning microscope (and/or its settings), the hypothetical spreading of the fluorescent compound from the cells.

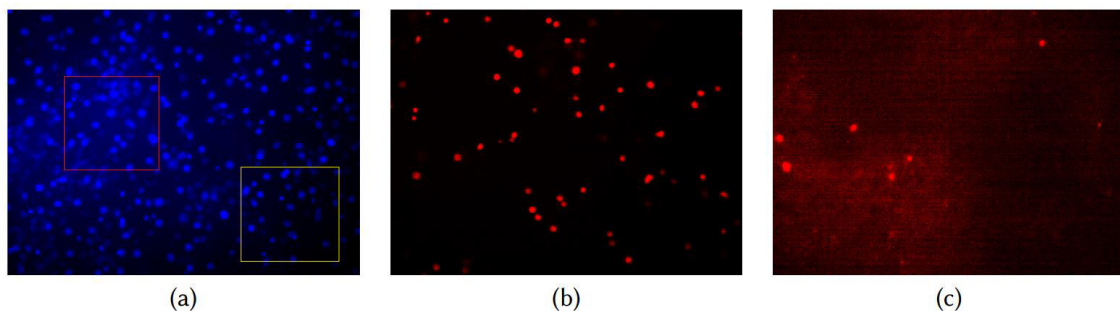


Figure 2. Images showing variations in conditions: in (a) the background inside the red rectangle is lighter than the background inside the yellow rectangle, in (b, c) shows the variability of the background

## Methods

### 1. YOLO's

YOLO (You Only Look Once), a renowned family of single-stage deep learning models, specializes in object detection. Introduced in 2015, YOLO has become widely adopted for its impressive balance of speed and accuracy [6]. Over time, YOLO has evolved through numerous iterations and enhancements, with subsequent versions being continually refined by the computer vision community [7, 8].

YOLOv8 [9]: YOLOv8, released in January 2023 by Ultralytics, introduced several innovations to improve flexibility and accuracy. This version utilized a "C2f module," a variant of the CSPNet module, which optimized feature extraction and model efficiency. It adopted an anchor-free design for bounding box predictions and integrated decoupled heads, independently addressing objectness, classification, and bounding box regression. YOLOv8 also employed CIoU and Distributional Focal Loss (DFL) for bounding box calculations, enhancing localization accuracy specifically for smaller objects like cells.

YOLOv9 [10]: YOLOv9 incorporated advanced methods such as Generalized Efficient Layer Aggregation Network (GELAN) and Programmable Gradient Information (PGI) to further reduce data loss during feature extraction and transformation. The PGI method preserved complete input information for robust gradient computations, ensuring more reliable weight updates and superior convergence.

YOLOv10 [11]: Released in mid-2024, YOLOv10 introduced improvements in computational efficiency and ensured higher accuracy for dense detection scenarios. It utilized multiscale anchor-free detection combined with advanced distributional loss functions like DFL. YOLOv10 addressed issues in detecting rare or clustered objects by modifying its confidence scoring approach and introducing enhanced training for small

object instances, which is critical in detecting both live and dead cells with significant morphological overlaps. However, when benchmarked, the accuracy of YOLOv10 was slightly lower than YOLOv9s on datasets involving rare objects, highlighting the importance of contextual model training.

YOLOv11 [12]: YOLOv11, one of the newest iteration of the YOLO series, boasts a streamlined design incorporating C3K2 blocks, SPFF (Spatial Pyramid Pooling Fast), and cutting-edge attention mechanisms such as C2PSA. This new architecture aims to bolster the detection of small objects and elevate overall accuracy, all while preserving the real-time inference speed that has become a hallmark of YOLO.

YOLOv12 [13]: YOLOv12 represents the current pinnacle of YOLO development as of late 2024. It integrates innovations such as self-supervised pretraining, transformer blocks, and enhanced attention mechanisms for improved feature representation. Enhanced quantization-aware training allowed YOLOv12 to operate efficiently even on edge devices, facilitating seamless integration into laboratory equipment for real-time analysis.

### 1.1 Implementation Details

We used models implemented in Ultralytics [14] with PyTorch backend and AGPL-3.0 licence. In the experiments with the YOLO architectures, it is important to highlight that fine-tuning was performed using the pretrained models YOLOv8-N, YOLOv9-T, YOLOv10-N, YOLOv11-N, and YOLOv12-N. Here, "N" represents the nano version (except "T" which indicates the tiny version), aiming to achieve a balance between speed and precision. Important characteristics of the models, such as the number of weights and operating speed, are shown in table 1.

Table 1. **Summary table of model characteristics**

Model	Weights (M)	FLOPs (G)	CPU Speed (FPS) ONNX	GPU Speed (FPS) TensorRT
YOLOv8-N	3.01	8.2	9.13	261.89
YOLOv9-T	2	7.8	7.8	133.66
YOLOv10-N	2.7	8.4	9.83	141.09
YOLOv11-N	2.59	6.4	9.98	221.77
YOLOv12-N	2.57	6.5	6.22	84.78

The input data shape for each model is  $3 \times 1280 \times 1280$  (CHW order), so the images are simply resized to this resolution. The batch size is set to 4. AdamW optimizer was used with an initial learning rate of  $1.667 \times 10^{-3}$  and a momentum of 0.9. Overall, AdamW is an improved version of the Adam optimizer that handles weight regularization more efficiently and, as a result, improves the quality and generalization ability of machine learning models [17]. We also saved the weights of the best performing version of the model for use in testing.

All the algorithms were trained using an NVIDIA RTX 3080 Ti (12 GB) with Python 3.11, PyTorch 2.7 and CUDA 11.8. CPU speeds measured with ONNX export on AMD Ryzen 7 5800X3D, GPU speeds measured with TensorRT export on NVIDIA RTX 3080 Ti. YOLOv12-N was used without FlashAttention [15, 16], which may negatively affect the performance.

### 1.2 Data Augmentation

Data augmentation is a vital step in enhancing the robustness and generalizability of deep learning models, particularly in complex tasks [18]. By artificially expanding the dataset through the application of various transformations, data augmentation ensures better model performance by exposing the model to diverse representations of the data.

In the context of YOLO training, several augmentation techniques can be applied to simulate varying conditions that might occur in real-world microscopy images. These techniques are particularly useful for addressing challenges like variability in cell morphology, differences in staining intensities, and diverse imaging conditions.

Ultralytics uses its own implementations of some augmentation techniques, and also uses the well-known Albumentations [19]. Unfortunately, the set of augmentations is already defined and cannot be changed directly. To overcome this, we use a monkey patch and define our own transformations.

The list of used augmentations, as well as their settings, is presented in the table 2

Table 2. Table of used augmentations, their probabilities and parameters

Transform	Probability (%)	Parameters
Mosaic	50	grid_size = 4 degrees $\in [-0.25, 0.25]$ translate $\in [-0.25, 0.25]$
Random Perspective	100	scale $\in [-0.25, 0.25]$ shear $\in [-0.25, 0.25]$ perspective = 0
Motion Blur	25	-
Median Blur	10	blur_limit = 3
Blur	10	blur_limit = 3
Image Compression	15	type = jpeg range $\in [85, 100]\%$
Vertical Random Flip	25	-
Horizontal Random Flip	25	-

## 2. Evaluation Metrics

Typically, quality metrics are based on correct and incorrect detection. A common method for this is the use of the intersection over union (IoU). According to [20] the IoU is a measure based on Jaccards index which measures the area of overlap between the predicted bounding box  $Bp$  and the true bounding box  $Bgt$  divided by the area of union between them:

$$IoU = \frac{area(Bp \cap Bgt)}{area(Bp \cup Bgt)}. \quad (1)$$

The value of IoU is then compared with a defined threshold  $t$  in order to establish when a detection is correct or incorrect. If  $IoU \geq t$ , then the detection is considered correct, otherwise the detection is considered incorrect.

Precision (P) is the ability of a model to identify only the relevant objects. It is the percentage of correct positive predictions and is given by:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{all\ detections}. \quad (2)$$

Recall (R) is the ability of a model to find all the relevant cases (all ground truths). It is the percentage of true positive detected among all relevant ground truths and is given by:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{all\ ground\ truths}. \quad (3)$$

Average Precision (AP) is a common evaluation metric used in object detection and instance segmentation tasks. It measures the accuracy and quality of the predicted bounding boxes or segmentation masks:

$$AP = \sum_n (R_n - R_{n-1})P_n, \quad (4)$$

where  $P_n$  and  $R_n$  are the precision and recall at the  $n$ -th threshold.

It is common to report two AP values:  $AP_{50}$  at a IoU threshold of 0.5 and  $AP_{50:95}$  which averages AP across IoU thresholds from 0.5 to 0.95 in steps of 0.05. We will focus only on the second metric, since it offers a more comprehensive and stringent assessment.

## Results

### 1. Detection Performance

All YOLO models trained for 150 epochs showed excellent detection results. All models show results on test data in the region of Precision  $\approx 95\%$ , Recall  $\approx 95\%$  and  $AP_{50:95} \approx 84\%$ , except for only YOLOv10-N. Detailed results of the quality testing are given in the table 3.

The YOLOv10 architecture, although having the relative low Recall, but high Precision on blue cell's detection. We can explain this by the fact that YOLOv10 is architecturally highly selective, which ensures high accuracy in measuring the counting sensitivity. Also, in our opinion, detecting blue cells was more difficult than red ones, due to the fact that there are significantly more of them and they can have different and more complex variations.

A high value for both Precision and Recall indicates that the model is not only good at identifying true positives (predictions of which classes are actually correct), but that it does so with a minimum number of errors, with virtually no positive cases missed. This means that the model is both accurate and sensitive, demonstrating balanced and high-quality behaviour in the detection task.

Table 3. **Performance evaluation on test data**

Model	Blue Cells			Red Cells		
	P (%)	R (%)	AP <sub>50:95</sub> (%)	P (%)	R (%)	AP <sub>50:95</sub> (%)
YOLOv8-N	95.2	96.2	85.2	96.2	95.1	83.8
YOLOv9-T	96.2	95.9	85	96.2	94.3	84.4
YOLOv10-N	97	91.8	84.4	94.3	94.4	83.6
YOLOv11-N	96.3	96.1	85.4	95.8	95	83.7
YOLOv12-N	96.4	95.5	85.5	96.5	95.5	84.2

It is worth noting that we used only the smallest versions of the models (nano or tiny). Considering that these networks are usually designed for devices with very limited computing power and usually do not imply the best quality, they showed excellent results. We also tried training larger models, but we could get either insignificant growth or even worse results. As a rule, larger models require more training data, which is not always possible. That is why we settled on the smaller versions of YOLO models.

## 2. Dataset Scalability

In the next experiment, we want to test the ability of YOLO models to learn on different amounts of training data. To do this, we vary the number of images in the training set and also fix the validation set (tab. 4). This experiment may be especially interesting because in biomedical applications, labelled datasets may be rare or labelling them may be difficult and time-consuming [21].

Table 4. **Comparative experiments on the scalability of a dataset for testing the learnability of models**

Experiment	Train (images)	Valid (images)
2%	5	72
5%	12	
10%	25	
25%	63	
50%	126	
100%	252	

We visualize the change in such quality metrics as Recall, Precision and AP50:95 over the epochs. An exponential moving average with  $\alpha = 0.6$  is pre-performed to smooth the curves (on fig. 3).

Of course, our experiment is not entirely fair, since in a real-world scenario a small amount of labelled data would be divided into a train-valid-test split (or train-valid). But still, we change the amount of available data in the training sample, and despite the extremely meagre sample, the neural network remains capable of learning and generalization, and can achieve relatively high quality results. Ultimately, only the training data has a direct impact on network learning and gradient propagation. The small size of the training sample means low diversity and variability of the data, which, however, does not interfere with the training and generalization of the network.

This is similar to the one-shot or few-shot learning technique, but this description is more applicable to some representative object of the class. In our case, several dozen or hundreds of class objects are located in one image. The smallest dataset size that we vary is 2% (or only 5 images out of 252), and in this sense it can be designated as few-shot image learning.

## Conclusions

This study demonstrates the potential of YOLO-based object detection models for cell localization and classification in live-dead cell viability assays. We introduced a fully annotated, domainspecific dataset that enables researchers to reframe cell viability analysis as an object detection task. The results of our experiments demonstrate the capacity of YOLO models to learn meaningful representations and generalize effectively even with a limited amount of training data, utilizing as little as (2-10% of the total dataset. Notably, the latest

YOLO models (v8 to v12) consistently achieved high-quality performance metrics, with Precision  $\approx 95\%$ , Recall  $\approx 95\%$  and AP<sub>50:95</sub>  $\approx 84\%$ , underscoring the robustness of these architectures for biomedical applications.

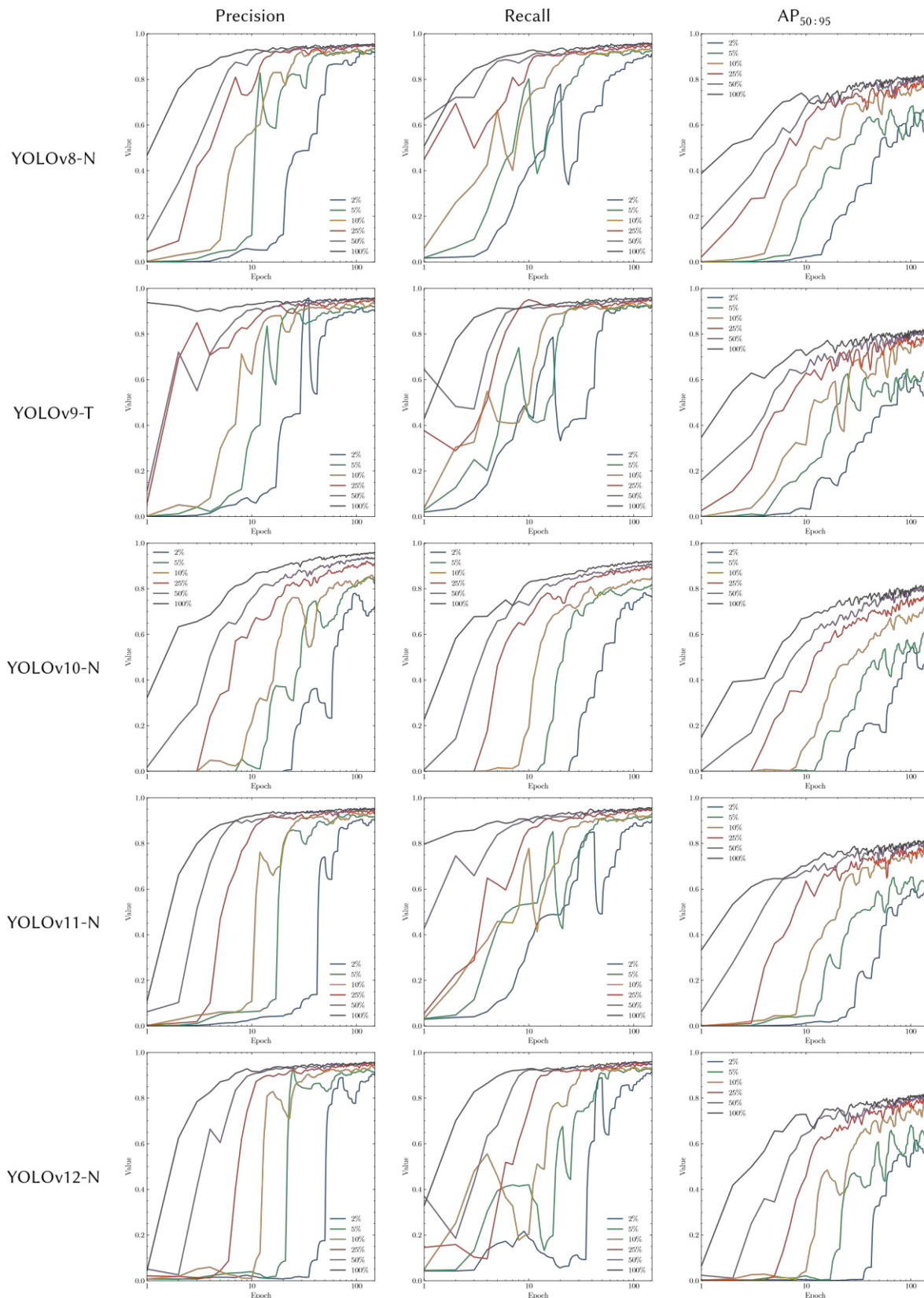


Figure 3. Validation metric curves. Each row presents the model validation curves. To smooth the curves, an exponential moving average with  $\alpha = 0.6$  is performed (logarithmic scale on x-axis)



Moreover, the models exhibit strong adaptability to small training samples, a critical feature for practical implementations where annotated data is often scarce. The introduction of this dataset, coupled with the demonstrated potential of modern YOLO architectures in this context, lays the foundation for further exploration of deep learning approaches in cell viability assays.

## References

1. B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, M. Humayun, YOLOv5-fpn: A robust framework for multi-sized cell counting in fluorescence images, *Diagnostics* 13 (2023) 2280. <http://dx.doi.org/10.3390/diagnostics13132280>.
2. C. Hu, S. He, Y. J. Lee, Y. He, E. M. Kong, H. Li, M. A. Anastasio, G. Popescu, Livedead assay on unlabeled cells using phase imaging with computational specificity, *Nature Communications* 13 (2022). <http://dx.doi.org/10.1038/s41467-022-28214-x>.
3. S. Lypez Flyrez, A. González-Briones, G. Hernández, C. Ramos, F. de la Prieta, Automatic cell counting with yolov5: A fluorescence microscopy approach, *International Journal of Interactive Multimedia and Artificial Intelligence* 8 (2023) 64. <http://dx.doi.org/10.9781/ijimai.2023.08.001>.
4. M. Lu, W. Shi, Z. Jiang, B. Li, D. Ta, X. Liu, Deep learning method for cell count from transmitted-light microscope, *Journal of Innovative Optical Health Sciences* 16 (2023). <http://dx.doi.org/10.1142/S1793545823500049>.
5. Q. Lin, G. Ye, J. Wang, H. Liu, Roboflow: a data-centric workflow management system for developing ai-enhanced robots, in: A. Faust, D. Hsu, G. Neumann (Eds.), *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, PMLR, 2022, pp. 1789–1794. <https://proceedings.mlr.press/v164/lin22c.html>.
6. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, p. 779788. <http://dx.doi.org/10.1109/cvpr.2016.91>.
7. J. Terven, D.-M. Cyrdova-Esparza, J.-A. Romero-González, A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas, *Machine Learning and Knowledge Extraction* 5 (2023) 16801716. <http://dx.doi.org/10.3390/make5040083>.
8. M. L. Ali, Z. Zhang, The yolo framework: A comprehensive review of evolution, applications, and benchmarks in object detection, *Computers* 13 (2024) 336. <http://dx.doi.org/10.3390/computers13120336>.
9. M. Yaseen, What is yolov8: An in-depth exploration of the internal features of the nextgeneration object detector, 2024. <https://arxiv.org/abs/2408.15857>.
10. C.-Y. Wang, I.-H. Yeh, H.-Y. Mark Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, Springer Nature Switzerland, 2024, p. 121. [http://dx.doi.org/10.1007/978-3-031-72751-1\\_1](http://dx.doi.org/10.1007/978-3-031-72751-1_1).
11. A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding, Yolov10: Real-time end-to-end object detection, 2024. <https://arxiv.org/abs/2405.14458>.
12. R. Khanam, M. Hussain, Yolov11: An overview of the key architectural enhancements, 2024. <https://arxiv.org/abs/2410.17725>.
13. Y. Tian, Q. Ye, D. Doermann, Yolov12: Attention-centric real-time object detectors, 2025. <https://arxiv.org/abs/2502.12524>.
14. G. Jocher, J. Qiu, A. Chaurasia, Ultralytics yolo, 2025. <https://doi.org/10.5281/zenodo.15514468>.
15. T. Dao, D. Y. Fu, S. Ermon, A. Rudra, C. R' e, FlashAttention: Fast and memory-efficient exact attention with IO-awareness, in: *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
16. T. Dao, FlashAttention-2: Faster attention with better parallelism and work partitioning, in: *International Conference on Learning Representations (ICLR)*, 2024.
17. I. Loshchilov, F. Hutter, Decoupled weight decay regularization, 2017. <https://arxiv.org/abs/1711.05101>.
18. T. Kumar, R. Brennan, A. Mileo, M. Bendeche, Image data augmentation approaches: A comprehensive survey and future directions, *IEEE Access* 12 (2024) 187536187571. <http://dx.doi.org/10.1109/ACCESS.2024.3470122>.
19. A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A. A. Kalinin, Albumentations: Fast and flexible image augmentations, *Information* 11 (2020). <https://www.mdpi.com/2078-2489/11/2/125>.
20. J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, E. A. Chávez-Urbola, A comprehensive survey of loss functions and metrics in deep learning, *Artificial Intelligence Review* 58 (2025). <http://dx.doi.org/10.1007/s10462-025-11198-7>.
21. J. Li, G. Zhu, C. Hua, M. Feng, B. Bennamoun, P. Li, X. Lu, J. Song, P. Shen, X. Xu, L. Mei, L. Zhang, S. A. A. Shah, M. Bennamoun, A systematic collection of medical image datasets for deep learning, *ACM Computing Surveys* 56 (2023) 151. <http://dx.doi.org/10.1145/3615862>.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-852-865

## Доверенная самообучающаяся полимасштабная сегментация полнослайдовых гистологических изображений для диагностики опухолей

В. Рехамния, А. М. Трыкин, Е. П. Васильев, В. Е. Турлапов

Исследовательский центр в области искусственного интеллекта, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

**Аннотация.** Гистологические полнослайдовые изображения (WSI) являются самым точным и информационно богатым источником для диагностики и исследования опухолей мозга. Они содержат иерархию изображений кратных разрешений, подобную MIP-текстурам, и создают возможность использовать в процессе сегментации признаки разного масштаба одновременно. Использование искусственного интеллекта (ИИ) в этой области тормозится высокой трудоемкостью разметки таких изображений и высокими требованиями доверенности результата ИИ в медицине. Предложен доверенный алгоритм конвейера обработки WSI, который решает указанные выше проблемы полностью на основе только самообучения, используя в своей структуре только объяснимые блоки: 1) автоэнкодер (АЕ) как инструмент выбора уровня WSI, соответствующего ткани и дополняющего клеточный уровень, имеющий максимальное разрешение; 2) кластеризации сегментируемых изображений по разномасштабным тканевым и клеточным признакам; 3) двухстадийный метод Оцу (TSMO) как инструмент выбора естественных порогов признака для сегментации объектов; 4) сегментации ядер клеток и оценки степеней поражения по геометрическим признакам формы. Исследование работоспособности алгоритма выполнено на неразмеченном датасете UPenn-GBM. Среднеквадратическая ошибка (MSE) АЕ составила 0,0016. Точность сегментации ядер клеток количественно оценена на размеченном датасете MoNuSeg. Она составила 77,46 % по метрике F1-score, что превосходит результаты аналогов, также использовавших только самообучение.

**Ключевые слова:** доверенный искусственный интеллект, диагностика опухолей, гистологические полнослайдовые изображения, полимасштабная сегментация WSI, кластеризация, сегментация тканей, сегментация ядра клетки, самообучаемая сегментация, автоэнкодер

## Trusted unsupevised multiscale segmentation of whole-slide histological images for tumor diagnosis

W. Rehamnia, A. M. Trykin, E. P. Vasiliev, V. E. Turlapov

Research Center for Artificial Intelligence, Lobachevsky State University, Nizhny Novgorod, Russia

**Abstract.** Histological whole-slide images (WSI) are the most accurate and information-rich source for diagnostics and research of brain tumors. They contain a hierarchy of images of multiple resolutions, similar to MIP textures, and create the ability to use features of different scales simultaneously in the segmentation process. The use of artificial intelligence (AI) in this area is hampered by the high labor intensity of labeling such images and high requirements for the trustworthiness of the AI result in medicine. A trusted algorithm for the WSI processing pipeline is proposed, which solves the above problems entirely on the basis of self-training only, using only explainable blocks in its structure: 1) an autoencoder (AE) as a tool for selecting the WSI level corresponding to the tissue and complementing the cellular level with the maximum resolution; 2) clustering of segmented images by different-scale tissue and cellular features; 3) Two-Stage Multithreshold Otsu method as a tool for selecting natural thresholds of features for object segmentation. 4) segmentation of cell nuclei and assessment of damage degrees based on geometric shape features. The algorithm performance study was performed on the unlabeled UPenn-GBM dataset. The mean square error (MSE) AE was 0.0016. The accuracy of cell nucleus segmentation was quantitatively assessed on the labeled MoNuSeg dataset. It was 77.46 % by the F1-score metric, which exceeds the results of analogs that also used only unsupervised models.

**Keywords:** trusted artificial intelligence, tumor diagnostics, histological full-slide images, multiscale WSI segmentation, clustering, tissue segmentation, cell nucleus segmentation, self-learning segmentation, autoencoder

### Введение

Использование результатов гистологии в полнослайдовой (WSI) форме стало стандартом современной цифровой патологии, позволяя проводить детальный многомасштабный анализ образцов тканей с высоким разрешением (Wright, et al., 2013) (Gu, et al., 2018). Вместе с тем гигапиксельные размеры WSI сами по себе создают проблемы для эффективности сегментации (Ronneberger, et al., 2015; Gu, et al., 2018). Достижение высокой точности сегментации эффективных моделей глубокого обучения (DL) требует обширных аннотаций на уровне пикселей, выполняемых экспертами. При этом результаты разметки одного датасета часто не обобщаются на различные клинические образцы и тем



более датасеты (Peng and Wang, 2021), (Bejnordi, et al., 2017). Непрозрачность моделей DL также не соответствует требованиям объяснимости и доверенности, которые обязательны для медицины (Tajbakhsh, et al., 2020). Наличие шума или смещения данных могут существенно снизить точность модели (Brixteel, et al., 2022; Roy, et al., 2024).

В результате появился интерес к разработке алгоритмов сегментации без учителя. Так, ученые Tajbakhsh и др. (2020) разработали методы самообучения для сегментации ядра клетки, достигшие оценки Dice score 74,77 %. Javed и др. (2024) предложили использовать для сегментации WSI взаимное обучение трансформеров (transformer) без учителя. Предложены и различные подходы неглубокого обучения. Например, Kochetov и др. (2024), Zhang и др. (2024) предложили алгоритм сегментации ядра на основе неглубокого обучения с оценкой F1 66,70 % и 70-75 % соответственно. Хотя эти методы более вычислительно эффективны и интерпретируемы, чем модели глубокого обучения, они не достигли необходимого уровня устойчивости и точности, необходимых в клинической практике. Их важным недостатком является также ограниченность только сегментацией ядра на уровне самого высокого разрешения WSI.

Использование машинного обучения (МО) в поиске изображений по их содержанию стало темой интереса в недавних исследованиях. Ozturk (2020) предложили метод разметки высокоуровневых признаков на основе стекового автокодировщика для поиска медицинских изображений по содержанию. Guo и др. (2020) сосредоточились на поиске изображений с сохранением конфиденциальности (PPIS) с использованием сверточных нейронных сетей (CNN) для безопасной классификации и поиска по закодированным медицинским изображениям большого размера. Singh с соавторами (2021) предложили модель с четырьмя сверточными слоями для поиска изображений по содержанию в данных о знаменитостях с использованием глубоких CNN. Kanwal и др. (2020) предложили метод, который объединяет симметрию, оценки FAST, фильтрацию формы и положения с CNN для поиска изображений большого размера, и достигли высокоточных результатов на сложных датасетах. Komura и др. (2022) исследовали универсальное кодирование гистологии панкреатических заболеваний с помощью глубоких текстурных оценок. Samalan с соавторами (2020) разработали систему поиска изображений на основе контента для изображений барабанной перепонки с использованием глубокого обучения. Zhuo и Zhou (2020) предложили метод дискриминантного представления низкой размерности для поиска изображений дистанционного зондирования Земли с высоким разрешением. Деерак и Амеер (2020) исследовали поиск опухолей в изображениях магнитно-резонансной томографии (МРТ) мозга, используя оценку сходства по контрастным потерям при кодировании в GoogLeNet. Kumar и др. (2021) представили U-Net-подобную нейронную сеть, реализующую поиск изображений по объектам присутствующим на изображении. Улучшены характеристики точности, достоверности и полноты. Monowar с соавторами (2022) представили AutoRet, самообучаемую рекуррентную сеть для поиска изображений на основе контента. В совокупности эти исследования демонстрируют разнообразные применения и достижения в методах нейронных сетей для поиска изображений по содержанию в различных медицинских и немедицинских областях.

Сегментация WSI без учителя использует, как правило, передовые методы машинного обучения для анализа гистологических изображений без необходимости обширного ручного аннотирования. Предложенные методы варьируются от слабообучаемых, самообучаемых до полностью самообучаемых подходов. Gadermaier и др. (2019) исследовали четыре метода и их комбинации для образования нового метода сегментации без учителя. Первый метод, SDS, использует модель U-Net для точной обучаемой сегментации. Второй метод преобразует входное изображение из размеченного датасета, имитируя замену красителя на альтернативный перед подачей на U-Net, и таким образом решает проблему разметки датасета с другим красителем. Третий метод, SDU, обсуждавшийся в их более раннем исследовании (Gadermaier, et al.), включает создание синтетических датасетов (с готовой аннотацией). Далее некоторый размеченный датасет может быть аугментирован с помощью генеративно-состязательной сети (GAN). Одним из ограничений данного исследования является использование синтетических изображений, основанных на визуальных предубеждениях GAN.

В статье Sebai и др. (2020) представлен целый фреймворк глубокого обучения на основе RCNN для обнаружения митозов на гистопатологических изображениях, включающий спектр подходов: строго обучаемые, слабообучаемые и без учителя. В работе Li и др. (2022) предложен фреймворк сегментации раковых клеток без учителя для гистопатологических изображений, использующий архитектуру глубокой сети U-Net. Авторы использовали самообучение (контрастное обучение) для этапа аугментации данных и показали, что оно более эффективно для производительности модели и снижения сложности, чем альтернативные стратегии аугментации (по крайней мере, в среде без учителя).

В исследовании van Nierop и др. (2024) представлена модель сегментации клеток без учителя, которая оптимизирует взаимную информацию между парами данных, достигая точной сегментации клеток без разметки экспертом. Их модель превзошла SOTA обучаемых и самообучаемых моделей, достигнув оценки Жаккара 0,71 на собственном наборе данных. Cisternino с соавторами (2024) использовали самообучение на огромном наборе данных из 1,7 миллиона изображений для сегментации тканей и прогноза экспрессии РНК. Их стратегия превзошла существующие методологии, улучшив оценку силуэта на 43 %, коррелируя морфологию ткани с экспрессией генов. Этот метод превзошел несколько моделей с учителем, продемонстрировав свою эффективность в исследованиях живых клеток.

Хотя предыдущие исследования успешно решили проблему разметки данных, они по-прежнему страдают из-за высоких вычислительных ресурсов и проблем с доверенностью. Поэтому Zhang и др. (2024) предложили более быстрое и интерпретируемое решение, использующее для сегментации ядер несколько преобразований цветовых моделей с k-средними и нечеткими c-средними (FCM). Кроме того, в комплексном обзоре Wei и др. (2024) выявили несколько будущих пробелов, включая интерпретацию глубокой кластеризации, выбор гиперпараметров (DL и параметры кластеризации), многовидовую глубокую кластеризацию (как показано в работе Chen и др. (2023)). Эти результаты вдохновляют на дальнейшие исследования более совершенных алгоритмов для полностью самообучаемой сегментации гистопатологических данных без потери ее точности.

### Постановка задачи

Из Введения следует, что наилучшим современным решением для сегментации гистологических WSI-изображений является подход, применяющий объяснимый и доверенный искусственный интеллект, максимально использующий предобработку, учитывающую все уровни WSI. То есть подход должен быть и полимасштабным и не требующим ручной разметки (т.е. unsupervised). Использование низких (тканевых) уровней разрешения WSI для повышения точности сегментации ядер в клетках мотивировано достижениями Kochetov и др. (2024) и Zhang с соавторами (2024), а объяснимость и доверенность результатов конвейера может быть достигнута посредством использования наглядных высокоуровневых признаков сегментации.

Говоря о доверенности подхода мы полагаемся на наше системное видение понятия доверенного искусственного интеллекта (Trusted Artificial intelligence, TAI), удовлетворяющее, тем не менее, существующим стандартам. Предлагается системное определение понятия доверенного ИИ, базированное на трёх основных свойствах-требованиях, понимаемых как категории:

Доверенный ИИ – это искусственный интеллект, удовлетворяющий трем категориальным требованиям: 1) управляемость человеком; 2) безопасность; 3) объяснимость.

В случае медицины или здравоохранения эти три свойства должны быть реализованы для треугольника «Врач–ИИ–Пациент», требующего доверенных отношений.

Все остальные требования, раскрытые в стандартах, выводятся из указанных трёх, раскрывая категории. Например,

- Человек: врач, пациент, семья, коллектив, население, правовые акты предприятий, управлений, министерств, субъектов государства,...
- Врач: человек, бригада скорой помощи, медицинское отделение, клиника, функциональное подразделение здравоохранения, клинические рекомендации...

- Пациент: человек, семья, коллектив предприятия или отрасли, общественная организация, население территории, граждане государства.

В данной работе сделана попытка обеспечить решению задачи сегментации WSI объяснимость и безопасность для представителей всех участников тройки «Врач–ИИ–Пациент» и контролируемость результата человеком. Наши решения в направлении построения алгоритма сегментации опухолей по WSI-данным гистологических изображений, который мог бы существенно повысить доверенность результата, предлагаются в следующих разделах.

## Теория

### 1. Датасеты

В данном исследовании мы использовали набор данных UPenn-GBM, включающий гистологические, MPT-данные и данные пациентов с глиобластомой, предоставленные Bakas и др. (2022). Мы ограничились его гистологической частью, включающей 71 WSIs в формате NDPI. Эти слайды получены с 9 различными уровнями увеличения, хотя для повышения вычислительной эффективности библиотека OpenSlide Python интерпретирует их как 18 уровней. Примеры фрагментов, извлеченных из этих слайдов, показаны на рисунке 1.

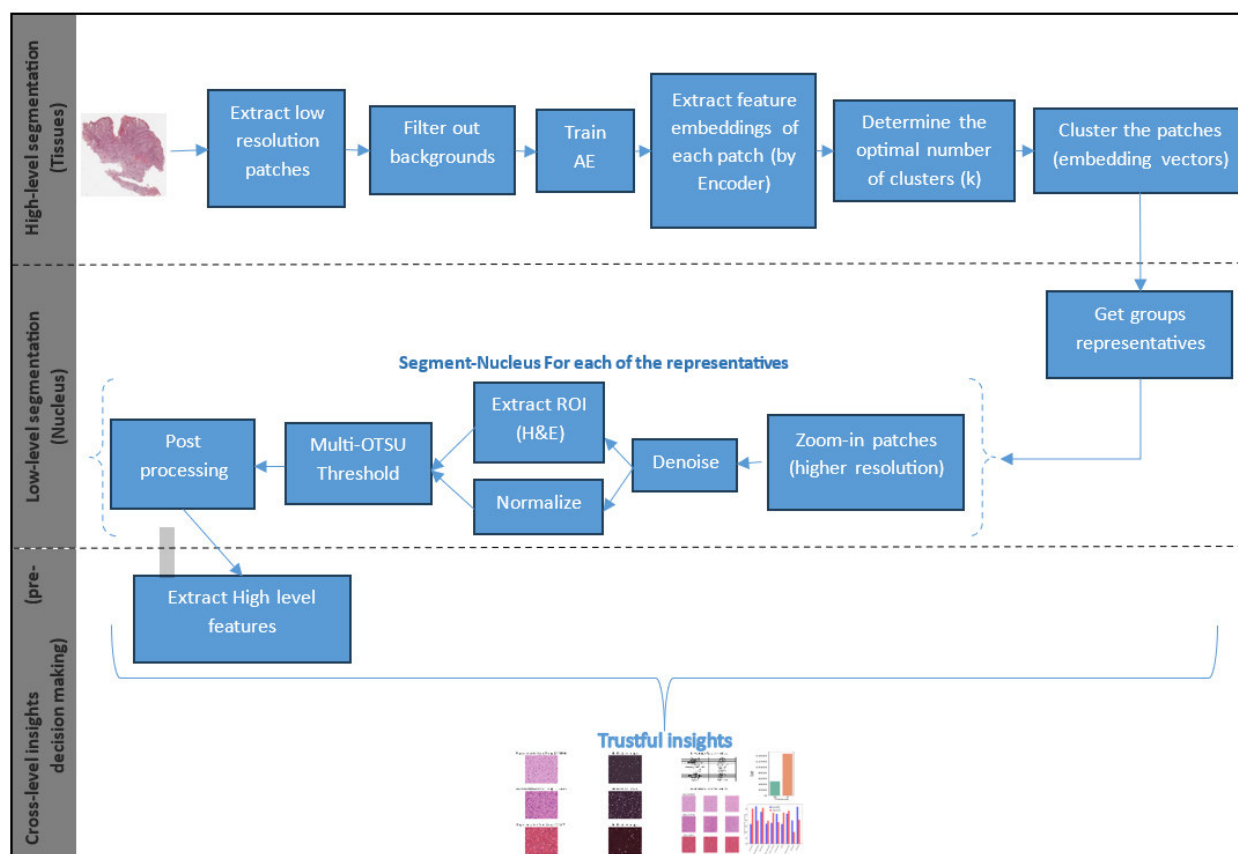


Рисунок 1. Структура алгоритма доверенной обработки: 1) уровень сегментации тканей (верхний ряд блоков); 2) уровень сегментации ядер (второй ряд блоков) – максимальное разрешение; 3) экстракция признаков высокого уровня, объединяющих оба уровня, для принятия доверенного решения

Учитывая, что набор данных UPenn-GBM не содержит какой-либо разметки, для оценки точности конвейера сегментации ядра и ее переносимости на другие датасеты дополнительно потребовался размеченный датасет. Для этой цели мы использовали датасет MoNuSeg, представленный Kumar и др. (2017). Этот набор данных содержит хорошо аннотированные изображения тканей, полученные из различных органов в различных больницах.

Тестовый набор состоит из 14 изображений TIF (по два образца гистопатологии из семи органов: молочной железы, почки, печени, предстательной железы, мочевого пузыря, толстой кишки, желудка), каждое размером  $1024 \times 1024 \times 3$ , с их аннотациями в формате XML.

## 2. Предлагаемый алгоритм

Предлагается алгоритм доверенной обработки WSI, структура которого представлена на рисунке 1. Эта структура должна обеспечивать решение трех групп задач: сегментация тканей; сегментация ядер клеток; кросс-уровневые решения.

**Сегментация тканей (низкое разрешение).** На этом этапе распознаются ткани, представленные в данных, и отображаются их области. Начинают с разбиения нижнего уровня WSI на более мелкие патчи, отфильтровывая их фоновую часть на основе статистических порогов, после чего автоэнкодер (АЕ) обучается реконструировать эти патчи для уровней тканей. После обучения кодирующая часть АЕ извлекает скрытое пространство признаков размером 256 (2x2x64) для каждого патча (встраивание признаков). Патчи кластеризуются с помощью алгоритма k-средних на основе этих признаков (гиперпараметр k определяется оптимальным срезом иерархической кластерной дендрограммы). Наконец, используя подход учета неопределенности (основанный на расстоянии до центроидов), для каждой ткани определяют N наиболее репрезентативных патчей (ближайших к центроиду по атрибутам).

**Сегментация ядер (высокое разрешение).** На данном этапе выполняется сегментация ядер клеток. Начинаем с шага шумоподавления, который удаляет артефакты из патчей. Восстановленный канал окрашивания гематоксилином объединяется с нормализованным патчем. Процесс нормализации описан в Масенко и др. (2009). После объединения выполняется пороговая пост-обработка (Post-processing) оптимизированным двухстадийным методом multi-Otsu (TSMO) (Huang, 2009), который дает вариант маски сегментации ядра, зашумленный квантованными порогами. Шум маски сегментации корректируется постобработкой операциями матморфологии (closing, opening, dilation). Для выбора лучшей маски исследованы три варианта пост-обработки результата TSMO, формирующие три разных маски сегментации: 1) первый использует операции closing, затем opening; 2) второй – одну операцию dilation; 3) третий – медианное размытие первого варианта маски.

**Кросс-уровневые решения** (принятие предварительного решения, подконтрольного специалисту согласно требованиям ТАИ). На этом этапе извлекаются суммарные признаки из выходных данных двух конвейеров (тканевого – текстурные признаки и ядерного – морфологические) и объединяются с визуализациями, чтобы сделать более быстрым и точным контроль со стороны врача-невропатолога.

## 3. Подготовка и детали реализации эксперимента

Размеры патчей и уровни WSI устанавливаются эмпирически для обеспечения пространственной согласованности на разных уровнях. Для получения четкой сегментации ядра начинаем с максимального разрешения (или MaxSize) 1024x1024 ( $2^{10} \times 2^{10}$ ) пикселя (MaxSize=1024). Затем уменьшаем разрешение, с каждым уровнем вдвое, до уровня, на котором становится различима только ткань, что занимает Nsteps=6 шагов. Линейный размер патча на этом уровне составляет 16 пикселей; он получается путём подстановки Nsteps=6 в формулу (1):

$$\text{LinearSize(Nsteps)} = \text{MaxSize} / (2^{\text{Nsteps}}). \quad (1)$$

Предлагаемая структура реализована на Python с использованием фреймворка PyTorch. Эксперименты проведены на персональном компьютере с графическим процессором NVIDIA GTX 1060. Для повышения эффективности обучения при высокоуровневой сегментации кодер включает несколько слоёв из предобученной модели ResNet18 (обученной на наборе данных ImageNet), а декодер представляет собой простую сверточную нейронную сеть (со случайными начальными весами). Сеть АЕ обучалась с использованием оптимизатора Adam с размером батча 128, скоростью обучения 0,001 и уменьшением веса  $1e-05$ , а также с использованием нормализации пакета и ранней остановки (после 10 шагов) в качестве стратегий регуляризации для повышения точности и обобщения. Кроме того, использовали асинхронный режим предварительной выборки патчей с 12 рабочими процессами для эффективного устранения узких мест ввода-вывода, что значительно ускоряет обучение, особенно при слабом графическом процессоре и большом размере пакета (Lin et al., 2018). Проиллюстрированные выше этапы тканевой сегментации исследованы экспериментально. Например, эксперимент показывает, что двухстадийный многоклассовый метод Оцу, как правило, быстрее и гораздо более устойчив к различию фрагментов, чем метод k-средних, который медленнее и иногда требует настройки гиперпараметра k для повышения точности. Набор данных UPenn-GBM разделён на

три подмножества: обучение (70 %) (124494 патча), валидация (15 %) (26678 патчей) и тестирование (15 %) (26678 патчей).

#### 4. Метрики оценки

*Среднеквадратическая ошибка* (MSE) оценивает ошибку прогнозирования модели для непрерывных результатов. Она рассчитывается по формуле (2), где более низкая MSE предполагает улучшение качества модели (James et al., 2013):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2)$$

*Оценка по Дайсу* (DSC), также известная как F1 в контексте классификации, представляет собой гармоническое среднее значение Precision и Recall, что обеспечивает баланс между этими двумя метриками, и особенно полезна в случаях неравномерного распределения классов. Она рассчитывается по формуле (3) или (4), где более высокая оценка предполагает улучшение качества модели (Sokolova and Lapalme, 2009):

$$Dice = 2 |A \cap B| / |A + B| = 2 \cdot TP / (2 \cdot TP + FP + FN) \quad (3)$$

или

$$F1 = 2 \text{ Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}) = 2 \cdot TP / (2 \cdot TP + FP + FN). \quad (4)$$

Коэффициент сходства Жаккара (JSC), также известный как «пересечение над объединением» (IoU), предсказанного  $|B|$  и истинного  $|A|$  объектов рассчитывается по формуле (5), где большее значение IoU свидетельствует об улучшении точности модели (Everingham et al., 2010):

$$J(A,B) = |A \cap B| / |A \cup B|. \quad (5)$$

MSE используется для оценки ошибки сегментации тканей автоэнкодером, тогда как Дайс и коэффициент Жаккара применяются для оценки сгенерированных масок ядер при сегментации ядер клеток.

#### 5. Признаки, экстрагированные из уровней WSI, для сегментации гистологических изображений

В качестве признаков сегментации исследовались две группы признаков: 1) текстурные – на основе Gray Level Co-occurrence Matrix (GLCM) плюс цвет окрашивания; 2) морфологические. Текстурные признаки эффективны как для сегментации тканей (включая их состояние), так и для детектирования ядра на фоне клетки. Морфологические признаки наиболее эффективны для определения состояния ядра по его форме. Список признаков получил в итоге следующие 11 позиций.

Текстурные признаки (вычисляются по значениям элементов GLCM-матрицы):

1. *Contrast*: измеряет контраст интенсивности между пикселем и его соседом по всему изображению. Высокий контраст указывает на большую разницу между значениями соседних пикселей.

2. *Homogeneity*: оценивает близость распределения элементов в GLCM-матрице к диагонали. Высокая однородность означает наличие существенных областей с однородной по интенсивности заливкой.

3. *Intensity*: средняя интенсивность пикселей в ядре, отражающая яркость.

4. *Staining Intensity*: интенсивность окрашивания в ядре, часто указывающая на концентрацию конкретного биомаркера или красителя.

5. *Number of Nuclei (num nuclei)*: количество отдельных ядер, обнаруженных в сегментированной области.

Морфологические признаки:

6. *Size*: площадь области ядра, измеренная в пикселях.

7. *Circularity*: мера того, насколько форма ядра близка к идеальному кругу; рассчитывается по формуле (6), где значение 1,0 соответствует идеальному кругу:

$$\text{Circularity} = 4\pi \times \text{Area} / \text{Perimeter}^2. \quad (6)$$

8. *Density*: количество ядер на единицу площади, отражающее плотность их упаковки, рассчитывается по формуле (7):

$$\text{Плотность} = (\text{Количество ядер}) / (\text{Площадь области интереса}). \quad (7)$$

9. *Eccentricity*: отношение расстояния между фокусами эллипса, который наилучшим образом соответствует ядру, к длине его большой оси. Значение близкое к 1 соответствует вытянутой форме, а 0 — идеальному кругу:

$$e = \sqrt{1 - (b/a)^2}, \quad (8)$$

где  $e$  — эксцентриситет эллипса;  $a$  и  $b$  — длина наибольшей/наименьшей оси эллипса.

10. *Spread*: пространственное распределение ядер, описывающее, насколько широко они разбросаны в области; рассчитывается как стандартное отклонение евклидовых попарных расстояний между центроидами ядер.

## Результаты экспериментов

### 1. Сегментация тканей

Обучение было прекращено с помощью техники раннего останова после 27 эпох, поскольку потери валидации перестали понижаться. На рисунке 2 показано, что 16-я эпоха дала наилучшую ошибку валидации 0,00166, в то время как наилучшая ошибка обучения, равная 0,00138, была достигнута последней эпохой. Фаза тестирования демонстрирует ту же потерю валидации.

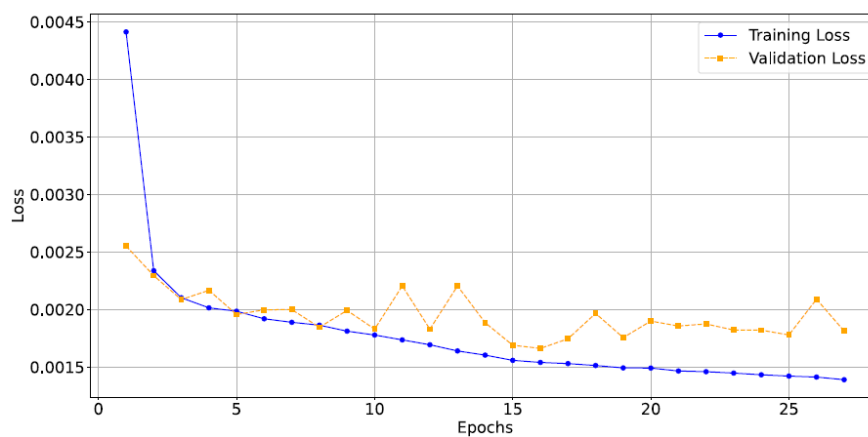


Рисунок 2. Графики потерь и валидации при обучении Training vs validation loss

В ходе обучения, всякий раз как ошибка валидации улучшается, сохраняется рисунок, аналогичный рисунку 3, на котором сравниваются 5-6 случайных участков (парами: исходное разрешение и его реконструкция) для визуальной оценки эффективности модели в задаче захвата важных характеристик ткани.

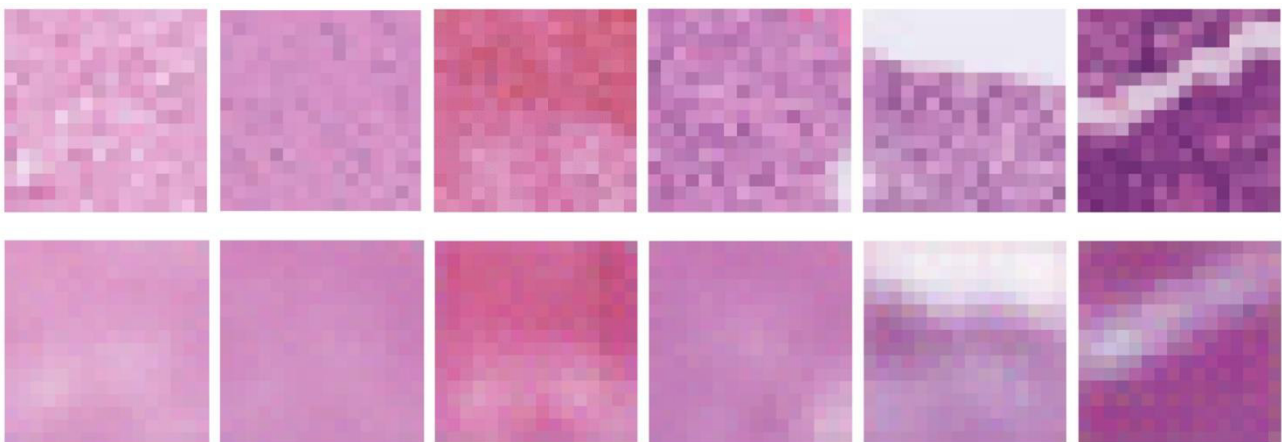


Рисунок 3. Сопоставление 6 пар патчей (фрагментов): исходное разрешение (*верхний ряд*) и его реконструкция (*нижний ряд*) патчей после одной эпохи обучения

После этого фаза глубокой кластеризации использует сглаженное латентное пространство, сформированное кодирующей частью 256 (2x2x64). Для определения подходящего количества групп,



как упоминалось, использована иерархическая дендрограмма кластеризации с использованием метода лучшего отсечения дерева, который предполагает выбор отсечения по наибольшему расстоянию на дендрограмме. На рисунке 4 представлен пример отсечения по наибольшему расстоянию с двумя тканями.

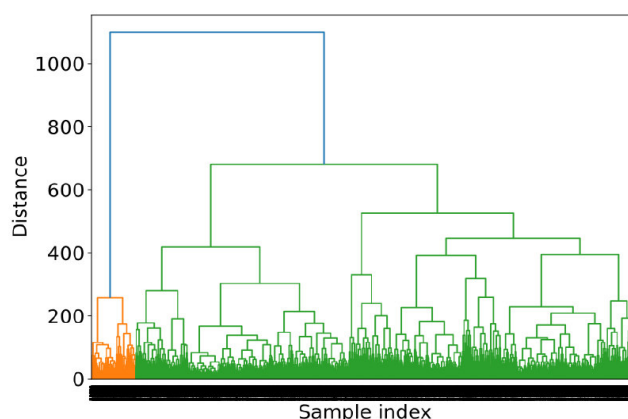


Рисунок 4. Дендрограмма иерархической кластеризации WSI с разделением тканей на два класса (выделены оранжевым и зеленым цветом) по признаку наибольшего расстояния

## 2. Сегментация ядер клеток

Как показано выше, сегментация тканей включает несколько этапов предварительной обработки и создает три различные маски сегментации. На рисунке 5 показаны этапы и результаты семантической сегментации ядер и три маски, сформированные в идеальном сценарии (без шума).

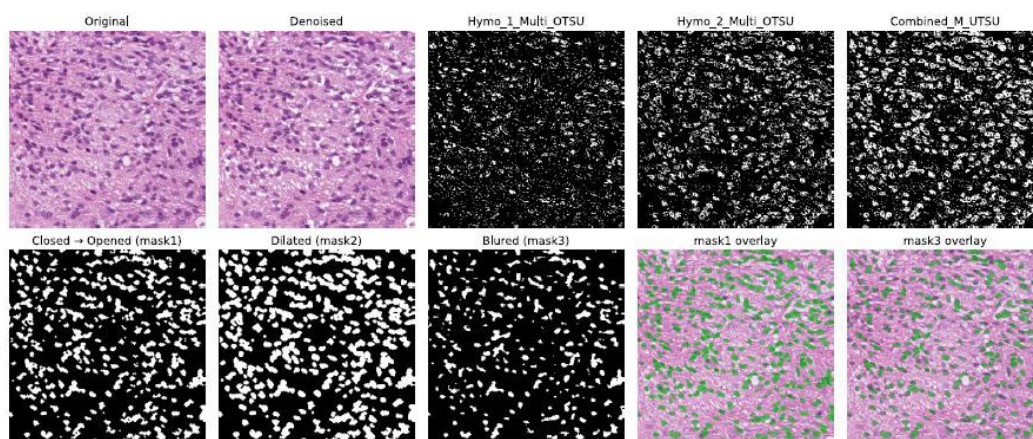


Рисунок 5. Визуализация этапов сегментации ядер клеток (без шума)

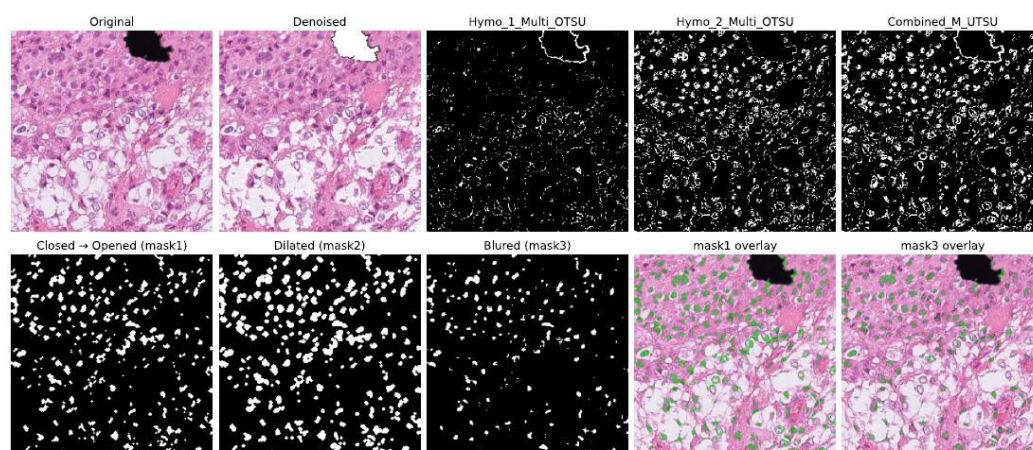


Рисунок 6. Визуализация этапов сегментации ядер клеток (с шумом 1-го типа)



Некоторые WSI из UPenn-GBM содержат значимый шум, поэтому перед дальнейшей обработкой к входному фрагменту применяется алгоритм шумоподавления. Анализ фрагментов выявил два вида шума: чёрную область (рис. 6) и серый шум (рис. 7).

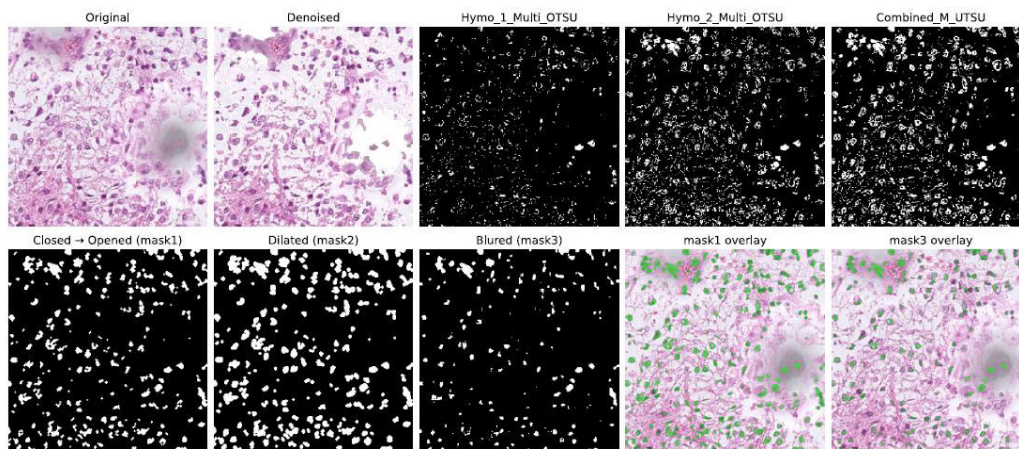


Рисунок 7. Визуализация этапов сегментации ядер клеток (с шумом 2-го типа)

Проведены эксперименты, определившие влияние такого шума на сегментацию ядер клеток и качество удаления шума. Результаты показали, что в отсутствие шумоподавления алгоритм может детектировать шумовые области как большие ядра. А шумоподавление удаляет оба вида шума, не жертвуя важными деталями изображения. Рисунок 7 демонстрирует пример, где метод успешно подавил шум и обнаружил на его месте два скрытых ядра.

### 3. Сравнение результатов алгоритма с недавними публикациями

Важно также сравнить достигнутую нами точность с точностью современных методов на популярных контролируемых наборах данных. Мы сделали это на датасете MoNuSeg. В таблице 1 показана точность сегментации ядер клеток каждого из 7 органов представленных MoNuSeg, достигнутая тремя нашими масками в метрике F1-score.

Таблица 1. Точность F1-сегментации ядер клеток для 7 органов из датасета monuseg тремя нашими масками

Organ	Mask1	Mask2	Mask3
Bladder	0.7943	0.7323	0.7622
Brain	0.8159	0.7696	0.8115
Breast	0.7418	0.6426	0.7426
Colon	0.7617	0.6408	0.7483
Kidney	0.7842	0.6927	0.7595
Lung	0.7428	0.5957	0.7605
Prostate	0.7703	0.7074	0.7398
<b>Average</b>	<b>0.7746</b>	<b>0.6867</b>	<b>0.7614</b>

Mask1 доминирует для большинства органов, и орган Brain имеет здесь лучшую оценку ( $F1=0,8159$ ), а орган Breast – самую низкую ( $F1=0,7418$ ). Точность Mask3 немного превосходит Mask1 для Breast (0.7426 против 0.7418) и более заметно для Lung (0.7605 против 0.7428). В итоге Mask1 является рекомендуемой.

На рисунке 8 показаны ground truth маски сегментации 7 органов из этого датасета, размеченные экспертами, и 3 маски, построенные нашим алгоритмом.

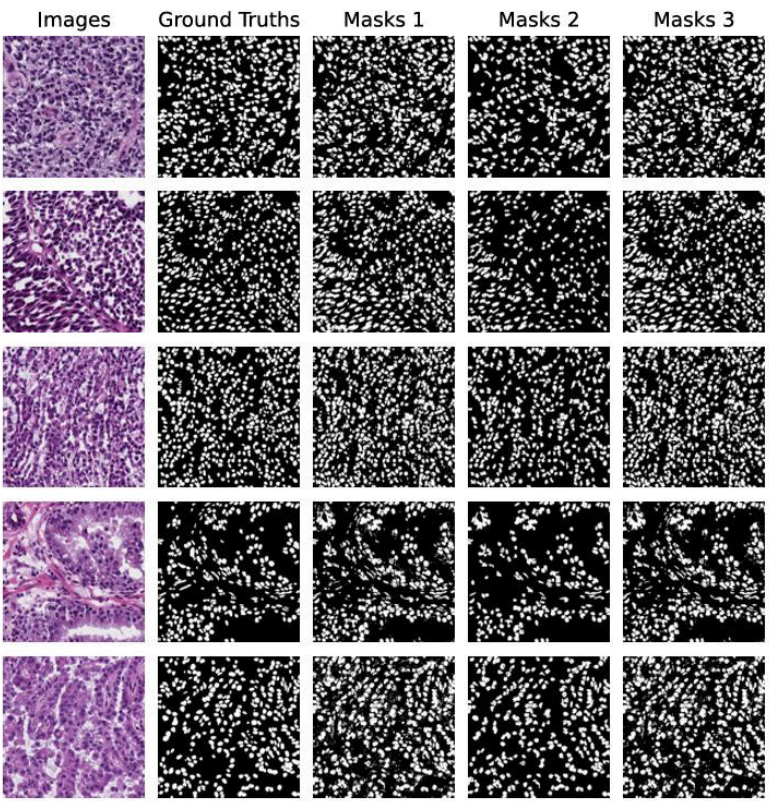


Рисунок 8. Сравнение на 5 образцах тканей датасета MoNuSeg, ручной разметки ядер (Ground Truth) и 3 масок сегментации, сгенерированных нашим алгоритмом

В таблице 2 представлены результаты недавних исследований сегментации ядра также на датасете MoNuSeg, за исключением одного (Kochetov et al., 2024), в котором для оценки использовался другой датасет. Исследования отсортированы в хронологическом порядке, но Supervised и Unsupervised раздельно. Для сравнения использована общая метрика Dice Score.

Таблица 2. Сравнение с недавними исследованиями сегментации ядра

Study	Method	Learning approach	F1(Dice)
Naylor, et al., 2018	max {Mask R-CNN, U-Net, FCN, <u>DIST</u> }	Supervised	78.63
Valanarasu, et al., 2021	LoGo: Local-Global training strategy	Supervised	79.56
Li, et al., 2023	LViT-LW: Language meets Vision Transformer in Medical Image Segmentation	Supervised	80.66
Roy, et al., 2024	GRU-Net: Supervised learning with Gaussian attention and dense skip connections	Supervised	80.35
Showrav and Hasan, 2024	Hi-gMISnet: Generalized segmentation using pGAN with dual-mode attention	Supervised	82.5
Wang, et al., 2024b	U-HRMLP: Multi-scale feature fusion to refine segmentation boundaries	Supervised	80.83
Wang, et al., 2024a	UDTransNet: Learnable skip connections with DAT and DRA modules	Supervised	79.47
Tajbakhsh, et al., 2020	Self-supervised approaches for nucleus segmentation	Self-supervised	74.77
Kochetov, et al., 2024	UNSEG: Bayesian-like framework to segment cells and their nuclei without requiring training data	Unsupervised	66.70
Zhang, et al., 2024	YCbCr color space with: k-means > fuzzy c-means	Unsupervised	75.53
The present study	Unsupervised Poly-scale Histology Segmentation	Unsupervised	77.46

Обсуждение результатов

1. Сравнение результатов UPHS-алгоритма и подходов с обучением

Поскольку наш алгоритм сегментации гистологических изображений может быть охарактеризован как алгоритм, реализованный полностью без учителя и обеспечивающий в процессе сегментации



полимасштабное использование WSI, введем для него краткое наименование UPHS (Unsupervised Poly-scale Histology Segmentation).

*Сравнение UPHS с подходами с учителем.* Значение оценки F1, полученное с помощью UPHS, сопоставимо с 7 представленными в таблице 2 supervised-подходами. Небольшое снижение оценки обусловлено неконтролируемым характером метода, не требующим разметки ядер для их сегментации. Кроме того, хотя решения на основе глубокого обучения (DL) высокоэффективны, однако часто обучение на одном датасете не переносится в полной мере на другие датасеты. Например, хотя модель GRU-Net (Roy, et al., 2024) превзошла все, кроме одного, решения SOTA на наборе данных MoNuSeg (80,35 %, табл. 2), но при обучении на наборе данных TNBC показала резкое снижение точности до 65,98 % в тестировании на MoNuSeg. В отличие от этого, настоящее исследование продемонстрировало большую устойчивость в сохранении точности.

*Сравнение UPHS с подходами без учителя.* Таблица 2 показывает, что UPHS превзошел все 3 приведенных в таблице подхода без учителя. Он превосходит байесовский фреймворк UNSEG (Kochetov, et al., 2024) примерно на 11 %, цветовое пространство YCbCr с k-means (Zhang, et al., 2024) примерно на 2 % и подходы с самообучением (Tajbakhsh, et al., 2020) примерно на 3 %.

2. Кроссуровневая аналитика

Обобщённая кроссуровневая WSI аналитика строится следующим образом. На этапе кроссуровневой аналитики для каждой обнаруженной ткани выбирается группа из  $N$  (в данном исследовании  $N$  принято равным 5) наиболее репрезентативных патчей, которая характеризуется центроидом тканевых признаков. Выбор определяется по критерию удаленности вектора признаков патча от центроида ткани. Выбранные патчи отображаются на соответствующие патчи уровня ядер (с помощью модуля «Open Slide» Python 3.x). Пример выбора для двух тканей показан на рисунке 9.

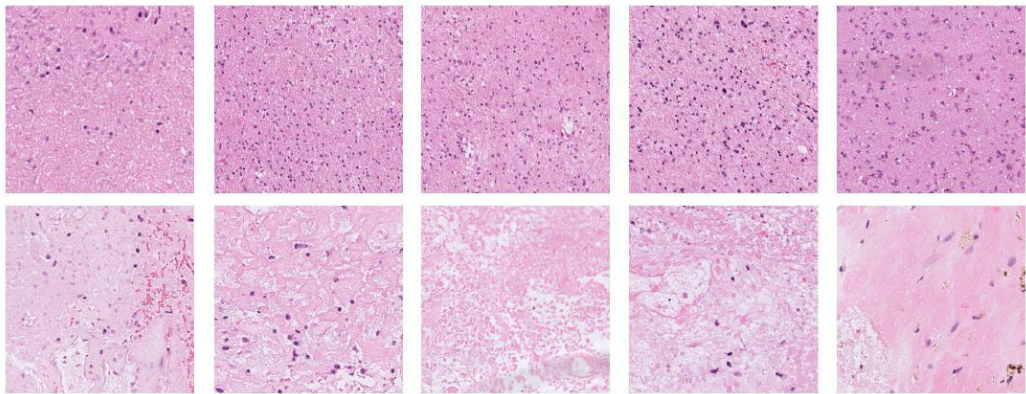


Рисунок 9. Пример 5 репрезентативных патчей для двух тканей

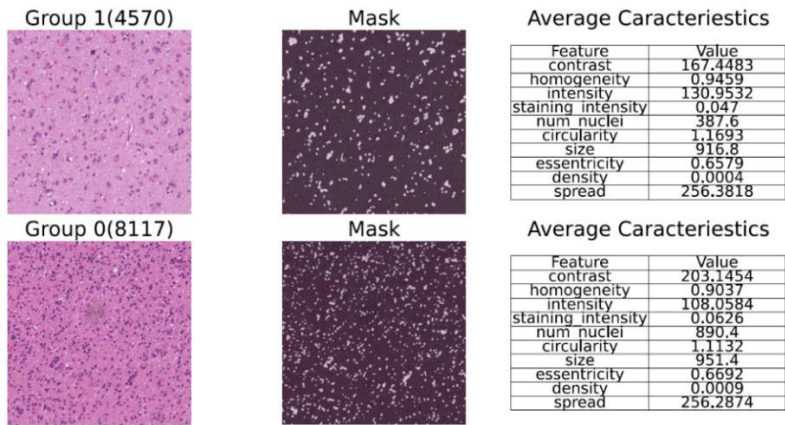


Рисунок 10. Пример репрезентативных патчей 2 тканей (по одному из группы), а также их маски сегментации ядер и векторы признаков, в которые добавлено число ядер и их параметры

Далее выполняются сегментация ядер и расчет их морфологических признаков. На рисунке 10 показаны по одному из пяти (группы) репрезентативных патчей для 2 тканей, а также маски сегментации ядер и вектор признаков каждого патча с указанием числа сегментированных ядер и их усредненных по результатам сегментации параметров.

Признаки патчей могут быть далее использованы для обучения детектированию типа опухоли и степени поражения.

### Заключение

Предложен доверенный (стабильный, безопасный, объяснимый и контролируемый человеком) метод сегментации без учителя для анализа гистологических WSIs (полнослайдовых изображений), названный UPHS (Unsupervised Polyscale Histology Segmentation). Отказ от использования учителя устраняет необходимость в ручной разметке датасетов, снижает требования к вычислительным ресурсам при сохранении конкурентоспособной точности семантической сегментации, предоставляет объяснимую информацию всего из двух репрезентативных уровней и включает учет наличия шума в высокоуровневую сегментацию. Эксперименты показали, что наш алгоритм конкурентоспособен, несмотря на свою простоту. Достигнута среднеквадратическая ошибка (MSE) реконструкции уровня тканей равная 0,0016, число признаков для дальнейшей классификации опухоли снижено до 10, получена наилучшая среди unsupervised подходов точность сегментации ядер с оценкой F1=77,46 %. Дальнейшие исследования будут сосредоточены на повышении точности сегментации ядер, исследовании эффективности признаков сегментации в задаче классификации опухоли, расширении тестирования надежности алгоритма.

Данные и код, подтверждающие результаты данного исследования, доступны по запросу.

### Источник финансирования

Работа выполнена при поддержке Министерства экономического развития Российской Федерации (соглашение о предоставлении гранта № 139-15-2025-004 от 17 апреля 2025 г., ИГК 000000Ц313925P3X0002)

### Список литературы

1. Bakas, S., Sako, C., Akbari, H., Bilello, M., Sotiras, A., Shukla, G., Rudie, J.D., Santamaría, N.F., Kazerooni, A.F., Pati, S., et al., **2022**. The university of pennsylvania glioblastoma (UPenn-GBM) cohort: advanced MRI, clinical, genomics, & radiomics. *Scientific data* 9, 453.
2. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermesen, M., Manson, Q.F., Balkenhol, M., et al., **2017**. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama* 318, 2199–2210.
3. Brixel, R., Bougleux, S., L'ezoray, O., Caillot, Y., Lemoine, B., Fontaine, M., Nebati, D., Renouf, A., **2022**. Whole slide image quality in digital pathology: review and perspectives. *IEEE Access* 10, 131005–131035.
4. Camalan, S., Niazi, M.K.K., Moberly, A.C., Teknos, T., Essig, G., Elmaraghy, C., Taj-Schaal, N., Gurcan, M.N., **2020**. Otomatch: Contentbased eardrum image retrieval using deep learning. *Plos one* 15, e0232776.
5. Chen, J., Mao, H., Woo, W.L., Peng, X., **2023**. Deep multiview clustering by contrasting cluster assignments, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16752–16761.
6. Cisternino, F., Ometto, S., Chatterjee, S., Giacomuzzi, E., Levine, A.P., Glastonbury, C.A., **2024**. Self-supervised learning for characterising histomorphological diversity and spatial rna expression prediction across 23 human tissue types. *Nature Communications* 15, 5906.
7. Deepak, S., Ameer, P., **2020**. Retrieval of brain mri with tumor using contrastive loss based similarity on googlenet encodings. *Computers in biology and medicine* 125, 103993.
8. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., **2010**. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88, 303–338.
9. Gadermayr, M., Gupta, L., Appel, V., Boor, P., Klinkhammer, B.M., Merhof, D., **2019**. Generative adversarial networks for facilitating stainindependent supervised and unsupervised segmentation: a study on kidney histology. *IEEE transactions on medical imaging* 38, 2293–2302.
10. Gadermayr, M., Gupta, L., Klinkhammer, B., Boor, P., Merhof, D. Unsupervisedly training gans for segmenting digital pathology with automatically generated annotations. *arxiv* **2018**. arXiv preprint arXiv:1805.10059.
11. Gu, F., Burlutskiy, N., Andersson, M., Wil'en, L.K., **2018**. Multi-resolution networks for semantic segmentation in whole slide images, in: *Computational Pathology and Ophthalmic Medical Image Analysis: First International Workshop, COMPAY 2018, and 5th International Workshop, OMIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16-20, 2018, Proceedings* 5, Springer, pp. 11–18.

12. Guo, C., Jia, J., Choo, K.K.R., Jie, Y., **2020**. Privacy-preserving image search (ppis): Secure classification and searching using convolutional neural network over large-scale encrypted medical images. *Computers & Security* 99, 102021.
13. Huang, D.-Y., & Wang, C.-H. (2009). Optimal multi-level thresholding using a two-stage Otsu optimization approach. *Pattern Recognition Letters* 30 (2009), 275-284.
14. James, G., Witten, D., Hastie, T., Tibshirani, R., et al., **2013**. An introduction to statistical learning. volume 112. Springer.
15. Javed, S., Mahmood, A., Qaiser, T., Werghi, N., Rajpoot, N., **2024**. Unsupervised mutual transformer learning for multi-gigapixel whole slide image classification. *Medical Image Analysis* 96, 103203.
16. Kanwal, K., Ahmad, K.T., Khan, R., Abbasi, A.T., Li, J., **2020**. Deep learning using symmetry, fast scores, shape-based filtering and spatial mapping integrated with cnn for large scale image retrieval. *Symmetry* 12, 612.
17. Kochetov, B., Bell, P.D., Garcia, P.S., Shalaby, A.S., Raphael, R., Raymond, B., Leibowitz, B.J., Schoedel, K., Brand, R.M., Brand, R.E., et al., **2024**. Unseg: unsupervised segmentation of cells and their nuclei in complex tissue samples. *Communications Biology* 7, 1062.
18. Komura, D., Kawabe, A., Fukuta, K., Sano, K., Umezaki, T., Koda, H., Suzuki, R., Tominaga, K., Ochi, M., Konishi, H., et al., **2022**. Universal encoding of pan-cancer histology by deep texture representations. *Cell Reports* 38.
19. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A., **2017**. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging* 36, 1550–1560.
20. Kumar, S., Jain, A., Kumar Agarwal, A., Rani, S., Ghimire, A., **2021**. [retracted] object-based image retrieval using the u-net-based neural network. *Computational Intelligence and Neuroscience* 2021, 4395646.
21. Li, Y., Wang, Y., Zhou, H., Wang, H., Jia, G., Zhang, Q., **2022**. Du-net based unsupervised contrastive learning for cancer segmentation in histology images, in: *International Conference on Intelligent Robotics and Applications*, Springer. pp. 201–210.
22. Li, Z., Li, Y., Li, Q., Wang, P., Guo, D., Lu, L., Jin, D., Zhang, Y., Hong, Q., **2023**. Lvit: language meets vision transformer in medical image segmentation. *IEEE transactions on medical imaging*.
23. Lin, H., Chen, H., Dou, Q., Wang, L., Qin, J., Heng, P.A., **2018**. Scannet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide image, in: *2018 IEEE winter conference on applications of computer vision (WACV)*, IEEE. pp. 539–546.
24. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E., **2009**. A method for normalizing histology slides for quantitative analysis, in: *2009 IEEE international symposium on biomedical imaging: from nano to macro*, IEEE. pp. 1107–1110.
25. Monowar, M.M., Hamid, M.A., Ohi, A.Q., Alassafi, M.O., Mridha, M.F., **2022**. Autoret: A self-supervised spatial recurrent network for contentbased image retrieval. *Sensors* 22, 2188.
26. Naylor, P., La’e, M., Reyat, F., Walter, T., **2018**. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging* 38, 448–459.
27. van Nierop, W.L., Schneider, J.N., van der Sommen, F., et al., **2024**. Unsupervised cell segmentation by invariant information clustering, in: *Medical Imaging 2024: Image Processing*, SPIE. pp. 195–206.
28. Ozturk, S., **2020**. Stacked auto-encoder based tagging with deep features for content-based medical image retrieval. *Expert Systems with Applications* 161, 113693.
29. Peng, J., Wang, Y., **2021**. Medical image segmentation with limited supervision: a review of deep network models. *IEEE Access* 9, 36827–36851.
30. Rehamnia, W., Turlapov, V., **2024**. XAI and weak annotation toward better accuracy and interpretability in the nested classification of brain tumors. URL: <http://dx.doi.org/10.25206/978-5-8149-3873-2-2024-675-689>.
31. Ronneberger, O., Fischer, P., Brox, T., **2015**. U-net: Convolutional networks for biomedical image segmentation, in: *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18, Springer. pp. 234–241.
32. Roy, A., Pramanik, P., Ghosal, S., Valenkova, D., Kaplun, D., Sarkar, R., **2024**. Gru-net: Gaussian attention aided dense skip connection based multiresnet for breast histopathology image segmentation, in: *Annual Conference on Medical Image Understanding and Analysis*, Springer. pp.300–313.
33. Sebai, M., Wang, X., Wang, T., **2020**. Maskmitosis: a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images. *Medical & Biological Engineering & Computing* 58, 1603–1623.
34. Showrav, T.T., Hasan, M.K., **2024**. Hi-gmisnet: generalized medical image segmentation using dwt based multilayer fusion and dual mode attention into high resolution pgan. *Physics in Medicine & Biology* 69, 115019.
35. Singh, P., Hrisheeksha, P., Singh, V.K., **2021**. Cbir-cnn: content-based image retrieval on celebrity data using deep convolution neural network. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)* 14, 257–272.
36. Sokolova, M., Lapalme, G., **2009**. A systematic analysis of performance measures for classification tasks. *Information processing & management* 45, 427–437.
37. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X., **2020**. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical image analysis* 63, 101693.

- 
38. Valanarasu, J.M.J., Oza, P., Hacıhaliloglu, I., Patel, V.M., **2021**. Medical transformer: Gated axial-attention for medical image segmentation, in: Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part I 24, Springer. pp. 36–46.
39. Wang, H., Cao, P., Yang, J., Zaiane, O., **2024a**. Narrowing the semantic gaps in u-net with learnable skip connections: The case of medical image segmentation. *Neural Networks* 178, 106546.
40. Wang, T., Zhang, K., Wang, W., Ma, M., Zhang, Y., Zhao, H., Liu, G., **2024b**. U-hrmlp: Refining segmentation boundaries in histopathology images, in: 2024 IEEE International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 1–5.
41. Wei, X., Zhang, Z., Huang, H., Zhou, Y., **2024**. An overview on deep clustering. *Neurocomputing* , 127761.
42. Wright, A.M., Smith, D., Dhurandhar, B., Fairley, T., Scheiber-Pacht, M., Chakraborty, S., Gorman, B.K., Mody, D., Coffey, D.M., **2013**. Digital slide imaging in cervicovaginal cytology: a pilot study. *Archives of pathology and laboratory medicine* 137, 618–624.



Секция 6

# ***ЦИФРОВАЯ ЗЕМЛЯ И БОЛЬШИЕ ДАННЫЕ***

---

DIGITAL EARTH AND BIG DATA



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия



## Динамическая визуализация в решении задач инжиниринга ГИС-данных экологического мониторинга атмосферного воздуха промышленных предприятий

А. В. Белова, И. А. Черницин, А. В. Кычкин

Центр инжиниринга данных и искусственного интеллекта Национального исследовательского университета «Высшая школа экономики», Пермь, Россия

**Аннотация.** В рамках задачи по контролю распространения загрязняющих веществ в атмосфере в целях нормирования выбросов промышленных предприятий рассмотрены аспекты инжиниринга ГИС-данных, включая моделирование концентраций на основе аппарата Гаусса и Лагранжа, их анализа, хранения и визуализации. На базе библиотеки GRAL, реализующей расчет рассеивания Лагранжа в динамике, то есть с учетом переходных процессов, разработано приложение, позволяющее учитывать метеорологические параметры и рельеф местности, строить краткосрочный прогноз распространения облака веществ на примере диоксида серы  $\text{SO}_4$  в виде контуров концентраций, а также строить динамические карты рассеиваний. Созданы вспомогательные программные модули, отвечающие за подготовку данных рельефа для использования моделью, модули трансформации форматов данных и преобразования координат ГИС в метрическую систему, а также модули динамической визуализации результирующего состояния концентраций вредных веществ в атмосфере. Формируемые графические образы динамических концентраций рассеиваний облаков обладают высокой степенью интерпретируемости и сопоставимы по достоверности с методиками ГОСТ, работающими в статике. Полученные графические результаты могут быть использованы операторами промышленных установок, экологами-специалистами и сотрудниками контрольно-надзорных органов для оперативного восприятия информации и оценки состояния атмосферы и рисков аварийных уровней загрязнений.

**Ключевые слова:** прогнозирование рассеивания, загрязняющее вещество, цифровая модель, атмосферный воздух, визуализация данных.

## Dynamic visualization within GIS data engineering for environmental monitoring of industrial enterprise atmospheric air

A. V. Belova, I. A. Chernitsin, A. V. Kychkin

Centre of Data and Artificial Intelligence Engineering National Research University  
Higher School of Economics Perm, Russia

**Abstract.** As part of the controlling the spread of pollutants in the atmosphere to regulate emissions from industrial enterprises, GIS data engineering tasks were considered, including modelling concentrations based on the Gaussian and Lagrange techniques, their analysis, storage and visualization. Based on the GRAL library, which implements the calculation of Lagrange dispersion in dynamics, i.e., considering transient processes, an application has been developed that allows meteorological parameters and terrain relief to be considered and a short-term forecast of the spread of a cloud of substances to be made using the example of  $\text{SO}_4$  in the form of concentration contours. Auxiliary software modules have been created that are responsible for preparing terrain data for use by the model, modules for transforming data formats and converting GIS coordinates to the metric system, as well as modules for visualizing the resulting state of harmful substance concentrations in the atmosphere. The graphic images of dynamic dispersion concentrations are highly interpretable and comparable in reliability to GOST methods operating in static conditions. The graphical results obtained can be used by industrial plant operators, environmental specialists and employees of control and supervisory authorities for the rapid perception of information and assessment of the state of the atmosphere and the risks of emergency pollution levels.

**Keywords:** dispersion prediction, pollutant, digital model, atmospheric air, data visualization.

### Введение

Стремительное повышение уровня загрязнения атмосферного воздуха оказывает сильное влияние на состояние окружающей среды и здоровье населения, что приводит к необходимости ужесточения правил промышленной безопасности и усиления средств контроля [1, 2]. Несмотря на прогресс в методах математического моделирования атмосферных процессов и технологиях контроля качества воздуха, многие промышленные объекты продолжают использовать устаревшие инструменты, не обеспечивающие оперативного прогнозирования и визуализации пространственно-временной динамики рассеивания выбросов [3, 4]. Особую актуальность приобретает задача динамической визуализации рассеивания загрязняющих веществ для интерактивного отображения пространственно-

временного распределения выбросов с учетом изменяющихся метеорологических параметров, характеристик источника и топографии местности.

Известные и устанавливаемые сегодня комплексы накапливают данные по качеству атмосферного воздуха, которые содержат состав токсичных компонентов атмосферы, и факт нарушения/соблюдения границ предельно допустимых концентраций в заданный момент времени. Определение источников выбросов и оценка динамики распространения загрязнений на практике реализуется уже после фиксации факта инцидента вручную специалистами надзорных органов или экспертами общественных организаций по защите окружающей среды с использованием простейших графических моделей в виде карт местности и контуров концентраций в статике. Подразумевается, что в статике процессы переноса загрязнений в атмосфере не связаны между собой по времени, то есть расчет делается на заданный фиксированный момент времени по средним значениям концентраций по году наблюдений для указанного направления и скорости ветра. Также не всегда учитывается рельеф местности.

В регионах расположения особо крупных промышленных предприятий и условий возрастающих требований к снижению их влияния на природу становится актуальным повышение качества анализа сложной по структуре информации об изменениях в окружающей среде и оценке воздействия на нее различных факторов. Сложившиеся противоречия между требованиями к современным системам экологического мониторинга и уровнем развития современных информационных технологий могут быть преодолены за счет развития практики инжиниринга ГИС-данных [5, 6]. Далее под инжинирингом данных будем подразумевать совокупность процессов создания, управления, анализа и визуализации пространственных данных экомониторинга с использованием специализированного программного обеспечения с целью более достоверного восприятия экологической обстановки и выявления взаимосвязей между данными.

Основные аспекты инжиниринга ГИС-данных экологического мониторинга атмосферного воздуха промышленных предприятий включают в себя:

- интеграцию данных из разных источников и систем (спутниковые снимки, съемка с дронов, измерения сети сенсоров интернета вещей, системы автоматического контроля – САК, устанавливаемые на дымовых трубах [7]) и их предварительную обработку для использования в ГИС;
- создание структуры данных, определение типов данных, их связей и организации слоев для эффективного хранения и управления ГИС-данными;
- создание и настройку программного обеспечения для вспомогательных задач;
- проведение пространственного анализа, геообработки и моделирования для выявления закономерностей;
- визуализацию данных, включая отрисовку контуров рассеиваний на карте в удобной для эколога форме.

Решение задач визуализации в рамках инжиниринга ГИС-данных экологического мониторинга позволит операторам технологических установок выполнять наблюдения за выбросами вредных веществ в динамике и своевременно корректировать их в случаях риска превышений допустимых норм.

### **Постановка задачи**

Целью данной работы является преодоление противоречий между требованиями к современным системам экологического мониторинга и уровнем развития современных информационных технологий за счет развития практики инжиниринга ГИС-данных, включая динамическую визуализацию результатов наблюдений. Параметры окружающей среды, такие как рельеф местности, а также изменяющиеся погодные условия (скорость и направление ветра, уровень влажности, наличие и характер осадков и др.), могут оказывать сильное влияние на траекторию и скорость распространения загрязнений, в связи с этим генерируемые графические изображения прогнозов изменений концентраций загрязнений должны их учитывать.

### **Теория**

Базой для построения графических карт загрязнений и динамики их распространения являются модели рассеивания. В рамках данной работы были рассмотрены две модели – модель Гаусса и модель на основе переменных Лагранжа и Эйлера.

### 1. Модель рассеивания Гаусса

Модель Гаусса является одной из самых изученных и наиболее часто используемых в мире моделей рассеивания. Она является достаточно простой и может быть использована при исследовании рассеивания на расстоянии 10 км по направлению ветра и высоте источника не выше 100 м. В основе этой модели лежит уравнение дисперсии для шлейфа от точечного источника. Гауссовы модели могут быть стационарными и нестационарными. Особенностью модели Гаусса является тот факт, что в расчетах используется так называемая «эффективная высота», поскольку при выбросе облако (шлейф) сначала приподнимается над трубой.

Модель основывается на достаточно большом количестве допущений, которые в реальности неверны (например, земля является идеальным отражателем). Однако из-за своей простоты и высокой скорости вычисления является достаточно популярной и основой для многих моделей, применяемых за рубежом [8].

Стационарная модель описывается следующим образом:

$$X = \frac{Q}{2\pi u_s \sigma_y \sigma_z} \left[ \exp \left\{ -0.5 \left( \frac{y}{\sigma_y} \right)^2 \right\} \right] \left[ \exp \left\{ -0.5 \left( \frac{H}{\sigma_z} \right)^2 \right\} \right],$$

где  $X$  – концентрация на расстоянии  $x$  в направлении ветра в определенный час, мкг/м<sup>3</sup>;

$u_s$  – средняя скорость ветра на высоте выброса загрязняющих веществ, м/с;

$Q$  – скорость выброса загрязняющих веществ, мкг/с;

$\sigma_y$  – стандартное отклонение поперечного распределения концентрации;

$\sigma_z$  – стандартное отклонение вертикального распределения концентрации;

$H$  – высота выброса загрязняющего вещества, м;

$Y$  – расстояние от источника до измерителя при боковом ветре, м.

Модель Гаусса основывается на следующих предположениях и допущениях:

- постоянство метеорологических условий на территории рассеивания, постоянство облачности, ветра и осадков;
- выбрасываемое загрязняющее вещество постоянно, то есть его характеристики, в том числе концентрация, не изменяются с течением выброса;
- загрязняющее вещество не поглощается землей и окружающими предметами;
- концентрация загрязняющего вещества распределяется в горизонтальном и вертикальном направлениях по закону Гаусса, а коэффициенты диффузии не зависят от координат [9].

К преимуществам гауссовой модели можно отнести довольно легкую формулу расчета, быструю скорость вычислений. Модель пригодна как для краткосрочных, так и долгосрочных прогнозов. Однако у данной модели достаточно много недостатков: она не учитывает изменение метеорологических условий, рельеф и не применима для описания вертикального распределения загрязняющего вещества. В основном модель Гаусса используется в статических расчетах по методикам, которые принято относить к ГОСТ. В этом случае карта выбросов считается для стационарных условий, когда направление и скорость ветра не меняются длительное время, при этом рассматривается плоская местность, здания и другие строительные объекты не учитываются. Специалисты промышленных предприятий широко используют такую модель для подготовки отчетов, а также планирования работы установок, поскольку ее результаты легко воспроизвести со стороны контрольно-надзорных органов в случае расследований аварий.

### 2. Модель рассеивания Лагранжа

С учетом выявленных недостатков модели Гаусса был рассмотрен способ описания движения жидкостей или воздушных масс в переменных Эйлера и Лагранжа. В математическом представлении в переменных Лагранжа скорости частиц представляют собой производные координат от времени, а в случае использования переменных Эйлера компоненты вектора скорости частиц в точке пространства являются функциями координат, которые в совокупности являются аргументом уравнения.

При описании движения частиц с помощью переменных Эйлера мы определяем, что наблюдается в точках рассматриваемого пространства, через которое движется среда. При представлении движения частиц по Лагранжу мы определяем, что происходит в каждой индивидуальной частице среды.

Основной принцип лагранжевых моделей — отслеживание множества фиктивных частиц, движущихся по траекториям в трехмерном поле ветра. Положение этих частиц рассчитывается по следующему уравнению:

$$x_{i,new} = x_{i,old} + (\bar{u}_i + u'_i) \cdot \Delta t,$$

где  $x_{i,new}$  — новое положение в пространстве (с  $i = 1, 2, 3$ );  $x_{i,old}$  обозначает предыдущее;

$\bar{u}_i$  — средняя значение скорости, а  $u'$  — флуктуирующая (случайная, стохастическая) часть, обусловленная турбулентностью движения частицы;

$\Delta t$  — приращение времени.

Частота прохождения частиц через счетную сетку связывает лагранжеву модель с эйлеровой.

Данный тип модели учитывает следующие параметры окружения при расчете:

- профиль ветра;
- вертикальную дисперсию частиц;
- горизонтальную дисперсию частиц;
- рост облака загрязняющих веществ при подъеме.

Стандартный профиль ветра в модели задается довольно просто и соответствует описанному в документации на модель GRAMM-GRAL [10]:

$$u(z) = u(z_a) \cdot \left(\frac{z}{z_a}\right)^{ex},$$

$$ex = \text{Max}(0.35 - 0.4 \cdot (|L|)^{-0.15}, 0.05) \text{ for } L < 0,$$

$$ex = 0.56 \cdot L^{-0.15} \text{ for } L \geq 0,$$

$$L: \text{Obukhov length}[m].$$

Таким образом, стандартные профили ветра непрерывно изменяются при стратификации. В нейтральных условиях и/или при больших длинах жесткости (городские условия) показатель профиля ветра близок к 0.20, а для сильно конвективных условий уменьшается до 0.05, для сильно устойчивых условий он увеличивается примерно до 0.40.

Для флуктуаций вертикальной составляющей компоненты ветра используется модель Franzese [11]:

$$dw = a(w, z) \cdot dt + [C_0 \cdot \varepsilon(z)]^{0.5} \cdot dW,$$

$$dz(t) = w(t) \cdot dt,$$

где  $dw$  — приращение вертикальной составляющей скорости частицы,  $C_0$  — универсальная константа равная 4.0 [12];  $\varepsilon(z)$  — средняя по ансамблю скорость диссипации турбулентной кинетической энергии;  $dW$  — случайное число с нулевым средним значением, дисперсией, равной  $dt$ , и гауссовой функцией плотности вероятности (ФПВ).

$$dt(z) = 0.01 \cdot \frac{2 \cdot \sigma_w^2}{C_0 \cdot \varepsilon(z)}.$$

Предполагается, что множитель  $a(w, z)$  является функцией вертикальной скорости:

$$a(w, z) = \alpha(z) \cdot w^2 + \beta(z) \cdot w + \gamma(z), \quad (1)$$

где  $\alpha(z), \beta(z), \gamma(z)$  — неизвестные параметры, которые задаются уравнением Фоккера–Планка:

$$w \cdot \frac{\partial P_E(w, z)}{\partial z} = - \frac{\partial [a(w, z) \cdot P_E(w, z)]}{\partial w} + \frac{C_0 \cdot \varepsilon(z)}{2} \cdot \frac{\partial^2 P_E(w, z)}{\partial w^2},$$

где  $P_E(w, z)$  — является Эйлеровой ФПВ вертикальной составляющей скорости турбулентности на высоте  $z$ .

Предполагая квадратичную форму зависимости ускорения, модель Franzese [11] не нуждается в какой-либо информации о форме  $P_E(w, z)$ , а требует только первый из четырех эйлеровых моментов вертикальной скорости. Коэффициенты в уравнении (1) можно выразить следующими формулами:

$$\alpha(z) = \frac{(1/3) \cdot \partial \bar{w}^4 / \partial z - \bar{w}^3 / (2 \cdot \bar{w}^2) \cdot [\partial \bar{w}^3 / \partial z - C_0 \cdot \varepsilon(z)] - \bar{w}^2 \cdot \partial \bar{w}^2 / \partial z}{\bar{w}^4 - (\bar{w}^3)^2 / \bar{w}^2 - (\bar{w}^2)^2}, \quad (2)$$

$$\beta(z) = \frac{1}{2 \cdot \bar{w}^2} \left[ \frac{\partial \bar{w}^3}{\partial z} - 2 \cdot \bar{w}^3 \cdot \alpha(z) - C_0 \cdot \varepsilon(z) \right], \quad (3)$$

$$\gamma(z) = \frac{\partial \bar{w}^2}{\partial z} - \bar{w}^2 \cdot \alpha(z). \quad (4)$$

В уравнениях (2)–(4)  $\overline{w^i}$  ( $i = 1, 2, 3, 4$ ) обозначает наибольший из эйлеровых моментов вертикальной скорости. Первый момент задает среднее значение вертикальной скорости, равной 0, второй – дисперсия – рассчитывается в стабильных и нейтральных условиях  $\overline{w^2} = 1.56 \cdot u_*^2(5)$ , связывающее уравнение:

$$\overline{w^2} = u_*^2 \cdot \left[ 1.15 + 0.1 \cdot \left( \frac{z_i}{-L} \right)^{0.67} \right]^2 \quad (6)$$

Третий момент был вычислен в стабильных и нейтральных условиях и для поверхностного слоя в целом  $\overline{w^3} = 0$ . В конвективных условиях [11]

$$\overline{w^3} = w_*^3 \cdot 1.1 \cdot \left( \frac{z}{z_i} \right) \cdot \left( 1 - \frac{z}{z_i} \right)^2,$$

где  $z_i$  – планетарный пограничный слой (ППС),  $u_*$  – скорость трения,  $w_*$  – множитель конвективной скорости, а  $h$  высота стабильного ППС рассчитанная для

$$L \geq 0: h = \min \left[ 0.4 \cdot \left( \frac{u_* \cdot L}{f} \right)^{1/2}, 800 \right],$$

$$L < 0: z_i = \min \left[ 0.4 \cdot \left( \frac{u_* \cdot L}{f} \right)^{1/2}, 800 \right] + 300 \cdot e^{0.01 \cdot L},$$

где  $L$  – длина Обухова, а  $f = 0.0001 \text{ s}^{-1}$  параметр Кориолиса [16].

Четвертый момент задается в конвективных условиях  $\overline{w^4} = 3.5 \cdot [\overline{w^2(z)}]^2$ , при стабильных и нейтральных условиях и поверхностным слоем  $\overline{w^4} = 3 \cdot [\overline{w^2(z)}]^2$ , являющимся Гауссовым предположением [11].

Скорость ветра и его направление имеют большие отклонения от среднего при низких средних скоростях из-за его извилистости в пространстве. Это приводит к измененной формуле распространения по вертикали, так как каждое направление считается основным и имеет различные высоты облака распространения [13].

Для учета особенностей рельефа при выполнении расчетов модель рассеивания Лагранжа требует данные рельефа местности в формате сетки значений с заданным шагом. Разработан алгоритм преобразования имеющихся открытых данных рельефа в виде контуров, представленных в формате GeoJSON, в необходимый формат данных, содержащий следующие шаги:

1. Загрузка и валидация данных контуров в формате GeoJSON;
2. Проецирование координат контуров из системы WGS84 в местную систему координат заданной области;
3. Сортировка контуров по высоте, создание внешних и внутренних контуров;
4. Преобразование контуров в набор непересекающихся заполненных полигонов для формирования корректного рельефа местности;
5. Формирование сетки высот с заданным шагом на основе полигонов.

Проецирование координат контуров из общей системы в местную необходимо для точного соответствия географическим данным, специфичным для выбранной области, что позволяет повысить точность прогнозирования рассеивания загрязняющих веществ.

На рисунке 1 представлен результат работы алгоритма для формирования карты высот, для примера взят фрагмент рельефа местности Липецкой области со стандартными долготой и широтой и контурами высот, далее было проведено преобразование в местную метрическую систему координат МСК-48, а затем контура высот (замкнутые линии) заменены на полигоны высот (области с заливкой).

Данные по рельефу совместно с метеоданными введены в библиотеку GRAL, которая может быть использована для расчета рассеиваний по Лагранжу. Дополнительно введены данные по источникам выбросов, в том числе высоты труб, диаметры, типы веществ и мощности выбросов. Расчет рассеивания загрязнений представляет собой сетку концентраций загрязняющих веществ в  $\text{кг}/\text{м}^2$ . Значения концентраций в разных точках сетки прогноза могут отличаться на тысячные доли килограмма и количество уникальных концентраций может достигать сотен тысяч значений, в зависимости от размера и разрешения сетки прогноза данные в сыром виде при создании графика

будут не читаемы, поэтому процесс визуализации результатов работы модели включает в себя этап предобработки значений.

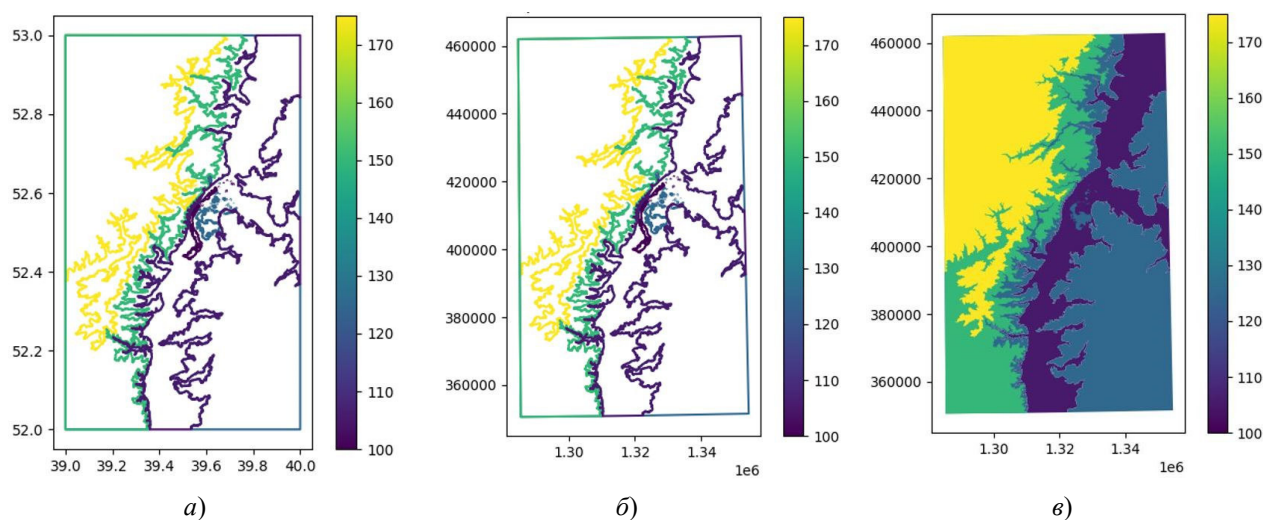


Рисунок 1. Карта высот выбранного участка местности в формате исходных данных по долготе/широте (а), проекция координат в метрическую систему МСК-48 (б), метрическая система МСК-48 с полигонами (в)

На данном этапе количество уникальных значений концентраций веществ сетки уменьшается для их последующего объединения в небольшое количество полигонов и визуализации на графике. Для этого выбирается пороговое значение концентрации, выше которого данные округляются до целого числа, а ниже – до выбранного количества промежутков. Выбор единицы в качестве порогового значения и количества промежутков для округления равного 5..10 позволяет сохранить информацию о высоких концентрациях веществ в эпицентрах источников распространения без потери очертаний облака вещества.

Для получения единого полигона с определенным значением концентрации вещества выбранные точки сетки преобразуются в квадратные полигоны с длиной стороны, равной шагу сетки, и центром полигона в координатах точки. Границы результирующих квадратных полигонов соприкасаются между собой, что позволяет выполнить их объединение в единый полигон или в мультиполигоны, что также важно при визуализации различных типов загрязнителей [13]. На рисунке 2 представлен результат визуализации результатов работы модели расчета рассеиваний с описанным алгоритмом обработки значений.

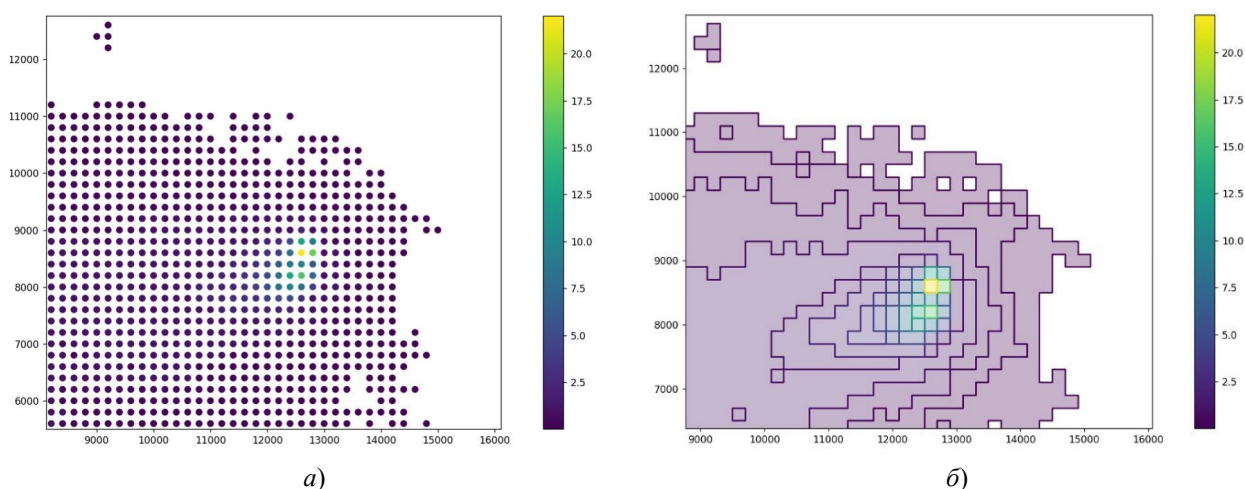


Рисунок 2. Результат визуализации результатов работы модели расчета рассеиваний в виде облака точек (а) и виде полигонов (б)

Параметры и пороговые значения могут быть скорректированы для индивидуальных условий задачи и типа загрязняющего вещества.

### Результаты экспериментов

Для тестирования модели Лагранжа была проведена серия экспериментальных расчетов по суткам, а затем посчитанные средние концентрации за зимний сезон 2024–2025 г. В качестве источника загрязнения выбрана одна труба (почти точечный источник) с постоянным значением выброса диоксида серы ( $\text{SO}_4$ ) и фиксированными скоростью и температурой выброса (см. табл. 1). Оценка качества модели проводилась путем выбора ряда точек на карте, усреднения значений концентраций вещества в них и сравнения прогноза модели со значениями, рассчитанными программой «ЭКОцентр–РРВА», версия 2.0 (см. табл. 2).

Таблица 1. Параметры тестового источника загрязнения атмосферы

Высота, м	Диаметр, м	Координаты		Параметры газовой смеси			Загрязняющее вещество	
		X1, X2	Y1, Y2	Скорость, м/с	Объем, м <sup>3</sup> /с	Темп., °C	Код	Выброс, г/с
42.0	1.1	2236807, 7	513435, 9	11.4286	10.861	268.6	0330	6.944

Таблица 2. Результаты экспериментальных расчетов

Модель	Среднее значение, доля ПДК			
Шаг сетки, м	100	50	10	5
Размер экспериментальной области – 1000 м на 1000 м				
ГОСТ	0.036	0.036	0.036	0.036
Модель Лагранжа	0.0425	0.0399	0.0338	0.0341
Размер экспериментальной области – 2000 м на 2000 м				
ГОСТ	0.030	0.030	0.030	0.030
Модель Лагранжа	0.0293	0.0277	0.0231	0.0228
Размер экспериментальной области – 5000 м на 5000 м				
ГОСТ	0.024	0.024	0.024	0.024
Модель Лагранжа	0.0124	0.0150	0.0137	0.0149

Результаты прогнозирования сравнивались с моделью, работающей на основе методики расчета государственного стандарта – с программой расчёта рассеивания для ЭВМ «ЭКОцентр–РРВА», версия 2.0 (положительное заключение экспертизы Росгидромета от 10.11.2020 № 140-08474/20И).

Расчетная область является квадратом и выбирается так, чтобы точечный источник был в её центре. Расчетная область представляла собой квадрат с длиной стороны от 1000 м до 5000 м (рис. 3–5). Во всех расчетах метеорологические условия остаются неизменными:

- средняя температура наружного воздуха –16.6 °C;
- скорость ветра ( $u^*$ ) 0.5..7 м/с;
- параметры перебора направления ветров 0..360 метео °.

По результатам тестирования можно заметить, что средние значения выбросов для всех расчетов в области 1000 м на 1000 м отличаются от значения, рассчитанного ГОСТ, не более чем на 15 % – модель Лагранжа приближает расчетные данные довольно хорошо к реальным. Как видно из графического представления, с уменьшением шага увеличивается разрешение контуров, но они не являются радиальными. Кроме того, в модели Лагранжа наибольшее влияние на результат оказывают отклонения от среднего направления ветра и его скорости. Средние значения выбросов для всех расчетов в области 5000 м на 5000 м на 50 % отличаются от значения, рассчитанного ГОСТ за счет влияния рельефа местности, но при этом все результаты являются заниженными, что допустимо для этого типа моделей, так как в них используется множество упрощений. Занижение прогнозных данных встречается по всей расчетной области, оно может быть исправлено с помощью поправочного множителя либо входных, либо выходных данных. Как видно из графического представления, тренд на мелкие локальные зоны сохраняется. Общая картина похожа на рассчитанную способом ГОСТа, но с некоторыми поправками на рельеф и метеорологические условия.



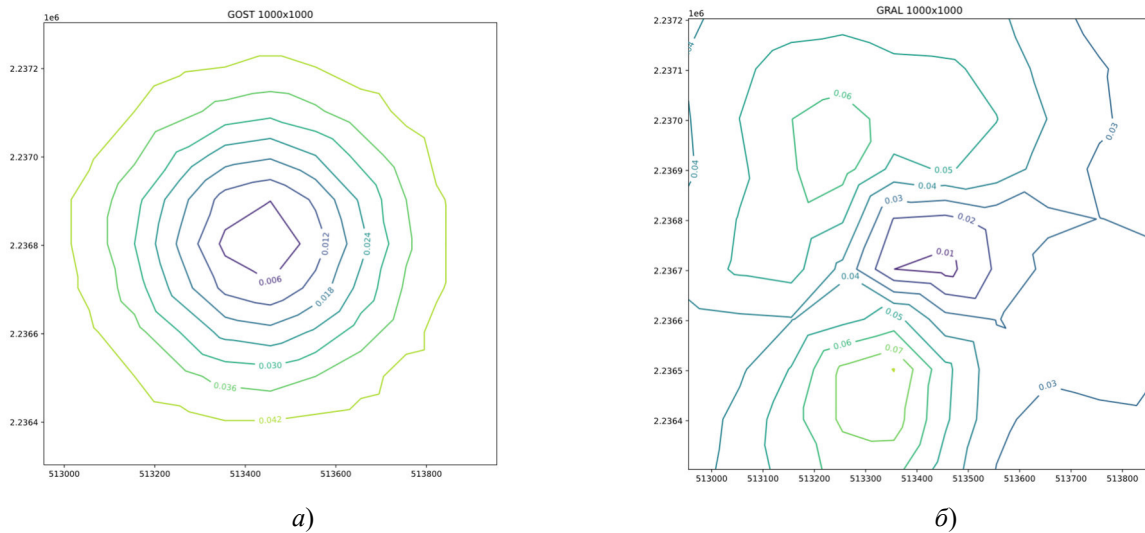


Рисунок 3. Визуальное представление расчетов на сетке 1000 \* 1000 м. Точечный источник находится в центре графика: а) расчет ГОСТ; б) расчет с шагом 10 м

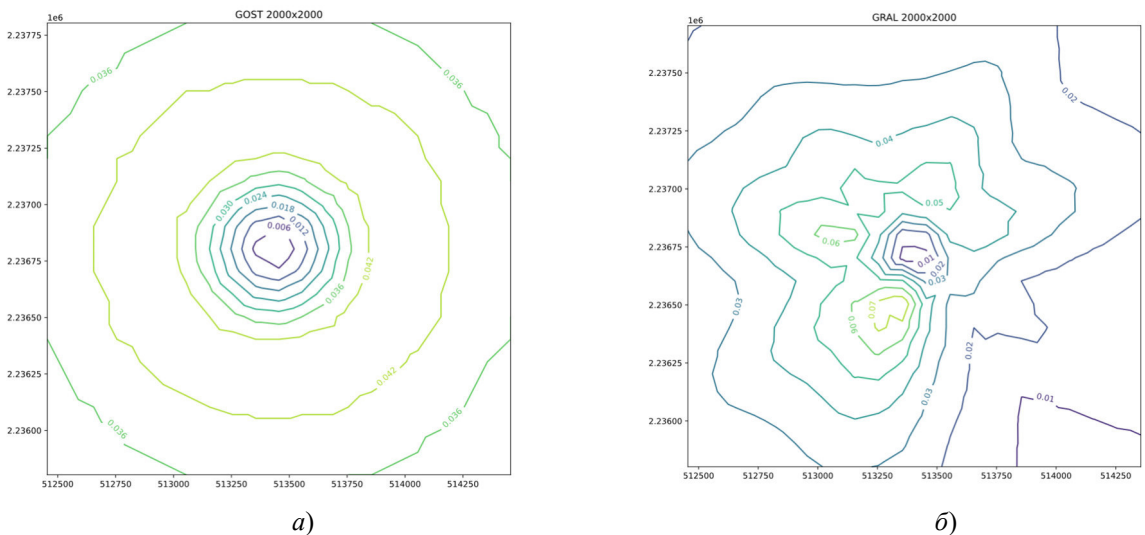


Рисунок 4. Визуальное представление расчетов на сетке 2000 \* 2000 м. Точечный источник находится в центре графика: а) расчет ГОСТ; б) расчет с шагом 50 м

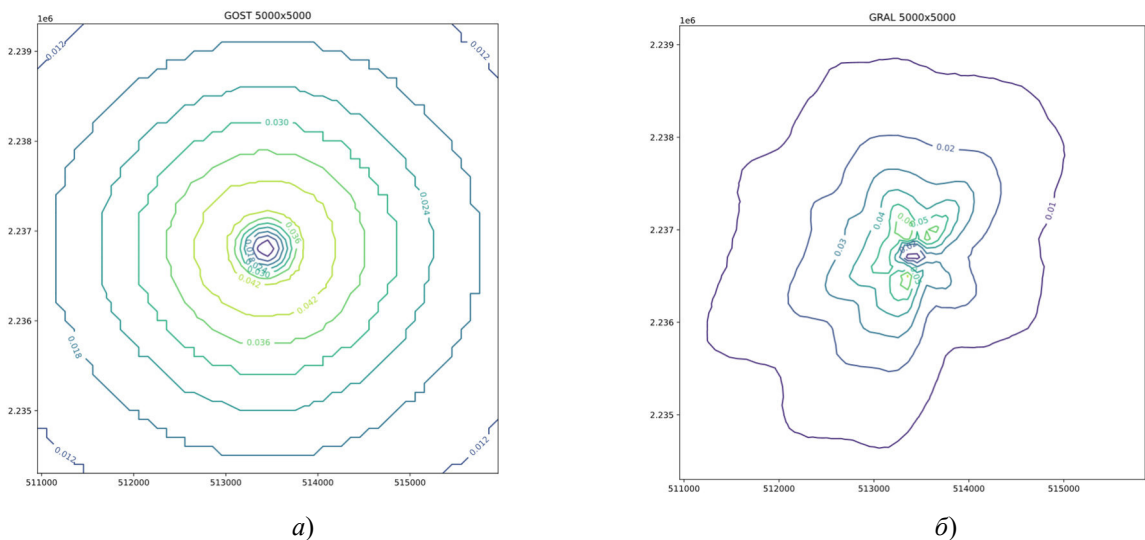


Рисунок 5. Визуальное представление расчетов на сетке 5000 \* 5000 м. Точечный источник находится в центре графика: а) расчет ГОСТ; б) расчет с шагом 100 м

### Обсуждение результатов

Представленная модель Лагранжа, в отличие от модели Гаусса, обладает рядом преимуществ, которые позволяют использовать ее для решения задач динамической визуализации рассеиваний загрязнений:

- позволяет учесть нестационарность источника выброса;
- учитывает влияние погоды, метеорологические параметры;
- учитывает строения и ландшафт;
- учитывает вертикальное распространение загрязнения.

Применение модели Лагранжа не противоречит расчетам по ГОСТ, при этом обеспечивает базу для использования более сложных моделей анализа данных с применением технологий машинного обучения [14].

Разработанный алгоритм визуализации прогнозов модели, использующий методы группировки значений концентраций, позволяет создать более информативный график рассеивания, по сравнению с простым отображением необработанных данных. Настройка параметров и порогов обработки концентраций помогает учесть особенности разных загрязняющих веществ и их максимальные безопасные значения.

### Заключение

Исходя из сравнительных расчетов по средним значениям видно, что модель Лагранжа позволяет получить расчетные данные, не противоречащие результатам ГОСТ, рассчитанным посредством ЭВМ «ЭКОцентр–РРВА», версия 2.0 (положительное заключение экспертизы Росгидромета от 10.11.2020 № 140-08474/20И). При увеличении количества ячеек или уменьшении размера ячейки сетки увеличивается точность определения характера распределения выброса, но частично уменьшается точность самих вычисленных данных. Наиболее точные результаты расчета выбросов моделью могут быть получены при некоторой корректировке входных данных и большом размере расчетной сетки.

Полученная модель рассеивания на основе переменных Лагранжа может быть использована для формирования динамических визуализаций распределения загрязняющих веществ в атмосфере с шагом обновления, например, 20 минут, при этом учитывая рельеф местности (рис. 6).

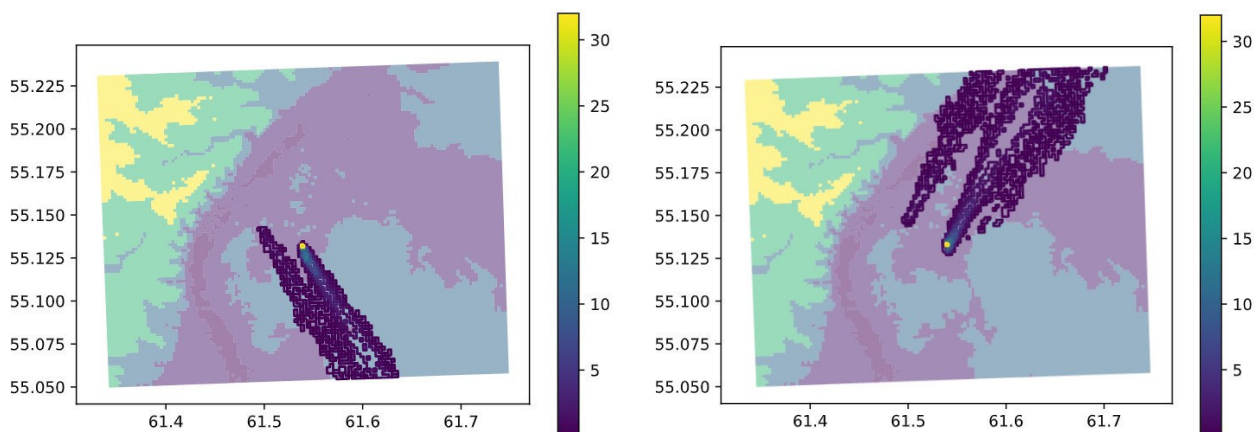


Рисунок 6. Пример прогнозирования распределения загрязняющих веществ с изменением параметров ветра в течение суток. В качестве фонового рисунка отображается карта рельефа местности

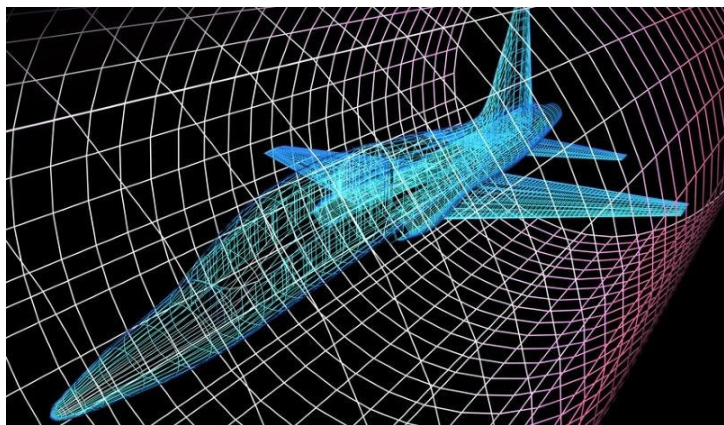
Набор полигонов концентраций загрязняющих веществ, получаемый при динамической визуализации результатов работы, может быть использован для онлайн-наблюдения за выбросами источников промышленных предприятий, а также оценки риска аварийных выбросов и их влияния на экономику региона [15]. Приведение координат полигонов к местной системе координат области упрощает процесс анализа рисков опасных уровней концентраций, а введение рельефа позволяет более достоверно оценивать динамику выбросов, что в целом соответствует аспектам инжиниринга ГИС-данных для систем экологического мониторинга.

**Источник финансирования**

Данная работа является результатом исследовательского проекта, реализованного в рамках Программы фундаментальных исследований НИУ ВШЭ.

**Список литературы**

1. Use of Machine Learning in Air Pollution Research: A Bibliographic Perspective / Jain S., Kaur N., Verma S., et al. // *Electronics*. 2022. Vol. 11, no. 21. P. 3621.
2. Air quality index and air pollutant concentration prediction based on machine learning algorithms / Liu H., Li Q., Yu D., Yu G. // *Applied Sciences*. 2019. Vol. 9, no. 19. P. 4069.
3. Prediction and analysis of pollution levels in Delhi using multilayer perceptron / Akhtar A., Masood S., Gupta C., Masood A. // *Data engineering and intelligent computing*. Springer, Singapore. 2018. Vol. 542. P. 563–572.
4. Internet of Things mobile – air pollution monitoring system (IoT-mobair) / Dhingra S., Babu Madda R., Gandomi A. H., et al. // *IEEE Internet of Things Journal*. 2019. Vol. 6, no. 3. P. 5577–5584.
5. A Review of Air Quality Modeling / Karroum K., Lin Y., Chiang Y.Y., et al. // *MAPAN*. 2020. Vol. 35. P. 287–300.
6. AiCareAir: Hybrid-ensemble Internet-of-Things sensing unit model for air pollutant control / Borah J., Nadzir M. S. M., Cayetano M. G., et al. // *IEEE Sensors Journal*. 2024. Vol. 24 (13). P. 21558–21565.
7. О состоянии и об охране окружающей среды Российской Федерации в 2019 году (Государственный доклад). М: Министерство природных ресурсов и экологии Российской Федерации, 2020. 1000 с.
8. Jeong S.-J. CALPUFF and AERMOD dispersion models for estimating odor emissions from Industrial Complex Area sources // *Asian Journal of Atmospheric Environment*. 2010. Vol. 5(1). P. 1–7.
9. Завгороднев А. В., Аكوпова Г. С., Мельников А. В. Теоретические основы рассеивания в атмосфере организованных нестационарных выбросов газа на объектах газотранспортных предприятий // *Территория Нефтегаз*. 2011. № 10. С. 68–73.
10. Evaluation of high-resolution GRAMM–GRAL (v15.12/v14.8) simulations over the city of Zürich, Switzerland / Berchet A., Zink K., Oetli D., et al. // *Geosci. Model Dev*. 2017. Vol. 10, no. 9. P. 3441–3459.
11. An efficient Lagrangian stochastic model of vertical dispersion in the convective boundary layer / Franzese P., Luhar A. K., Borgas M. S. // *Atmospheric Environment*. 1999. Vol. 33(15) P. 2337–2345.
12. Estimation of the Lagrangian structure function constant C0 from surface layer wind data / Anfossi D., Degrazia G., Ferrero E., et al. // *Boundary-Layer Meteorol*. 2000. Vol. 95. P. 249–270.
13. Stationary and portable multipollutant monitors for high-spatiotemporal-resolution air quality studies including online calibration / Buehler C., Xiong F., Zamora M. L., et al. // *Atmospheric Measurement Techniques*. 2021. Vol. 14(2). P. 995–1013.
14. Modeling P M2.5 forecast using a self-weighted ensemble GRU network: Method optimization and evaluation / Huang H., Qian C. // *Ecological Indicators*. 2023. Vol. 156. Article. 111138.
15. Air pollution control or economic development? Empirical evidence from enterprises with production restrictions / Feng T., Chen X., Ma J., et al. // *Journal of Environmental Management*. 2023. Vol. 336. Article. 117611.



Секция 7

***ГЕОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ.  
КОМПЬЮТЕРНАЯ ГРАФИКА***

---

**GEOMETRIC MODELING.  
COMPUTER GRAPHICS**



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия

УДК 004.93'11 + 519.173 + 515.124.4

DOI: 10.25686/978-5-8158-2474-4-2025-878-898

## Извлечение геометрических объектов из пространства чертежа CAD-модели посредством изоморфизма графов

**И. А. Федотов<sup>1,2</sup>**<sup>1</sup>Navio, Москва, Россия<sup>2</sup>Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

**Аннотация.** Работа направлена на решение проблемы логического распознавания объектов и их образов посредством задач изоморфизма графа совместно с дискретной метрической геометрией. В качестве постановки выбрана задача нахождения известной геометрической формы из словаря, в пространстве CAD модели, где известны координаты точек и связи между ними. От более общей задачи распознавания по проекции трехмерного объекта на сетчатку (ПЗС-матрицу) постановка отличается тем, что нет проекционных искажений, связи представляют собой прямые без функционального пространства кривых и параметрических поверхностей. Уровни детализации объекта опущены ввиду пока излишней сложности задачи гомеоморфизма подграфа. Допущение на знание координат не принципиально, достаточно расстояний между точками. С учетом допущений можно трактовать постановку как распознавание образов геометрических объектов по изометрической проекции на сетчатке.

С практической точки зрения предложенный метод позволяет решать новые классы задач: поиск конкретной или подобной детали в сборке, поиск максимально близкой детали или общей части/границы, версионирование CAD-модели (выделение максимально общей части), поиск разных классов симметрий, анализ деформации объекта. Обеспечивает унификацию процесса распознавания, сводя разные классы задач к типовой задаче сопоставления графов, исключительно в машинной обработке. Например, определение предмета и его ориентации относительно наблюдателя по силуэту или его части - как задачу поиска атрибутированного по ребрам цикла/цепи через поиск неиндуцированного изоморфного подграфа или его максимального вхождения.

**Ключевые слова:** изоморфизм графов, сопоставления графов, дискретная метрическая геометрия, геометрические формы, распознавание образов, системы компьютерного проектирования.

## Geometric forms extraction from CAD model space through graph isomorphism approaches

**I. A. Fedotov<sup>1,2</sup>**<sup>1</sup>Navio, Moscow, Russia<sup>2</sup>Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

**Abstract.** This work aims to solve the problem of logical recognition of objects and their patterns using graph isomorphism problems combined with discrete metric geometry. The chosen problem setting is finding a known geometric shape from a dictionary in the model space of a CAD model, where the coordinates of points and the relationships between them are known. This setting differs from the more general problem of recognition based on the projection of a three-dimensional object onto the retina (CCD matrix) in that there are no projection distortions; the relationships are straight lines, without the functional space of curves and parametric surfaces. Levels of object detail are omitted due to the currently excessive complexity of the subgraph homeomorphism problem. The assumption of knowing the coordinates is not essential; the distances between points are sufficient. Given these assumptions, the setting can be interpreted as pattern recognition of geometric objects based on an isometric projection onto the retina.

From a practical perspective, the proposed method enables solving new classes of problems: searching for a specific or similar part in an assembly, searching for the closest part or common part/boundary, versioning a CAD model (identifying the most common part), searching for different symmetry classes, and analyzing object deformation. It unifies the recognition process, reducing various classes of problems to a typical graph matching task, exclusively for machine processing. For example, determining an object and its orientation relative to the observer based on its silhouette or part of it is like searching for an edge-attributed cycle/chain by searching for a non-induced isomorphic subgraph or its maximum common subgraph.

**Keywords:** graph isomorphism, graph matching, distance discrete geometry, geometry forms, pattern recognition, CAD.

## Введение

They drew all manner of things— everything that begins with an M—.

---

*Lewis Carroll Alice's Adventures in Wonderland.*  
United Kingdom 1865 like Matching, Морф, and the Cayley-Menger determinant - you know, you say things are 'a lot alike' – have you ever seen such a thing as a graph similarity drawing?

Геометрические фигуры. Всякая ограниченная часть пространства называется геометрическим телом.

Геометрическое тело можно подразделять на части; каждая часть геометрического тела есть также геометрическое тело.

Граница геометрического тела, т.е. то, чем оно отделяется от остального пространства, наз. поверхностью.

Граница поверхности называется линией.

Граница линии называется точкой.

---

*А. Киселев*

*Элементарная геометрия. Москва 1914 г.*

Используем определение геометрической фигуры Киселева, вынесенное во второй эпиграф для отображения геометрических понятий на теорию графов и дискретную геометрию. Так, поверхность (грань) будем отождествлять с гиперребром гиперграфа, линию – с ребром графа, а точку – с узлом. Такое обобщение достаточно мощно и может в будущем использоваться для сложных поверхностей и кривых, в нашей статье рассматриваются связи в виде линий, и приставка гипер не требуется, разве что для описания граней.

- Геом. тело/фигура (совокупность граней/поверхностей) → Гиперграф,  $H(V, E, L)$
- Поверхность/грань → Гиперребро,  $E$
- Линия/прямая (пересечение поверхностей) → Ребро,  $e \in E(H)$
- Параметры линии (длина, кривизна, ...) → Атрибуты ребер мультиграфа,  $L_i^e$
- Точка (пересечение линий) → Узел,  $v \in V(H)$
- Параметры точки (координаты, тип, ...) → Атрибуты узла,  $L_i^v$

При описании движения и эволюции графа в будущем, возможно, потребуется привлечь время. Графы, которые эволюционируют во времени, называют темпоральными (temporal) или time-evolving, общая запись для них:  $H(V(t), E(t), L(t))$ , т.е. эволюционировать могут вершины, ребра и метки независимо друг от друга. Некоторые авторы такой вид графов сводят к мультиграфам, считая время дискретным и добавляя новую размерность к меткам, тем самым вводя тензора для общности с машинным обучением. Есть другой путь: оставить время аналоговым (дискретным), а эволюцию множеств  $V(t)$ ,  $E(t)$ ,  $L(t)$  выразить в виде дифференциальных (разностных) уравнений, тем самым соединив дискретную математику и аналитическую механику, если мы говорим про динамические системы.

Пока умозрительно видится, что геометрическую фигуру можно описать методами теории графов (дискретной геометрии), сложности возникают в технике описания: как точно описать форму фигуры, чем можно пренебречь, а что действительно необходимо для последующего распознавания? Далее представим своё видение проблемы и способы её решения на примере нескольких практических задач.

Отдельно стоит сказать о графовых алгоритмах на изоморфизм (табл. 1), они не являются основным предметом статьи, хотя и играют ключевую роль, потенциально способные свести на нет все потуги теоретических соображений. Действительно, задачи являются NP полными, с волнующим факториалом в оценке производительности, но на сегодняшний день существуют эффективные алгоритмы, способные работать в реальном времени (за исключением пока разве что алгоритма SHOM и частично MCS) на больших плотных графах. Автор как раз является разработчиком нескольких из них, в том числе и специально спроектированных для геометрических задач, и поэтому дальнейшее повествование пойдет в предположении, что проблема с алгоритмами не является краеугольной. Она может быть эффективно решена даже на относительно скромных по нынешним меркам системах уровня desktop и даже на embedded уровне. Ссылки и некие обзоры алгоритмов можно найти в библиографии работ [18, 19] или на странице автора.

В статье для простоты, а также будущего перехода к проекциям на сетчатке, сосредоточимся на плоских фигурах, состоящих из четырех точек, называемых 4-угольниками, и попытаемся представить их описание, опираясь исключительно на графовое представление. Покажем, что с помощью понятия изоморфизмов и введенной матрицы формы можно находить объекты на чертеже после аффинных и общих (даже разрушающих) преобразований. Помимо этого с помощью расстояний (мировой функции [4–7]) и понятий из дискретной геометрии [1–3] в виде длин отрезков и полного графа из них, станет возможен переход в инвариантное пространство, не зависящее ни от координат (положения в пространстве), ни от реальных размеров предмета. Такой переход в будущем позволит создавать антропоцентрическую систему координат (систему, создаваемую субъектом познания, машиной или человеком) по ходу процесса познания окружающего мира, что является главным моментом в распознавании и отождествлении образов.

Текст пронизан сквозным примером четырехугольника типа kite, этот тип относительно прост (среди всех типов Quadrilaterals), но уже сложнее квадрата и ромба.

Таблица 1. Задачи сопоставления графов (graph matching)

Task name	Canonical name	Description
MCP	Maximal Clique Problem	Поиск полного подграфа, разложение на независимые клики. Структурный анализ графа, выделение основных структурных компонент и их взаимосвязей
AUT	Automorphism	Поиск автоморфизма графа. Разложение на симметрии графа, только структурные или с учетом семантики
GI	Graph Isomorphism	Сравнение графов на изоморфность, эквивалентность форм
SI	Subgraph Isomorphism	Нахождение подграфов изоморфных образцу из сложного графа чертежа
MCS	Maximum Common Subgraph	Нахождение максимально общей изоморфной части двух графов (пересечение графов)
SHOM	Subgraph Homeomorphism	Выделение подграфов или их максимально общей изоморфной части с учетом уровня детализации (абстракции) на большом чертеже

### Задача дискретной геометрии: Дельтоид (Ромбоид) / Kite

Представим kite в безразмерном параметрическом виде с помощью соотношений диагоналей

$$\frac{|bc|}{|ad|} = \beta = \frac{2}{4} = 0.5 \text{ и разбиением большой диагонали } \frac{|od|}{|ao|} = \gamma = \frac{1}{3}.$$

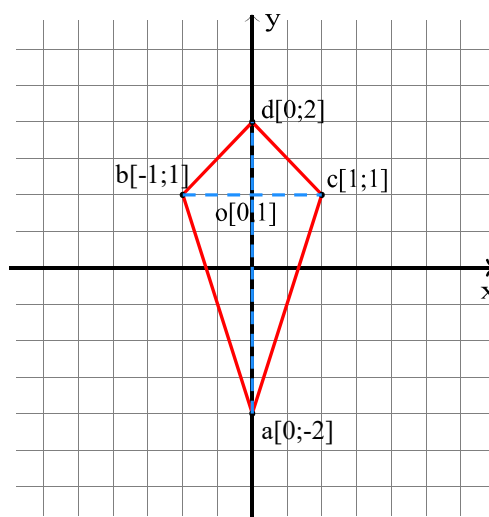


Рисунок 1. Kite



Это не единственный способ представления, но он нам будет удобен в последующем, хотя на самом деле главное, что заданы все координаты точек, а параметрический вид нужен для аналитического описания фигуры. Т.е., расположив четыре точки на плоскости, мы однозначно задаём четырех угольник, его форму определяют опять же координаты точек, т.е. их взаимное расположение. С точки зрения теории графов нашу фигуру можно описать неориентированным графом с четырьмя вершинами и четырьмя ребрами (матрицей смежности) и весовой матрицей. Некоторые авторы представляют такой граф исключительно весовой матрицей, но мы для удобства будем выделять структурную компоненту – булеву матрицу смежности  $\mathbf{A}$  и семантическую – вещественную матрицу весов  $\mathbf{W}$ . Тогда

$$A = \begin{pmatrix} a & b & c & d \\ a & 0 & 1 & 1 & 0 \\ b & 1 & 0 & 0 & 1 \\ c & 1 & 0 & 0 & 1 \\ d & 0 & 1 & 1 & 0 \end{pmatrix}. \quad (1)$$

Поскольку граф не ориентирован ( $A = A^T$ ) и без петель ( $\text{diag}(A) = 0$ ), нам для однозначного задания достаточно только верхнего надтреугольника. Вместе с тем наряду с матрицей  $A$  незримо присутствует матрица  $\bar{A} - I$ , где  $\bar{A}$  – логическое отрицание,  $I$  – единичная матрица (нужна для устранения несуществующих петель, которые неизбежно появляются при инверсии), для краткости такую матрицу можно также обозначать как  $\bar{A}$ , подразумевая, что петли задачей не предусмотрены, такая матрица называется дополнением графа и всегда незримо существует, т.е. имманентна основному графу.

$$\bar{A} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Видно, что при принятой индексации вершин диагонали ромбоида  $|ad|, |bc|$  появляются на антидиагонали матрицы  $\bar{A}$ .

Описанное выше элементарно, некоторый интерес представляет весовая матрица, её представим как полную матрицу длин всех отрезков (существующих и несуществующих ребер) и нормированную на максимальное значение, пропустив (оставив пытливым читателям) упражнения по математической геометрии, по Рылову [8], имеем

$$\mathbf{M} = \begin{pmatrix} 0 & \sqrt{\frac{1}{(1+\gamma)^2} + \frac{\beta^2}{4}} & \sqrt{\frac{1}{(1+\gamma)^2} + \frac{\beta^2}{4}} & 1 \\ \sqrt{\frac{1}{(1+\gamma)^2} + \frac{\beta^2}{4}} & 0 & \beta & \sqrt{\frac{\gamma^2}{(1+\gamma)^2} + \frac{\beta^2}{4}} \\ \sqrt{\frac{1}{(1+\gamma)^2} + \frac{\beta^2}{4}} & \beta & 0 & \sqrt{\frac{\gamma^2}{(1+\gamma)^2} + \frac{\beta^2}{4}} \\ 1 & \sqrt{\frac{\gamma^2}{(1+\gamma)^2} + \frac{\beta^2}{4}} & \sqrt{\frac{\gamma^2}{(1+\gamma)^2} + \frac{\beta^2}{4}} & 0 \end{pmatrix}, \quad (2)$$

где параметры  $\beta, \gamma$  определены в начале раздела. Видно, что матрица у нас симметрическая, положительная, с элементами, не превышающими 1. На главной диагонали стоят нули, что в вычислительном плане не очень хорошо, в дальнейшем главную диагональ можно заполнить весами

узлов, метками или другими атрибутами. На самом деле мы получили гораздо больше, чем матрицу весов, полученная нами матрица имеет физический смысл – формы фигуры, она задана в отношениях, т.е. независима относительно исходных координат и положения фигуры в пространстве, и её абсолютных размеров. Сам её вид специфичен, т.е. заполнение числами уникально для каждой фигуры, поэтому переобозначим её через  $\mathbf{M}$  от греческого слова *Μορφή* (форма). На самом деле  $\mathbf{M}$  есть ни что иное, как нормированный (обезразмеренный) (1,1)-минор определителя Кэли-Менгера [1–3].

Тогда существующие ребра (длины сторон) извлекаются как

$$M = A \circ \mathbf{M},$$

а несуществующие (диагонали) как

$$\overline{M} = \overline{A} \circ \mathbf{M},$$

где  $\circ$  – покомпонентное матричное произведение (Адамара, Шура) [13].

Для всех четырехугольников матрица  $\mathbf{A}$  будет неизменна, меняется только матрица формы  $\mathbf{M}$ . Что, собственно говоря, логично, поскольку в математическом плане это всего лишь изоморфизмы некоего образа. В вычислительном плане исходная задача изоморфизма дополняется матрицей формы и на языке задач сопоставления графов называется как изоморфизм атрибутированного по ребрам подграфа.

Теперь представим численное задание матрицы формы, поскольку очевидно, что аналитическое достаточно трудозатратно, благо можно получить матрицу  $\mathbf{M}$ , используя только координаты вершин.

$$a=(0.0;-2.0); \quad b=(-1.0;1.0); \quad c=(1.0;1.0); \quad d=(0.0;2.0);$$

$$\begin{aligned} e_{ab} &= \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} = \sqrt{(-1-0)^2 + (1+2)^2} = \sqrt{10} \approx 3.1623; \\ e_{bd} &= \sqrt{(x_d - x_b)^2 + (y_d - y_b)^2} = \sqrt{(0+1)^2 + (2-1)^2} = \sqrt{2} \approx 1.4142; \\ e_{dc} &= \sqrt{(x_c - x_d)^2 + (y_c - y_d)^2} = \sqrt{(1-0)^2 + (1-2)^2} = \sqrt{2} \approx 1.4142; \\ e_{ca} &= \sqrt{(x_a - x_c)^2 + (y_a - y_c)^2} = \sqrt{(0-1)^2 + (-2-1)^2} = \sqrt{10} \approx 3.1623; \\ grayblack e_{ad} &= \sqrt{(x_d - x_a)^2 + (y_d - y_a)^2} = \sqrt{(0-0)^2 + (2+2)^2} = \sqrt{4^2} = 4; \\ e_{bc} &= \sqrt{(x_c - x_b)^2 + (y_c - y_b)^2} = \sqrt{(1+1)^2 + (1-1)^2} = \sqrt{2^2} = 2. \end{aligned} \quad (3)$$

Найдем максимальное ребро: очевидно, это неявное ребро - диагональ:  $e^e = e_{ad} = e_{bc} = 4$ , тогда матрица формы

$$\mathbf{M} = \begin{pmatrix} a & b & c & d \\ 0 & \frac{e_{ab}}{e^e} & \frac{e_{ac}}{e^e} & \frac{e_{ad}}{e^e} \\ \frac{e_{ba}}{e^e} & 0 & \frac{e_{bc}}{e^e} & \frac{e_{bd}}{e^e} \\ \frac{e_{ca}}{e^e} & \frac{e_{cb}}{e^e} & 0 & \frac{e_{cd}}{e^e} \\ \frac{e_{da}}{e^e} & \frac{e_{db}}{e^e} & \frac{e_{dc}}{e^e} & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{\sqrt{10}}{e^e} & \frac{\sqrt{10}}{e^e} & \frac{4}{e^e} \\ \frac{\sqrt{10}}{e^e} & 0 & \frac{2}{e^e} & \frac{\sqrt{2}}{e^e} \\ \frac{\sqrt{10}}{e^e} & \frac{2}{e^e} & 0 & \frac{\sqrt{2}}{e^e} \\ \frac{4}{e^e} & \frac{\sqrt{2}}{e^e} & \frac{\sqrt{2}}{e^e} & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{\sqrt{5}}{2\sqrt{2}} & \frac{\sqrt{5}}{2\sqrt{2}} & 1 \\ \frac{\sqrt{5}}{2\sqrt{2}} & 0 & \frac{1}{2} & \frac{1}{2\sqrt{2}} \\ \frac{\sqrt{5}}{2\sqrt{2}} & \frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} \\ 1 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & 0 \end{pmatrix}.$$

При произвольной индексации вершин матрица будет иметь другой вид, но значения будут те же, только стоять они будут в других местах. На языке теории графов это называется изоморфизм, т.е. сохраняющий/независимый от формы.

Матрица смежности  $\mathbf{A}$  задаётся видимыми ребрами, например, через команду polyline (autocad) для любой индексации вершин

$$polyline [0;2],[1;1],[0;-2],[-1;1],[0;2]; \quad //d \rightarrow c \rightarrow a \rightarrow b \rightarrow d.$$

### 1. Построение матрицы формы $\mathbf{M}$

Приведем конструктивную процедуру построения матрицы формы: пусть у нас имеются множество точек фигуры, для нашего примера это  $a, b, c, d$ , сформируем из него матрицу точек  $G$

$$G = \begin{pmatrix} a & b & c & d \end{pmatrix} = \begin{pmatrix} x_a & x_b & x_c & x_d \\ y_a & y_b & y_c & y_d \end{pmatrix} = \begin{pmatrix} 0 & -1 & 1 & 0 \\ -2 & 1 & 1 & 2 \end{pmatrix}. \quad (4)$$

Мы записали матрицу  $G$  в привычном представлении, где вектора точек располагаются по столбцам, при обработке, однако, удобнее иметь транспонированное представление –  $G^T$ , поскольку согласно Fortran нотации  $x$  координаты будут лежать в одном столбце,  $y$  – в другом и т.д., при таком хранении увеличивается скорость работы за счет равномерного доступа к данным и векторизации, в нашем случае по столбцам.

Из матрицы (4) точек получим матрицу векторов (6) (полный граф межточечных соединений), но поскольку это будет уже многомерная матрица (тензор), для привычного двумерного вида разобьём её на разные компоненты (одну – с координатами  $x$ , другую – с  $y$  и т.д.). Тогда

$$\begin{aligned} G_x &= \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot G; \\ G_y &= \begin{pmatrix} 0 & 1 \end{pmatrix} \cdot G. \end{aligned} \quad (5)$$

$$\begin{aligned} V_x &= j \otimes G_x - G_x^T \otimes j^T; \\ V_y &= j \otimes G_y - G_y^T \otimes j^T, \end{aligned} \quad (6)$$

где  $\otimes$  – произведение Кронекера, а  $j$  – вектор-столбец, состоящий из единиц (столбец матрицы  $J$ )  $j^T = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$

$$\begin{aligned} V &= \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} a & b & c & d \end{pmatrix} - \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} a-a & b-a & c-a & d-a \\ a-b & b-b & c-b & d-b \\ a-c & b-c & c-c & d-c \\ a-d & b-d & c-d & d-d \end{pmatrix}; \\ V &= \begin{pmatrix} \overline{aa} & \overline{ab} & \overline{ac} & \overline{ad} \\ \overline{ba} & \overline{bb} & \overline{bc} & \overline{bd} \\ \overline{ca} & \overline{cb} & \overline{cc} & \overline{cd} \\ \overline{da} & \overline{db} & \overline{dc} & \overline{dd} \end{pmatrix} = \begin{pmatrix} 0 & \overline{ab} & \overline{ac} & \overline{ad} \\ \overline{ba} & 0 & \overline{bc} & \overline{bd} \\ \overline{ca} & \overline{cb} & 0 & \overline{cd} \\ \overline{da} & \overline{db} & \overline{dc} & 0 \end{pmatrix}, \end{aligned}$$

где  $\overline{ab} = \begin{bmatrix} x_b - x_a \\ y_b - y_a \end{bmatrix}$ , например. Тогда матрица длин векторов будет очевидно симметрична  $L = L^T$  и иметь вид

$$L = \sqrt{\sum_{i=1}^{dim} (V_i \circ V_i)}, \quad (7)$$

где  $\sqrt{\phantom{x}}$  – поэлементный квадратный корень (matlab: sqrt()/realsqrt(), julia: sqrt.()) или в записи через нормы

$$L = \begin{pmatrix} 0 & P\overline{ab}P_2 & P\overline{ac}P_2 & P\overline{ad}P_2 \\ P\overline{ba}P_2 & 0 & P\overline{bc}P_2 & P\overline{bd}P_2 \\ P\overline{ca}P_2 & P\overline{cb}P_2 & 0 & P\overline{cd}P_2 \\ P\overline{da}P_2 & P\overline{db}P_2 & P\overline{dc}P_2 & 0 \end{pmatrix}, \quad (8)$$

и матрица формы уже выражается как

$$\mathbf{M} = \frac{L}{Pvec(L)P_\infty} = \frac{1}{Pvec(L)P_\infty} I \cdot L = \frac{1}{k^*} I \cdot L. \quad (9)$$

Из (9) нам нужен только верхний треугольник, и для его получения не обязательно выполнять все матричные операции, достаточно только тех что его определяют. Для этого можно сразу сделать  $V$  в нужном виде:

$$vech(V)^T = \begin{pmatrix} \vec{ab} \\ \vec{ac} \\ \vec{ad} \\ \vec{bc} \\ \vec{bd} \\ \vec{cd} \end{pmatrix} = \left( G \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \right)^T - \left( G \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \right)^T, \quad (10)$$

а  $L$  рассчитывать по той же формуле (7) уже для матрицы  $vech(V)$ .

Хотя мы сознательно избавились от углов в задании фигуры, они всё равно присутствуют неявно через теорему косинусов  $\cos \gamma = \frac{a^2 + b^2 - c^2}{2ab}$ , равно как объём и площадь фигуры можно вычислить по

матрице длин (8) через формулу Герона:  $S = \frac{1}{4} \sqrt{4a^2b^2 - (a^2 + b^2 - c^2)^2}$ , триангулируя предварительно

фигуру на элементарные треугольники, что собственно и было сделано Артуром Кэли совместно с Карлом Менгером. Совокупность углов будет являться обобщенными координатами (понятие из аналитической механики), наравне с длинами и координатами точек и использовать их можно в тех приложениях, где углы вычислить проще, чем расстояния (например, в сферической астрономии).

В силу специфики статьи для CAD-систем мы ориентировались на естественный в данной области способ задания фигуры посредством координат вершин, но он не единственный. Из записи (9) можно видеть, что нам достаточно знать только длины отрезков, вообще говоря, всех явных и неявных. Можно обойтись только явными, а неявные выразить через них, этот процесс тут опустим, поскольку самое главное, на чем надо акцентировать внимание, – это то, что нам нужны только длины ребер (без углов и координат, и даже не сами длины, а их соотношения), а значит данную технику можно использовать и в сфере реального распознавания образов, т.е. посредством камер с реального мира. Ребра и длины можно выделить и посчитать из данных с ПЗС-матрицы или (для краткости) сетчатки. Постановка задачи там гораздо сложнее, поскольку имеются искажения и вообще нужно уметь распознавать движения и выделять инерциальные системы координат.

## 2. Преобразования фигуры

При проведении преобразований над точками, матрица длин (7) изменится. Так, вектор по двум точкам (допустим,  $a, b$ ) будет

$$\vec{ab} = \begin{bmatrix} x_b - x_a \\ y_b - y_a \end{bmatrix} = \begin{bmatrix} x_{ab} \\ y_{ab} \end{bmatrix}.$$

Его длину можно выразить через внутреннее произведение векторов

$$\langle \vec{ab}, \vec{ab} \rangle = \vec{ab}^T \cdot \vec{ab};$$

$$P \vec{ab} P_2 = \sqrt{\langle \vec{ab}, \vec{ab} \rangle} = \rho = \sqrt{2\sigma},$$

где  $\rho$  – расстояние;  $\sigma$  – мировая функция, введенная Рыловым и независимо Сингом [10]. Тогда если мы произведём линейное (или нет) преобразование над точками  $a' = Pa, b' = Pb$ , где  $P = \begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix}$ , то

$$\vec{ab}' = \begin{bmatrix} p_1(x_b - x_a) + p_2(y_b - y_a) \\ p_3(x_b - x_a) + p_4(y_b - y_a) \end{bmatrix} = \begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix} \cdot \begin{bmatrix} x_b - x_a \\ y_b - y_a \end{bmatrix} = P \cdot \vec{ab},$$

поэтому внутреннее произведение преобразованного вектора будет

$$\langle \vec{ab}', \vec{ab}' \rangle = \vec{ab}'^T P^T \cdot P \vec{ab}.$$

При ортогональных преобразованиях  $P = Q$ , где  $Q^T = Q^{-1}$ , произведение  $Q^T \cdot Q = I$  и соответственно матрица длин остаётся без изменений, при деформирующих преобразованиях это в общем виде не так. Хотя матрица  $L$  не сохраняется при деформирующем равномерном масштабировании  $P_s = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$

$$L_s = k \sqrt{(V_x \circ V_x + V_y \circ V_y)}, \quad (11)$$

но матрица формы остаётся без изменений, поскольку

$$\mathbf{M}_s = \frac{L_s}{\text{Pvec}(L_s) \text{P}_\infty} = \frac{kL}{k \text{Pvec}(L) \text{P}_\infty} = \frac{L}{\text{Pvec}(L) \text{P}_\infty}. \quad (12)$$

Это особенно важно при реальном распознавании образов с перспективных проекций. Таким образом, можно сделать следующую классификацию преобразований:

1. Перестановка (Переиндексация вершин)  $A' = P^{-1}AP = P^TAP$ .
2. Параллельный перенос  $\Delta = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$ .
3. Вращение (Поворот)  $\mathbf{P} = \mathbf{Q} = \begin{pmatrix} \cos(\varphi) & \mp \sin(\varphi) \\ \pm \sin(\varphi) & \cos(\varphi) \end{pmatrix}$ .
4. Отражение  $\mathbf{P} = \mathbf{Q} = I - 2uu^T$ , где  $u$  – вектор единичной длины, ортогональный некоторой прямой  $\ell$ , относительно которой происходит отражение.
5. Равномерное масштабирование  $\mathbf{P} = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} = kI$ .
6. Растяжение / Сжатие  $\mathbf{P} = \begin{pmatrix} k_x & 0 \\ 0 & k_y \end{pmatrix}$ .
7. Разрушающие преобразования.
8. Последовательность преобразований  $\mathbf{P} = \mathbf{P}_1 \cdot \mathbf{P}_2 \neq \mathbf{P}_2 \cdot \mathbf{P}_1$  в общем виде важна, однако если матрицы коммутирующие, то порядок не важен.
9. Комбинация ортогональных и деформирующих преобразований:
  - $x^T \mathbf{P}^T \mathbf{Q}^T \mathbf{Q} \mathbf{P} x = x^T \mathbf{P}^T \mathbf{P} x$  при порядке преобразований: сначала деформирующие, потом ортогональные, последние можно не учитывать;
  - $x^T \mathbf{Q}^T \mathbf{P}^T \mathbf{P} \mathbf{Q} x$  Сначала ортогональные, потом деформирующие: в этом случае учитывать нужно все.

На матрицу  $L$  не влияют преобразования 1-4, матрица формы  $\mathbf{M}$  не изменяется преобразованиями 1-5, матрица смежности  $A$  нечувствительна (в смысле сохранения смежности – изоморфизма) к преобразованиям 1-6. В общем случае в результате разрушающих преобразований меняется матрица смежности и матрица формы, но можно искать максимальное вхождение подграфа MCS, т.е. искать неразрушенные части фигуры и анализировать их деформации.

### 3. Переиндексация (Перестановка) графа

Хотя в геометрии редко применяют процедуры переиндексации вершин (точек), для алгоритмов на изоморфизм графа это, можно сказать, основной инструмент, поэтому подробнее распишем и зафиксируем эту процедуру. Суть переиндексации (перестановки) графа – генерация нового изоморфного графа. За отправную точку положим, что нам известна новая последовательность индексов вершин графа, а именно следующая таблица 2 (биекция), или перестановка [5].

Таблица 2. Переиндексация вершин графа

$i$	$j, P^T$	$j, P$
<b>0</b> (a)	<b>1</b> (b)	<b>2</b>
<b>1</b> (b)	<b>2</b> (c)	<b>0</b>

Окончание таблицы 2

$i$	$j, P^T$	$j, P$
<b>2 (c)</b>	<b>0 (a)</b>	<b>1</b>
<b>3 (d)</b>	<b>3 (d)</b>	<b>3</b>

$$\pi : \begin{pmatrix} 0 & 1 & 2 & 3 \\ 2 & 0 & 1 & 3 \end{pmatrix}$$

или в виде циклической перестановки

$$\pi : (0, 2, 1)(3).$$

Само преобразование подобия над матрицей смежности  $A$  выражается в следующем виде:

$$A' = P^{-1}AP = P^TAP, \quad (13)$$

где  $P$  – матрица перестановок порождается перестановкой  $\pi$ . Учитывая, что множитель  $P$  переставляет столбцы, а  $P^T$  – строки матрицы смежности, тогда в ячейках  $P[i, j] = 1$ , в остальных – нули:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Обычно в решателях обходятся без этой дорогостоящей процедуры, достаточно лишь таблицы 2, полная перестановка нужна в случае, если необходимо действительно переставить вершины в графе, перед пересохранением, например.

Полный пакет преобразований над основными матрицами  $A, G, M$  будет

$$\begin{aligned} A' &= P^{-1}AP = P^TAP; \\ G' &= GP; \\ M' &= P^TMP. \end{aligned} \quad (14)$$

Эти три матрицы в теории графов играют каждая свою роль:

- $A$  – матрица смежности, определяет топологию (структуру фигуры), размерность  $n \times n$ ;
- $G$  – матрица атрибутов при вершинах, характеризует свойства вершин графа, размерность  $3n/2n/n = k \times n$ ;
- $M$  – матрица реберных атрибутов, задаёт параметры связей между объектами, размерность  $n \times n$ .

#### 4. Построения фигуры по матрицам $A, M$

Опишем процесс построения фигуры воздушного змея (Kite), заданной матрицами структурной связи  $A$  и матрицей формы  $M$ . Фигура или граф, получаемый при этом (рис. 2), является некой разновидностью графа единичных кругов [14, 15] с той разницей, что круги не единичного диаметра а задаются матрицей формы, а в 3D-случае вместо кругов будут сферы.

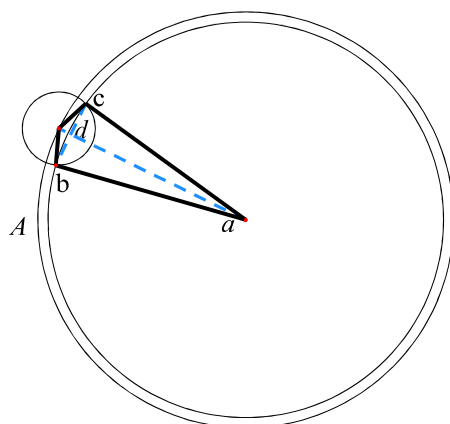


Рисунок 2. Метод окружностей



1. Зафиксируем параметры  $\beta = 0.3$  и  $\gamma = 0.1$ , построим матрицу  $M$  и выберем нужную длину самого большого ребра  $|ad| = 5$ .
2. Выберем точку, от которой отходит самая длинная сторона (и/или точку с самым большим количеством ребер)  $a$ , и разместим её на плоскости (в любом месте, если не оговорено иное).
3. Для каждого ребра (и неявного тоже) точки  $a$  раствором циркуля, равным соответствующей длине, чертим окружности.
4. На окружности наибольшего радиуса (самая длинная сторона) опять же произвольно выбираем точку, у нас это будет точка  $d$ , внутри этой окружности и будет лежать наша фигура.
5. Из точки  $d$  раствором циркуля, равным длинам исходящих из этой точек ребер, чертим окружности. В нашем случае радиусы этих окружностей будут равны.
6. Точки пересечений выше построенных окружностей с окружностями (меньшего чем  $|ad|$  радиуса) с центром в точке  $a$  и будут искомыми точками  $b$  и  $c$ .
7. При размещении точек  $b$  и  $c$  нужно, если требуется, соблюдать хиральность, т.е. сохранять знак векторного произведения векторов.

#### Несколько замечаний

Из процесса построения видно, что система координат (как и размерные величины) никоим образом не присутствуют, напротив, сама фигура обладает характеристиками, играющими роль внутренней системы координат. Так например, определяющей величиной является размер нормирующего вектора, от которого откладываются все остальные. Основные точки можно найти, исследуя структуру матрицы  $A$  – граф связи, такими, скажем, могут являться максимальная клика графа, узел с максимальной степенью и т.д. В нашем случае это  $a$  и  $d$  – точки, являющиеся концами вектора наибольшей длины. Но точка  $a$  предпочтительней, поскольку от неё отходят наибольшие отрезки, длинные стороны ромбоида и нормирующая диагональ.

#### Поиск симметрий фигуры через автоморфизм графа (AUT)

```

0-----
1 #Create Autocad figure from pseudo command-line example
2  # nodes setting area on graph notations
3  a = [ 0.0,-2.0]; b = [-1.0, 1.0]; c = [ 1.0, 1.0]; d = [ 0.0, 2.0];
4
5  # edges setting area on graph notations
6  line(a,b); line(b,d); line(d,c); line(c,a);
7-----
8 # Mapping to graph (kite_graph.grf format example)
9 # Pattern graph
10 # Graph with vertex attributes
11  # nodes creating area -> define $M matrix
12  4      # nodes count
13  0 0.0 -2.0 # a mapping (id + v.attr)
14  1 -1.0 1.0 # b
15  2 1.0 1.0 # c
16  3 0.0 2.0 # d
17
18 # edges creating area -> define A matrix
19 # necessary minimum for edge storage:
20 # 0 1 2  # node 0 have connection with 1 and 2 nodes
21 # 3 1 2  # node 3 have connection with 1 and 2 nodes
22 2      # edges count from node 0
23 0 1    # line(a,b) (id1 id2)
24 0 2    # line(a,c)
25 1      # edges count from node 1
26 1 3    # line(b,d)
27 1      # edges count from node 2

```

```

28 2 3    # line(c,d)
29 0      # edges count from node 3
30-----

```

Листинг 1. Отображение создания CAD-модели на граф

Для простоты обозначим  $l = \sqrt{\frac{1}{(1+\gamma)^2} + \frac{\beta^2}{4}}$ ,  $s = \sqrt{\frac{\gamma^2}{(1+\gamma)^2} + \frac{\beta^2}{4}}$  – соответственно длинные и короткие стороны ромбоида. Тогда исходные матрицы будут

$$A = \begin{pmatrix} a & b & c & d \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} 0 & l & l & 1 \\ l & 0 & \beta & s \\ l & \beta & 0 & s \\ 1 & s & s & 0 \end{pmatrix}.$$

Условия разрешимости можно записать в табличном виде (табл. 3), в виде матрицы или на естественном языке (табл. 4) [16].

Таблица 3. Условия разрешимости задачи

level	node	V.cond	E.cond
0	a	$d[0 2] \wedge \mathbf{M}[0 l,l]$	–
1	b	$d[1 2] \wedge \mathbf{M}[1 l,s]$	$A[0] \wedge \mathbf{M}[0 l]$
2	c	$d[2 2] \wedge \mathbf{M}[2 l,s]$	$(A[0] \wedge \mathbf{M}[0 l]) \wedge (\overline{A[1]} \wedge \mathbf{M}[1 \beta])$
3	d	$d[3 2] \wedge \mathbf{M}[3 s,s]$	$(\overline{A[0]} \wedge \mathbf{M}[0 1]) \wedge (A[1] \wedge \mathbf{M}[1 s]) \wedge (A[2] \wedge \mathbf{M}[2 s])$

При решении этой задачи посредством какого-либо алгоритма (см. обзор и библиографию [18]) на изоморфизм (с поддержкой атрибутов на ребра) получим таблицу биекций (табл. 5). Откуда формулу симметрии (без учета хиральности) можно записать в виде, схожим по своей структуре с циклической перестановкой:

$$S_e = \{a\}, \{b, c\}, \{d\} \longrightarrow \pi : (a)(b, c)(d).$$

Таблица 4. Предикаты разрешимости на естественном языке

level	node	Natural Language Predicates
0	a	для заполнения 0-го слоя нужны вершины с 2 соседями <b>и</b> расстоянием $l$ до них
1	b	вершины с двумя соседями <b>и</b> с расстояниями $l, s$ <b>и</b> вершина 0-го слоя имеет связь размера $l$ .
2	c	вершины с двумя соседями <b>и</b> с расстояниями $l, s$ <b>и</b> вершина 0-го слоя имеет связь размера $l$ <b>и</b> вершина 1-го слоя имеет виртуальную связь ( <b>не</b> имеет видимой связи) размера $\beta$ .
3	d	вершины с двумя соседями <b>и</b> с расстоянием $s, s$ <b>и</b> вершина 0-го слоя имеет виртуальную связь ( <b>не</b> имеет видимой связи) размера 1 <b>и</b> вершины 1-го <b>и</b> 2-го слоёв имеют связи размеров $s$ .

Отметим, что с точки зрения топологической симметрии (изоморфизма графа без атрибутов), симметричными будут все вершины  $\{a, b, c, d\}$ , а вот при учете хиральности все вершины будут уникальные  $\{a\}, \{b\}, \{c\}, \{d\}$ .

Таблица 5. Биекции задачи на автоморфизм графа

level	pattern node	#1	#2
0	a	a	a
1	b	b	c

Окончание таблицы 5

level	pattern node	#1	#2
2	c	c	b
3	d	d	d

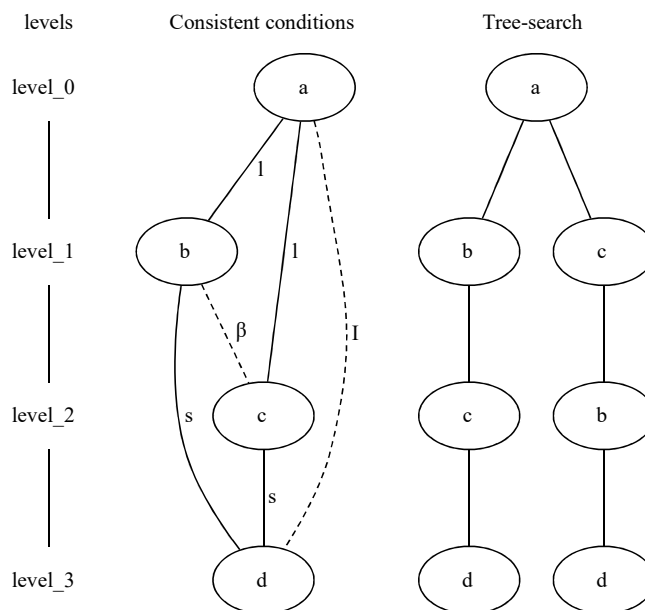


Рисунок 3. Условия разрешимости для задачи симметрии и результирующее дерево поиска

### Извлечение равномерно деформированной фигуры на сложном чертеже через изоморфизм подграфа (SI)

Приведем пример извлечения заданной фигуры из неструктурированного чертежа (рис.4). В качестве искомого образца возьмем фигуру kite (см. рис. 1) с матрицей координат

$$G_{kite} = \begin{pmatrix} a & b & c & d \\ 0 & -1 & 1 & 0 \\ -2 & 1 & 1 & 2 \end{pmatrix}.$$

На рисунке этот образец разместим с помощью следующих преобразований:

$$G_{kite}^{SI} = \begin{pmatrix} 11 \\ 15 \end{pmatrix} + Q \cdot P \cdot \left( \begin{pmatrix} 0 \\ 2 \end{pmatrix} + G_{kite} \right) = \begin{pmatrix} 11.0 & 11.5072 & 12.8928 & 12.6 \\ 15.0 & 17.4785 & 16.6785 & 17.7713 \end{pmatrix},$$

где матрица вращения  $Q = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix} = \begin{pmatrix} 0.866025 & 0.5 \\ -0.5 & 0.866025 \end{pmatrix}$  при  $\varphi = -30^\circ$  и матрица

равномерного скалирования  $P = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.8 \end{pmatrix}$ . И с помощью алгоритма на изоморфизм подграфа мы

должны найти эту комбинацию точек их взаимного расположения среди всех точек и линий чертежа. Размещение фигуры можно производить не по координатам вершин, а например, с помощью графового метода дисков заданного радиуса [14, 15], где в качестве точки отсчета можно взять точку  $pt17 = [11.0; 15.0]$ .

Сам чертеж имеет простую структуру (рис. 5), состоящую из 9 компонент связности, хотя, по правилам, принятым в конструкторских фирмах, структура может несколько меняться, всё зависит от того, как рисуют пересечение линий и точки на линиях. Если точка делит линию, то задача становится сложной из-за возникающего гомеоморфизма.

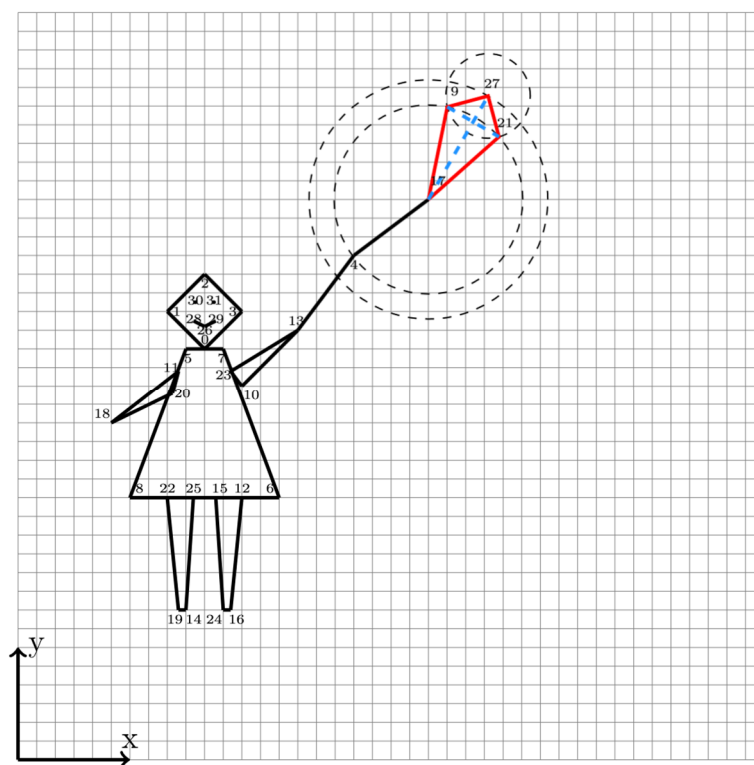


Рисунок 4. Эпическое полотно 'Kate and Kite'

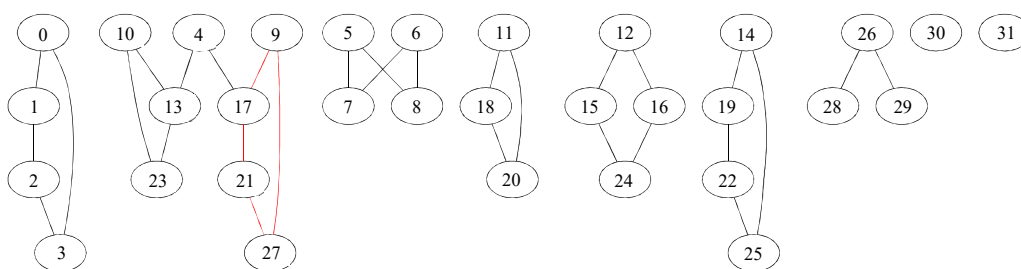


Рисунок 5. Графовое (структурное) представление чертежа без атрибутов

```

0 Solution count: 40
1 pattern size: 4
2 -----
3 Total states exploration: 134
4 Total states handled: 174
5 --- Breeding factor statistics ---
6 0: |28: 1 | width: 28, average value: 28 disp: 0
7 1: |0: 1 |2: 25 |3: 2 | width: 56, average value: 2 disp: 0.21428573
8 2: |0: 10 |1: 42 |2: 4 | width: 50, average value: 0.89285713 disp: 0.23852043
9 3: |0: 10 |1: 40 | width: 40, average value: 0.80000001 disp: 0.16
10 Total Width = 174
11 0:27|27|25|25|24|24|22|22|21|21|19|19|17|17|16|16|15|15|14|14|12|12|9|9|8|8|7|7|6|6|5|5|3|3|2|2|1|1|0|0
12 1:21|9|22|14|16|15|25|19|27|17|22|14|21|9|24|12|24|12|25|19|16|15|27|17|6|5|6|5|8|7|8|7|2|0|3|1|2|0|3|1
13 2:9|21|14|22|15|16|19|25|17|27|14|22|9|21|12|24|12|24|19|25|15|16|17|27|5|6|5|6|7|8|7|8|0|2|1|3|0|2|1|3
14 3:17|17|19|19|12|12|14|14|9|9|25|25|27|27|15|15|16|16|22|22|24|24|21|21|7|7|8|8|5|5|6|6|1|1|0|0|3|3|2|2
15 Solutions after symmetry: 5
16 1: {27,17,21,9}
17 2: {25,22,14,19}
18 3: {24,16,15,12}
19 4: {8,6,5,7}
20 5: {3,2,0,1}

```

Листинг 2. Вывод алгоритма изоморфизма подграфа

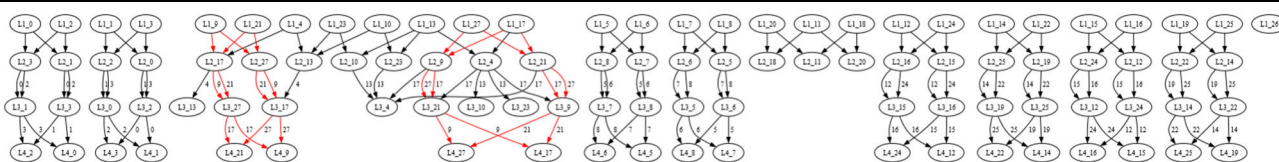


Рисунок 6. Типовой дамп дерева поиска решателями 1-го поколения (алг.1)

Таблица 6. Список найденных четырехугольников и их матрицы формы

#	Quadrilaterals	M matrix	Shape
0	{27,17,21,9}	$\begin{pmatrix} 0.0 & 1.0 & 0.3536 & 0.3536 \\ 1.0 & 0.0 & 0.7906 & 0.7906 \\ 0.3536 & 0.7906 & 0.0 & 0.5 \\ 0.3536 & 0.7906 & 0.5 & 0.0 \end{pmatrix}$	Kite
1	{25,22,14,19}	$\begin{pmatrix} 0.0 & 0.3972 & 0.9827 & 1.0 \\ 0.3972 & 0.0 & 0.9665 & 0.9581 \\ 0.9827 & 0.9665 & 0.0 & 0.0636 \\ 1.0 & 0.9581 & 0.0636 & 0.0 \end{pmatrix}$	Trapezoid
2	{24,16,15,12}	$\begin{pmatrix} 0.0 & 0.0636 & 0.9827 & 0.9665 \\ 0.0636 & 0.0 & 1.0 & 0.9581 \\ 0.9827 & 1.0 & 0.0 & 0.3972 \\ 0.9665 & 0.9581 & 0.3972 & 0.0 \end{pmatrix}$	Trapezoid
3	{8,6,5,7}	$\begin{pmatrix} 0.0 & 0.848 & 0.9057 & 1.0 \\ 0.848 & 0.0 & 1.0 & 0.9057 \\ 0.9057 & 1.0 & 0.0 & 0.212 \\ 1.0 & 0.9057 & 0.212 & 0.0 \end{pmatrix}$	Trapezoid
4	{3,2,0,1}	$\begin{pmatrix} 0.0 & 0.7071 & 0.7071 & 1.0 \\ 0.7071 & 0.0 & 1.0 & 0.7071 \\ 0.7071 & 1.0 & 0.0 & 0.7071 \\ 1.0 & 0.7071 & 0.7071 & 0.0 \end{pmatrix}$	Square

### 1. Пример анализа преобразований фигуры

В качестве примера опишем процедуру восстановления к. масштабирования (в нашем случае он взят 0.8 по построению), поскольку он наиболее важен, в перспективных проекциях, через него определяется расстояние до объекта. Так, после задачи GI с атрибутами на ребра алгоритма 1 и учета симметрии графа шаблона  $S_e = \{a\} \{b,c\} \{d\}$  получим искомую таблицу 7 биекций с выровненным порядком индексации вершин. Матрица формы фигуры на чертеже, очевидно, будет равна аналогичной матрице паттерн графа, а вот матрица длин в случае скалирования будет другая.

Таблица 7. Искомые биекции алгоритма 1

levels	ptr	trg
0	a	17
1,2	b,c	9,21
3	d	27

$$L_{kite}^{SI} = \rho\_matrix([pt17, pt9, pt21, pt27]) = \begin{pmatrix} 0.0 & 2.5298 & 2.5298 & 3.2 \\ 2.5298 & 0.0 & 1.6 & 1.1314 \\ 2.5298 & 1.6 & 0.0 & 1.1314 \\ 3.2 & 1.1314 & 1.1314 & 0.0 \end{pmatrix}.$$

Видно, что матрица длин отличается от исходной (3) на множитель 0.8, для его нахождения достаточно взять отношение максимальной дины в найденной форме к максимальной длине оригинальной фигуры  $\frac{3.2}{4} = 0.8$  или любых других соответствующих элементов, кроме диагональных.

```

0 -----
1 #Create AutoCAD figure from pseudo command-line example
2 # nodes setting area on graph notations
3 pt0=[5.0,11.0]; pt1=[4.0,12.0]; pt2=[5.0,13.0]; pt3=[6.0,12.0];
4 pt4=[9.0,13.5]; pt5=[4.5,11.0]; pt6=[7.0,7.0]; pt7=[5.5,11.0];
5 pt8=[3.0,7.0]; pt9=[11.50718,17.47846]; pt10=[6.0,10.0]; pt11=[4.3,10.4];
6 pt12=[6.0,7.0]; pt13=[7.5,11.5]; pt14=[4.5,4.0]; pt15=[4.75,7.0];
7 pt16=[5.7,4.0]; pt17=[11.0,15.0]; pt18=[2.5,9.0]; pt19=[4.3,4.0];
8 pt20=[4.14,9.8]; pt21=[12.89282,16.67846]; pt22=[4.0,7.0]; pt23=[5.7,10.4];
9 pt24=[5.5,4.0]; pt25=[5.25,7.0]; pt26=[5.0,11.6]; pt27=[12.6,17.77128];
10 pt28=[4.7,11.75]; pt29=[5.3,11.75]; pt30=[4.75,12.25]; pt31=[5.25,12.25];
11
12 # edges setting area on graph notations
13 # head
14
15 line(pt0,pt1);line(pt1,pt2);line(pt2,pt3);
16 line(pt3,pt0);line(pt28,pt26);line(pt26,pt29);
17 # body
18 line(pt5,pt7);line(pt7,pt6);line(pt6,pt8);line(pt8,pt5);
19 # right hand
20 line(pt18,pt11);line(pt11,pt20);line(pt20,pt18);
21 # left hand
22 line(pt13,pt10);line(pt10,pt23);line(pt23,pt13);
23 # right leg
24 line(pt19,pt14);line(pt14,pt25);line(pt25,pt22);line(pt22,pt19);
25 # left leg
26 line(pt24,pt16);line(pt16,pt12);line(pt12,pt15);line(pt15,pt24);
27 # rope
28 line(pt13,pt4);line(pt4,pt17);
29 # kite
30 line(pt17,pt21);line(pt21,pt27);line(pt27,pt9);line(pt9,pt17);
31 -----

```

Листинг 3. CAD-модель и её граф

```

Data:  $H_p(A_p, G_p), H_t(A_t, G_t)$ 
Result:  $B_{(V_p \rightarrow V_t)}^{A,M}, B_{(V_p \rightarrow V_t)}^A$  ; /* Struct+Semantic and only Struct Bijection tables */
1  $M_p \leftarrow M\_matrix(G_p)$ ;
2  $H_p(A_p, G_p) \leftarrow M_p$ ;
3  $C \leftarrow ConsistentConditions(H_p(A_p, G_p, M_p))$ ;
4  $Matchings \leftarrow SI(C, H_t(A_t))$ ; /* Subgraph Isomorphism (not e.attributed) task */
5 while  $size(Matchings) \neq 0$  do
6    $H_s(A_s, G_s) \leftarrow Matchings.pop()$ ;
7    $M_s \leftarrow M\_matrix(G_s)$ ;
8    $H_s(A_s, G_s) \leftarrow M_s$ ;
9    $S \leftarrow GI(C, H_s(A_s, M_s))$ ; /* Graph Isomorphism (e.attributed) task */
10  if  $S \neq 0$  then
11     $B_{(V_p \rightarrow V_t)}^{A,M}.push(S)$ ; /* Struct & Semantic solutions */
12  else
13     $B_{(V_p \rightarrow V_t)}^A.push(H_s(A_s))$ ; /* Struct only solutions */
14  end
15 end

```

Алгоритм 1. Алгоритм разделённого решения

## 2. Графы отношений высших порядков

Матрица  $M$  не единственная, что может описать форму, в некоторых задачах может применяться матрица  $L$ : например, когда размеры объекта известны и не меняются. Но всё-таки больший интерес вызывает задача классификации форм, т.е. выявление отличий, например, ромба от kite или от квадрата и распознавания именно формы фигуры (например kit'ы разных пропорций). Тематика этой статьи ограничена постановкой задачи извлечения форм, а то, что сказано ниже, уже относится к сфере искусственного интеллекта, т.е. алгоритмам логической классификации объектов и выявления общих черт. Приведем здесь только методику построения графов отношений, которые уже применимы для автоматического построения логических предикатов и последующего их решения.

Для фигуры из четырех углов ( $n = 4$ ) мы имеем  $e = \frac{n(n-1)}{2} = 6$  ребер, 4 видимых и 2 неявных

(диагонали). Сформируем из них также матрицу отношений (записанную в лексикографическом порядке):

$$\Lambda = \begin{pmatrix} 1 & P\bar{a}c P_2 P\mathbf{a}b P_2 & P\bar{a}d P_2 P\mathbf{a}b P_2 & P\bar{b}c P_2 P\mathbf{a}b P_2 & P\bar{b}d P_2 P\mathbf{a}b P_2 & P\bar{c}d P_2 P\mathbf{a}b P_2 \\ P\bar{a}b P_2 P\bar{a}c P_2 & 1 & P\bar{a}d P_2 P\bar{a}c P_2 & P\bar{b}c P_2 P\bar{a}c P_2 & P\bar{b}d P_2 P\bar{a}c P_2 & P\bar{c}d P_2 P\bar{a}c P_2 \\ P\bar{a}b P_2 P\mathbf{a}d P_2 & P\bar{a}c P_2 P\mathbf{a}d P_2 & 1 & P\bar{b}c P_2 P\mathbf{a}d P_2 & P\bar{b}d P_2 P\mathbf{a}d P_2 & P\bar{c}d P_2 P\mathbf{a}d P_2 \\ P\bar{a}b P_2 P\bar{b}c P_2 & P\bar{a}c P_2 P\bar{b}c P_2 & P\bar{a}d P_2 P\bar{b}c P_2 & 1 & P\bar{b}d P_2 P\bar{b}c P_2 & P\bar{c}d P_2 P\bar{b}c P_2 \\ P\bar{a}b P_2 P\bar{b}d P_2 & P\bar{a}c P_2 P\bar{b}d P_2 & P\bar{a}d P_2 P\bar{b}d P_2 & P\bar{b}c P_2 P\bar{b}d P_2 & 1 & P\bar{c}d P_2 P\bar{b}d P_2 \\ P\bar{a}b P_2 P\bar{c}d P_2 & P\bar{a}c P_2 P\bar{c}d P_2 & P\bar{a}d P_2 P\bar{c}d P_2 & P\bar{b}c P_2 P\bar{c}d P_2 & P\bar{b}d P_2 P\bar{c}d P_2 & 1 \end{pmatrix} \quad (15)$$

и конкретное численное воплощение kite-матрицы:

$$\Lambda_{\diamond} = \begin{pmatrix} 1.0 & \mathbf{1.0} & 1.2649 & 0.6325 & 0.4472 & 0.4472 \\ \mathbf{1.0} & 1.0 & 1.2649 & 0.6325 & 0.4472 & 0.4472 \\ 0.7906 & 0.7906 & 1.0 & \mathbf{0.5} & 0.3536 & 0.3536 \\ 1.5811 & 1.5811 & \mathbf{2.0} & 1.0 & \mathbf{0.7071} & \mathbf{0.7071} \\ 2.2361 & 2.2361 & 2.8284 & \mathbf{1.4142} & \mathbf{1.0} & \mathbf{1.0} \\ 2.2361 & 2.2361 & 2.8284 & \mathbf{1.4142} & \mathbf{1.0} & 1.0 \end{pmatrix}. \quad (16)$$

По числам видно, какие стороны равны (1 в позиции), какие имеют целые отношения ( $\frac{1}{\beta} = 2$ ),

а также скрытые отношения, которые сразу не видно – ( $1.41421 \approx \sqrt{2} = \frac{P\bar{b}c P_2}{P\bar{c}d P_2}$ ) – гипотенуза

равнобедренного прямоугольного треугольника  $\triangle bdc$ . Для квадрата, например,  $\Lambda_{\square}$  будет куда более выразительна:

$$\Lambda_{\square} = \begin{pmatrix} 1.0 & 1.0 & 1.4142 & 1.4142 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.4142 & 1.4142 & 1.0 & 1.0 \\ 0.7071 & 0.7071 & 1.0 & 1.0 & 0.7071 & 0.7071 \\ 0.7071 & 0.7071 & 1.0 & 1.0 & 0.7071 & 0.7071 \\ 1.0 & 1.0 & 1.4142 & 1.4142 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.4142 & 1.4142 & 1.0 & 1.0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sqrt{2} & \sqrt{2} & 1 & 1 \\ 1 & 1 & \sqrt{2} & \sqrt{2} & 1 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 & 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 & 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 1 & \sqrt{2} & \sqrt{2} & 1 & 1 \\ 1 & 1 & \sqrt{2} & \sqrt{2} & 1 & 1 \end{pmatrix} \quad (17)$$

Видно, что матрица  $\Lambda$  уже не симметрическая, верхний надтреугольник состоит из прямых отношений, нижний – из обратных (такую матрицу можно назвать, например, обратно-симметрической,



$\Lambda \circ \Lambda^T = \Lambda^T \circ \Lambda = J$ ,  $J$  -матрица единиц/единичный элемент в произведении Адамара). Также видно, что наша матрица  $M$  (делитель в элементах выделен синим) является лишь подмножеством более общей матрицы  $\Lambda$ . Основным свойством матрицы является то, что она состоит из соотношений всех размеров в фигуре (отсылая к пифагорейской математике, уходящей вглубь вавилонской алгебры [9]) и единичные элементы будут отвечать равным отрезкам, а если мы будем использовать удвоенную мировую функцию, т.е.  $2\sigma = \rho^2$ , то 2 (двойка) в определенной позиции означает гипотенузу равнобедренного прямоугольного треугольника (диагональ квадрата). Матрицы  $M, \Lambda$  пригодны для использования в системах компьютерной алгебры – CAS. С последней ассоциирована своя матрица смежности  $A_\Lambda$ , но её появление для текущих задач несколько преждевременно.

### 2.1. Построение матрицы $\Lambda$

$$L_h = L_n \text{vec}(L^T) = \text{vech}(L) = \left( P \bar{a} b P_2 \quad P \bar{a} c P_2 \quad P \bar{a} d P_2 \quad P \bar{b} c P_2 \quad P \bar{b} d P_2 \quad P \bar{c} d P_2 \right)^T \quad (18)$$

где  $L_n = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$  – матрица исключения

(elimination matrix) [17]. Тогда матрица  $\Lambda$  строится уже через матрицу  $L_h$  и вектор единиц  $j$  соответствующего размера

$$\Lambda = \text{diag}(L_h)^{-1} \cdot (j \otimes L_h^T). \quad (19)$$

### 3. Совместное решение и динамическая нормировка

При решении задачи выделения формы посредством SI алгоритма с одновременным поиском нужной формы, нам наперед не известен нормирующий множитель, чтобы из размерной матрицы  $L$  получить безразмерную  $M$ , поэтому можно проводить нормировку динамически, используя для этого первую длину ребра от слоя 0 к 1 (множитель  $\alpha$ ), особенно это актуально при решении задачи MCS (пересечения графов). Основу для этого как раз обеспечивает матрица  $\Lambda$  паттерн графа.

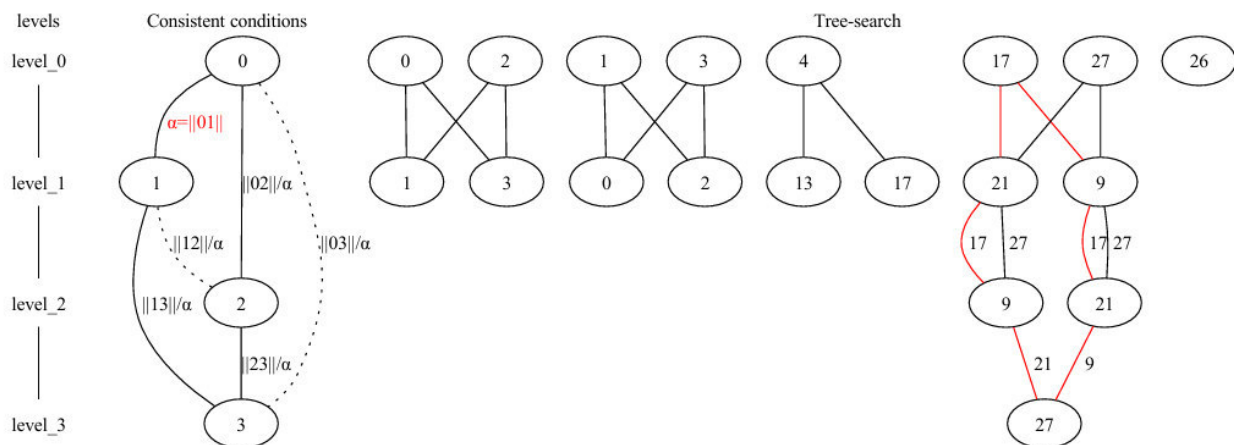


Рисунок 7. Условия разрешимости для обобщенно-взвешенного SI алгоритма и его дерево поиска

```

Data:  $H_p(A_p, G_p), H_t(A_t, G_t)$ 
Result:  $B_{(V_p \rightarrow V_t)}^{A,M}$ ; /* Bijection table or reduction tree */
1  $C \leftarrow \text{ConsistentConditions}(H_p(A_p, G_p));$ 
2  $\alpha \leftarrow C.\text{getLeaderEdgeAttribute}();$ 
3  $\rho_p \leftarrow \rho\_matrix(G_p);$ 
4  $M_p \leftarrow M\_matrix(\rho_p, \alpha);$ 
5  $H_p(A_p, G_p) \leftarrow M_p;$ 
6  $C \leftarrow C.\text{update}(H_p(A_p, G_p, M_p));$ 
7  $\rho_t \leftarrow \rho\_matrix(G_t);$ 
8  $H_t(A_t, G_t) \leftarrow \rho_t;$ 
9  $B_{(V_p \rightarrow V_t)}^{A,M} \leftarrow SI(C, H_t(A_t, \rho_t));$  /* Subgraph Isomorphism (e.attributed) task */

```

Алгоритм 2. Алгоритм совместного решения

Таблица 8. Сравнения алгоритмов решения

	Алгоритм 1	Алгоритм 2
Память	Не нужна большая матрица $L/(M)$	Строит большую матрицу $L$
Дерево поиска	Велико	Малое
Быстродействие	Умеренное	Велико
Обход дерева	Bfs/Dfs	Dfs like
Модель параллелизма	SI:parallel for/GI:Tasks по формам	Tasks по деревьям леса
Вложенный параллелизм	-/+	+
Недостатки	Множественные решения	Сложности с SIMD-конвейером
Достоинства	Декомпозирует модель на формы	Быстро находит решение
Использование	Анализ и исследование	Промышленное

## А. Сравнение графовых подходов и методов машинного обучения

Таблица 9. Сравнение методов

	Машинное обучение	Графовые предикаты
Типы мышления	Рефлексивное мышление	Абстрактно-логическое
Скорость работы	Высокая	В общем случае низкая
Точность результатов	Вероятностная оценка	Точный метод
Детерминизм	Нет	Строгий
Предпочтительный способ реализации	Аналоговый (мозг)	Цифровой (ЭВМ)
Предсказательная способность	Низкая	Высокая
Обучение	Необходимо	Нет
Способ развития	Дообучение	Добавления/вывод предиката
Передача знаний	Обучение с учителем	Копирование свода предикатов
Абстракция	Слабая	Абстрактно по сути
Энергозатраты	Низкие	Высокие
Параллелизм	Высокий	В общем случае низкий
Место в когнитивном каскаде	Первичная обработка	Уточнение выхода нейронной сети

## В. Julia код для построения основных матриц статьи

```

# form.jl Functions for form matrix calculation
# include("c:\\LOCAL\\julia\\form.jl")
using LinearAlgebra

function  $\sigma\_matrix$ (G)
    dim = size(G)
     $\sigma$  = zeros(dim[2], dim[2])

```

---

```

Vi = zeros(dim[2],dim[2])
for i in 1:dim[1]
    for j in 1:dim[2] # Vi = kron(j,G[i,:]) analog
        Vi[j,:] = G[i,:]
    end
    Vi -= Vi'
    Vi .*= Vi
     $\sigma$  += Vi
end
return  $\sigma$ 
end

function  $\rho$ _matrix(G)
     $\rho$  = sqrt. ( $\sigma$ _matrix(G))
    return  $\rho$ 
end
function  $M$ _matrix(G)
     $\rho$  =  $\rho$ _matrix(G)
    # k = norm(vec( $\rho$ ), Inf)
    k = maximum(vec( $\rho$ ))
    M = (1/k) .*  $\rho$ 
    return M
end

function  $M$ _matrix( $\rho$ ,  $\alpha$ )
    M = (1/ $\alpha$ ) .*  $\rho$ 
    return M
end

function  $\Lambda$ _matrix(G)
     $\rho$  =  $\rho$ _matrix(G)
    dim = size(G)
    n = dim[2]
    k = convert(Integer, (n*(n-1))/2)
     $\Lambda$  = zeros(k,k)
    Lh = zeros(k,1)
    iter = 1
    for i in 1:n
        for j in i+1:n
            Lh[iter,] =  $\rho$ [i,j]
            iter += 1
        end
    end
    D = (inv(Diagonal(vec(Lh))))
     $\Lambda$  = D*kron(ones(k,1), Lh')
    return  $\Lambda$ 
end
function  $M$ _matrix( $\Lambda$ , line)
    n = convert(Integer, (1+sqrt(1+8*size( $\Lambda$ )[1]))/2)
    M = zeros(n,n)
    iter = 1
    for i in 1:n
        for j in i+1:n
            if i != j
                M[i,j] =  $\Lambda$ [line,iter]
                M[j,i] = M[i,j]
                iter += 1
            end
        end
    end
end

```

```

    end
    return M
end

function permutation(new_order)
    n = size(new_order)[1]
    P = zeros(n,n)
    for i in 1:n
        P[i,new_order[i]] = 1;
    end

    return P
end

# calling example:
julia> include("c:\\LOCAL\\julia\\form.jl")
M_matrix (generic function with 1 method)
julia> G=[0 -1 1 0; -2 1 1 2];
julia> M = M_matrix(G);
julia> display(M)
0.0      0.790569  0.790569  1.0
0.790569  0.0      0.5      0.353553
0.790569  0.5      0.0      0.353553
1.0      0.353553  0.353553  0.0

```

Листинг 4: Листинги на языке Julia

### С. Онлайн-ресурсы

- Bit Twiddling Hacks
- Permutations as a Product of Transpositions
- TUDataset: A collection of benchmark datasets for learning with graphs
- The Network Data Repository with Interactive Graph Analytics and Visualization

### Благодарности

Автор благодарит компанию *Navio* за финансирование участия в конференции и лично *Наталию Анатольевну Францеву* за вклад в подготовку доклада.

### Список литературы

1. Menger K. Untersuchen uber allgemeine Metrik // Mathematische Annalen. Vol. 100. P. 75-113, 1928.
2. Blumenthal L.M. Theory and applications of Distance Geometry. Oxford: Clarendon Press, 1953. 368 p.
3. Liberti L., Lavor C. Six mathematical gems from the history of Distance Geometry // Special Issue: Many Faces of Distances. 2016. Vol. 23. iss. 5. arXiv:1502.02816
4. Рылов Ю.А. Дискретная геометрия пространства-времени и каркасная концепция динамики частиц. URL: [ggdgp3rw.pdf](#)
5. Rotman J.J. A First Course in Abstract Algebra with Applications. 3rd edition. Upper Saddle River: Prentice Hall, 2006. 581 p.
6. Rylov Yu. A. (2011) Discrete space-time geometry and skeleton conception of particle dynamics // International Journal of Theoretical Physics. 2012. Vol. 51, iss. 6. P. 1847-1865. DOI: 10.1007/s10773-011-1061-y. Available at: [arXiv:1110.3399v1](#)
7. Rylov Yu.A. (2001) Geometry without topology as a new conception of geometry // Int. Jour. Mat. & Mat. Sci. 2002. Vol. 30, iss. 12. P. 733-760. Available at: [arXiv:math/0103002](#).
8. Рылов Ю.А. Что такое геометрия? Математическая геометрия и физическая геометрия как два аспекта собственно евклидовой геометрии. URL: [tg1rw.pdf](#)
9. Ван Дер Варден Б.Л. Пробуждающаяся наука I. Математика древнего Египта, Вавилона и Греции. М.: ГИФМЛ, 1959. 460 с.
10. Synge J.L. Relativity: the General Theory. Amsterdam: North-Holland Publishing Company, 1960
11. Mac Duffee C.C. The Theory of Matrices. Springer, Berlin 1933.

12. Ланкастер П. Теория матриц. М.: Наука, 1973. 280 с.
13. Хорн Р., Джонсон Ч. Матричный анализ. М.: Мир, 1989. 656 с.
14. Clark B. N. Unit disk graphs, M. Math. Thesis, University of Waterloo, 1985.
15. Clark B. N., Colbourn C. J., Johnson D. S. Unit disk graphs //Discrete mathematics. 1990. Vol. 86, no. 1-3. P. 165-177.
16. Boole G. An Investigation of the Laws of Thought. New York: Dover Publications, 1854.
17. Magnus, J. R., Neudecker, H. The Elimination Matrix: Some Lemmas and Applications // SIAM Journal on Algebraic Discrete Methods. 1980. Vol. 1, no. 4. P. 422–449. doi:10.1137/0601049
18. Федотов И.А. Особенности алгоритмической и архитектурной оптимизации алгоритмов сопоставления графов на примере задачи изоморфного вхождения подграфа// Графы, игры и модели: материалы междунар. науч. конф., 14 окт. 2022 г. Майкоп: АГУ, 2022. С. 42-46. URL: <https://ggm.cmcagu.ru/>
19. Федотов И.А. Вероятностная оценка ширины дерева поиска в задаче изоморфного вхождения подграфа// Вероятностные методы в дискретной математике: материалы XI междунар. науч. конф., 29 мая 2024 г. Петрозаводск: ИПМИ КарНЦ РАН, 2024. С. 118-121. URL: <http://vmdm2024.krc.karelia.ru>

УДК 621.01

DOI: 10.25686/978-5-8158-2474-4-2025-899-905

## Определение рабочих зон механизмов рук антропоморфных роботов с учётом синтеза движений по вектору скоростей

Д. Б. Фёдоров

Омский государственный технический университет, г. Омск, Россия

**Аннотация.** Рабочая зона механизма руки антропоморфного робота представляет собой множество положений центров захвата, определяющих множество достижимых точек при заданных ограничениях на значения обобщённых координат. В статье проведены исследования, которые позволили определить совокупность кривых линий уровня, задающих рабочие зоны механизмов рук антропоморфных роботов. Данные исследования могут быть использованы при проектировании различных перемещений объектов манипулирования с помощью антропоморфных роботов.

**Ключевые слова:** механизмы манипуляторов, мгновенные состояния механизмов, вектор обобщённых скоростей, компьютерное моделирование движений антропоморфных роботов, выходное звено, рабочая зона манипулятора

## Determination of working zones of arm mechanisms of anthropomorphic robots taking into account the synthesis of movements along the velocity vector

D. B. Fedorov

Omsk State Technical University, Omsk, Russia

**Abstract.** The working area of the anthropomorphic robot arm mechanism is a set of positions of the capture centers that determine a set of attainable points under given constraints on the values of the generalized coordinates. The article contains studies that allowed us to determine a set of curved level lines that define the working areas of anthropomorphic robot arm mechanisms. These studies can be used in designing various movements of objects of manipulation using android robots.

**Keywords:** manipulator mechanisms, instantaneous states of mechanisms, vector of generalized velocities, computer modeling of movements of anthropomorphic robots, output link, working area of the manipulator.

### Введение

Антропоморфные роботы представляют собой человекоподобных роботов, способных взаимодействовать с окружающим миром подобно человеку, освобождая его от выполнения рутинных операций на производстве. Большинство роботов представляют собой механизмы манипуляторов, которые управляются посредством микропроцессорных контроллеров. Выходное звено (ВЗ) механизма манипулятора позволяет перемещать объекты в рабочем пространстве и обеспечивает выполнение технологических операций. Данные роботы играют ключевую роль в автоматизации технологических процессов и производств [1]. Такие роботы применяются, например, при автоматизации работ в складских помещениях или в медицине при разработке протезов [2].

Рабочая зона механизма манипулятора может быть определена как набор положений, которые могут быть достигнуты ВЗ механизма, взаимодействующего с окружающей средой. Важным критерием для проектирования и планирования перемещений робота является изучение его рабочей зоны [3]. Аналитическое задание рабочей зоны механизма руки антропоморфного робота на начальном этапе реализации его движения позволяет проводить анализ положения целевых точек синтезируемой траектории ВЗ и определять их достижимость [4].

В настоящее время актуальной задачей является разработка интеллектуальных систем управления механизмами рук антропоморфных роботов [5-7].

### Постановка задачи

При компьютерном моделировании движений роботов используется метод, основанный на матрицах частных передаточных отношений, позволяющих установить связь между скоростями выходного звена и обобщёнными скоростями. Обычно в рамках данного метода определяются позиции антропоморфных роботов с целью достижения механизмом руки заданных целевых точек. Для этого необходимо разработать алгоритм, позволяющий заранее предотвращать возникновение тупиковых ситуаций при компьютерном управлении движением антропоморфных роботов.

Определение рабочих зон механизмов манипуляторов способствует созданию эффективных алгоритмов управления движением рук антропоморфных роботов, что является актуальной проблемой для робототехнических систем и автоматизации технологических процессов. Так как управление перемещением роботов основано на построении движений по вектору скоростей, то рационально определять границы рабочих зон с учётом возникающих погрешностей линеаризации. Поэтому определим рабочие зоны механизмов рук антропоморфных роботов на основе синтеза малых движений по вектору скоростей.

### Теория

На рисунках 1 и 2 представлены кинематические схемы и туловища двух антропоморфных роботов. На рисунках точки  $O^1(O^1_1, O^1_2)$ ,  $O^2(O^2_1, O^2_2)$ , ... и т.п. задают узловые точки, число которых равно числу систем координат используемых для задания модели механизма. Указанные точки определяют начала систем координат, используемых при задании модели кинематической цепи. Положения точек  $O^1$ ,  $O^2$  и т.п. задаются значениями списков массивов, характеризующих структурную модель исполнительных механизмов рук антропоморфных роботов. Значения списков заданы в таблицах 1 и 2. Размерность указанных массивов является одинаковой для каждого механизма и определяется значением параметра  $n_m$ . Параметр  $n_m$  задаёт количество систем координат, используемых для задания структурной модели механизмов манипуляторов.

Указанные массивы дают возможность организовывать циклы при вычислениях матриц  $M_{0,k}$ , задающих положение систем координат, связанных со звеньями механизма. Параметр  $l_{sm}$  определяет смещение (линейное или угловое) по направлению осей систем координат, неподвижно связанных со звеньями механизма. Линейное смещение обозначается  $l_i$ , угловое –  $\varphi$ . Значения длин  $l_k$  задают размеры звеньев механизмов.

Линейные смещения  $l_i$  вдоль осей координат для антропоморфного робота AR-600E определяют отрезки  $l_1 = O_1O_2$ ,  $l_2 = O_2O_3$ ,  $l_3 = O_3O_6$ ,  $l_4 = O_6O_8$ ,  $l_5 = O_8O_{10}$  и  $l_6 = O_{10}O_{12}$  (рис. 1а). Параметр  $l_6$  определяет расстояние до характерной точки кисти – центра ладони (данная точка принимается за центр ВЗ). Соответственно для руки андроидного робота  $l_1 = O_3O_5$ ,  $l_2 = O_5O_6$ ,  $l_3 = O_7O_8$ ,  $l_4 = O_9O_{10}$ ,  $l_5 = O_{12}O_{13}$ . В таблицах 1, 2 параметр  $n_{kod}$  определяет значение кода преобразований систем координат при переходе от системы  $O_k$  к системе  $O_{k-1}$  [5]. Предельные значения обобщённых координат для двух механизмов заданы следующими значениями:  $q_i^{max} = 180^\circ$  и  $q_i^{min} = -180^\circ$ . Для андроидного робота максимальное значение первой обобщённой координаты  $q_1$  было принято равным 50 мм. Данная координата определяет поступательное перемещение вдоль оси  $x^0$ .

Определим графическим методом границы рабочих областей механизмов антропоморфных роботов. Для решения поставленной задачи целесообразно использовать метод компьютерного геометрического моделирования.

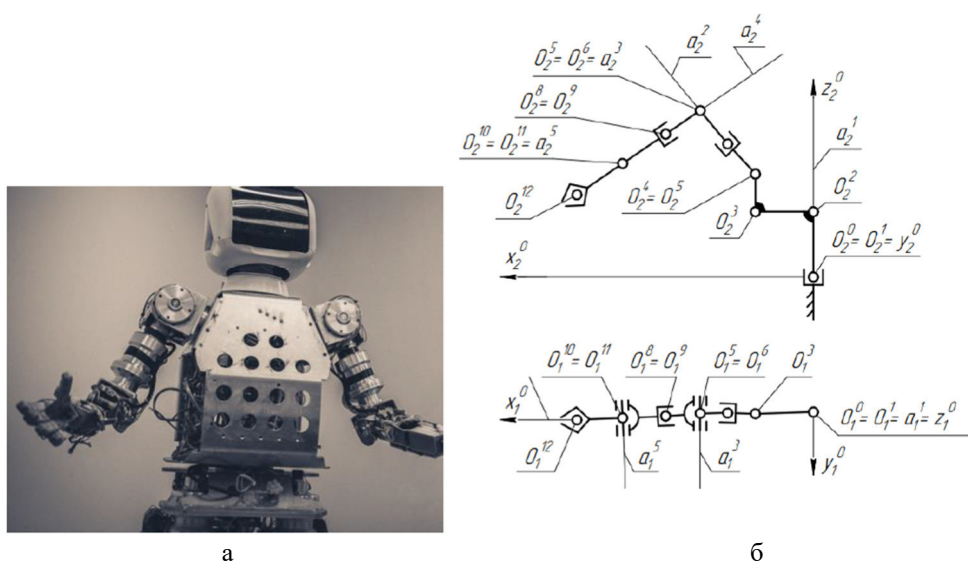
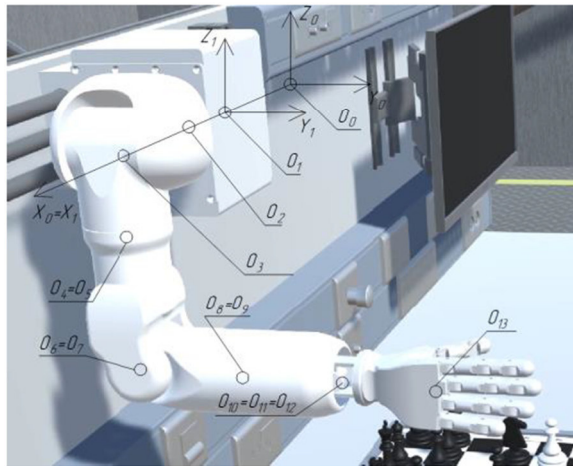
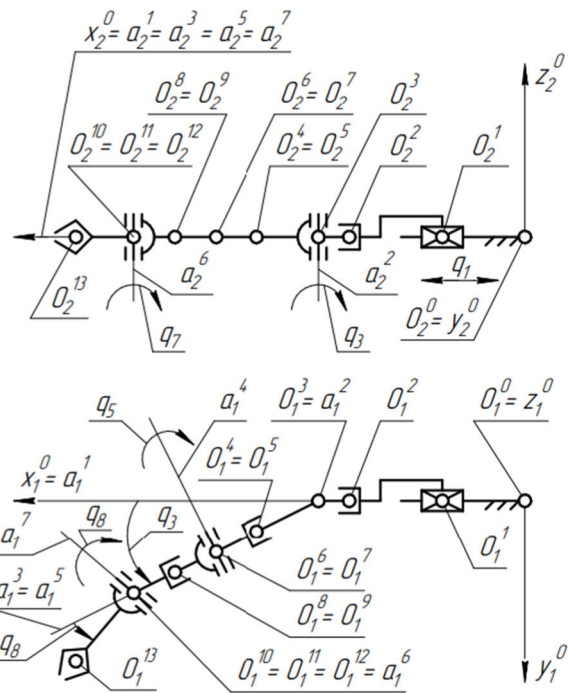


Рисунок 1. Антропоморфный робот AR-600E: а) общий вид туловища (рисунок заимствован с сайта НПО «Андроидная техника»); б) кинематическая схема механизма руки





a



6

Рисунок 2. Андроидный робот: *а)* общий вид руки; *б)* кинематическая схема механизма руки

**Таблица 1. Значения массивов и кодов, определяющих структурную модель механизма руки антропоморфного робота AR-600E,  $m_n = 12$**

Массивы	Номер преобразования систем координат											
	1	2	3	4	5	6	7	8	9	10	11	12
$q_i(град.)$	$q_1 = 25^{\circ}$	0	0	0	$q_2 = 20^{\circ}$	0	$q_3 = 65^{\circ}$	0	$q_4 = 25^{\circ}$	0	$q_5 = 65^{\circ}$	0
$l_{sm}(мм)$	0	$l_1 = 300$	$l_2 = -120$	$\varphi = 80^{\circ}$	0	$l_3 = 100$	0	$l_4 = 80$	0	$l_5 = 150$	0	$l_6 = 250$
$n_{kod}$	3	12	11	7	3	12	2	12	3	12	2	12

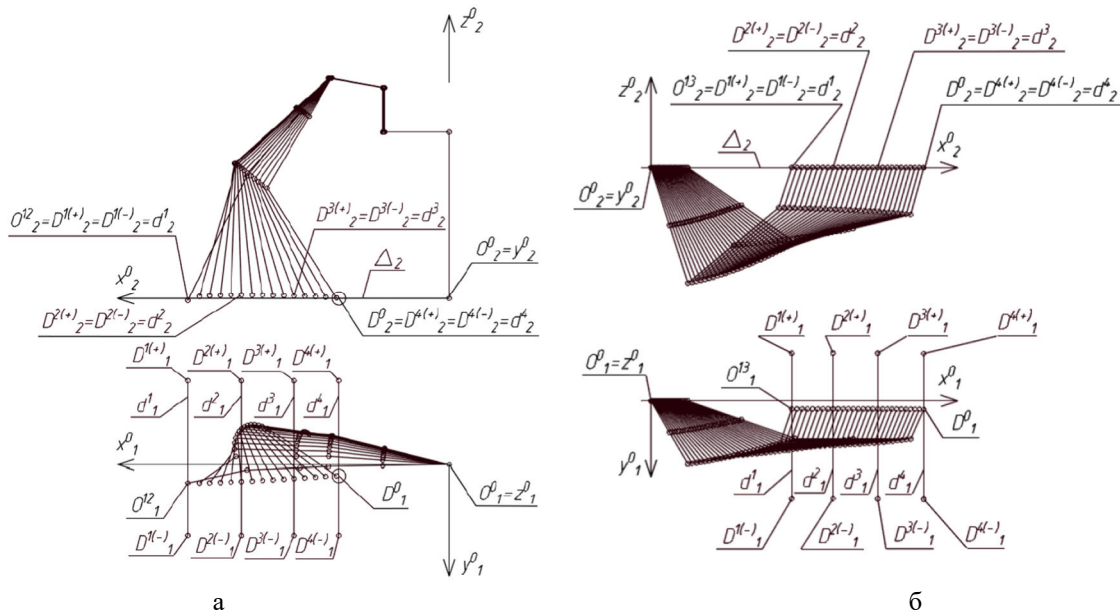
Таблица 2. Значения массивов и кодов, определяющих структурную модель механизма руки андроидного робота,  $m n = 13$

Массивы	Номер преобразования систем координат												
	1	2	3	4	5	6	7	8	9	10	11	12	13
$q_i, (град.мм)$	$q_1 = 0$	$q_2 = 180^\circ$	$q_3 = 0^\circ$	0	$q_4 = 0^\circ$	0	$q_5 = 90^\circ$	0	$q_6 = 0^\circ$	0	$q_7 = 0^\circ$	$q_8 = 0^\circ$	0
$smi (мм)$	0	0	0	$l_1 = 103$	0	$l_2 = -130$	0	$l_3 = 104$	0	$l_4 = 104$	0	0	$l_5 = 98$
$n_{kod}$	4	1	2	12	3	12	1	12	3	12	1	2	12

### Результаты исследований

На рисунке 3 представлены синтезы движений, при котором центры захвата  $O^{12}(O^{12}_1, O^{12}_2)$  и  $O^{13}(O^{13}_1, O^{13}_2)$  перемещаются по отрезку прямой, параллельной оси  $x_0$  до целевой точки  $D^0(D^0_1, D^0_2)$ . Начальные значения обобщённых координат для первого и второго механизмов были соответственно приняты равными  $q_1 = 1^\circ, q_2 = 1^\circ, q_3 = -60^\circ, q_4 = 15^\circ, q_5 = -15^\circ$  и  $q_1 = 16,69^\circ, q_2 = 151,25^\circ, q_3 = 15,77^\circ, q_4 = 101,31^\circ, q_5 = 112,89^\circ, q_6 = -160,32^\circ, q_7 = -33,14^\circ, q_8 = 31,8^\circ$ .

При указанном движении были сохранены промежуточные значения обобщённых координат, задающих промежуточные конфигурации  $L^{Di}(q_1 \div q_5)$  и  $L^{Di}(q_1 \div q_8)$ . На рисунке 2 показаны положения целевых точек  $D^{i(+)}(D^{i(+)}_1, D^{i(+)}_2), D^{i(-)}(D^{i(-)}_1, D^{i(-)}_2)$ , которых заведомо невозможно достичь, и прямых  $d^1, d^2, \dots$ , задающих траектории движения центра ВЗ.



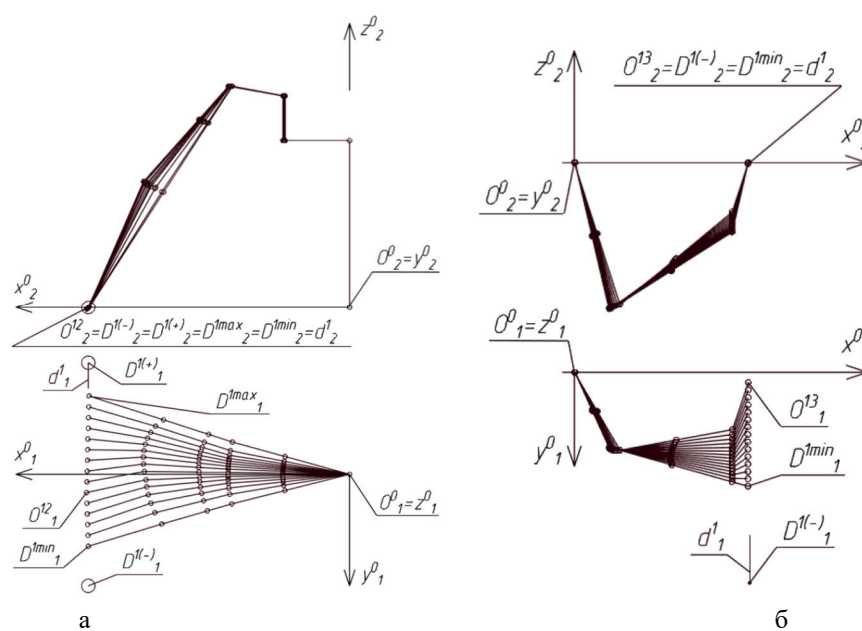


Рисунок 4. Построение движения механизмов рук роботов по прямой  $d^1$ , параллельной оси  $y^0$ :  
а) AR-600E; б) андроидного робота

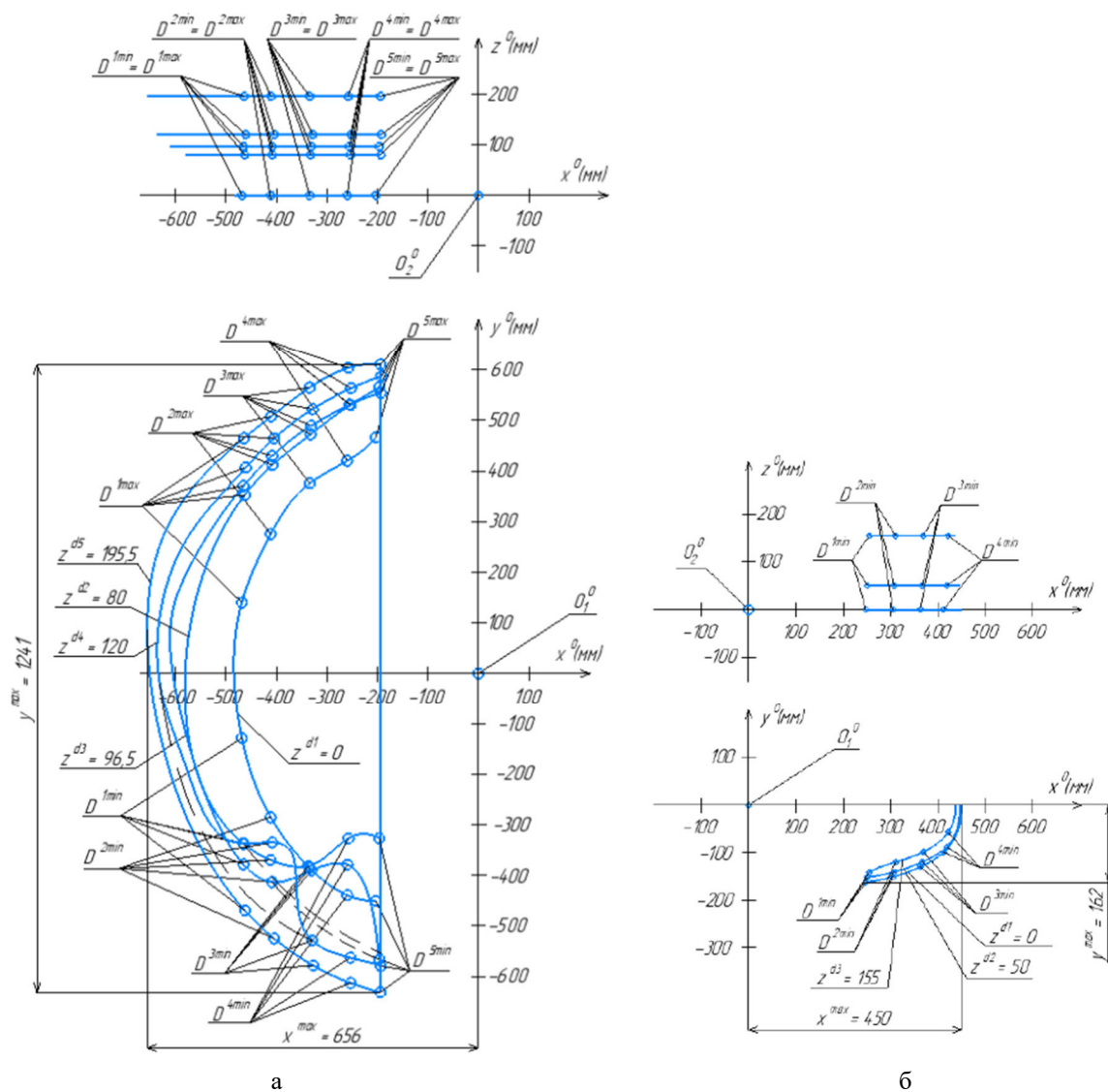


Рисунок 5. Линии уровня, задающие рабочие зоны механизмов рук антропоморфных роботов:  
а) AR-600E; б) андроидного робота

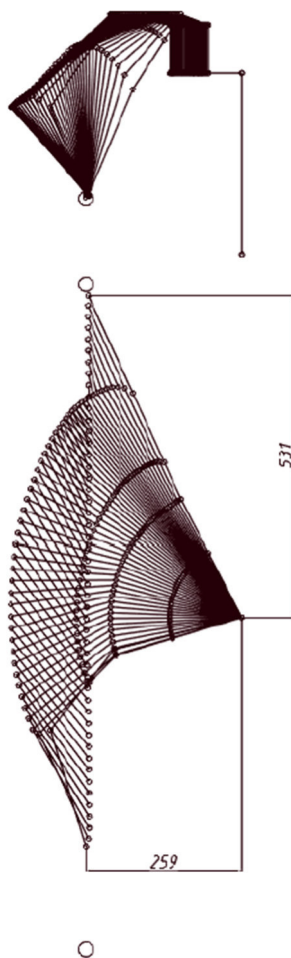


Рисунок 6. Возникновение тупиковой ситуации у антропоморфного робота AR-600E при синтезе движений по траектории, параллельной оси  $y^0$

### Выводы и заключение

Проведённые исследования позволили определить линии уровня, ограничивающие рабочие зоны механизмов рук антропоморфных роботов на основе синтеза движения по вектору скоростей.

Максимальные смещения по направлению осей  $y^0$  и  $x^0$  для робота AR-600E составляют  $y_{max} = 1241$  мм и  $x_{max} = 656$  мм, а для андроидного робота –  $y_{max} = 162$  мм и  $x_{max} = 450$  мм. Синтез движений робота AR-600E указал на наличие в рабочей зоне этого механизма области, где возникают тупиковые ситуации.

Границы этой области определены с помощью линий уровня на рисунке 5а. При синтезе движений антропоморфного робота, представленного на рисунке 2, тупиковых ситуаций не возникает.

Результаты проведённых исследований могут быть использованы при разработке информационно-управляющих систем подвижных объектов, в частности при управлении движением рук антропоморфных роботов.

### Список литературы

1. Азизов И.Р. Обзор существующих конструкций антропоморфных роботов-манипуляторов // Инновации природообустройства и защиты окружающей среды. Саратов, 2019. С. 341-344.
2. Шадеркина В.А., Шадеркина А.И., Петрушин А.А. Антропоморфные роботы в медицине: варианты технологий и перспективы // Российский журнал телемедицины и электронного здравоохранения. 2023. №9(1). С. 49-59.
3. Шахматов Е.В., Илюхин В.Н., Мезенцев Д.А. Исследование статических характеристик рабочего пространства робота AR600E // Динамика и виброакустика. 2019. Т.5, №4. С. 6- 12;
4. Притыкин Ф.Н., Небритов В.И. Построение рабочей зоны механизма руки андроидного робота с учётом положения запретных зон // Омский научный вестник. 2017. №1. С. 5-8.

- 
5. Притыкин Ф.Н. Исследование кинематических параметров андроидного робота при автоматизированном синтезе движений по вектору скоростей / Ф.Н. Притыкин // Омский научный вестник. 2022. №2(182). С. 5-9.
  6. Притыкин Ф.Н. Преодоление тупиковых ситуаций синтеза движений антропоморфных роботов на основе использования перемещений оси схватоносителя по линейчатой поверхности, ограничивающей телесный угол сервиса // Омский научный вестник. 2024. №3(191). С. 14-21.
  7. Воробьёв Е.И., Михеева А.В., Моргуненко К.О. Механизм ориентирующих движений антропоморфных роботов и протезов рук с линейными и вращательным приводами // Вестник машиностроения. 2019. №11. С. 54-56.

УДК 004.942.84+624.042.041

DOI: 10.25686/978-5-8158-2474-4-2025-906-911

## Генерация параметрических поверхностей для моделирования рельефа с псевдослучайными геометрическими характеристиками

А. В. Симонов<sup>1</sup>, М. В. Симонов<sup>2</sup>

<sup>1</sup>Нижегородский государственный архитектурно-строительный университет, Нижний Новгород, Россия

<sup>2</sup>Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

**Аннотация.** В статье представлен инновационный алгоритм компьютерной генерации рельефных поверхностей, разработанный для задач цифровизации аэродинамических исследований в строительной отрасли. Разработана двухуровневая методика, основанная на применении профессионального программного обеспечения и инструментов прикладной информатики и программирования. Исследование направлено на решение актуальной проблемы многократного параметрического моделирования сложных ландшафтов, необходимого для анализа аэродинамических параметров местности. Методика зарегистрирована как программа для ЭВМ и может применяться в различных задачах параметрического моделирования, в т.ч. отличных от обозначенной.

**Ключевые слова:** цифровая поддержка, компьютерное моделирование, численные расчеты, программная инженерия, программирование, архитектурно-строительная аэродинамика.

## Generation of parametric surfaces for terrain modeling with pseudo-random geometric characteristics

A. V. Simonov<sup>1</sup>, M. V. Simonov<sup>2</sup>

<sup>1</sup>Nizhny Novgorod State University of Architecture and Civil Engineering, Nizhny Novgorod, Russia

<sup>2</sup>National Research Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

**Abstract.** This paper presents an innovative algorithm for computer-generated terrain surfaces, developed for digitalizing aerodynamic research in the construction industry. The study introduces a two-level methodology based on professional software and applied computer science tools. It addresses the pressing challenge of parametric modeling for complex landscapes required to analyze terrain aerodynamic parameters. The developed method is registered as a computer program and can be applied to various parametric modeling tasks, including those beyond the specified scope.

**Keywords:** digital support, computer modeling, numerical calculations, software engineering, programming, architectural and construction aerodynamics

### Введение

Современный этап развития научных исследований характеризуется активным внедрением цифровых технологий во все сферы познавательной деятельности [1, 2]. Особенно ярко эта тенденция проявляется в технических науках, где компьютерное моделирование стало неотъемлемой частью исследовательского процесса [3-6]. В контексте строительной аэродинамики, находящейся на стыке фундаментальной гидрогазодинамики и прикладных строительных дисциплин, цифровые методы открывают новые возможности для изучения взаимодействия архитектурных объектов с ветровыми потоками [7-10].

Актуальность данного исследования обусловлена необходимостью разработки эффективных инструментов параметрического моделирования рельефных поверхностей, которые позволяют проводить масштабные численные эксперименты по оценке аэродинамических характеристик строительных объектов в различных ландшафтных условиях. Традиционные подходы к созданию таких моделей, основанные на ручном вводе данных или использовании простейших генераторов случайных чисел, не обеспечивают необходимой гибкости и достоверности результатов, в частности имеющих итерационный характер.

### Постановка задачи

В докладе представлен вариант решения задачи по разработке алгоритма для многократной параметрической генерации условно-случайных рельефов, пригодных для аэродинамического анализа методами компьютерного моделирования и прикладного программирования.



## Теория

В рамках проведенной работы был разработан комплексный подход, сочетающий:

- применение специализированного программного обеспечения для трехмерного моделирования;
- разработку собственного алгоритма генерации рельефных поверхностей по входным геометрическим параметрам посредством использования среды программирования;
- интеграцию с системами вычислительной гидрогазодинамики.

Основные этапы исследования включали:

- анализ существующих методов параметрического моделирования;
- разработку математического аппарата для описания условно-случайных поверхностей;
- создание программной реализации алгоритма на языке C#;
- тестирование и верификацию полученных результатов.

## Реализация алгоритма

Ядром разработанной методики стал двухуровневый алгоритм генерации рельефных поверхностей.

Первый уровень представляет собой оригинальную разработку – программный модуль на языке C#, выполняющий следующие функции:

- генерация координат точек рельефа по заданным параметрам: количество точек поверхностей по осям  $OX$  и  $OY$  ( $x$ ,  $y$  соответственно), шаг сетки координат ( $step$ , м), максимальная высота точки поверхности ( $max$ , м), максимальный высотный перепад между двумя соседними узлами сетки ( $max\_diff$ , м);
- контроль геометрических характеристик создаваемой поверхности;
- экспорт данных в форматы, совместимых с используемым ПО: для данного примера используется вариант вывода координат в формате электронной таблицы *.CSV*.

Отмечается, что геометрические параметры модели, являющиеся входными данными для генерации рельефной поверхности, выбираются исходя из цели исследований в области архитектурно-строительной аэродинамики, т.е. являются аргументами функции, отображающей зависимость аэродинамических параметров среды от параметров геометрических.

Алгоритм реализован в двух вариантах в зависимости от объема исходных параметров (рис. 1): базовая версия и расширенная с дополнительными возможностями управления характеристиками поверхности.



Рисунок 1. Интерфейс разрабатываемого программного обеспечения в «базовом» (а) и «продвинутом» (б) режимах работы

Основной принцип работы программы основан на построчном заполнении массивов координат и их записи в формате электронной таблицы для последующего импорта в специализированное ПО для компьютерного моделирования.



На втором уровне осуществляется создание цифровой модели с использованием профессиональных САПР-решений и последующая ее интеграция в вычислительный комплекс *Ansys*, модуль *Fluid Flow (CFX)* для проведения аэродинамических численных экспериментов.

*Autodesk Revit* используется для построения рельефа на основе координат точек генерируемой топоповерхности с использованием результирующего файла, полученного программным путем (рис. 2).

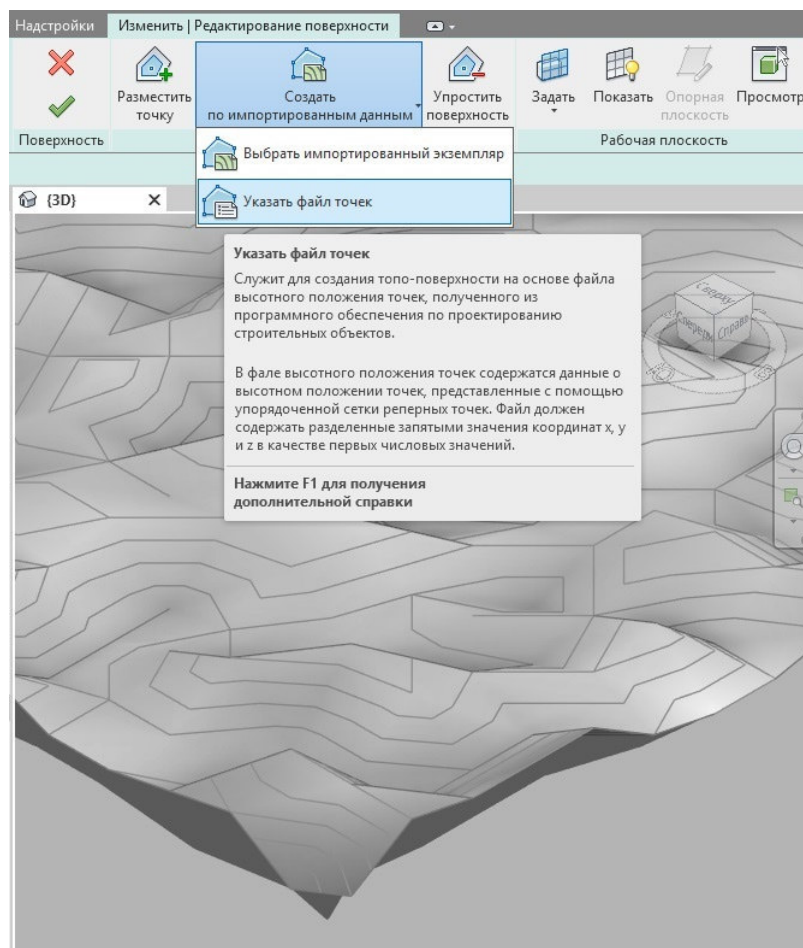


Рисунок 2. Трехмерное моделирование рельефной поверхности в программной среде

На следующем этапе производятся геометрические преобразования модели, включающие в себя экспорт поверхности в графический комплекс трехмерного моделирования для ее переформатирования в 3D-тело по последовательности, описанной на рисунке 3, а также конвертация полученной модели воздушного пространства в формат *ACIS* (.SAT), являющийся универсальным для прочтения программно-вычислительными комплексами для расчета аэродинамических параметров. При необходимости модель может быть дополнена геометрическими объектами разного рода, расположенными на поверхности топосолида.

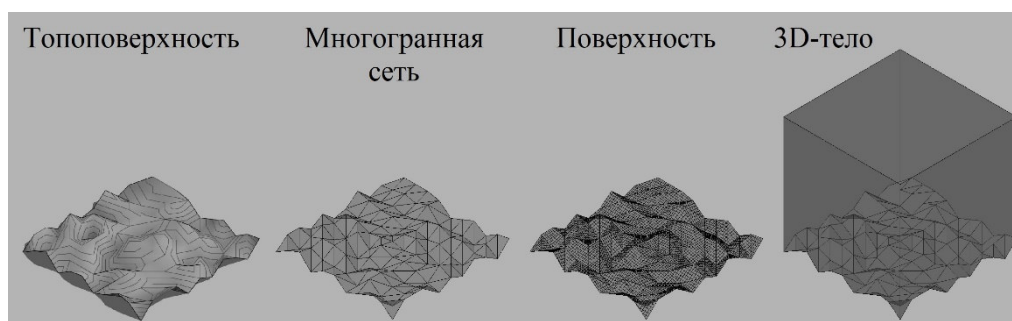


Рисунок 3. Геометрические преобразования модели в среде графического комплекса

Процессом, завершающим описываемый алгоритм, является сам аэродинамический эксперимент, включающий в себя преобразование модели в конечно-элементную посредством применения препроцессоров *DesignModeler* и *Mesh* (рис. 4), наложение на модель граничных условий в препроцессоре *Setup*, решение и визуализация расчета в постпроцессоре *Results* (рис. 5).

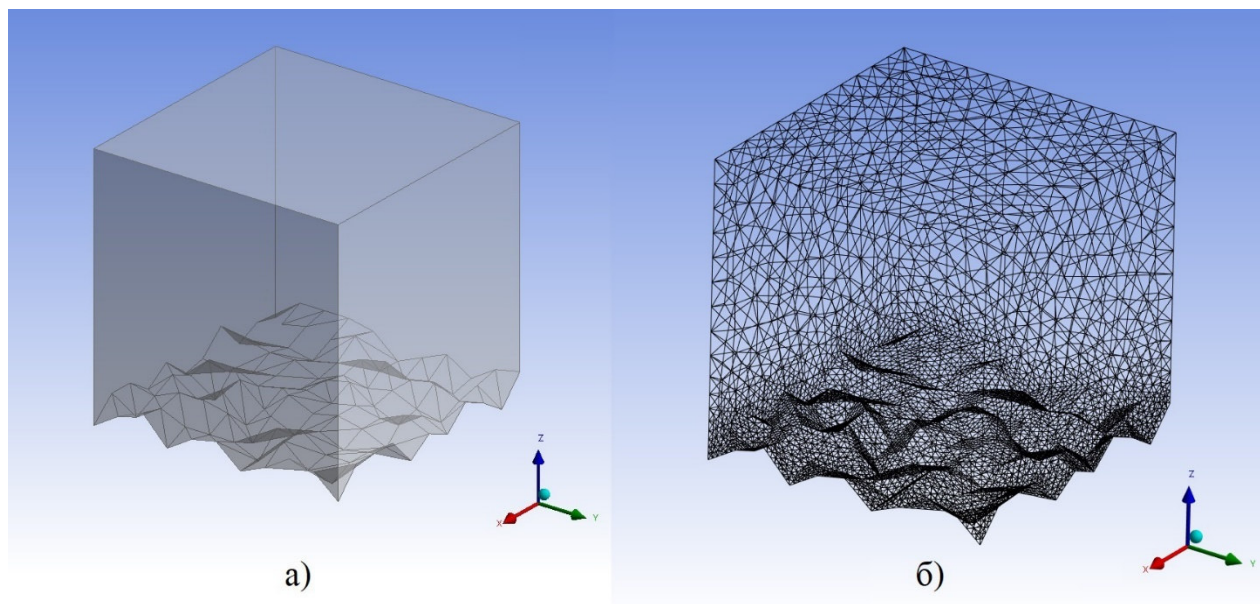


Рисунок 4. Геометрическая модель в среде препроцессора *DesignModeler* (а) и конечно-элементная расчетная модель, сформированная препроцессором *Mesh* (б)

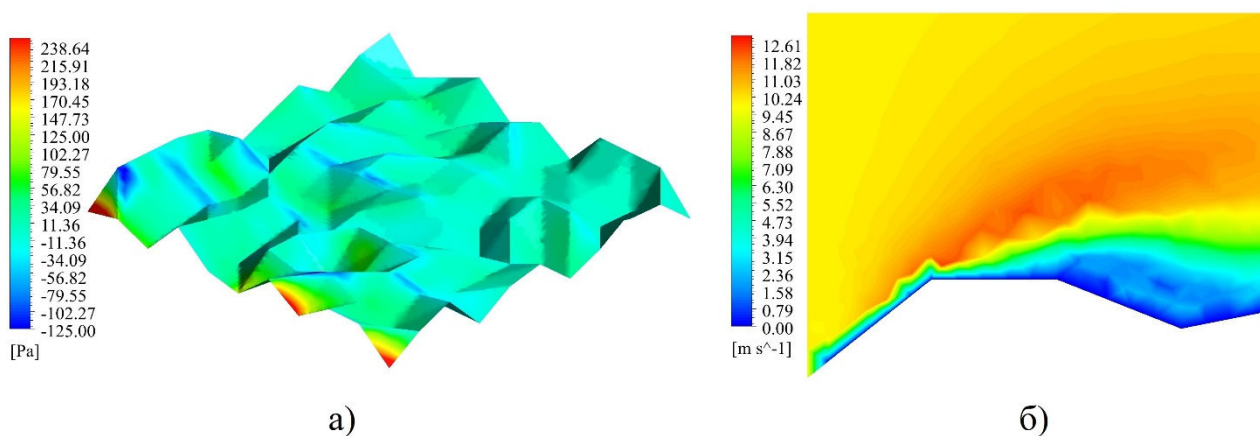


Рисунок 5. Изополя распределения давления ветровых потоков по рельефной поверхности (а) и профиль срыва ветрового потока вдоль ее излома (б) в среде постпроцессора *Results*.

### Обсуждение результатов

Общий процесс генерации условно-случайной рельефной поверхности, поддерживаемый методами компьютерного моделирования и геометрических преобразований, описан в виде графологической блок-схемы на рисунке 6.

Разработанная методика была успешно апробирована в серии вычислительных экспериментов по моделированию аэродинамических процессов. Основные преимущества предложенного подхода:

- высокая степень автоматизации процесса создания моделей;
- возможность генерации большого количества вариантов рельефа для цифровой поддержки исследований, имеющих итерационный характер;
- гибкость в настройке параметров поверхности, недостижимая при ручном моделировании;
- совместимость с распространенными программными платформами, в т.ч. отечественными.

Перспективные направления дальнейшего развития включают:

- интеграцию алгоритма в *BIM*-среду через библиотеки *API*;
- разработку адаптивных методов генерации рельефа;
- создание библиотеки типовых рельефных поверхностей в зависимости от аэродинамических характеристик формируемой ими воздушной среды.

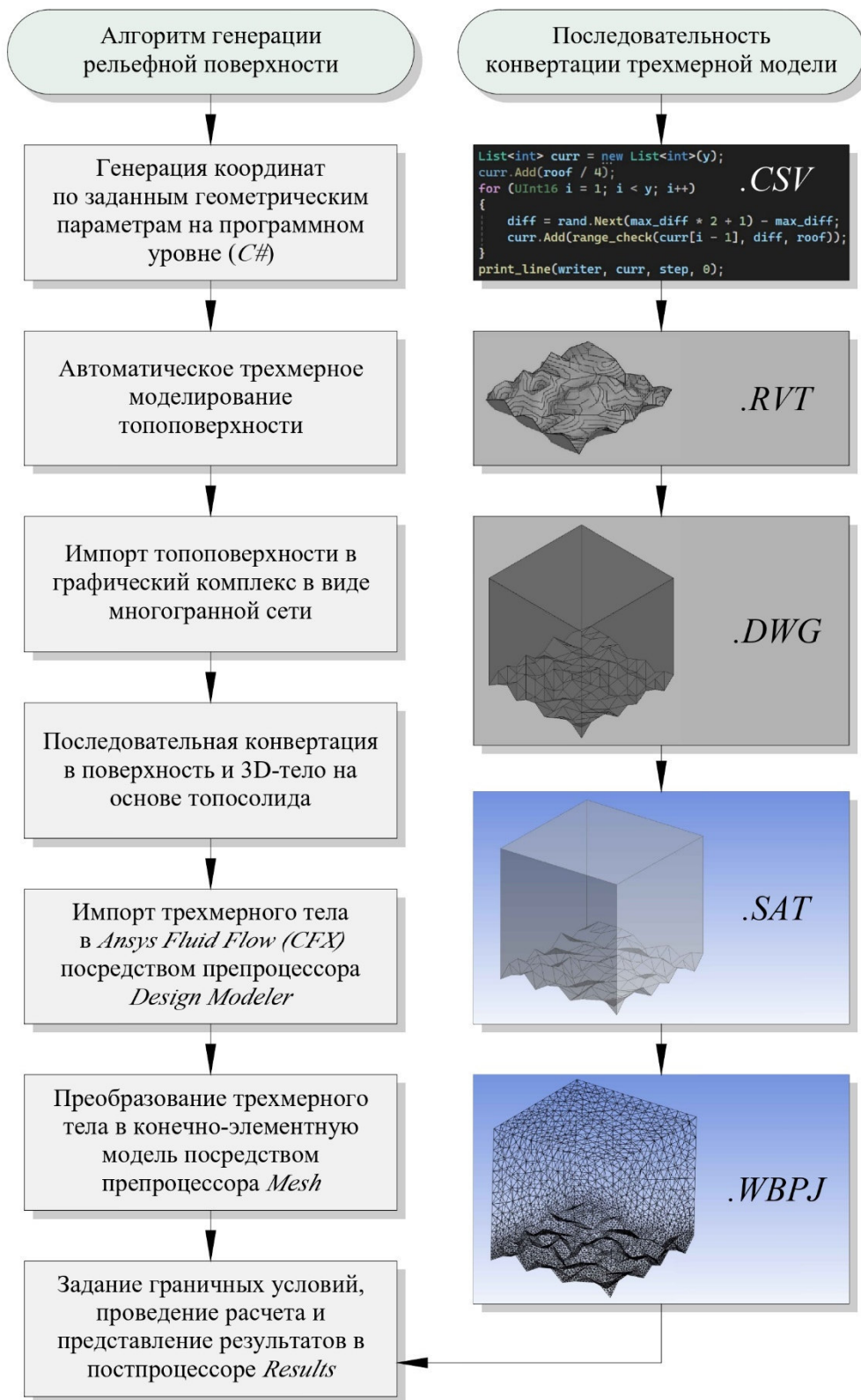


Рисунок 6. Алгоритм генерации рельефной поверхности методами компьютерного моделирования и программирования

### Заключение

Представленный алгоритм компьютерной генерации рельефных поверхностей демонстрирует высокую эффективность при решении задач строительной аэродинамики. Его применение позволяет существенно сократить временные затраты на подготовку расчетных моделей и повысить достоверность результатов за счет возможности проведения многовариантных исследований.

Разработанная методика имеет значительный потенциал для использования не только в аэродинамических расчетах, но и в других областях, требующих параметрического моделирования сложных поверхностей.

В процессе разработки алгоритма цифровой поддержки аэродинамических исследований авторами было получено свидетельство о государственной регистрации программы для ЭВМ «Генератор координат параметрических поверхностей для компьютерного моделирования рельефа» (*Номер свидетельства: RU 2025662278. Дата регистрации: 24.04.2025*).

**Научный руководитель** – Хазов Павел Алексеевич, Федеральное государственное бюджетное образовательное учреждение высшего образования «Нижегородский государственный архитектурно-строительный университет»

### Список литературы

1. Floridi L. The 4th Revolution: How the Infosphere is Reshaping Human Reality. Oxford: Oxford University Press, 2014. 248 p.
2. Manovich L. Cultural Analytics. Cambridge, MA: MIT Press, 2020. 320 p.
3. Белоцерковский О.М. Численное моделирование в механике сплошных сред. М.: Наука, 1994. 440 с. ISBN 5-02-014894-8.
4. Smith J., Doe A. Development of a C#-Based Framework for Finite Element Analysis in Mechanical Engineering // Journal of Computational Engineering. 2020. Vol. 15, no. 3. P. 45–60.
5. C#-Based Software for Automated Structural Design Optimization / R. Garcia, et al. // Engineering Structures. 2022. Vol. 45. P. 203–215.
6. Тюленев А.И., Голубкин В.А. Цифровые двойники в аэродинамических исследованиях // Ученые записки ЦАГИ. 2020. Т. 51, № 3. С. 45–58.
7. Вальгер С.А., Федоров А.В., Федорова Н.Н. Моделирование несжимаемых турбулентных течений в окрестности плохообтекаемых тел с использованием ПК ANSYS Fluent // Вычислительные технологии. 2013. Т. 18, № 5. С. 27–40.
8. Хазов П.А. Компьютерное моделирование аэродинамического обтекания комплекса уникальных зданий // Вестник компьютерных и информационных технологий. 2024. Т. 21, № 6(240). С. 23–30.
9. Хазов П.А., Ведяйкина О.И. Компьютерное моделирование аэродинамического обтекания и оценка пешеходной аэродинамической комфортности комплекса зданий // Омский научный вестник. 2024. № 3(191). С. 56–63.
10. Хазов П.А., Симонов А.В., Чибаква Е.А. Влияние уровня геометрической детализации компьютерной модели на результаты анализа аэродинамических процессов // Вестник компьютерных и информационных технологий. 2024. Т. 21, № 11(245). С. 9–18.



УДК 514.122.3

DOI: 10.25686/978-5-8158-2474-4-2025-912-919

## Гладкие преобразования *B*-сплайнов

С. Н. Ганчук

Российский федеральный ядерный центр – Всероссийский научно-исследовательский институт экспериментальной физики (РФЯЦ-ВНИИЭФ) – федеральное государственное унитарное предприятие Государственной корпорации по атомной энергии «Росатом», г. Саров, Россия

**Аннотация.** *B*-сплайны являются одним из основных геометрических объектов современных CAD-систем. Различные преобразования *B*-сплайнов широко применяются при построении практически каждого геометрического объекта. В работе рассматриваются вопросы гладкого сопряжения *B*-сплайнов на примере двух задач. Первая задача – гладкое сопряжение двух сплайнов произвольной степени, при этом под «гладким сопряжением» понимается выполнение в точке сопряжения условий непрерывности производных всех порядков, вплоть до порядка  $(p - 1)$ , где « $p$ » – степень исходных сплайнов. После гладкого сопряжения два сплайна без погрешности заменяются одним сплайном, определенным на меньшем числе контрольных точек. Вторая задача – понижение степени *B*-сплайна, одним из шагов которой является гладкое сопряжение. Решение задач гладкого сопряжения сводится к оптимизационной задаче при наличии ограничений, которая решается методом неопределенных множителей Лагранжа.

**Ключевые слова:** *B*-сплайн, система автоматизированного проектирования, гладкое сопряжение.

## Smooth B-spline transformations

S. N. Ganchuk

Russian Federal Nuclear Center - All-Russian Research Institute of Experimental Physics (RFNC- VNIIEPh) – Federal State Unitary Enterprise of the State Atomic Energy Corporation “Rosatom”, Sarov, Russian Federation

**Abstract.** B-splines are one of the main geometric objects of modern CAD systems. Various transformations of B-splines are widely used in the construction of almost every geometric object. The paper considers the issues of smooth conjugation of B-splines using two problems as an example. The first problem is a smooth conjugation of two splines of an arbitrary degree, where "smooth conjugation" is understood as fulfillment of the conditions of continuity of derivatives of all orders at the conjugation point, up to the order  $(p - 1)$ , where "p" is the degree of the original splines. After the smooth conjugation, the two splines are replaced without an error by one spline defined with the smaller number of control points. The second problem is the reducing of the degree of the B-spline, one of the steps of which is smooth conjugation. The solution of smooth conjugation problems is reduced to an optimization problem in the presence of constraints, which is solved by the method of undetermined Lagrange multipliers.

**Keywords:** B-spline, computer-aided design system, smooth conjugation.

### Введение

Сложность современных изделий машиностроения такова, что их проектирование невозможно без использования CAD-систем, в которых одной из основных функциональных операций является сопряжение *B*-сплайнов с заданным порядком непрерывности при наложении дополнительных ограничений на аппроксимирующую линию и понижение степени *B*-сплайнов.

Сопряжения *B*-сплайнов с заданным порядком гладкости с математической точки зрения является довольно сложной задачей [1], и, к примеру, в ставших к настоящему времени классическими монографиях [2, 3] вопросы сопряжения не рассматриваются. Разнообразным вопросам сопряжения линий и поверхностей с невысоким порядком гладкости (вплоть до второго) в отечественной и иностранной научной литературе уделяется достаточное внимание. При этом вопросы сопряжения *B*-сплайнов рассматриваются в небольшом количестве работ и в них изучаются сплайны 3-й степени и без дополнительных ограничений, что делает задачу гладкого сопряжения *B*-сплайнов с дополнительными ограничениями актуальной. Кроме этого, замена без погрешности нескольких сопряженных между собой сплайнов одним сплайном весьма востребована, так как позволяет представить несколько геометрических объектов одним.

Задача понижения степени *B*-сплайна является как самостоятельной, так и развитием задачи аппроксимации NURBS-кривой *B*-сплайном. На практике при работе геометрического ядра современных CAD-систем при некоторых преобразованиях появляются сплайны очень высоких степеней (например, 12-й степени). Естественно, возникает задача понизить степень этих сплайнов

и проконтролировать точность отклонения от исходной кривой. Таким образом, задача понижения степени сплайнов с контролем точности, также является актуальной.

### Постановка задачи

В настоящей работе рассматривается постановка задачи для сопряжения  $B$ -сплайнов, у которых в реальной части узлового вектора отсутствуют кратные узлы.

Заданы два  $B$ -сплайна степени  $p$   $P(u)$  и  $Q(v)$  (если степени сплайнов не совпадают, то у сплайна меньшей степени увеличивается степень до  $p$ , используя процедуру, описанную в [2], которая не приводит к искажению сплайна):

$$P(u) = \sum_{i=0}^n P_i N_{i,p}(u), \quad u_p \leq u \leq u_{n+1}, \quad (1)$$

$$Q(v) = \sum_{j=0}^m Q_j N_{j,p}(v), \quad v_p \leq v \leq v_{m+1}, \quad (2)$$

где  $P_i, Q_j, i = \overline{0, n}, j = \overline{0, m}$  – заданные контрольные точки для сплайнов  $P(u)$  и  $Q(v)$  соответственно;  $u, v$  – параметрические переменные, определенные на узловых векторах  $U$  и  $V$  соответственно.

$$U = \{\underbrace{0, \dots, 0}_{p+1}, u_{p+1}, \dots, u_n, \underbrace{1, \dots, 1}_{p+1}\}, \quad (3)$$

$$V = \{\underbrace{0, \dots, 0}_{p+1}, v_{p+1}, \dots, v_m, \underbrace{1, \dots, 1}_{p+1}\}. \quad (4)$$

Потребуем, чтобы для  $B$ -сплайнов (1) и (2) выполнялось условие  $n, m > p$  (при  $n = m = p$  получаем кривые Безье, для которых полностью гладкое сопряжение рассмотрено ниже). Так как единая терминология для узловых векторов в современной научной литературе отсутствует (см. примеры различных терминологий в монографиях [2–4]), станем использовать терминологию классической монографии [2]. Тогда узловой вектор вида (3) и (4) будем называть *Clamping* (зажатый) узловой вектор.

Сформируем *Clamping* (зажатый) узловой вектор  $T$ :

$$T = \{u_0, u_1, \dots, u_n, u_{n+1} = v'_p, v'_{p+1}, \dots, v'_{m+p+1}\}, \quad (5)$$

где  $u_i, i = \overline{0, (n+1)}$  – соответствующие элементы узлового вектора (3) сплайна (1);  $v'_i = f(v_i), i = \overline{p, (m+p+1)}$ ;  $f$  – линейная функция;  $v_i, i = \overline{p, (m+p+1)}$  – соответствующие элементы узлового вектора (4) сплайна (2).

Функция  $f$  имеет вид

$$v'_i = v_{i+p-n-1} + u_{n+1}, \quad i = \overline{p, (m+p+1)}. \quad (6)$$

Вектор (5) представляет собой объединение векторов (3), (4), при этом от вектора (3) отбрасываем последние  $p$  кратных узлов, от вектора (4) отбрасываем первые  $(p+1)$  кратные узлы и к оставшимся узлам вектора (4) применяем преобразование (6).

Для программной реализации узловой вектор (5) удобнее представить в следующем виде:

$$T = \{t_0, t_1, \dots, t_{n+m+2}\}, \quad (7)$$

$$\text{где } t_i = \begin{cases} u_i, & i = \overline{0, (n+1)}, \\ v_{i+p-n-1} + u_{n+1}, & i = \overline{(n+1), (n+m+2)}. \end{cases}$$

Определим сплайны (1), (2) как один составной сплайн  $\bar{R}(t)$ , определённый на узловом векторе (7):

$$\bar{R}(t) = \begin{cases} P(w), & t_p \leq w \leq t_{n+1}, \\ Q(f^{-1}(w)), & t_{n+1} \leq w \leq t_{n+m+2-p}. \end{cases} \quad (8)$$

Сплайн (8) в точках  $P(t_{n+1})$  и  $Q(f^{-1}(t_{n+1}))$  в общем случае не обладает непрерывностью.

Необходимо выполнить полностью гладкое сопряжение сплайнов (1) и (2), то есть определить контрольные точки  $R_i, i = \overline{0, (n+m-p+1)}$  для аппроксимирующего сплайна  $R(t)$  степени  $p$ :

$$R(t) = \sum_{i=0}^{n+m-p+1} R_i N_{i,p}(t), \quad t_p \leq t \leq t_{n+m-p+2}, \quad (9)$$

где  $t$  – параметрическая переменная, определенная на узловом векторе (7).

Сплайн (9) должен минимизировать метрику  $d(R, \bar{R})$  между сплайнами (9) и (8), которую примем в виде дискретной коэффициентной нормы, минимизирующей отклонение контрольных точек  $\hat{P}_i, \hat{Q}_j$ ,  $i = \overline{0, n}, j = \overline{0, m}$  от исходных  $P_i, Q_j$ ,  $i = \overline{0, n}, j = \overline{0, m}$ :

$$d(R, \bar{R}) = O(\varepsilon_i, \delta_j) = \left( \sum_{i=0}^n \|\varepsilon_i\|^2 + \sum_{j=0}^m \|\delta_j\|^2 \right) \rightarrow \min. \quad (10)$$

Нахождение аппроксимирующего сплайна  $R(t)$  будем выполнять в два этапа.

**Этап 1.** Выполняется полностью гладкое сопряжение сплайнов (1) и (2), для чего рассчитываются новые контрольные точки  $\hat{P}_i, \hat{Q}_j$ ,  $i = \overline{0, n}, j = \overline{0, m}$ , которые используются для построения новых сопряженных  $B$ -сплайнов  $\hat{P}(u)$  и  $\hat{Q}(v)$ . При этом узловые вектора исходных и сопряженных сплайнов не изменяются и соответствуют (3) и (4). Сплайны  $\hat{P}(u)$  и  $\hat{Q}(v)$  должны удовлетворять условиям гладкого сопряжения

$$\hat{P}^{(r)}(u)|_{u=1} = \hat{Q}^{(r)}(v)|_{v=0}, \quad r = \overline{0, (p-1)}, \quad (11)$$

где  $r$  – порядок производной.

**Этап 2.** Преобразование без погрешности (абсолютно точно) сопряженных сплайнов  $\hat{P}(u)$  и  $\hat{Q}(v)$  в один аппроксимирующий  $B$ -сплайн  $R(t)$  (9), который определен на узловом векторе (7), для чего необходимо выполнить расчет контрольных точек  $R_i$ ,  $i = \overline{0, (n+m-p+1)}$ .

Кроме ограничений (11) на  $B$ -сплайны  $\hat{P}(u)$  и  $\hat{Q}(v)$  (и соответственно на  $R(t)$ ) могут быть наложены дополнительные ограничения:

1) полного совпадения с одним из исходных  $B$ -сплайнов:

$$\hat{P}(u) = P(u), \quad u_p \leq u \leq u_{n+1}, \quad (12)$$

или

$$\hat{Q}(v) = Q(v), \quad v_p \leq v \leq v_{m+1}; \quad (13)$$

2) сопряженные  $B$ -сплайны  $\hat{P}(u)$  и  $\hat{Q}(v)$  должны проходить через точки, заданные на исходных сплайнах, с сохранением значений производных в этих точках:

$$\hat{P}^{(r1)}(u)|_{u=u_{kp}} = P^{(r1)}(u)|_{u=u_{kp}}, \quad kp = \overline{0, l}, \quad r1 = \overline{0, rp}, \quad (14)$$

$$\hat{Q}^{(r2)}(v)|_{v=v_{kq}} = Q^{(r2)}(v)|_{v=v_{kq}}, \quad kq = \overline{0, m}, \quad r2 = \overline{0, rq}, \quad (15)$$

где  $(l+1), (m+1)$  – количество точек, в которых заданы ограничения.

Таким образом, задача полностью гладкого сопряжения  $B$ -сплайнов (1), (2) состоит в оптимизационной задаче (10), при ограничениях (11), кроме этого, могут присутствовать дополнительные ограничения (12), (13) и (14), (15).

### Используемый метод

Оптимизационную задачу (10) при ограничениях (11) будем решать методом неопределенных множителей Лагранжа [5]. Функция Лагранжа имеет вид

$$L = \sum_{i=n-p+1}^n \|\varepsilon_i\|^2 + \sum_{j=0}^{p-1} \|\delta_j\|^2 + \sum_{r=0}^{p-1} \lambda_r \left[ \sum_{i=0}^r (P_{n-i} + \varepsilon_{n-i}) N_{n-i,p}^{(r)}(1) - \sum_{j=0}^r (Q_j + \delta_j) N_{j,p}^{(r)}(0) \right]. \quad (16)$$

Вычисляя частные производные функции (16) по  $\varepsilon_{n-i}, \delta_j$  и  $\lambda_r$ ,  $i, j, r = \overline{0, (p-1)}$  и приравнявая их к нулю, получаем систему из  $3p$  линейных уравнений относительно неизвестных  $\varepsilon_{n-i}, \delta_j$  и  $\lambda_r$ ,  $i, j, r = \overline{0, (p-1)}$ .

Дополнительные ограничения (12) или (13) тривиальны, они выполняются исключением из оптимизационной задачи соответствующих переменных ( $\varepsilon_{n-i}$  для (12) и  $\delta_j$  для (13)). Ограничения (14) и (15) добавляются к оптимизационной задаче по аналогии с ограничениями (11), как это выполнено в [5].



### Переход к аппроксимирующему сплайну $R(t)$

Преобразование двух полностью сопряженных  $B$ -сплайнов  $\hat{P}(u)$  и  $\hat{Q}(v)$  в один  $B$ -сплайн  $R(t)$  той же степени  $p$  производится без какой-либо погрешности с помощью расчета новых контрольных точек  $R_k$ ,  $k = \overline{0, (n + m - p + 1)}$  и нового узлового вектора для этого сплайна:

$$T = \left\{ \underbrace{0, \dots, 0}_{p+1}, t_{p+1}, \dots, t_{n+m-p+1}, \underbrace{t_{n+m-p+2}, \dots, t_{n+m-p+2}}_{p+1} \right\}. \quad (17)$$

Отметим, что  $B$ -сплайны  $\hat{P}(u)$  и  $\hat{Q}(v)$  после гладкого сопряжения определены на *Clamping* (зажатых) узловых векторах.

Введем следующие обозначения:  $\hat{P}_{UnClR}$ ,  $\hat{Q}_{UnClL}$  – сплайны, полностью совпадающие со сплайнами  $\hat{P}(u)$ ,  $\hat{Q}(v)$  соответственно, при этом сплайн  $\hat{P}_{UnClR}$  (в соответствии с принятой терминологией) определен на *UnClamping\_Rigth* (не зажатом справа) узловом векторе, сплайн  $\hat{Q}_{UnClL}$  определен на *UnClamping\_Left* (не зажатом слева) узловом векторе.

Переход от *Clamping* (зажатого) узлового вектора к не зажатому справа или слева векторам выполняется по следующим соотношениям:

$$u_{n+i+2} = u_{n+i+1} + (u_{p+i+1} - u_{p+i}), \quad i = \overline{0, (p-1)}, \quad (18)$$

$$u_{p-i-1} = u_{p-i} + (u_{n-i+1} - u_{n-i}), \quad i = \overline{0, (p-1)}. \quad (19)$$

По (18) переходим к *UnClamping\_Rigth* (не зажатый справа), по (19) – к *UnClamping\_Left* (не зажатый слева).

Для «объединения» *UnClamping\_Rigth* (не зажатый справа) вектора с *UnClamping\_Left* (не зажатый слева) в один *Clamping* (зажатый) узловой вектор вида (17) необходимо воспользоваться процедурой (7).

Преобразование сплайнов  $\hat{P}$ ,  $\hat{Q}$  в  $\hat{P}_{UnClR}$ ,  $\hat{Q}_{UnClL}$  соответственно выполняется с помощью следующих равенств:

$$\hat{P}_{UnClR}(u)^{(r)}|_{u=1} = \hat{P}(u)^{(r)}|_{u=1}, \quad r = \overline{0, (p-2)}, \quad (20)$$

$$\hat{Q}_{UnClL}(v)^{(r)}|_{v=0} = \hat{Q}(v)^{(r)}|_{v=0}, \quad r = \overline{0, (p-2)}. \quad (21)$$

В развернутом виде условия (20), (21) имеют вид

$$\sum_{i=0}^p \hat{P}_{UnClR_{n-i}} N_{UnCl_{n-i,p}}^{(r)}(1) = \sum_{i=0}^p \hat{P}_{Cl_{n-i}} N_{Cl_{n-i,p}}^{(r)}(1), \quad r = \overline{0, (p-2)}, \quad (22)$$

$$\sum_{j=0}^p \hat{Q}_{UnClL_j} N_{UnCl_{j,p}}^{(r)}(0) = \sum_{j=0}^p \hat{Q}_{Cl_j} N_{Cl_{j,p}}^{(r)}(0), \quad r = \overline{0, (p-2)}. \quad (23)$$

Система линейных уравнений (22) решается относительно  $\hat{P}_{UnClR_{n-i}}$ ,  $i = \overline{0, (p-2)}$ , система (23) – относительно  $\hat{Q}_{UnClL_j}$ ,  $j = \overline{0, (p-2)}$ . Остальные контрольные точки сплайнов  $\hat{P}_{UnClR}$ ,  $\hat{Q}_{UnClL}$  совпадают с соответствующими контрольными точками сплайнов  $\hat{P}_{Cl}$ ,  $\hat{Q}_{Cl}$ .

В точке сопряжения сплайнов  $\hat{P}_{UnClR}$ ,  $\hat{Q}_{UnClL}$  совпадают производные вплоть до порядка  $(p-1)$ , поэтому контрольные точки  $\hat{P}_{UnClR_{n-i}}$ ,  $i = \overline{0, p}$  совпадают с контрольными точками  $\hat{Q}_{UnClL_j}$ ,  $j = \overline{0, p}$ . Объединив контрольные точки  $\hat{P}_{UnClR_i}$ ,  $i = \overline{0, n}$  и  $\hat{Q}_{UnClL_j}$ ,  $j = \overline{0, m}$ , получаем контрольные точки  $R_i$ ,  $i = \overline{0, (n + m - p + 1)}$  аппроксимирующего сплайна  $R(t)$ .

### Результаты и их обсуждение

На рисунке 1, а представлены исходные и сопряженные  $B$ -сплайны степени  $p = 3$  ( $n = m = 4$ ).

Исходные сплайны имеют узловые векторы:

$$U = V = \{0; 0; 0; 0; 0,5; 1; 1; 1; 1\}.$$

На рисунке 1, б, в представлены результаты преобразования сплайнов  $\hat{P}(u)$ ,  $\hat{Q}(v)$  в сплайны  $\hat{P}_{UnClR}(u)$  и  $\hat{Q}_{UnClL}(v)$  с узловыми векторами:

$$U_{UnClR} = \{0; 0; 0; 0; 0,5; 1; 1,5; 2; 2,5\}, \\ V_{UnClL} = \{-1,5; -1; -0,5; 0; 0,5; 1; 1; 1; 1\}.$$

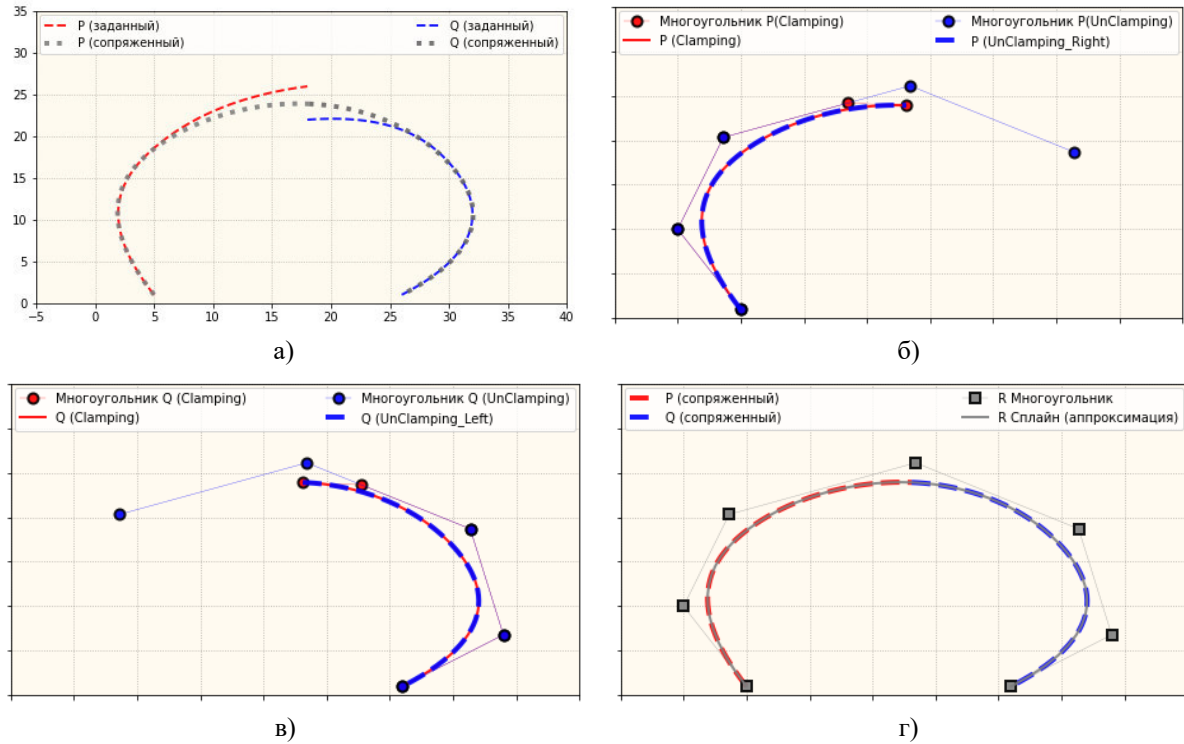


Рисунок 1. Сопряжение сплайнов: а – исходные B-сплайны  $P(u)$ ,  $Q(v)$  и сопряженные  $\hat{P}(u)$ ,  $\hat{Q}(v)$ ; б – сплайн  $\hat{P}_{UnClR}(u)$ ; в – сплайн  $\hat{Q}_{UnClL}(v)$ ; г – «объединение» сплайнов  $\hat{P}_{UnClR}(u)$  и  $\hat{Q}_{UnClL}(v)$  в аппроксимирующий сплайн  $R(t)$

Из рисунка 1, б, в видно, что последние три контрольные точки сплайна  $\hat{P}_{UnClR}(u)$  совпадают с первыми тремя контрольными точками сплайна  $\hat{Q}_{UnClL}(v)$ . На рисунке 1, г иллюстрируется «объединение» сплайнов  $\hat{P}_{UnClR}(u)$  и  $\hat{Q}_{UnClL}(v)$  в аппроксимирующий сплайн  $R(t)$  с узловым вектором

$$T = \{0; 0; 0; 0; 0,5; 1; 1,5; 2; 2; 2; 2\}.$$

На рисунке 2 приводятся результаты сопряжения при последовательном добавлении следующих дополнительных ограничений для точки  $u_0 = 0,75$ :

$$\hat{P}(u_0) = P(u_0), \quad (25)$$

$$\hat{P}^{(1)}(u_0) = P^{(1)}(u_0), \quad (26)$$

$$\hat{P}^{(2)}(u_0) = P^{(2)}(u_0). \quad (27)$$

Дополнительное ограничение (25) соответствует прохождению аппроксимирующего сплайна  $R(t)$  через точку  $u_0$  сплайна  $P(u)$ . Совместные ограничения (25) и (26) соответствуют требованию прохождения  $R(t)$  через точку  $P(u_0)$  с сохранением первой производной. Ограничения (25)-(27) устанавливают требования прохождения через точку  $P(u_0)$  с сохранением первой и второй производных.

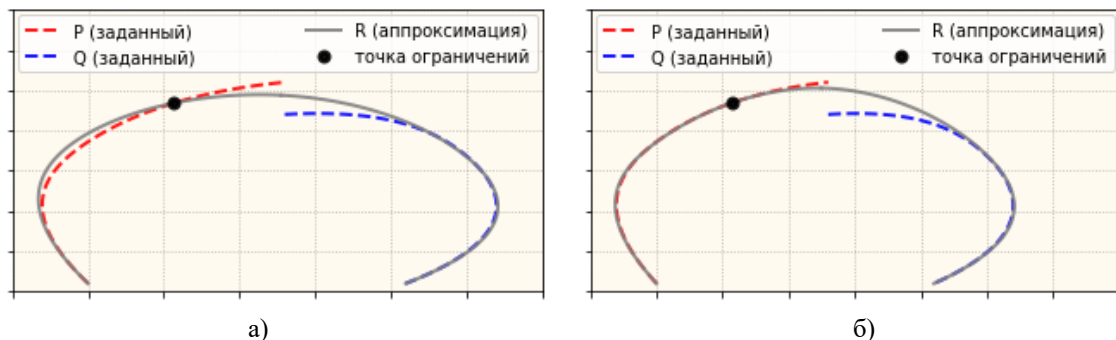


Рисунок 2. Сопряжение с дополнительными ограничениями: а – ограничение (25) – прохождение через точку, б – ограничения (25)-(26) – прохождение через точку с сохранением 1-й производной

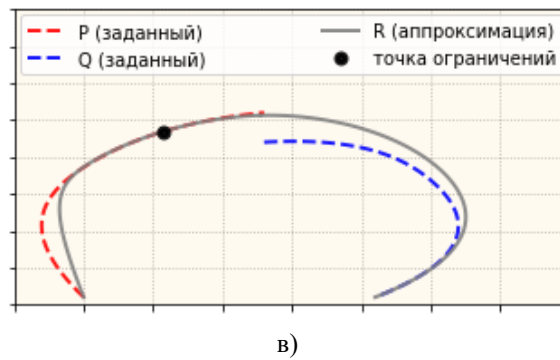


Рисунок 2. Сопряжение с дополнительными ограничениями:  $\phi$  – ограничения (25)-(27) – прохождение через точку с сохранением 1-й и 2-й производных

### Понижение степени B-сплайна

Задан сплайн  $C_0(t)$  степени  $p_0$ :

$$C_0(t) = \sum_{i=0}^{n_0} P_{0i} N_{i,p_0}(t),$$

где  $(n_0 + 1)$  – количество контрольных точек сплайна  $C_0(t)$ .

Необходимо аппроксимировать сплайн (28) сплайном  $C_1(t)$  степени  $p_1$ , где  $p_1 < p_0$ :

$$C_1(t) = \sum_{i=0}^{n_1} P_{1i} N_{i,p_1}(t),$$

где  $(n_1 + 1) \leq (n_0 + 1)$  – количество контрольных точек сплайна  $C_1(t)$ .

Алгоритм включает в себя последовательность четырех следующих вычислительных процедур.

#### 1. Представление сплайна $C_0(t)$ как набора сопряженных кривых Безье

Для выполнения этой процедуры в узловом векторе сплайна  $C_0(t)$  все внутренние узлы (узлы в реальной части вектора) делаются кратными степени сплайна  $p_0$ . Это выполняется с помощью процедуры вставки дополнительных узлов в узловый вектор [2]. Пример представления сплайна как набора сопряженных кривых Безье методом вставки кратных узлов приведен на рисунке 3.

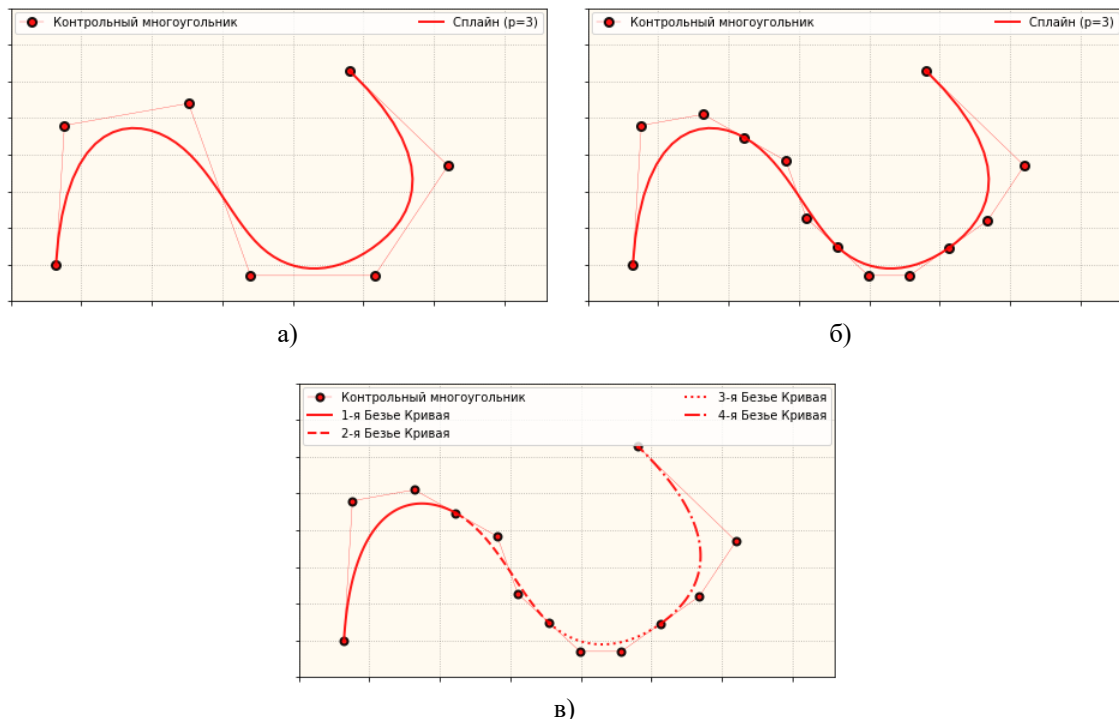


Рисунок 3. Представления сплайна как набора сопряженных кривых Безье:  $a$  – исходный сплайн 3-й степени, узловый вектор  $U = (0,0,0,0, 0.25, 0.5, 0.75, 1,1,1,1)$ ;  $б$  – сплайн после вставки кратных узлов,  $U = (0,0,0,0, 0.25, 0.25, 0.25, 0.5, 0.5, 0.5, 0.75, 0.75, 1,1,1,1)$ ;  $в$  – 4 кривых Безье 3-й степени

## 2. Понижение степени кривой Безье

Запишем исходную кривую Безье степени  $p$  в виде  $B$ -сплайна:

$$C(t) = \sum_{i=0}^p P_i N_{i,p}(t), \quad (28)$$

где  $P_i, i = \overline{0, p}$  – контрольные точки.

Запишем кривую Безье степени  $(p-1)$ :

$$F(t) = \sum_{i=0}^{p-1} F_i N_{i,p-1}(t), \quad (29)$$

где  $F_i, i = \overline{0, (p-1)}$  – контрольные точки.

Рассмотрим интегральную метрику разности кривых (28) и (29) и минимизируем эту метрику по контрольным точкам кривой  $F(t)$ :

$$J(F_0, \dots, F_{p-1}) = \int_0^1 (C(t) - F(t))^2 dt \rightarrow \min_{F_0, \dots, F_{p-1}}. \quad (30)$$

Дополним оптимизационную задачу (30) четырьмя ограничениями:

$$F(t|_{t=0}) = C(t|_{t=0}), \quad (31)$$

$$F^{(1)}(t|_{t=0}) = C^{(1)}(t|_{t=0}), \quad (32)$$

$$F(t|_{t=1}) = C(t|_{t=1}), \quad (33)$$

$$F^{(1)}(t|_{t=1}) = C^{(1)}(t|_{t=1}). \quad (34)$$

Ограничения (31), (32) задают неизменность начальной точки и производной в этой точке. Ограничения (33), (34) – аналогичные ограничения для конечной точки. Оптимизационная задача (30) при ограничениях (31)-(32) решается методом неопределенных множителей Лагранжа, который подробно описан в работе [5].

Пример понижения степени различных кривых Безье алгоритмом (30)-(34) приведен на рисунке 4.

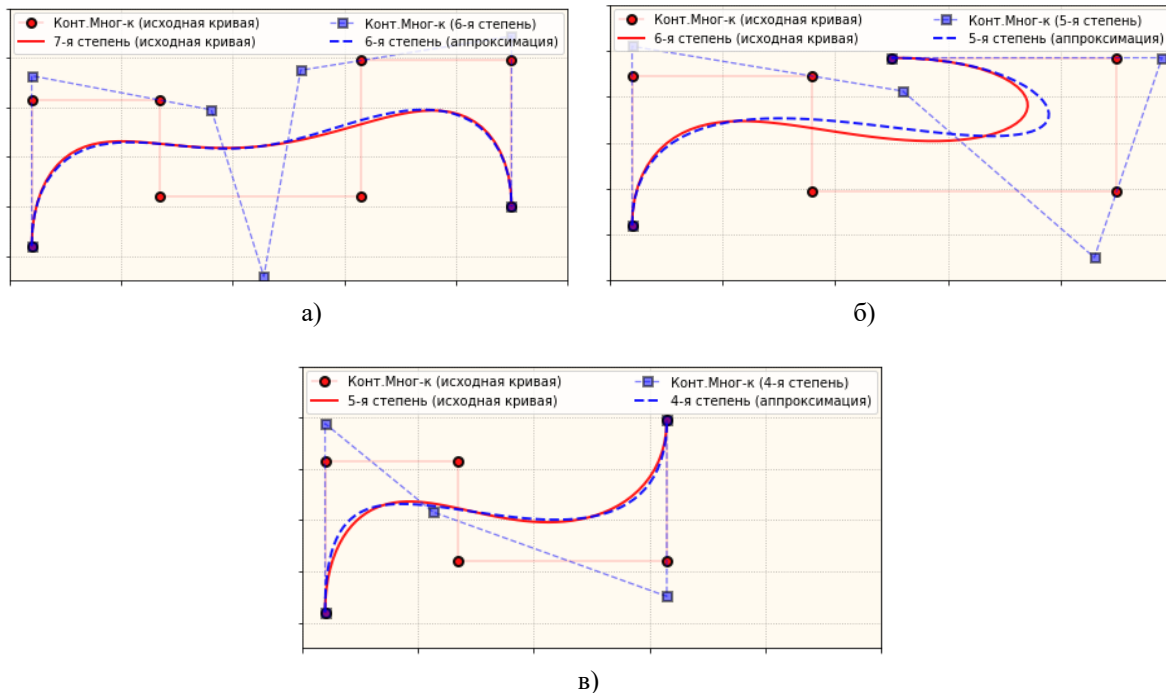


Рисунок 4. Понижение степени кривых Безье: а – понижение с 7-й до 6-й степени; б – понижение с 6-й до 5-й степени; в – понижение с 5-й до 4-й степени

Следует отметить, что результат работы нашего алгоритма довольно близок к результатам алгоритма понижения степени кривых Безье, приведённого в [120].

## 3. Гладкое сопряжение кривых Безье

На предыдущем шаге алгоритма (шаг 2 «Понижение степени кривой Безье») кривые Безье сопряжены между собой с порядком  $C^1$ . На текущем шаге кривые Безье сопрягаются с порядком

гладкости, равным степени кривых, для чего используется алгоритм (программный модуль), приведенный в работе [5].

#### 4. Расчет метрик

Рассчитываются метрики Хаусдорфа, кривизны и квадратичной разности между исходным сплайном и полученным после шага 3 сплайном (этот сплайн представляет собой набор гладко сопряженных кривых Безье).

На рисунке 5, а приведен результат работы алгоритма по понижению степени с 24-й до 12-й. На рисунке 5, б показано возрастание метрики Хаусдорфа по мере понижения степени кривой, приведенной на рисунке 5, а. На рисунке 5, б представлены метрики для предлагаемого алгоритма и алгоритма, основанного на интегральной норме (то есть минимизации функционала (30)).

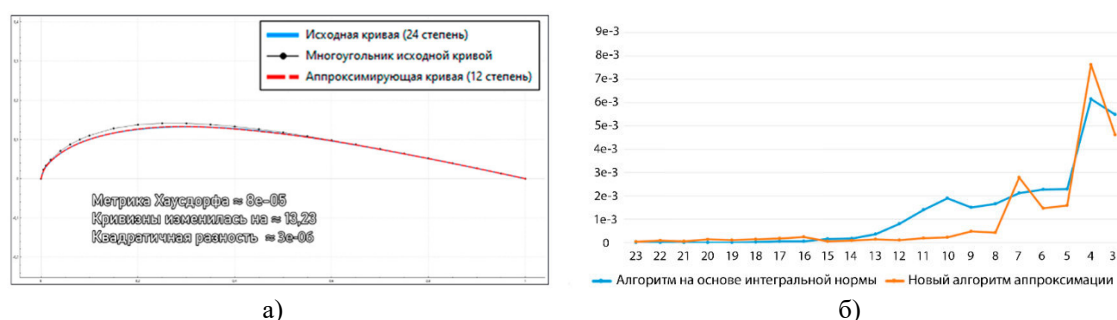


Рисунок 5. Результат работы алгоритма

#### Выводы

1. Решена задача гладкого сопряжения двух сплайнов произвольной степени, при этом под «гладким сопряжением» понимается выполнение в точке сопряжения условий непрерывности производных всех порядков, вплоть до порядка  $(p - 1)$ , где « $p$ » – степень исходных сплайнов.

2. Решена задача замены без погрешности двух гладко сопряженных сплайнов, определенных на меньшем числе контрольных точек.

3. Решена задача понижения степени  $B$ -сплайна, одним из шагов которой является гладкое сопряжение.

4. Решение задач гладкого сопряжения сводится к оптимизационной задаче при наличии ограничений, которая решается методом неопределенных множителей Лагранжа.

#### Список литературы

1. Голованов Н. Н. Геометрическое моделирование / Н. Н. Голованов. – М.: Изд-во Физико-математической литературы, 2002. 472 с. ISBN 5-94052-048-0.
2. Piegl L., Tiller W. The NURBS book. 2 ed. Springer, 1997. 646 p. ISBN 978-3-540-61545-3.
3. Rogers D. F. An introduction to NURBS: with historical perspective. Elsevier, 2001. 324p. ISBN 1-55860-669-6.
4. Salomon D. Curves and Surfaces for Computer Graphics. Springer, 2006. 461 p. ISBN 0-387-24196-5.
5. Аппроксимация сопряжения кривых Безье с сохранением порядка гладкости и дополнительными ограничениями / С. Н. Ганчук, О. В. Кривошеев, С. В. Маврин [и др.] // Вестник Южно-Уральского государственного университета. Серия: Строительство и архитектура. 2024. Т. 24, № 1. С. 59-69. DOI: 10.14529/build240108.

УДК 515:145

DOI: 10.25686/978-5-8158-2474-4-2025-920-929

## Развитие теории $\alpha\beta$ -триангуляции в трехмерном евклидовом пространстве

**В. В. Рустамян**

МИРЭА – Российский технологический университет, Москва, Россия

**Аннотация.** Исследование расширяет теорию  $\alpha\beta$ -триангуляции в трёхмерном евклидовом пространстве. Автор выделяет характерные особенности трёхмерного случая, такие как наличие замкнутых дискретных поверхностей и необходимость уточнения существующих определений и свойств. Проводится анализ минимальных конфигураций замкнутых  $\alpha\beta$ -триангуляций, примером которых служит треугольный диэдр. Анализ позволяет выявить критическое отношение приблизительно 16,67 % между количеством  $\alpha$ - и  $\beta$ -граней, что имеет практическое значение. Предложен эффективный механизм рекомбинации  $\beta$ -ребер в трёхмерном пространстве, позволяющий сохранять целостность триангуляции при работе с поверхностями свободной формы. Потенциальные направления будущих исследований включают расширение предложенного подхода на поверхности рода выше 0, увеличивая область его применения. Данная работа вносит существенный вклад в развитие вычислительной геометрии, архитектурного проектирования и цифровой обработки поверхностей.

**Ключевые слова:** оптимальная триангуляция,  $\alpha\beta$ -триангуляция, свойства  $\alpha\beta$ -триангуляции, поверхности свободной формы, дискретные поверхности

## Development of the theory of $\alpha\beta$ -triangulation in three-dimensional Euclidean space

**V. V. Rustamyan**

MIREA – Russian Technological University, Moscow, Russia

**Abstract.** This study expands the theory of  $\alpha\beta$ -triangulation into three-dimensional Euclidean space. The author identifies characteristic features of the three-dimensional case, such as the presence of closed discrete surfaces and the need to refine existing definitions and properties. Analysis of minimal configurations of closed  $\alpha\beta$ -triangulations, exemplified by the triangular dihedron, allows revealing a critical ratio of approximately 16.67 % between the number of  $\alpha$ - and  $\beta$ -faces, which holds practical relevance. An efficient mechanism for recombination of  $\beta$ -edges in three-dimensional space is proposed, ensuring the integrity of the triangulation when handling free-form surfaces. Future research directions include extending the approach to higher-genus surfaces, thereby broadening its application scope. This work makes a valuable contribution to the development of computational geometry, architectural design, and digital surface processing.

**Keywords:** optimal triangulation,  $\alpha\beta$ -triangulation, properties of  $\alpha\beta$ -triangulation, free-form surfaces, discrete surfaces

### Введение

Одной из задач генеративного проектирования [1, 2] в архитектуре является задача аппроксимации поверхностей с использованием дискретных классов эквивалентных элементов. Основной стратегией является разработка методов, позволяющих аппроксимировать заданную поверхность небольшим набором плиточных элементов ( $n$ -угольников), что позволяет минимизировать количество уникальных компонентов и снизить затраты на производство строительных конструкций. В терминах дискретной дифференциальной геометрии [3] оптимизации подлежат дискретные поверхности – аналоги (если точнее, то «дискретные аналоги», по А.И. Бобенко) аналитических или поверхностей свободных форм. Работы ведутся над дискретными поверхностями с плиточными элементами разной природы. Одним из распространенных типов плиток является четырехугольник. Дискретные поверхности, образованные четырёхугольниками, принято называть сетями [1]. Одни публикации [4-8] посвящены оптимизации по признаку конгруэнтности сетей. Другие публикации касаются гексагональных плиток [9] и дискретных поверхностей со смещенными типами плиток [10, 11]. Значительное внимание уделяют исследованию методов оптимизации триангуляций [12-16]. Несмотря на различия подходов в исследованиях, существует общая логика: сначала определяются объекты исследования, затем проводится их предварительная обработка, после чего применяются соответствующие методы оптимизации.

В рамках решения задачи аппроксимации замкнутых поверхностей свободной формы полиэдрами с группами конгруэнтных граней была создана программа компьютерного геометрического моделирования [17]. Объект исследования – замкнутая поверхность свободной формы, заданная в виде

stl-файла с высоким числом граней. Предварительная обработка – получение триангуляции с нужным числом граней из заданной. Для оптимизации триангуляции применялся генетический алгоритм. В программе используется кандидатное решение (хромосома, индивид) в виде вектора с координатами всех вершин, разбитый на аллели по трем координатам каждой вершины. Для получения первого индивида используется алгоритм локального упрощения триангуляции [18, 19] с помощью коллапса по наименьшему ребру и с ограничениями валентностей вершин от 4 до 7. В процессе оптимизации производится перемещение вершин триангуляции полиэдра пока не будет найдено частичное или в идеальном случае полное решение. Полным считается решение, в котором отсутствуют одиночные по признаку конгруэнтности треугольники. При этом не ставилось задачи ограничения численности групп. Например, если триангуляция содержит 256 граней, то обнаружение триангуляции со 128-ю парами конгруэнтных треугольников считается полным решением [20].

Такой подход имеет три основных недостатка: первый – слишком широкое поисковое пространство, так как первое поколение индивидов не содержит в себе ни одной пары конгруэнтных граней с точностью до случайного совпадения; второй – отсутствие ориентира в виде эталонного треугольника или группы эталонных треугольников; третий и наиболее существенный недостаток – статичность структуры триангуляции первого индивида. Если триангуляцию рассматривать как граф, то под статичностью структуры подразумевается отсутствие возможности изменения валентностей вершин. Это ограничивает поисковое пространство задачи оптимизации в качественном направлении. Так как возможно возникновение ситуации невозможности полного решения при определенной структуре первого индивида [21].

В публикации [22] выдвинута гипотеза о том, что до совершения оптимизационных действий над триангуляцией, телом которого является заявленный полиэдр, возможно задать для любой триангуляции одну или несколько групп конгруэнтных треугольников без потери подвижности математической модели для механизмов оптимизации. В связи с этим была создана математическая модель « $\alpha\beta$ -триангуляция». Определены  $\alpha\beta$ -триангуляция в пространстве  $E^2$ , операции сшивки и разреза и описан алгоритм образования  $\alpha\beta$ -триангуляции из произвольной плоской триангуляции. По сути своей  $\alpha\beta$ -триангуляция является оптимальной триангуляцией с ограничениями по критерию минимизации суммы длин ребер.

В данной статье определяются  $\alpha\beta$ -триангуляция и ее свойства в пространстве  $E^3$ . Рассматриваются примеры  $\alpha\beta$ -триангуляции для триангуляций известных полиэдров, гомеоморфных сфере. Определяется механизм рекомбинации  $\beta$ -ребер в пространстве  $E^3$ .

### Постановка задачи

В данном исследовании ставились следующие цели:

- 1) определить  $\alpha\beta$ -триангуляцию в пространстве  $E^3$ ;
- 2) рассмотреть свойства  $\alpha\beta$ -триангуляции;
- 3) определить механизм рекомбинации  $\beta$ -ребер.

### Теория

#### 1. $\alpha\beta$ -триангуляция в пространстве $E^3$

**Определение.** Сильно связную триангуляцию  $K$  размерности 2 в пространстве  $E^3$ , в которой возможно выделить некоторое множество граней, их замыкания обозначить  $\alpha$ -треугольниками и дополнение к ним до  $K$  –  $\beta$ -элементами, назовем  $\alpha\beta$ -триангуляцией ( $T_{\alpha\beta}$ ), если она удовлетворяет условиям:

- 1) тела  $\alpha$ -треугольников не имеют общих точек;
- 2) триангуляция  $K$  содержит только вершины остовов  $\alpha$ -треугольников;
- 3) сумма длин  $\beta$ -ребер минимальна.

Комплексом в данной статье будем называть триангуляции и всевозможные их подмножества. Телом комплекса будем называть точечное множество данного пространства, являющееся соединением всех элементов комплекса [23, с. 99].

Все свойства  $T_{\alpha\beta}$  [22] работают и в трехмерном евклидовом пространстве. Основным отличием от плоской триангуляции является существование замкнутых дискретных поверхностей. В связи с этим дополним некоторые свойства. Нумерация свойств будет соблюдаться как в источнике.



«Свойство 6. Минимальной  $\alpha\beta$ -триангуляцией является  $\alpha$ -треугольник» [22].

Рассмотрим замкнутые дискретные сферические поверхности. На рисунке 1в показана триангуляция первого тела Платона – тетраэдра, в которой выделен  $\alpha$ -треугольник. Оставшаяся противоположная вершина не входит в остов ни одного треугольника не имеющего общих вершин с  $\alpha$ -треугольником. Из определения ясно, что тетраэдр не может быть  $\alpha\beta$ -триангуляцией. Замкнутой  $\alpha\beta$ -триангуляцией с двумя  $\alpha$ -треугольниками будет являться триангуляция октаэдра. На рисунке 1б можно наблюдать две выделенные грани –  $\alpha$ -треугольники. Шесть оставшихся граней являются  $\beta$ -гранями.

Минимальной замкнутой  $\alpha\beta$ -триангуляцией является триангуляция, которая содержит 3 вершины, 3 ребра и 2 грани (треугольный диэдр). Одна из граней будет  $\alpha$ -гранью, другая  $\beta$ -гранью. На рисунке 1а представлена сферическая мозаика [24, 25] треугольного диэдра.

**Свойство 6.1.** Минимальной замкнутой  $\alpha\beta$ -триангуляцией является треугольный диэдр.

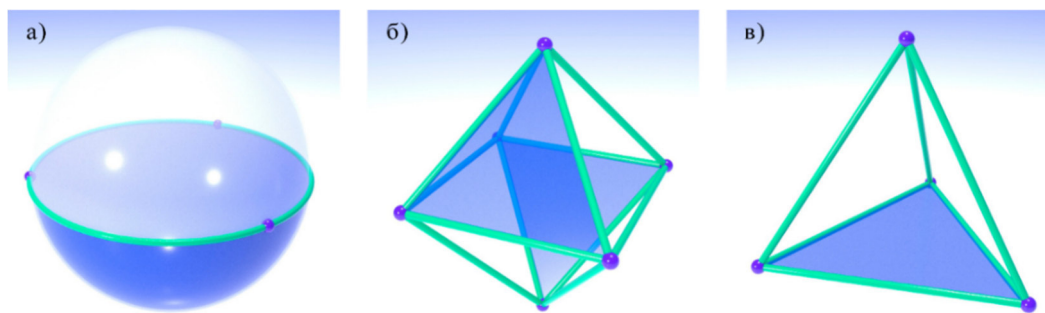


Рис. 1. Представление в виде сферической мозаики треугольного диэдра ( $\alpha\beta$ -триангуляции с одним  $\alpha$ -треугольником) (а);  $\alpha\beta$ -триангуляция с двумя  $\alpha$ -треугольниками и 6-ю  $\beta$ -треугольниками (б); триангуляция тетраэдра с выделенным  $\alpha$ -треугольником, которая не является  $\alpha\beta$ -триангуляцией (в)

«Свойство 7. Число  $\beta$ -граней в  $\alpha\beta$ -триангуляции находится на отрезке  $\{b \in \mathbb{N}: 2(a-1) \leq b \leq 5(a-1)\}$ ,  $a = |K_\alpha^2|$  – число  $\alpha$ -граней,  $b = |K_\beta^2|$  – число  $\beta$ -граней» [22].

Рассмотрим замкнутые дискретные поверхности рода 0 [23, с. 139]. На рисунке 2а приведен максимальный планарный граф замкнутой  $\alpha\beta$ -триангуляции, где  $a = 2$ . Видно, что граф гомеоморфен октаэдру (рис. 1б). При этом к числу  $\beta$ -граней незамкнутой поверхности добавится одна грань  $AB'C'$ . Если к комплексу без последней грани добавить  $\alpha$ -треугольник  $A''B''C''$  (рис. 2б), то образуется комплекс, в котором добавится еще 5  $\beta$ -граней и  $AB'C'$  заменится на  $A''B''C'$ . По индукции легко определяется число  $\beta$ -граней для триангуляций рода 0.

**Свойство 7.1.**  $b = 5(a-1) + 1$ , где  $b$  – число  $\beta$ -граней в замкнутой  $\alpha\beta$ -триангуляции рода 0.

Свойство 7.1 подтверждает свойство 6.1, при  $a = 1$ ,  $b = 1$ .

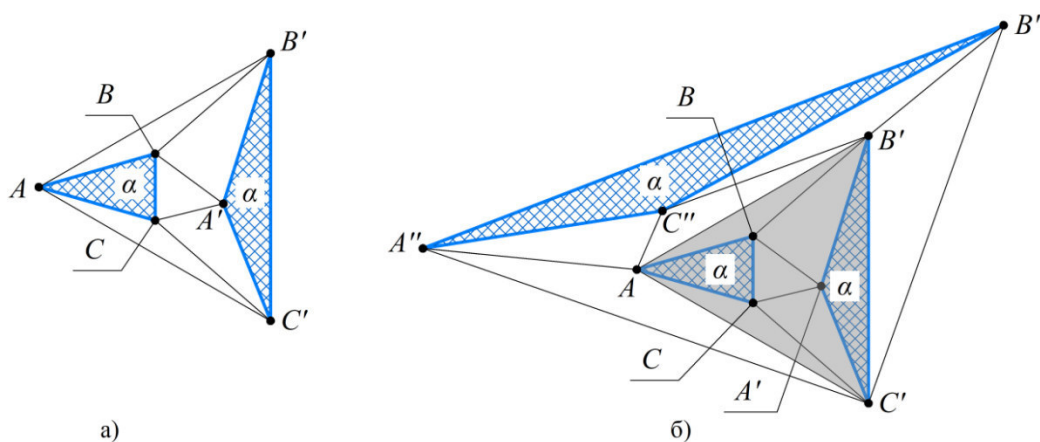


Рис. 2. Число  $\beta$ -граней замкнутой  $\alpha\beta$ -триангуляции

«Свойство 9.  $\alpha\beta$ -триангуляция с  $a = 3$  является минимальной из комплексов, границы которых – непрерывная последовательность  $\beta$ -отрезков, тогда и только тогда, когда хотя бы один из трех  $\alpha$ -треугольников неправильный» [22].

Свойство 9 в трехмерном евклидовом пространстве останется неизменным по числу  $\alpha$ -треугольников и справедливо для любых треугольников.

**Свойство 9.1.**  $\alpha\beta$ -триангуляция с  $a = 3$  является минимальной из комплексов, границы которых – непрерывная последовательность  $\beta$ -отрезков.

Общий алгоритм образования  $\alpha\beta$ -триангуляций из произвольной незамкнутой триангуляции представлен в статье [22]. Приведем пример образования  $\alpha\beta$ -триангуляций на основе триангуляции куба. Рассмотрим произвольную триангуляцию куба (рис. 3). В такой триангуляции возможно выделить два  $\alpha$ -треугольника и независимо от их расположения останутся две свободные вершины. Мысль о добавлении еще одной вершины для того, чтобы перевести триангуляцию в  $\alpha\beta$ -триангуляцию, является естественной (свойство 1 [22]), но куда именно добавить вершину и каков механизм добавления, не является очевидным. И что делать при одной свободной вершине? Потому каждую свободную вершину и ее звезду [23 с. 101] стоит рассматривать отдельно. Есть два способа реконструирования звезд вершин: первый – применить метод коллапса вершины, который приведет к понижению числа граней на 2; второй – применить метод порождения грани. Если механизм порождения грани будет обратен механизму коллапса грани, то замыкание порождаемой грани будет являться  $\alpha$ -треугольником. Число граней триангуляции увеличивается на 4. Тем самым для триангуляции куба есть три возможных исхода. Коллапс двух вершин приведет к  $\alpha\beta$ -триангуляции с двумя  $\alpha$ -треугольниками (рис. 4а), что комбинаторно эквивалентно триангуляции октаэдра с точностью до рекомбинации  $\beta$ -ребер (рис. 1б). Коллапс одной вершины и порождение грани для второй вершины приведет к  $\alpha\beta$ -триангуляции с тремя  $\alpha$ -треугольниками (рис. 4б). Например, такая триангуляция эквивалентна триангуляции без добавления вершин трехскатного купала (одно из тел Джонсона) и графу на рисунке 2б. Порождение грани для обеих вершин приведет к  $\alpha\beta$ -триангуляции с четырьмя  $\alpha$ -треугольниками (рис. 4в), что эквивалентно  $\alpha\beta$ -триангуляции икосаэдра.

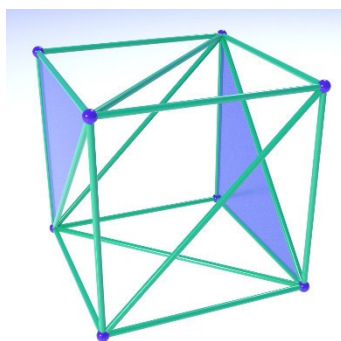


Рисунок 3. Произвольная триангуляция куба

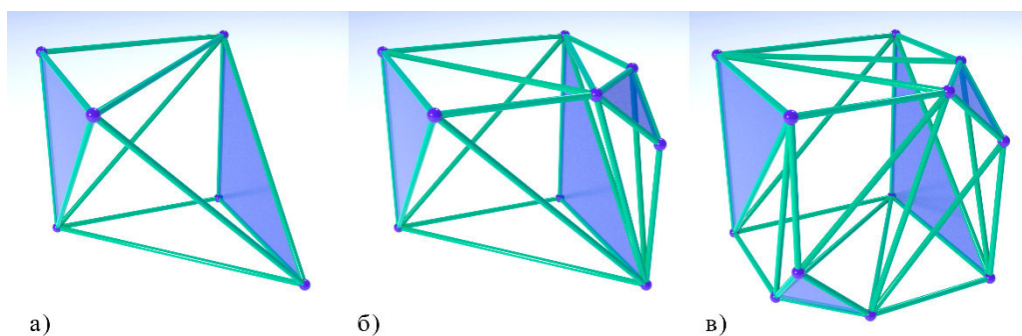

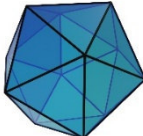
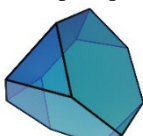

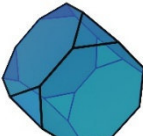
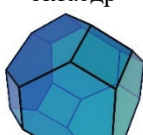


Рисунок 4. Реконструкция триангуляции куба

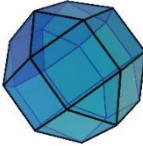
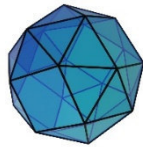
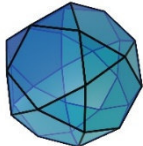
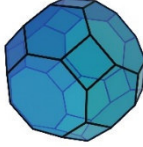
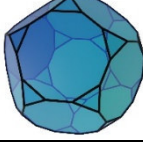
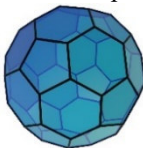
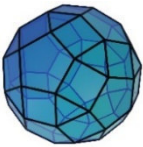
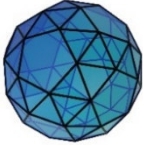
При реконструкции триангуляции с одной свободной вершиной, очевидно, может быть только два результата. Например, в случае триангуляции тетраэдра (рис. 1в) может получиться или  $\alpha\beta$ -триангуляция с одним  $\alpha$ -треугольником (рис. 1а), или  $\alpha\beta$ -триангуляции с двумя  $\alpha$ -треугольниками (рис. 1б). Если придерживаться рекомендациям из общего алгоритма [22] образования  $\alpha\beta$ -триангуляции, то в случае с тетраэдром образуется минимальная  $T_{\alpha\beta}$  (рис. 1а), в случае куба –  $T_{\alpha\beta}$  на рисунке 4б.

Опираясь на свойство 7.1 и свойство 1 [22], составим таблицу  $\alpha\beta$ -триангуляций известных полиэдров (табл. 1). Выполним расчет средней степени вершин и процента содержания  $\alpha$ -граней в триангуляции. Шифр граней полиэдра показывает количество многоугольников, содержащихся в полиэдре. Например,  $12^4 8^6 6^8$  – в усеченном кубооктаэдре 12 треугольников, 8 шестиугольников, 6 восьмиугольников. Количество граней триангуляции  $|T^2|$  – показатель количества треугольников после триангуляции полиэдра без добавления вершин. Такими же показателями являются количество ребер  $|T^1|$  и вершин  $|T^0|$  триангуляции. Можно заметить, что для разных полиэдров с одинаковым числом вершин все показатели одинаковы. Эти триангуляции на самом деле отличаются распределением степеней вершин. Но в следствии свойства 1 [22] все триангуляции с одним показателем  $|T^0|$  комбинаторно эквивалентны с точностью до рекомбинации  $\beta$ -ребер. Средняя степень вершин рассчитывается как удвоенное число ребер, деленное на число вершин триангуляции. Содержание  $\alpha$ -граней рассчитывается как соотношение  $\alpha$ -граней к общему числу граней триангуляции.

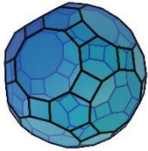
Таблица 1. Параметры  $\alpha\beta$ -триангуляций известных полиэдров

№	Наименование полиэдра	Шифр граней полиэдра	Кол-во граней тр-ции $ T^2 $	Кол-во ребер тр-ции $ T^1 $	Кол-во вершин тр-ции $ T^0  =  T_\alpha^0 $	Кол-во $\alpha$ -граней $ T_\alpha^2 $	Средняя степень вершин $2 T^1 / T^0 $	Содержание $\alpha$ -граней в тр-ции $ T_\alpha^2 / T^2 , \%$
1	Октаэдр 	$8^3$	8	12	6	2	4	25%
2	Икосаэдр 	$20^3$	20	30	12	4	5	20%
	Усеченный тетраэдр 	$4^3 4^6$						
	Кубооктаэдр 	$8^3 6^4$						
3	Усеченный куб 	$8^3 6^8$	44	66	24	8	5,5	18,18%
	Усеченный октаэдр 	$6^4 8^6$						

Продолжение таблицы 1

№	Наименование полиэдра	Шифр граней полиэдра	Кол-во граней тр-ции $ T^2 $	Кол-во ребер тр-ции $ T^1 $	Кол-во вершин тр-ции $ T^0  =  T_\alpha^0 $	Кол-во $\alpha$ - граней $ T_\alpha^2 $	Средняя степень вершин $2 T^1 / T^0 $	Содержание $\alpha$ -граней в тр-ции $ T_\alpha^2 / T^2 , \%$
	Ромбокубо- октаэдр 	$8^3 18^4$						
	Курносый куб 	$32^3 6^4$						
4	Икосододекаэдр 	$20^3 12^5$	56	84	30	10	5,6	17,86%
5	Усеченный кубооктаэдр 	$12^4 8^6 6^8$	92	138	48	16	5,75	17,39%
6	Усеченный додекаэдр 	$20^3 12^{10}$	116	174	60	20	5,8	17,24%
	Усеченный икосаэдр 	$12^5 20^6$						
	Ромбоикосо- додекаэдр 	$20^3 30^4 12^5$						
	Курносый додекаэдр 	$80^3 12^5$						

Окончание таблицы 1

№	Наименование полиэдра	Шифр граней полиэдра	Кол-во граней тр-ции $ T^2 $	Кол-во ребер тр-ции $ T^1 $	Кол-во вершин тр-ции $ T^0  =  T_\alpha^0 $	Кол-во $\alpha$ - граней $ T_\alpha^2 $	Средняя степень вершин $2 T^1 / T^0 $	Содержание $\alpha$ -граней в тр-ции $ T_\alpha^2 / T^2 $ , %
7	Ромбоусеченный икосододекаэдр 	30 <sup>4</sup> 20 <sup>6</sup> 12 <sup>10</sup>	236	354	120	40	5,9	16,95%

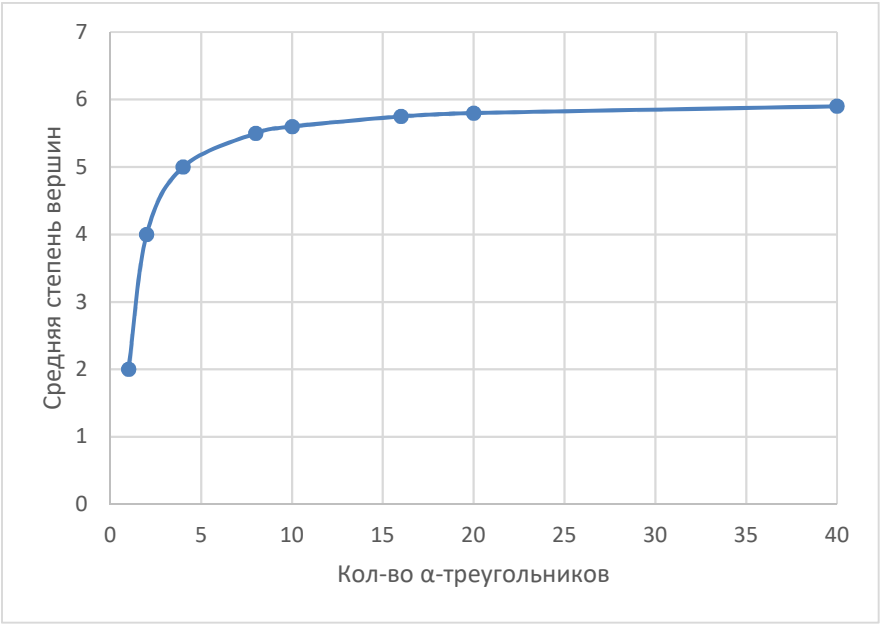


Рисунок 5. График зависимости показателя  $2|T^1|/|T^0|$  от количества  $\alpha$ -треугольников

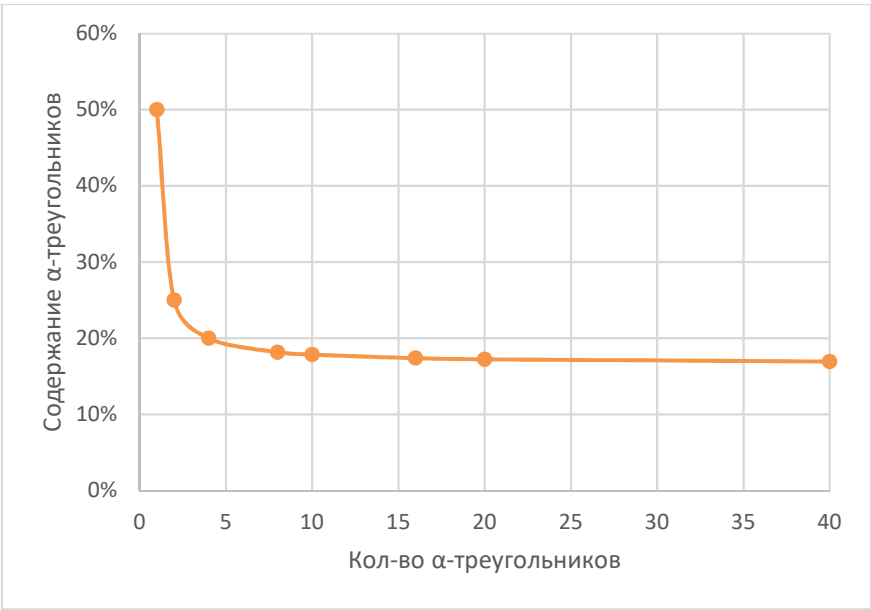


Рисунок 6. График зависимости показателя  $|T_\alpha^2|/|T^2|$  от количества  $\alpha$ -треугольников

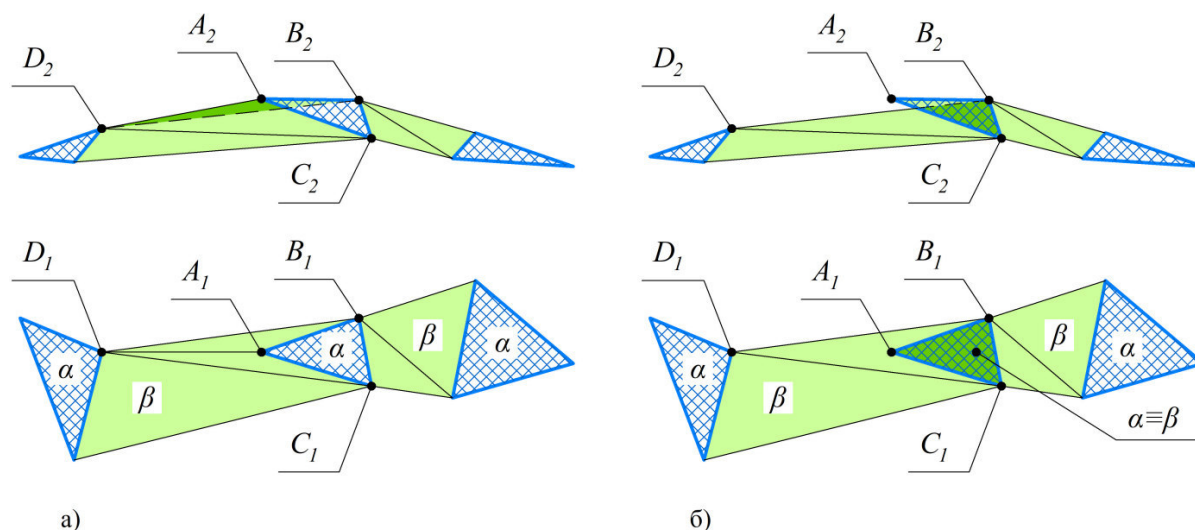
К показателям таблицы 1 добавим показатели  $\alpha\beta$ -триангуляции с одним  $\alpha$ -треугольником (рис. 1а) и построим графики зависимости показателя  $2|T^1|/|T^0|$  (рис. 5) и показателя  $|T_\alpha^2|/|T^2|$  (рис. 6) от количества  $\alpha$ -треугольников. Из графиков видно, что при увеличении числа  $\alpha$ -треугольников показатели асимптотически стремятся к определенным значениям. Так как данными триангуляциями мы аппроксимируем сферу, то бесконечное увеличение  $\alpha$ -треугольников в пределе приведет к триангуляции плоскости. На плоскости регулярной является триангуляция со степенью вершин равной 6 (рис. 7). При этом  $\alpha$ -треугольники также располагаются регулярно. Если разбить триангуляцию плоскости на равные множества, то окажется, что на каждый  $\alpha$ -треугольник приходится 5  $\beta$ -граней. Это означает, что показатель  $|T_\alpha^2|/|T^2|$  равен  $1/6$ , то есть 16,67 %. Средняя степень вершин  $2|T^1|/|T^0| = 6$ .



Рисунок 7. Регулярная триангуляция плоскости

## 2. Рекомбинация $\beta$ -ребер $T_{\alpha\beta}$ в $E^3$

В отличие от плоской триангуляции, где требуется проверка пересечений  $\beta$ -ребер с  $\alpha$ -треугольниками, для осуществления третьего условия определения рекомбинация  $\beta$ -ребер в пространстве  $E^3$  теоретически не имеет ограничений. Это следует из особенностей дискретных поверхностей в общем случае. Например, поверхность способна иметь самопересечения. Если многообразие поверхностей свободной формы ограничить, например рассматривать гладкие поверхности без самопересечений, тогда потребуется вводить условия проверки на самопересечения.

Рисунок 8. Рекомбинация  $\beta$ -ребер

Рассмотрим пример (рис. 8), где возникает самопересечение. На рисунке 8а приведен фрагмент  $\alpha\beta$ -триангуляции, где в четырехугольнике  $ABDC$  диагонали  $AD$  и  $BC$  являются конкурентными по длине. Диагональ  $BC$  оказывается короче  $AD$ , вследствие чего происходит рекомбинация (flip). На рисунке 8б показан результат такой рекомбинации. Видно, что из треугольников  $DBA$  и  $DAC$  образовались



треугольники  $DBC$  и  $ABC$ . При этом в треугольнике  $ABC$  совпадают две грани –  $\alpha$  и  $\beta$ . Образуется эдакий «карман»  $BAC$ . Дискретная поверхность с такими элементами не может аппроксимировать гладкую поверхность. Соответственно, при создании алгоритма рекомбинации в пространстве  $E^3$  потребуется ввести ограничение на создание  $\beta$ -ребра, если оно совпало с  $\alpha$ -ребром.

### Выводы

Определено, что свойства математической модели « $\alpha\beta$ -триангуляция» в пространстве  $E^2$  выполнимы и в пространстве  $E^3$ . Триангуляции в трехмерном пространстве способны быть замкнутыми, в связи с чем дополнены некоторые свойства. Механизм образования  $T_{\alpha\beta}$  из произвольной триангуляции и операции разреза и сшивки останутся неизменными.

Установлены предельные случаи замкнутой  $\alpha\beta$ -триангуляции: с одним  $\alpha$ -треугольником объект, показанный на рисунке 1а; с бесконечным числом  $\alpha$ -треугольников – на рисунке 7. Эмпирически выявлено соотношение числа  $\alpha$ -граней к  $\beta$ -граням, равное 16,67 %. Это свидетельствует о том, что при решении задачи аппроксимации замкнутых поверхностей свободной формы полиэдрами с группами конгруэнтных граней посредством  $\alpha\beta$ -триангуляции возможно задать, как минимум, полученный процент  $\alpha$ -треугольников конгруэнтными или группами конгруэнтных. Так как  $\alpha$ -треугольники напрямую независимы друг от друга, то эта математическая модель обеспечивает отсутствие потери подвижности триангуляции для механизмов оптимизации.

Интересным к дальнейшему изучению представляется обобщение свойства 7.1 для  $\alpha\beta$ -триангуляций рода, отличного от 0.

Таким образом, исследование позволило развить теорию  $\alpha\beta$ -триангуляции в трёхмерном пространстве, установить важную пропорцию между  $\alpha$ - и  $\beta$ -гранями, предложить эффективные механизмы рекомбинации  $\beta$ -ребер и определить направление дальнейших работ.

### Список литературы

1. Бешшар А. А., Клименко П. Я. Сравнение методов вычислительного проектирования в архитектуре // Молодой исследователь Дона. 2022. № 3(36). С. 31-36.
2. Акшов Э. А. Использование вычислительного проектирования и искусственного интеллекта при моделировании архитектурных объектов // Архитектура и современные информационные технологии. 2023. № 2(63). С. 298-315. DOI: 10.24412/1998-4839-2023-2-298-315.
3. Бобенко А. И., Сурис Ю. Б. Дискретная дифференциальная геометрия. Интегрируемая структура / пер. с англ. В. Э. Адлера. М.: Ижевск: Ин-т компьютерных исслед, 2010. 488 с. ISBN 978-5-93972-798-3.
4. K-set tilable surfaces / Fu C. W., Lai C. F., He Y., Cohen-Or D. // ACM Transactions on Graphics. 2010. Vol. 29, no. 4. P. 1-6. DOI: <https://doi.org/10.1145/1778765.1778781>.
5. Jimenez M., Müller C., Pottmann H. Discretizations of Surfaces with Constant Ratio of Principal Curvatures // Discrete & Computational Geometry. 2020. Vol. 63, no. 3. P. 670-704. DOI: 10.1007/s00454-019-00098-7.
6. Architectural freeform surfaces designed for cost-effective paneling through mold re-use / Pellis D., Kilian M., Wang H., et al. // Advances in Architectural Geometry. 2020. P. 2-16.
7. Tsiliakos M. Fast Geometry Optimizations for Architectural Workflows: A Case Study on the Computational Geometry of High-Rises // ACADIA 2024: Designing Change-Proceedings for the 2024 Association for Computer Aided Design in Architecture Conference. 2024. Vol. 1. P. 635-644.
8. Free-form Surface Approximation Using Rotational Patches / Liu Y., Xie Y. M., Lee T. U., et al. // ACM Transactions on Graphics. 2025. Vol. 44, no. 5. P. 1-14. DOI: 10.1145/3744707
9. Freeform honeycomb structures / Jiang C., Wang J., Wallner J., Pottmann H. // Computer Graphics Forum. 2014. Vol. 33, no. 5. P. 185-194.
10. Masonry shell structures with discrete equivalence classes / Chen R., Qiu P., Song P., et al. // ACM Transactions on Graphics. 2023. Vol. 42, no. 4. P. 1-12. DOI: 10.1145/3592095
11. Reducing the Number of Different Faces in Free-Form Surface Approximations Through Clustering and Optimization / Liu Y., Lee T.-U., Rezaee J. A., et al. // Computer-Aided Design. 2023. Vol. 166. P. 103633. DOI 10.1016/j.cad.2023.103633.
12. Singh M., Schaefer S. Triangle Surfaces with Discrete Equivalence Classes // ACM Transactions on Graphics. 2010. Vol. 29. DOI: 10.1145/1833351.1778783.
13. Lee T. U., Liu Y., Xie Y. M. Planar-thick panels and 3D-printed gap fillers: A hybrid digital fabrication approach to curved surface approximations // Composite Structures. 2024. Vol. 331. P. 117875. DOI: <https://doi.org/10.1016/j.compstruct.2024.117875>.



14. Bi M., Liu Y., Xie Y. M. Clustering and optimisation of nodes, beams and panels for cost-effective fabrication of free-form surfaces // *Proceedings of IASS Annual Symposia / International Association for Shell and Spatial Structures (IASS)*. 2023. No. 19. P. 1-12. <https://doi.org/10.1016/j.engstruct.2024.117912>.
15. Design and optimisation of grid shells for maximising both structural performance and panel similarity / Zhong Y., Liu Y., Lu H., et al. // *Engineering Structures*. 2025. Vol. 329. P. 119820. <https://doi.org/10.1016/j.engstruct.2025.119820>.
16. Modeling and fabrication with specified discrete equivalence classes / Liu Z., Zhang Z., Zhang D., et al. // *ACM Transactions on Graphics*. 2021. Vol. 40, no. 4. P. 1-12. <https://doi.org/10.1145/3450626.3459843>.
17. Свидетельство о государственной регистрации программы для ЭВМ № 2024612472 Российская Федерация. Программа создания триангулированных полиэдров, грани которых группируются по свойству конгруэнтности треугольников, аппроксимирующих замкнутые не самопересекающиеся поверхности свободной формы с помощью генетического алгоритма: № 2023685692; заявл. 24.11.2023; опублик. 01.02.2024 / В. В. Рустамян.
18. Скворцов А. В., Мирза Н. С. Алгоритмы построения и анализа триангуляции. Томск: Изд-во Томского университета, 2006. 168 с. ISBN 5-7511-2028-0.
19. Лебединская Н. А., Лебединский Д. М. Преобразование триангуляций при помощи элементарных операций // *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*. 2009. № 1. С. 84-86.
20. Рустамян В. В. Анализ основных параметров генетического алгоритма при решении задачи аппроксимации замкнутых поверхностей свободной формы полиэдрами с группами конгруэнтных треугольников // *Геометрия и графика*. 2024. Т. 12, № 2. С. 13-25. DOI: 10.12737/2308-4898-2024-12-3-13-25.
21. Рустамян В. В. Анализ топологии полиэдров в задаче аппроксимации замкнутых поверхностей полиэдрами с группами конгруэнтных граней // *GraphiCon 2024: материалы 34-й Международной конференции по компьютерной графике и машинному зрению*, Омск, 17–19 сентября 2024 года. Омск: Омский государственный технический университет. 2024. С. 827-836. DOI: 10.25206/978-5-8149-3873-2-2024-827-836.
22. Рустамян В. В.  $\alpha\beta$ -триангуляция на евклидовой плоскости // *Геометрия и графика*. 2025. Т. 13, № 1. С. 15-25. DOI: 10.12737/2308-4898-2025-13-1-15-25.
23. Александров П. С. Комбинаторная топология. 2-е изд. М.: Ленанд, 2020. 664 с. ISBN 978-5-9710-7311-6.
24. Sarhangi R. An introduction to medieval spherical geometry for artists and artisans // *Bridges London: Mathematics, Music, Art, Architecture, Culture*. 2006. P. 551-560.
25. Sarhangi R. Illustrating Abu al-Wafā' Būzjānī: Flat Images, Spherical Constructions // *Iranian Studies*. 2008. Vol. 41, no. 4. P. 511-523.

УДК 004.946

DOI: 10.25686/978-5-8158-2474-4-2025-930-933

## Динамическая персонализация интерфейсов с использованием эргономического анализа, закона Фиттса и обучения с подкреплением

С. Ченарани

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

**Аннотация.** В статье рассматривается подход к динамической персонализации пользовательских интерфейсов, основанный на интеграции методов эргономической оценки движений рук, расширенной модели закона Фиттса и алгоритмов обучения с подкреплением (Q-learning). Предложенный метод обеспечивает повышение эффективности и комфорта взаимодействия пользователя с интерфейсом за счёт автоматической адаптации геометрии и параметров расположения элементов меню. Экспериментальные результаты подтверждают существенные количественные улучшения в скорости выполнения задач, снижении ошибок и повышении субъективной удовлетворённости пользователей.

**Ключевые слова:** VR-интерфейсы, закон Фиттса, обучение с подкреплением, эргономика, персонализация

## Dynamic personalization of interfaces using ergonomic analysis, Fitts' law, and reinforcement learning

Chenarani Sasan

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

**Abstract.** The article presents an approach to dynamic personalization of user interfaces that integrates ergonomic hand-movement assessment, an extended model of Fitts's law, and reinforcement-learning algorithms (Q-learning). The proposed method enhances both efficiency and comfort of user interaction by automatically adapting the geometry and spatial layout of menu elements. Experimental results confirm notable quantitative improvements in task-completion speed, error rate reduction, and user satisfaction.

**Keywords:** VR interfaces, Fitts's law, reinforcement learning, ergonomics, personalization.

Современные интерфейсы виртуальной и дополненной реальности требуют высокой степени адаптивности для обеспечения удобства и эффективности взаимодействия [1-4]. Традиционные методы проектирования интерфейсов не всегда учитывают индивидуальные особенности пользователей, что приводит к росту времени выполнения задач, ошибкам и снижению комфорта. Настоящее исследование направлено на решение этой проблемы путём разработки алгоритмического комплекса, сочетающего эргономическую оценку движений рук, количественную валидацию через закон Фиттса [5] и динамическую персонализацию с использованием Q-learning [6].

Предлагаемая методология исследования основана на интеграции эргономического анализа движений рук, математической валидации закона Фиттса в условиях трёхмерного пространства и применения методов обучения с подкреплением для динамической персонализации интерфейса.

Для количественной оценки эргономичности движений была разработана методика нормализации векторов и вычисления коэффициентов комфорта. Пусть движение описывается вектором

$$v_i = (x_i, y_i, z_i), \quad i = 1, \dots, N,$$

где  $N$  – количество элементарных движений руки;  $(x_i, y_i, z_i)$  – координаты конечной позиции в трёхмерном пространстве.

Каждому движению сопоставляется коэффициент комфорта  $C_i$ , нормированный в диапазон  $[0; 1]$ . Значение  $C_i$  определяется как обратная функция от углового отклонения и длины вектора:

$$C_i = \frac{1}{1 + \alpha \cdot |v_i| + \beta \cdot \theta_i},$$

где  $|v_i|$  – длина вектора (амплитуда движения);  $\theta_i$  – угол отклонения от нейтрального положения руки;  $\alpha, \beta$  – весовые коэффициенты.

Индивидуальная метрика эргономичности движения пользователя вычисляется как среднее по всем движениям:

$$C_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N C_i.$$

Таким образом, методика позволяет объективно сравнивать различные сценарии расположения элементов интерфейса с точки зрения эргономики.

Для анализа временных характеристик взаимодействия в VR применялось расширение классического закона Фиттса на трёхмерное пространство. Классическая модель Фиттса задаётся уравнением

$$T = a + b \cdot \log_2 \left( 1 + \frac{D}{W} \right),$$

где  $T$  – время достижения цели;  $D$  – расстояние до цели;  $W$  – ширина цели;  $a, b$  – эмпирические коэффициенты.

В условиях трёхмерного взаимодействия необходимо учитывать не только расстояние в плоскости, но и глубину цели. Для этого вводится дополнительный параметр  $\sigma$ , учитывающий пространственный фактор:

$$T_{3D} = a + b \cdot \log_2 \left( 1 + \frac{D}{W} \right) + \sigma \cdot \phi,$$

где  $\phi$  – угол между направлением взгляда и вектором движения руки.

Соответствие модели фактическим данным проверялось с использованием коэффициента детерминации и анализа остаточных ошибок.

Для динамической адаптации интерфейса применялся алгоритм Q-learning. Пусть состояние среды описывается вектором параметров

$$s_t = (W_t, R_t, L_t, S_t),$$

где  $W_t$  – ширина элементов меню;  $R_t$  – радиус расположения;  $L_t$  – тип раскладки (круговая, кластерная, дуговая, силовая);  $S_t$  – форма элементов.

Действие  $a_t$  изменяет один или несколько параметров интерфейса. Обновление функции ценности  $Q(s, a)$  осуществляется по стандартному правилу:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[ r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right],$$

где  $\alpha$  – скорость обучения;  $\gamma$  – коэффициент дисконтирования;  $r_t$  – вознаграждение за текущее действие.

Функция вознаграждения агрегирует несколько показателей:

$$r_t = \sum_{k=1}^M w_k \cdot \widetilde{R}_k,$$

где  $\widetilde{R}_k$  – нормализованные метрики;  $w_k$  – веса важности.

Для учёта неопределённости в прогнозах GPR использовалась модифицированная стратегия выбора действий:

$$a^* = \arg \max_{a \in \mathcal{A}} (Q(s, a) + \beta \sigma_{\text{pred}}(s)),$$

где  $\sigma_{\text{pred}}(s)$  – оценка дисперсии предсказания модели;  $\beta$  – коэффициент баланса между эксплуатацией и исследованием.

Экспериментальная часть была проведена с участием 24 респондентов (12 мужчин и 12 женщин в возрасте 20–22 лет, студенты бакалавриата технических специальностей). Все участники имели нормальное или скорректированное зрение и не сообщали о нарушениях опорно-двигательного аппарата. Опыт работы с VR-системами варьировался: 8 новичков, 8 пользователей со средним опытом, 8 экспертов.

Среднее время выполнения задач сокращалось на протяжении первых серий эксперимента и стабилизировалось после 40–50 итераций. Было зафиксировано снижение на 22.2 % в среднем по всем пользователям при переходе от случайной раскладки к оптимизированной (рис. 1).

Ошибка клика ( $E_{\text{click}}$ ) оставалась в пределах допустимых значений и не увеличивалась при динамических изменениях интерфейса. Это подтверждает, что ускорение взаимодействия не сопровождалось потерей точности (рис. 2).

Эффективность траектории движения ( $\eta_{\text{path}}$ ) варьировалась в пределах 0,7–0,9, что указывает на преимущественно прямолинейные и экономичные движения. У экспертов показатель стабильно выше, у новичков наблюдался рост на протяжении серий (рис. 3).

Интегральная метрика вознаграждения ( $r$ ) демонстрировала положительную динамику. Колебания вокруг среднего уровня указывали на процесс адаптации системы, при этом в долгосрочной перспективе наблюдалось устойчивое повышение (рис. 4).

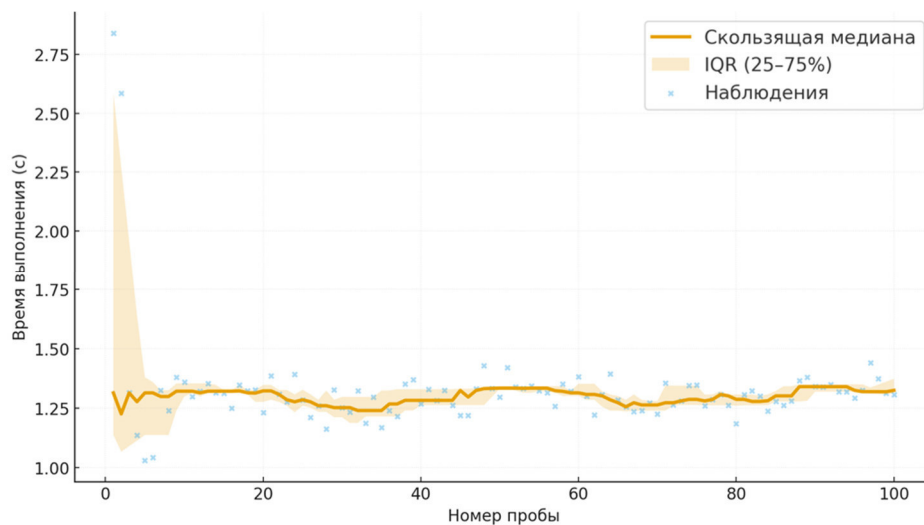


Рисунок 1. Динамика среднего времени выполнения задачи

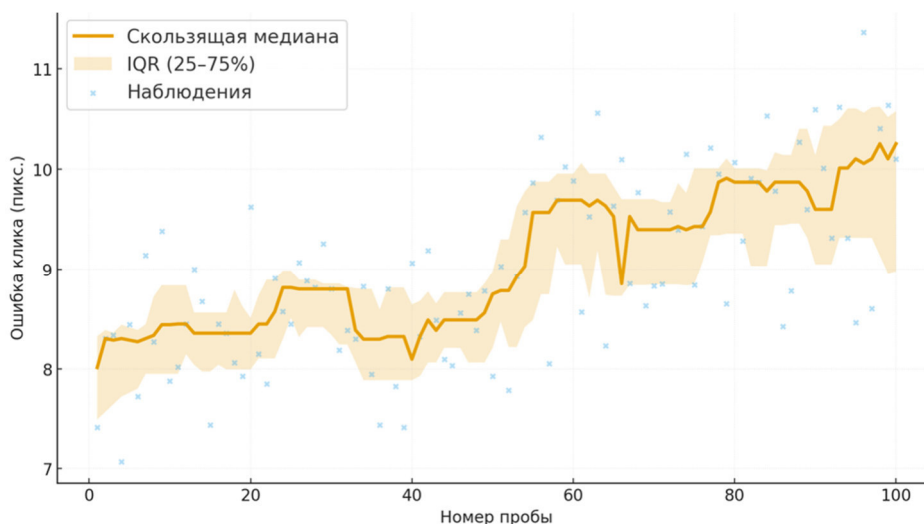


Рисунок 2. Динамика средней ошибки клика

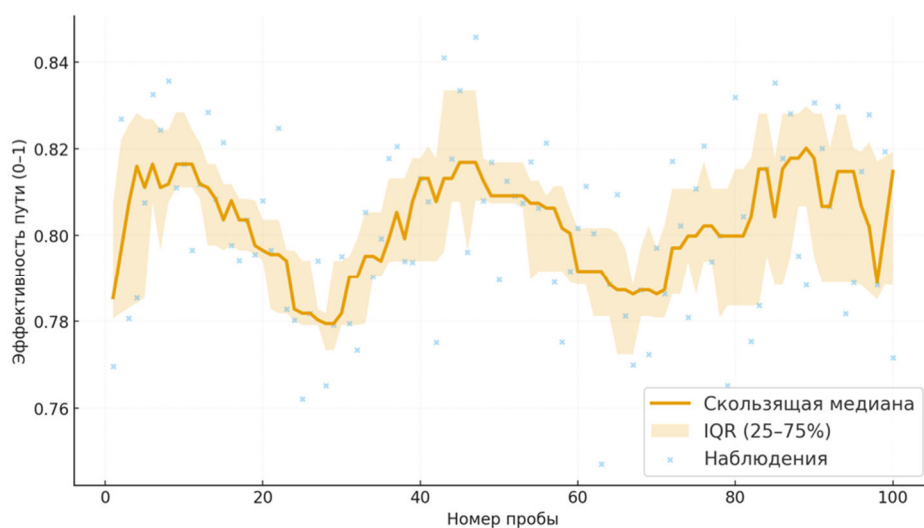


Рисунок 3. Динамика средней эффективности пути

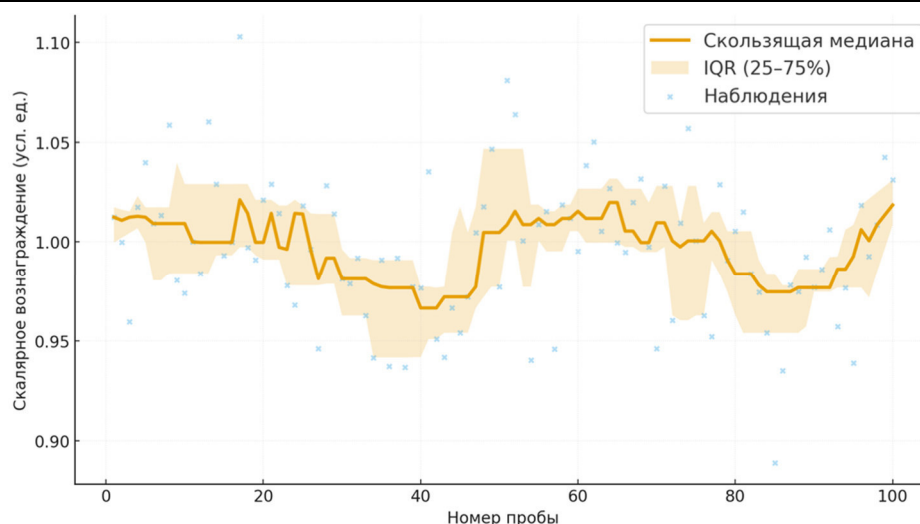


Рисунок 4. Динамика среднего скалярного вознаграждения

### Выводы

В работе предложен комплексный алгоритм динамической адаптации интерфейсов, включающий эргономическую оценку движений рук, расширение закона Фиттса на 3D-пространство и применение обучения с подкреплением. Экспериментальные исследования с участием 24 респондентов показали значительное улучшение временных характеристик (сокращение времени выполнения задач на 22.2 %), стабильность точности и рост субъективной удовлетворённости пользователей.

Разработанный подход можно рассматривать как эффективное средство повышения эргономичности и персонализации интерфейсов, применимых как в образовательных, так и в прикладных VR/AR-системах.

### Список литературы

1. Ченарани С., Щур, С. Ю. Исследование условий иммерсивности интеллектуального интерфейса симбиотической системы с применением VR-технологий // Неделя науки ИСИ: материалы Всероссийской научно-технической конференции. 2021. С. 331-333.
2. Ченарани С., Щур С. Исследование условий иммерсивности интеллектуального интерфейса симбиотической системы с применением VR-технологий // Неделя науки ИСИ / Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого». 2021. С. 331–333.
3. Ченарани С., Щур С. Разработка прототипа интеллектуального интерфейса в среде VR // Неделя науки ИСИ: сборник материалов Всероссийской конференции. 2022. С. 249–251.
4. Макейкова М.С. Создание виртуального пространства для учебных заведений: аннотация к дипломной работе / БГУ, ФРФКТ, Кафедра интеллектуальных систем. 2024.
5. Chakraborty P., Yadav S. Applicability of Fitts' law to interaction with touchscreen: review of experimental results // Theoretical Issues in Ergonomics Science. 2023. Vol. 24, no. 1. 5. Pp. 532–546. DOI: 10.1080/1463922X.2022.2114034.
6. Towards spatial computing: recent advances in multimodal natural interaction for XR headsets / Z. Wang, M. Rao, S. Ye, et al. // arXiv preprint. 2025.

УДК 004.925.8

DOI: 10.25686/978-5-8158-2474-4-2025-934-941

## Моделирование структуры «звезда Паскаля» с применением генеративного метода

С. Ю. Щур

Санкт-Петербургский политехнический университет Петра Великого,  
Санкт-Петербург, Россия

**Аннотация.** В работе рассмотрены вопросы применения декларативного программирования в создании геометрических структур, генеративный характер построения структуры «звезда Паскаля». Представлен алгоритм создания структуры с использованием инструментов логического программирования – предикатов. Результаты исследования могут лечь в основу создания новых алгоритмов и поиска решения фундаментальных задач геометрического моделирования. Описываемый метод является более наглядным и оптимизированным по сравнению с созданием структуры «вручную». Результаты работы могут лечь в основу решения задач в области многомерного проектирования с анализом сложных геометрических образов, а также поиском соответствия между имеющимися алгоритмами. Определена возможность генерации геометрических образов с помощью логических функций в программе «Симплекс».

**Ключевые слова:** геометрическое моделирование, генеративное проектирование, язык логического программирования, пролог, предикат

## Modeling of the Pascal's Star structure using the generative method

S. Yu. Shchur

Peter the Great St. Petersburg Polytechnic University,  
St. Petersburg, Russia

**Abstract.** The paper discusses the application of declarative programming in the creation of geometric structures. The generative nature of the construction of the Pascal's star structure is considered. An algorithm for creating a structure using predicate logic programming tools is presented. The results of the research can form the basis for the creation of new algorithms and the search for solutions to fundamental problems of geometric modeling. The described method is more visual and optimized than creating a structure "manually". The results of the work can form the basis for solving problems in the field of multidimensional design with the analysis of complex geometric patterns, as well as the search for a match between existing algorithms. The possibility of generating geometric images using logical functions in the Simplex program is determined.

**Keywords:** geometric modeling, generative design, logical programming language, prolog, predicate.

### Введение

В данной работе рассматриваются методы генеративного проектирования в создании структуры «звезда Паскаля». Оценка эффективности генеративной политики представлена в работе [1]. Идея применения методов генеративного проектирования изложена ранее [2] и заключается в широких возможностях генерации вариантов проектирования на основе заданных параметров и ограничений.

Для визуального анализа геометрических структур графоаналитические алгоритмы стали наиболее развитыми и удобными за последнее время [3].

Объект исследования – генеративное проектирование.

Предмет исследования – моделирование геометрических структур на основе методов генеративного проектирования.

Связь декларативного программирования и геометрического моделирования отображается в серии работ [4, 5], в частности описывается использование языка PROLOG для доказательства теорем евклидовой геометрии. В качестве основной программы была выбрана система «Симплекс» [6].

### Постановка задачи

**Целью работы** является описание алгоритма генеративного проектирования геометрической структуры «звезда Паскаля» при помощи инструмента декларативного программирования. Актуальность исследования основывается на недостатке программных средств автоматизировать процесс построения геометрических структур при решении задач подобного рода.

Одной из задач данного исследования является демонстрация возможностей программного обеспечения «Симплекс», в частности программного модуля «Логика». Методы генеративного проектирования обсуждались ранее. Возможности модуля «Логика», основанного на принципах декларативного программирования, также были обозначены ранее [2].

Возможности программного комплекса автоматизированного проектирования «Симплекс» в решении задач генеративного проектирования определены. Это позволяет рассуждать о поиске решения геометрических задач многомерного пространства. Работа базируется на использовании языка декларативного программирования «Пролог» [7].

### Теория

В качестве основы предлагается рассмотреть структуру «звезда Паскаля». Построение фигуры происходит следующим образом: в окружность или коническое сечение вписывается шестиугольник или определяются 6 точек A, B, C, D, E, F, принадлежащих данной конике  $f$ . Точки соединяются последовательно: AB, BC, CD, DE, EF, FA. В результате пересечения противоположных сторон получаем точки M, P, N (рис. 1).

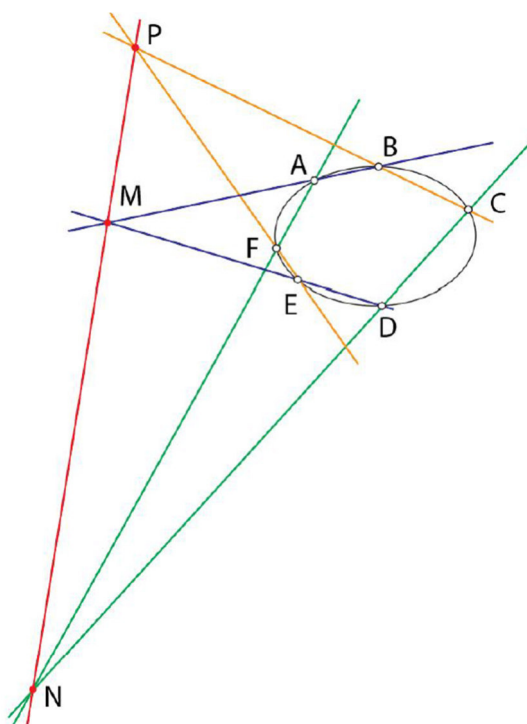


Рисунок 1. Звезда Паскаля

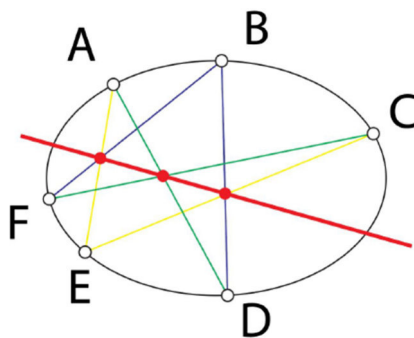


Рисунок 2. Структура «звезда Паскаля» с самопересекающимся контуром

Точки M, P и N по теореме Паскаля обязательно будут лежать на одной прямой - прямой Паскаля. Теорема остается верной даже при произвольном расположении точек на конике и самопересечении контура (рис. 2).



С точки зрения практического значения данные свойства структуры используются при решении задач проективной геометрии, в теории конических сечений, построении геометрических фигур и т.д. Свойства звезды Паскаля используются в улучшении точности классификации изображений, а также в процессе распознавания рукописного текста, а также обработки изображений с геометрическими фигурами.

Создание геометрических структур на основе декларативного программирования позволяет вести дальнейшую работу в рамках обнаружения объектов по геометрическим свойствам, классифицировать фигуры по их проективным характеристикам, а также проводить анализ симметрии и коллинеарности в изображениях.

В перспективе возможны разработка специализированных свёрточных нейронных сетей для работы с проективной геометрией, а также создание новых архитектур с учетом свойств конических сечений.

### Результаты экспериментов

В системе «Симплекс» была создана коника с расположенными на ней точками a, b, c, d, e, f (рис. 3).

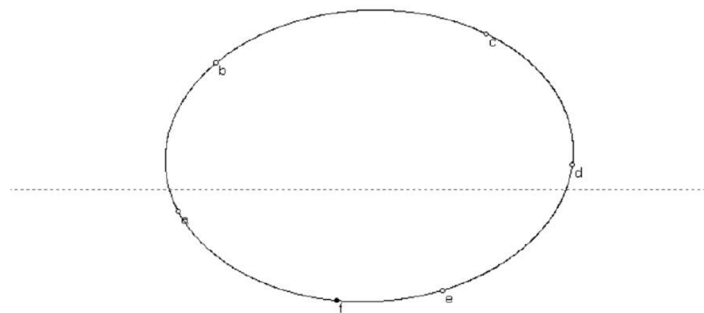


Рисунок 3. Геометрическая модель в Симплексе

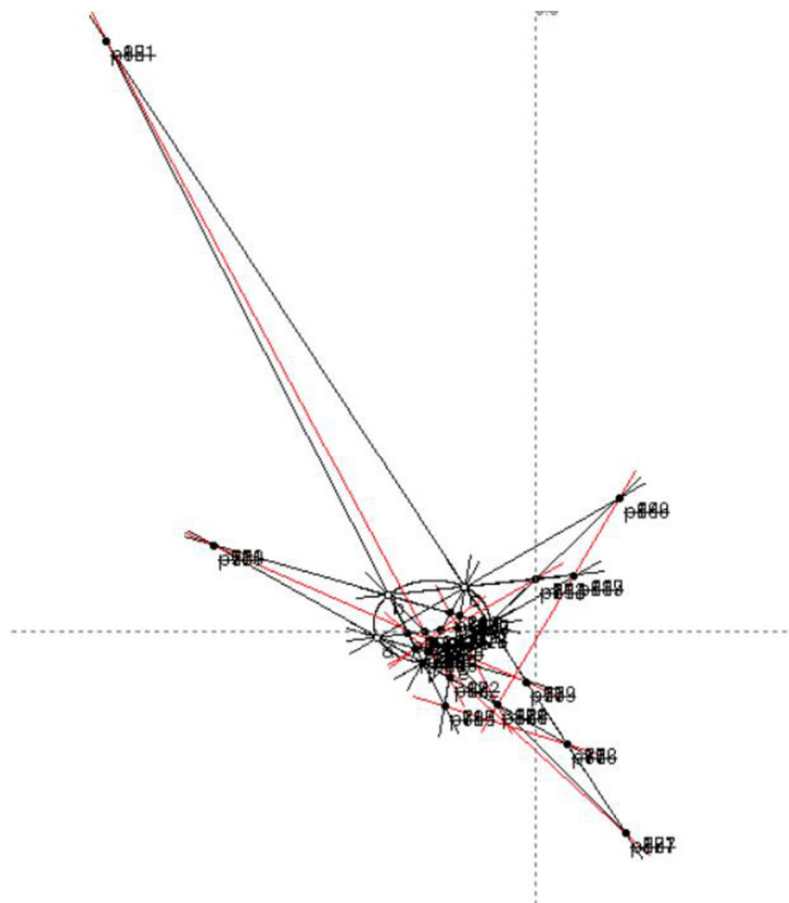


Рисунок 4. Общий вид результата генерации звезды в Симплексе

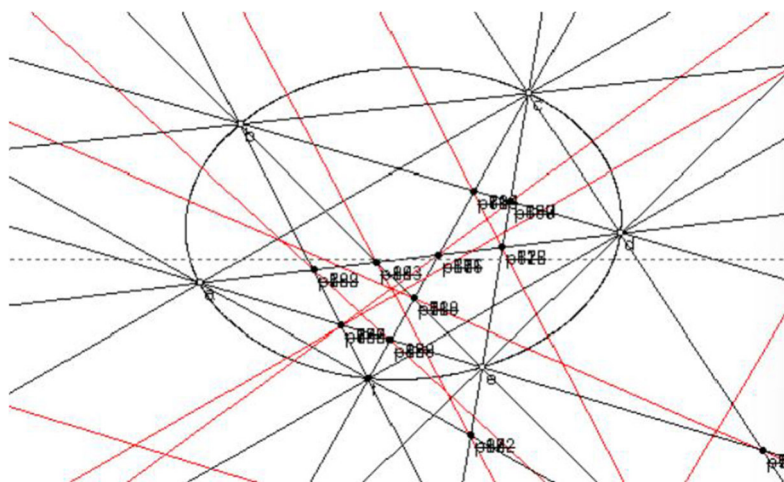


Рисунок 5. Результат генерации звезды в Симплексе

На рисунках 4, 5 представлена структура, полученная в результате генерации в Симплексе. Для изменения геометрического образа программа остается прежней. Программно задается изменение геометрии, без непосредственного влияния и доработки вручную.

Целью программы является определение `group(A,B,C,D,E,F),pascal(A,B,C,D,E,F),execute`. Последний в данном случае является оператором исполнения алгоритма.

Программа составлена следующим образом:

```
group(A,B,C,D,E,F) if
member(A,[a,b,c,d,e,f]),
member(B,[a,b,c,d,e,f]),
member(C,[a,b,c,d,e,f]),
member(D,[a,b,c,d,e,f]),
member(E,[a,b,c,d,e,f]),
member(F,[a,b,c,d,e,f]),
ne(A,B),
ne(A,C),
ne(A,D),
ne(A,E),
ne(A,F),
ne(B,C),
ne(B,D),
ne(B,E),
ne(B,F),
ne(C,D),
ne(C,E),
ne(C,F),
ne(D,E),
ne(D,F),
ne(E,F).
pascal(A,B,C,D,E,F) if
lt(A,E),
lt(A,D),
lt(B,F),
lt(B,D),
lt(C,F),
lt(C,E),
nnf(O1,'o','Главный'),
```

```

nnf(O2,'o','Главный'),
nnf(O3,'o','Главный'),
nnf(O4,'o','Главный'),
nnf(O5,'o','Главный'),
nnf(O6,'o','Главный'),
nnf(O7,'o','Главный'),
nnf(P1,'p','Главный'),
nnf(P2,'p','Главный'),
nnf(P3,'p','Главный'),
addline([[E], [A]], [O1], '0', 'Прямая задана двумя точками', 'Главный'),
addline([[D], [A]], [O2], '0', 'Прямая задана двумя точками', 'Главный'),
addline([[F], [B]], [O3], '0', 'Прямая задана двумя точками', 'Главный'),
addline([[D], [B]], [O4], '0', 'Прямая задана двумя точками', 'Главный'),
addline([[F], [C]], [O5], '0', 'Прямая задана двумя точками', 'Главный'),
addline([[E], [C]], [O6], '0', 'Прямая задана двумя точками', 'Главный'),
addline([[O1], [O3]], [P1], '0', 'Точка пересечения двух прямых', 'Главный'),
addline([[O2], [O5]], [P2], '0', 'Точка пересечения двух прямых', 'Главный'),
addline([[O4], [O6]], [P3], '0', 'Точка пересечения двух прямых', 'Главный'),
addline([[P1], [P2], [P3]], [O7], '0', 'Прямая задана триадой точек', 'Главный'),
attribute(O1,'Главный',[@alv(5)]),
attribute(O2,'Главный',[@alv(5)]),
attribute(O3,'Главный',[@alv(5)]),
attribute(O4,'Главный',[@alv(5)]),
attribute(O5,'Главный',[@alv(5)]),
attribute(O6,'Главный',[@alv(5)]),
attribute(O7,'Главный',[@alv(5)]),
attribute(O7,'Главный',[@acr(255)]).

```

В данной программе изначально задается группа  $\text{group}(A, B, C, D, E, F)$ , если выполняется ряд условий:  $\text{member}(A, [a, b, c, d, e, f]) \dots \text{member}(F, [a, b, c, d, e, f])$ . Являются предикатами проверки того, что A, B, C, D, E, F входят в состав списка, представленного в квадратных скобках.

Предикаты  $\text{ne}(A, B) \dots \text{ne}(E, F)$  являются предикатами проверки истинности утверждения, что аргументы, представленные в скобках, не унифицируемы.

$\text{pascal}(A, B, C, D, E, F)$  выполняется, если:

Выполняются предикаты  $\text{lt}(A, E) \dots \text{lt}(C, E)$ , которые служат для того, чтобы проверить, верно ли неравенство  $A < E \dots C < E$ .

Выполняются предикаты  $\text{nnf}(O1, 'o', \text{'Главный'}) \dots \text{nnf}(P3, 'p', \text{'Главный'})$ , которые определяют имя новой переменной по префиксу 'o' в алгоритме 'Главный'

Выполняются предикаты  $\text{addline}([E], [A], [O1], '0', \text{'Прямая задана двумя точками', 'Главный'}) \dots \text{addline}([E], [C], [O6], '0', \text{'Прямая задана двумя точками', 'Главный'})$  - добавляет отношения в алгоритм Симплекса, где  $[E], [A] \dots [E], [C]$  являются входными параметрами отношения;

$[O1] \dots [O6]$  являются выходными параметрами отношения;

'0' - тип согласования;

'Точка задана двумя точками' - функция;

'Главный' - алгоритм.

Предикаты  $\text{addline}([O1], [O3], [P1], '0', \text{'Точка пересечения двух прямых', 'Главный'}) \dots \text{addline}([O4], [O6], [P3], '0', \text{'Точка пересечения двух прямых', 'Главный'})$  добавляют отношения в алгоритм в виде точки пересечения прямых, заданных ранее.

Последний схожий предикат  $\text{addline}([P1], [P2], [P3], [O7], '0', \text{'Прямая задана триадой точек', 'Главный'})$  отвечает за отношение в виде прямой, определяемой найденными ранее тремя точками.

$\text{attribute}(O1, \text{'Главный'}, [@alv(5)]) \dots \text{attribute}(O7, \text{'Главный'}, [@acr(255)])$  назначает атрибуты объектам списочной переменной Симплекса.

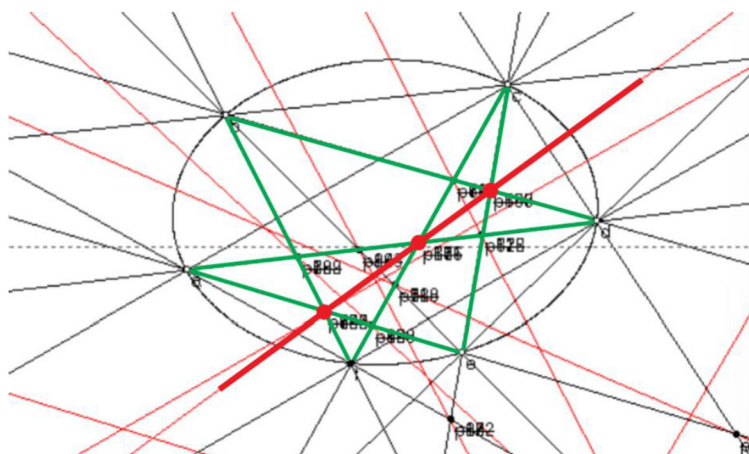


Рисунок 6. Геометрическая модель в Симплексе

В результате с помощью модуля «Логика» в Симплексе методом генеративного перебора вариантов соединения точек удалось получить различные варианты структуры «звезда Паскаля».

В таблице 1 представлены результат сгенерированных вариантов возможных структур, отвечающих запросу, реализованному в программе.

Таблица 1. Результат генерации структур

#	A	B	C	D	E	F
1	a	b	c	d	e	f
2	a	b	c	d	f	e
3	a	b	c	e	d	f
4	a	b	c	e	f	d
5	a	b	c	f	d	e
6	a	b	c	f	e	d
7	a	b	d	c	e	f
8	a	b	d	c	f	e
9	a	c	b	d	e	f
10	a	c	b	d	f	e
11	a	c	b	e	d	f
12	a	c	b	e	f	d
13	a	c	b	f	d	e
14	a	c	b	f	e	d
15	a	d	b	e	c	f
16	a	d	b	f	c	e
17	b	a	c	d	e	f
18	b	a	c	d	f	e
19	b	a	c	e	d	f
20	b	a	c	e	f	d
21	b	a	c	f	d	e
22	b	a	c	f	e	d
23	b	a	d	c	e	f
24	b	a	d	c	f	e
25	b	c	a	d	e	f
26	b	c	a	d	f	e
27	b	c	a	e	d	f
28	b	c	a	e	f	d
29	b	c	a	f	d	e
30	b	c	a	f	e	d
31	b	d	a	e	c	f
32	b	d	a	f	c	e
33	c	a	b	d	e	f
34	c	a	b	d	f	e
35	c	a	b	e	d	f

Окончание таблицы 1

#	A	B	C	D	E	F
36	c	a	b	e	f	d
37	c	a	b	f	d	e
38	c	a	b	f	e	d
39	c	b	a	d	e	f
40	c	b	a	d	f	e
41	c	b	a	e	d	f
42	c	b	a	e	f	d
43	c	b	a	f	d	e
44	c	b	a	f	e	d
45	d	a	b	e	f	c
46	d	a	b	f	e	c
47	d	b	a	e	f	c
48	d	b	a	f	e	c

### Обсуждение

Обоснование использования методов генеративного проектирования в вопросах геометрического моделирования лежит прежде всего в плоскости автоматизации проектирования. Однако этим оно не ограничивается. В работе представлены новые подходы в решении геометрических задач. Методы генеративного проектирования интересны прежде всего тем, что они могут подсказать новые пути решения уже устоявшихся фундаментальных задач.

Декларативный характер дает преимущества прежде всего в упрощении пользовательского взаимодействия с геометрической машиной. Работа является ярким примером того, как можно наглядно создать чертеж и динамически его редактировать, внося в него изменения без непосредственного внесения изменений в геометрию структур.

Помимо этого декларативное программирование позволяет создать на базе отлаженной системы некую экспертную систему, которая будет оценивать результаты по критериям. Это позволит сопоставлять их, находить взаимосвязи и отличия в структурах, перебирать множество вариантов для решения определенной задачи.

### Выводы

В результате полученной структуры удалось определить, что решение геометрических задач и построение модели проходит быстрее с использованием методов генеративного проектирования. Язык «Пролог» и основанный на нем модуль «Логика» в Симплексе являются мощной основой для использования в решении подобных задач. Сама структура логического программирования базируется на переборе возможных вариантов и определении правильных, подходящих к искомой форме.

Область геометрического моделирования не исчерпывается решением одной задачи. Данный метод может быть использован в дальнейшем в решении задач более сложных, например, в области многомерного проектирования. Он также позволяет анализировать сложные геометрические образы, находить соответствия между имеющимися алгоритмами. В работе определена возможность генерации геометрических образов с помощью логических функций в программе «Симплекс».

### Благодарность

Автор статьи благодарит научного руководителя Волошинова Дениса Вячеславовича, СПбГУТ им. Бонч-Бруевича, СПбПУ Петра Великого за помощь в подготовке работы.

### Список литературы

1. Алексеев В. Ф., Лихачевский Д. В., Пискун Г. А. Оценка эффективности генеративной политики при оптимизации проектных решений. 2025.
2. Щур С. Ю. Динамическое редактирование геометрических образов с использованием логических инструментов в программной среде Симплекс // GraphiCon 2024 : материалы 34-й Международной конференции по компьютерной графике и машинному зрению, Омск, 17–19 сентября 2024 года. Омск: Омский

государственный технический университет, 2024. С. 837-842. DOI: 10.25206/978-5-8149-3873-2- 2024-837-842. EDN FXJSXX.

3. Волошинов Д. В. Теория автоматизации проектирования объектов и процессов на основе методов конструктивного геометрического моделирования. СПб., 2010.

4. Coelho H., Pereira L. M. Automated reasoning in geometry theorem proving with Prolog // Journal of Automated Reasoning. 1986. Vol. 2. P. 329-390.

5. Extracting isovolumes from three-dimensional torso geometry using PROLOG / Replogle J. A., et al. // IEEE transactions on information technology in biomedicine. 1998. Vol. 2, no. 1. P. 10-19.

6. Волошинов Д. В., Соловьева А. В. Разработка алгоритмического комплекса для решения задач конструктивной геометрии // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2020) : сборник научных статей IX Международной научно-технической и научно-методической конференции: в 4 т., Санкт-Петербург, 26–27 февраля 2020 г. Том 2. Санкт-Петербург: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2020. С. 182-186. EDN: VRYFMK.

7. Bratko I. Prolog programming for artificial intelligence. Pearson education, 2001.

УДК 004.921

DOI: 10.25686/978-5-8158-2474-4-2025-942-944

## Алгоритм построения рельефа с учётом затенения функционально-воксельным методом для моделирования зон обнаружения радиолокационных станций

А. Б. Яров<sup>1</sup>, А. В. Толлок<sup>2</sup>

<sup>1</sup>Московский государственный технологический университет «СТАНКИН», Россия, Москва.

<sup>2</sup>Институт проблем управления им. В.А. Трапезникова РАН, Россия, Москва

**Аннотация.** При моделировании зон обнаружения радиолокационных станций функционально-воксельным методом с учётом рельефа невозможно учитывать пространство, находящееся вне доступа для электромагнитных лучей за экранированными областями модели. Применение R-функций в рамках функционально-воксельного метода не учитывает области, находящиеся за препятствиями. Решением является применение специального алгоритма, который основывается на существующих программах трассировки лучей для воксельных сцен. Создаётся массив, который и заполняется данными существующего рельефа. При помощи метода трассировки лучей (Voxel Ray Traversal) алгоритм определяет необходимую территорию, которую нужно идентифицировать на существующей функционально-воксельной модели рельефа посредством изменения её локальных функций в найденных точках. Такое решение позволяет учесть углы затенения уже существующих моделей на сложных рельефах и использовать модель зоны обнаружения при решении различных задач радиолокации.

**Ключевые слова:** функционально-воксельный метод, моделирование РЛС, метод бросания лучей, трассировка лучей.

## An algorithm for constructing terrain with shadowing using the functional-voxel method for radar detection zone modeling

A. B. Yarov<sup>1</sup>, A.V. Tolok<sup>2</sup>

<sup>1</sup>Moscow State Technological University 'STANKIN', Russia, Moscow

<sup>2</sup>V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Russia, Moscow

**Abstract.** When modeling radar detection zones using the functional-voxel method with terrain, it is impossible to account for the space beyond the shielded areas of the model that is inaccessible to electromagnetic rays. The use of R-functions within the functional-voxel method does not account for areas located behind obstacles. The solution is a specialized algorithm based on existing ray tracing programs for voxel scenes. An array is created and populated with existing terrain data. Using ray tracing (Voxel Ray Traversal), the algorithm determines the required area to be identified on the existing functional-voxel terrain model by modifying its local functions at the determined points. This solution allows for the shading angles of existing models on complex terrain and the use of the detection zone model in solving various radar problems.

**Keywords:** Functional-voxel method, radar modeling, ray casting, ray tracing.

### Введение

При построении зон обнаружения радиолокационных станций функционально-воксельным методом с заданным рельефом невозможно учитывать углы закрытия этой модели. Применение R-функций в рамках функционально-воксельного метода не учитывает области, находящиеся за препятствиями. В связи с этим возникает необходимость в разработке специализированного алгоритма, способного находить это затенение при построении трёхмерных зон видимости функционально-воксельным методом.

### Постановка задачи

Актуальной задачей в области моделирования радиолокационного наблюдения является точное определение трёхмерных зон видимости с учётом влияния рельефа местности. Такое ограничение не позволяет точно смоделировать зону обнаружения радиолокационных систем в сложно пересечённой местности.

### Теория

В основе разработанного алгоритма трассировки лучей лежат современные подходы компьютерной графики, которые могут подходить для задач функционально-воксельного моделирования. Базовыми



из них являются бросание лучей (Ray Casting), пошаговая трассировка (Ray Marching) с использованием функций знаковых расстояний (SDF), воксельная трассировка (Voxel Ray Traversal, DDA).

Так как для моделирования используется воксельная сетка, то целесообразно для каждого луча применять алгоритм DDA, который часто используют в воксельных моделях. Метод определяет последовательность вокселей, через которые проходит луч, обеспечивая полный охват траектории без пропусков.

Прокладка лучей в модели происходит при помощи создания трёхмерного массива сцены  $S(i,j,k)$ , в который записывается информация о рельефе.

Для заполнения данного массива используются локальные функции, получаемые в результате считывания нормалей заданной поверхности рельефа в форме М-образов. Сетка массива заполняется в каждой точке, где локальная функция  $n1x+n2y+n3z+n4 \geq 0$ . Такое условие позволяет точно определить наличие рельефа в каждой точке массива для последующего анализа трассировки лучей в моделируемой сцене.

В месте расположения радиолокационной станции инициируется набор лучей, направленных ко всем точкам всех границ сцены. Таким образом обеспечивается полное покрытие пространства за счёт перебора всех возможных направлений распространения электромагнитных волн.

На каждом шаге трассировки выполняется проверка текущего вокселя по предварительно сформированному массиву  $S(i,j,k)$ . Обнаружение препятствия (рельефа) интерпретируется как факт экранирования всех последующих вокселей вдоль данного луча. Эти воксели корректируют существующие М-образы рельефа, добавляя к ней теневую составляющую.

### Результаты экспериментов

Результатом обработки всех лучей является обновлённый трёхмерный воксельный рельеф в виде М-образов для каждого слоя объекта. Этот рельеф является обновлённым V-представлением модели (рис. 1), в котором локальные функции в теневых местах дополняются данными о препятствии.

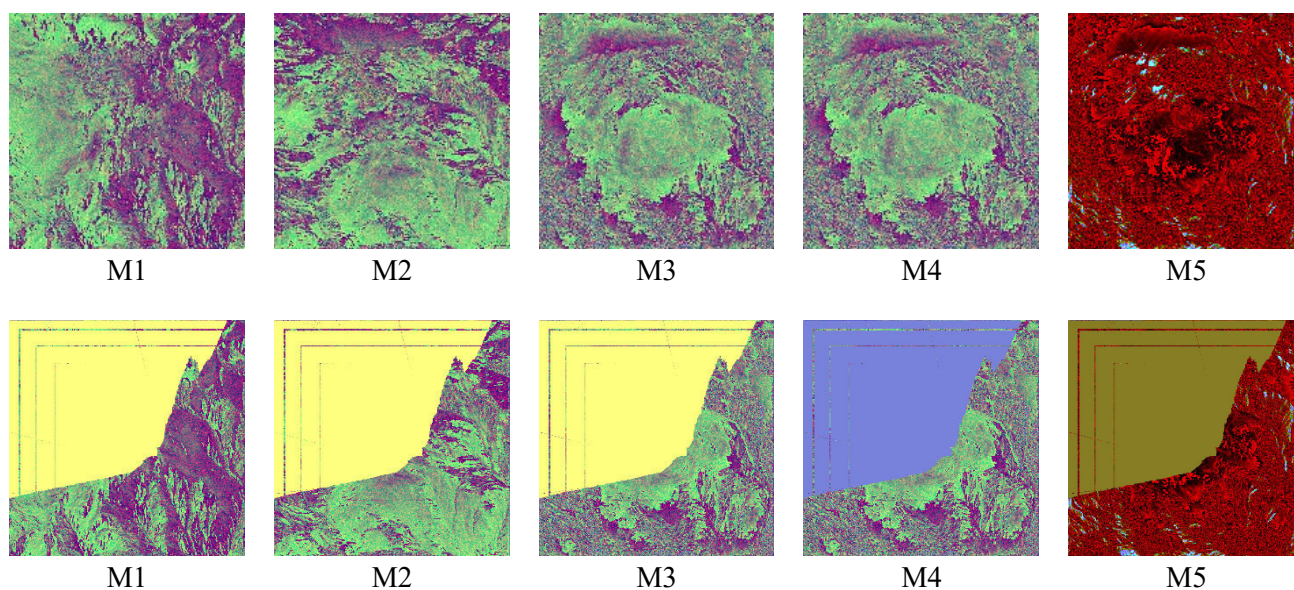


Рисунок 1. Результат работы алгоритма в одном из сечении, где демонстрируется отличие М-образов до и после его использования.

### Обсуждение результатов

Визуальный анализ результатов демонстрирует корректное построение теневых зон за препятствиями. На М-образях рельефа (рис. 1) чётко идентифицируются области, соответствующие лучам, остановленным преградами (M1-M3 – жёлтый цвет, M4 – голубой цвет, M5 – коричневый цвет).

Важным наблюдением является алиасинг (ступенчатость), который является следствием дискретной природы воксельного представления и применяемого алгоритма трассировки. Данный артефакт может быть существенно уменьшен с применением методов сглаживания компьютерной графики.

## Выводы

Полученная модель позволяет откорректировать зону обнаружения в условиях сложно пересечённой местности при использовании теоретико-множественных операций R-функционального моделирования, так как применяемый алгоритм позволяет учесть экранированную от лучей область. Дальнейшие исследования предлагается направить на оптимизацию производительности алгоритма, устранение создаваемых артефактов и внедрение его для функционально-воксельной модели.

## Список литературы

1. Толок А.В. Функционально-воксельный метод в компьютерном моделировании / под ред. академика РАН С.Н. Васильева. М.: Физматлит, 2016. 112 с.
2. Толок А.В. Графические образы-модели в информационных технологиях // Прикладная информатика. 2009. № 4. С. 31-40.
3. Толок А.В., Толок Н.Б. Построение функционально-воксельной модели рельефа методом билинейной интерполяции триангулированной сетки // Труды 13-го Всероссийского совещания по проблемам управления (ВСПУ XIII, Москва, 2019). М.: ИПУ РАН, 2019. С. 3191-3196.
4. R-функции и обратная задача аналитической геометрии в трехмерном пространстве / Максименко-Шейко К.В., Мацевитый А.М., Толок А.В., Шейко Т.И. // Информационные технологии. 2007. № 10. С. 23-32.
5. Толок А.В. Функционально-воксельный метод в компьютерном моделировании. М.: Физматлит, 2016. 112 с.
6. Рвачев В.Л. Теория R-функций и некоторые ее приложения. Киев: Наукова думка, 1982. 552 с.
7. Теоретические основы радиолокации / под ред. Я. Д. Ширмана. Москва: Советское радио, 1970. 560 с.

## Обоснование эффективности разработки игровых локаций с помощью процедурной генерации

В. Б. Головкина, М. А. Бобров

Национальный исследовательский технологический университет «МИСИС», г. Москва, Россия

**Аннотация.** Популярность компьютерных игровых продуктов обусловлена возможностью погружать пользователя в виртуальный игровой мир, воплощая эффект полного присутствия. Важнейшим элементом большинства компьютерных игр являются трехмерные локации. Возрастающие требования к визуальной составляющей оказывают непосредственное влияние на разработку игрового места, приводящее к увеличению времени и человеческих ресурсов, а следовательно, стоимости готового продукта. Решением проблемы является использование процедурно генерируемой составляющей, позволяющей создавать детализированные виртуальные миры без необходимости ручной разработки каждого элемента, затрачивая основной промежуток времени на разработку условий и функций генерации. В итоге возможность изменения процедурных объектов в реальном времени влияет на сокращение человеческих и временных ресурсов

**Ключевые слова:** IT-технологии, локация, цифровая модель, текстура, процедурная генерация, нодовая система, Side FX Houdini, Blender 3D.

## Justification of the effectiveness of developing game locations using procedural generation

V. B. Golovkina, M. A. Bobrov

National University of Science and Technology MISIS, Moscow, Russia

**Abstract.** The popularity of computer game products is due to the ability to immerse the user in a virtual game world, creating a full-presence effect. The most important element of most computer games is the 3D location. The increasing requirements for the visual component have a direct impact on the development of the game environment, leading to an increase in time and human resources, and consequently, the cost of the finished product. The solution to this problem is the use of a procedurally generated component, which allows for the creation of detailed virtual worlds without the need for manual development of each element, focusing instead on the development of generation conditions and functions. As a result, the ability to change procedural objects in real time reduces human and time resources.

**Keywords:** IT technologies, location, digital model, texture, simulations, procedural generation, node system, Side FX Houdini, Blender 3D.

### Введение

В связи с развитием цифровых программных инструментов область 3D-графики в игровой индустрии постоянно эволюционирует. Согласно отчету «Newzoo», в 2023 году мировой рынок компьютерных игровых продуктов достиг объема в 200 миллиардов долларов. Рост популярности игровых прямых эфиров на интернет-сервисах и сервисах киберспорта привлекает новых потребителей и способствует развитию игрового сообщества. В отчете компании DFC Intelligence отмечается, что сегодня в мире насчитывается более 3,7 миллиарда пользователей компьютерных игровых продуктов [1].

В региональном разрезе самым крупным рынком остается Азиатско-Тихоокеанский регион. Основными странами, развивающими гейм-индустрию, являются Китай и Япония. На втором месте находится Северная Америка, где основным участником рынка выступает США. В свою очередь, Восточная Европа относится к слабо развитым компьютерным игровым рынкам [2].

Что касается Российской Федерации, то процесс создания компьютерных игр, успешно преодолевая возникающие трудности, демонстрирует устойчивый подъем. На фоне поддержки ИТ-отрасли государством отечественные разработки постепенно заменяют ушедшие с рынка программы, а крупные игроки объединяются в ассоциации для поддержания отрасли [3].

Важнейшим элементом компьютерной игры является игровая локация, занимающая основную часть экрана, которую пользователь видит большую часть игрового времени и благодаря чему передается визуальный стиль и атмосфера продукта [4, 5].

Игровая локация – это определенная местность, в которой проходят разные этапы игры. Во время создания игрового мира для простоты разработки его делят на несколько локаций – определенных зон, в которых будут разворачиваться игровые события. В то же время каждая локация представляет собой изолированный участок игрового мира [6].

Современные игровые пространства отличаются достаточно высоким качеством проработки деталей и достоверностью фона, будь то морское дно, улица города, заснеженная горная вершина или тайная пещера. В то же время авторы публикации [7] отмечают проблемы, касающиеся реалистичности локаций, например рельефа местности, природных элементов, интеграции различных объектов и пр. Здесь же возникают вопросы обеспечения высокой производительности при использовании сложной графики и значительной детализации. С одной стороны, современные компьютеры становятся все мощнее, с другой – возрастают требования к качеству графики и производительности игр.

Определенную техническую сложность с затратой дополнительного времени и усилий представляет инвариантность локаций, что непосредственно влияет на повышение стоимости разработки.

В данной работе предпринята попытка выявить преимущества использования процедурной генерации на примере создания локации в жанре «стратегия».

### **Постановка задачи**

Процедурная генерация контента (Procedural Content Generation) в игровой индустрии не является новым явлением [8, 9]. Её применение способствует сокращению времени на создание продукта, так как основной промежуток затрачивается на разработку условий и функций генерации. В итоге возможность изменения процедурных объектов в реальном времени влияет на сокращение ресурсов и исключает необходимость создания элементов с нуля или трудоемкого изменения существующих.

Гипотеза позволяет утверждать, что применение процедурной генерации на примере создания игровой локации промышленной зоны в жанре «стратегия» способствует оптимизации процесса разработки и существенно влияет на сокращение времени по сравнению с традиционным методом, что положительно сказывается на финансовых затратах.

### **Теория**

Упрощенно существует два типа игровых локаций: экстерьерные, открытые пространства, такие как поле, лес, город, пляжи, и интерьерные, в качестве которых могут выступать комнаты и коридоры зданий, пещеры, тоннели метро [6].

С учетом разнообразия игровых продуктов трехмерные игровые локации создаются под тип игровых миров, имеющих подразделения: открытые миры, как в данной работе, полуоткрытые и линейные [10].

Бриф включает ряд требований на разработку продукта: жанром является стратегия в реальном времени. Трехмерная локация имеет симметричную компоновку с разделением по центру диагональю, на которой находятся два прохода, являющиеся точками интереса. Визуальный стиль локации должен отображать промышленную зону. Диагональ, расположенная в центре локации, разделяет цветовой тон архитектуры на сторону с холодными и сторону с теплыми цветами. Сама диагональ представляет собой канал, при разработке которого рекомендовано опираться на местность под названием «Los Angeles River». Для производительности предпочтительна условно-реалистичная стилизация, а также слабая детализация, учитывая вид сверху и планируемую возможность запустить компьютерный игровой продукт на современных бюджетных игровых компьютерах. Выбран формат трехмерных моделей fbx и abc с двумерными текстурными картами в формате png. Проект разрабатывается с использованием игрового движка Unreal Engine 5.

Разработка компьютерных игровых продуктов требует соблюдения ряда предписаний, содержащихся в нормативных документах. Так, в статье 1240 ГК РФ «Использование результата интеллектуальной деятельности в составе сложного объекта» указано на необходимость соблюдения авторского права по отношению к разработке и внедрению трехмерных и двумерных компьютерных объектов в трехмерных игровых локациях [11].

Компьютерные игровые продукты, в том числе локации, попадают под действие Федерального закона «О защите детей от информации, причиняющей вред их здоровью и развитию» (ФЗ-436) [12, 13].

На составляющую компьютерных игровых продуктов влияет положение Федерального закона «О противодействии экстремистской деятельности» (ФЗ-114) [12, 14].

Разрабатываемые локации должны содержать местность для конкретных целей и задач, связанных с организацией процесса разработки компьютерных игр. При разработке необходимо проектировать пространство и навигацию с учетом использования шаблонного дизайна частей локации для контролирования поведения игрока и его решений. Также не рекомендуется использовать высокую детализацию, учитывая подразумеваемый вид сверху, и использовать levels of detail для избежания высоких аппаратных нагрузок [15,16].

На рисунке 1 изображена блок-схема, созданная про помощи модели AS-IS, в нотации eEPC (событийной цепочки процессов), демонстрирующая процесс разработки трехмерной игровой локации. Целью моделирования является выявление возможных недостатков процесса разработки на этапе планирования.

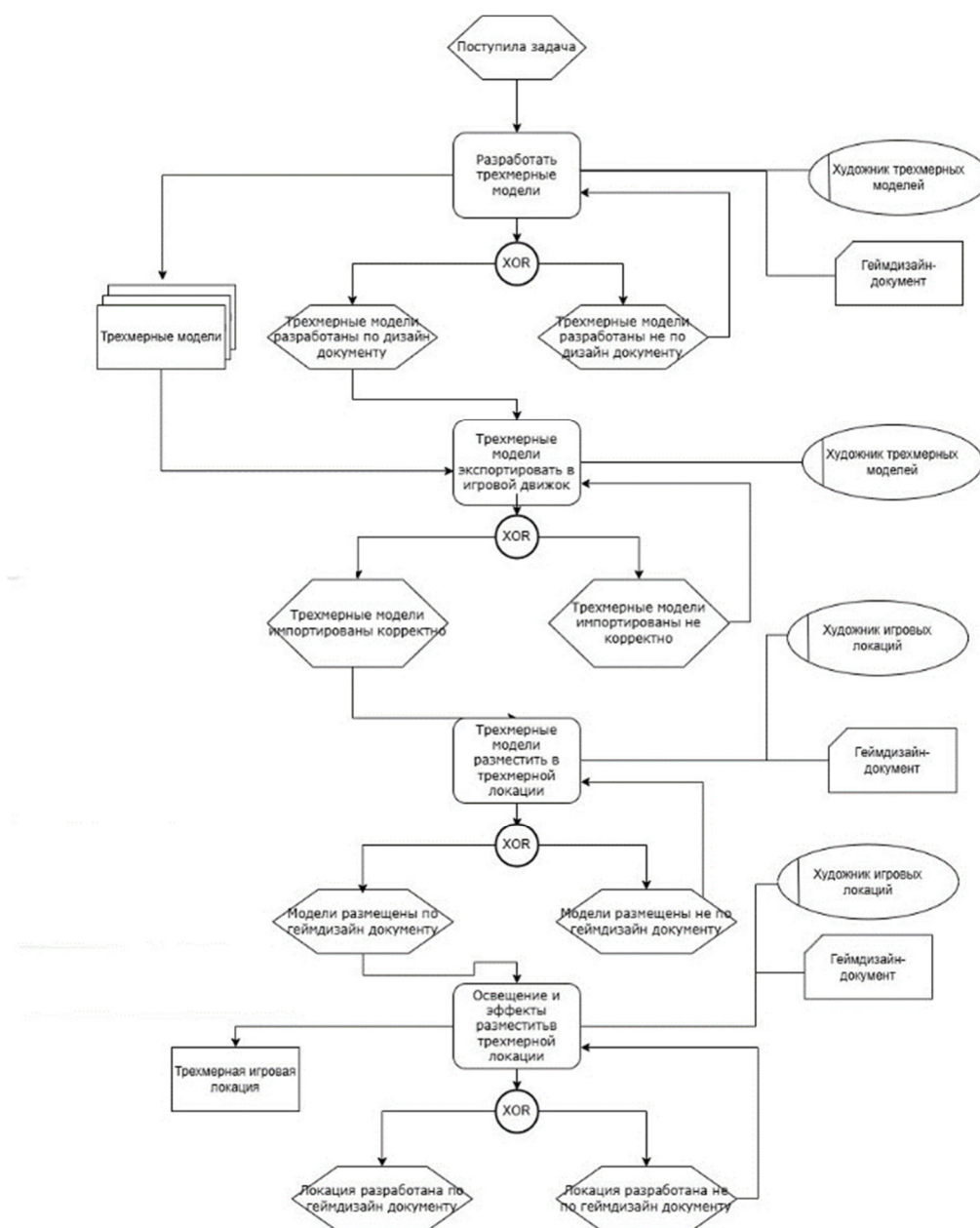


Рисунок 1. Блок-схема процесса разработки трехмерной игровой локации

Аналогичным образом рассмотрен процесс разработки трехмерных моделей, состоящий из следующих этапов: разработка низкополигональной и высокополигональной моделей для дальнейшего создания карт нормалей и текстурирования, разработки UV-развертки.

Референсы зданий промзоны сгенерированы программой искусственного интеллекта «Kandinsky». Всего получено десять изображений. В качестве примера на рисунке 2а представлено изображение,



сгенерированное по промту «складское здание, концепт-арт, вид издали сверху», а на рисунке 2б – по промту «завод, красный кирпич, аэрофотоснимок».



Рисунок 2. Изображения, созданные искусственным интеллектом: а – здание склада, б – здание завода

Работа с искусственным интеллектом позволила получить необходимую информацию о материалах, ландшафте, расположении светового потока и физических особенностях объектов.

Этап разработки трехмерных моделей реализован с применением программного обеспечения (ПО) Blender 3D, так как этот графический редактор может портить файлы проектов в любом из возможных расширений.

В качестве основного программного обеспечения для работы с процедурной составляющей, использующей нодовые системы, языки VEX и Python, выбрано Side FX Houdini, которое имеет связь с игровым движком Unreal Engine 5, в котором производится демонстрация финальной версии продукта.

Разработка первой итерации локации состоит из обычных примитивов, из-за чего процесс часто называют блокинг, grayboxing и whiteboxing, однако whiteboxing часто рассматривается с позиции размещения силуэтов еще не готовых моделей. В основном это происходит для определения габаритов объектов и теста геймплея – игровой и функциональной составляющей [17, 18]. Составные трехмерные модели разделены на три категории зданий: склад, логистический центр, завод. Каждая категория содержит составные трехмерные модели для генерации отдельно первого этажа здания и последующих этажей.

Для генерации здания склада разработаны четыре составные трехмерные модели: стена второго и последующих этажей, стена первого этажа с окнами, стена первого этажа с дверями, стена первого этажа с воротами для транспорта.

Для генерации здания логистического центра разработано пять составных трехмерных моделей: стена второго и последующих этажей, стена первого этажа, стена первого этажа с окнами, стена первого этажа с дверями, стена первого этажа с разгрузочно-загрузочными воротами для транспорта. Вышеупомянутые составные трехмерные модели приведены на рисунке 3 в соответствующем порядке.

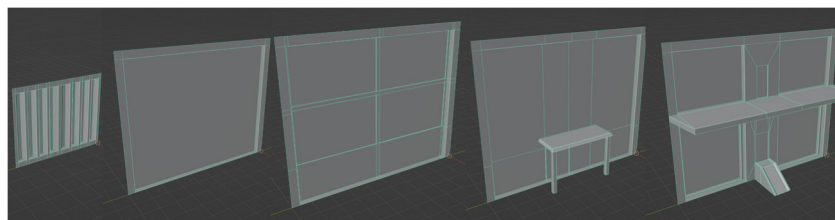


Рисунок 3. Составные трехмерные модели для процедурной генерации логистического центра

Для генерации здания завода разработано восемь составных трехмерных моделей: стена второго и последующих этажей основной части здания, стена с окнами второго и последующих этажей основной части здания, стена первого этажа основной части здания, стена первого этажа с окнами основной части здания, стена первого этажа с дверями основной части здания, стена первого этажа с разгрузочно-загрузочными воротами для транспорта, стена верхней части здания, стена с окнами верхней части здания.

Программным обеспечением для разработки текстурных карт трехмерных объектов являются Adobe Substance 3D Painter и Adobe Substance 3D Designer.

Разработаны текстурные карты для составных трехмерных моделей для процедурной генерации с использованием программного обеспечения Substance Painter. Текстурные карты для составных трехмерных моделей здания склада используют теплые цветовые оттенки, для составных трехмерных моделей логистического центра – холодные цветовые оттенки. Для составных моделей здания завода разработаны две версии с теплыми и холодными цветовыми оттенками.

Для здания склада основным материалом является металл. В то время как стены имеют просто металлический цвет, каркас стен имеет оранжевый оттенок. Двери и проемы окон имитируют пластик. Ворота и дверь в воротах являются более темным вариантом оранжевого оттенка каркаса.

У здания логистического центра схожая ситуация, где основным материалом является металл. Каркас стен имеет синий оттенок. Проемы окон имитируют черный пластик. Дверь является менее насыщенным синим оттенком каркаса. Навесы имеют черный шероховатый материал.

Для здания завода разработано два набора схожих текстурных карт с единственным различием в цвете имитации кирпича, где в теплых тонах он является ярко-оранжевым и темно-зеленым в холодных тонах. Основным материалом является имитация кирпича. Фундамент является имитацией бетона.

Двери и поддержка навеса имеют серый металлический материал. Крыша навеса и ворота имеют темный шероховатый материал. Пример разработки текстурных карт представлен на рисунке 4.

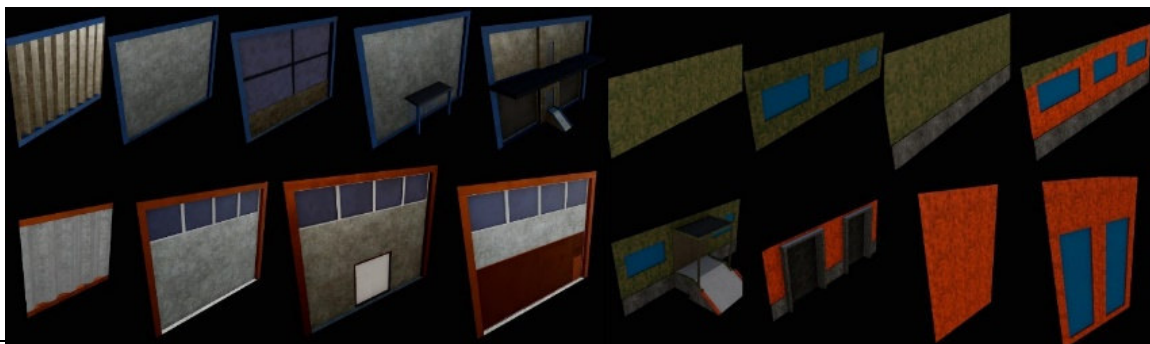


Рисунок 4. Пример разработки текстурных карт

Процесс импортирования трехмерных моделей в игровой движок состоит из следующих этапов: импорт модели и текстурных карт в игровой движок и их распределение по файловой системе игрового движка, разработка материала с использованием импортированных двумерных карт. Процесс размещения моделей в трехмерной игровой локации включает разработку функционального объекта для некоторых трехмерных моделей, а также размещение импортированных и подготовленных моделей. Далее следует размещение освещения и визуальных эффектов.

Геометрия, сгенерированная в Side FX Houdini, автоматически передается в игровой движок посредством полного переноса или отдельных точек со ссылками на трехмерные модели внутри проекта. Наложение материалов в основном происходит посредством ссылки на материал с текстурными картами в проекте, который был присвоен точке или геометрии. В исключительных случаях может потребоваться ручное наложение материала на геометрию. Рассмотрим последовательность выполнения работы в нодовой системе.

Процедурная генерация зданий состоит из разделения примитивной геометрии на этажи, наполнения краев этажей составными трехмерными моделями с заменой составных блоков в области, указанной дополнительной примитивной геометрией. Разбиение на этажи происходит при помощи разделения вертикальных граней примитивной геометрии по высоте. Далее нижняя плоскость примитивной геометрии дублируется на данные точки, тем самым образуя грани для заполнения каждого этажа стенами при помощи составных моделей (рис. 5).

Правила заполнения работают по регулярным выражениям, которые состоят из названий стен, повторение которых задается при помощи типа скобок, в которые заключено название стены. К примеру, первому этажу здания склада задана логика генерации стен в виде строки с правилом `[g_w01wgd01][g_w01wgd01]<g_w01wgd01>[g_w01wgd01][g_w01wgd01]`.



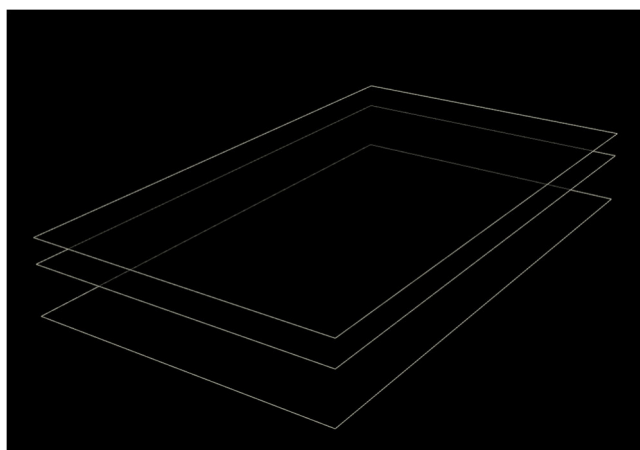


Рисунок 5. Результат разбиения примитивной геометрии на этажи

Итогом заполнения являются точки, расположенные на гранях этажей и содержащие информацию о модели, которую необходимо подставить на место точки, и об относительной масштабности, которую необходимо задать составной модели. Данные точки экспортируются в игровой движок.

Дополнительно разработан функционал замены типа составной модели при помощи второстепенных примитивов, где составные модели, входящие в область второстепенного примитива, заменяются на заранее определенную иную составную модель.

Таким образом разработаны три генератора разных зданий: здания склада, здания логистического центра, здания завода с возможностью выбора используемых текстурных карт на геометрии.

Генерация ландшафта выполнена при помощи пустой карты высоты, которая в последствии деформируется примитивной геометрией, расставленной в игровом движке. Деформированная карта высоты далее получает обработку размытием для сглаживания грубых углов и последующую деформацию при помощи наложения шумов Manhattan Cellular F1 и Sparse Convulsion. В итоге поле высоты получает слабое искажение и последующую слабую деформацию при помощи симуляции эрозии земли, представленной на рисунке 6а. Перед финальной отправкой в игровой движок ландшафт подвергается выравниванию в местах нахождения геометрии зданий и дорог (рис. 6б).

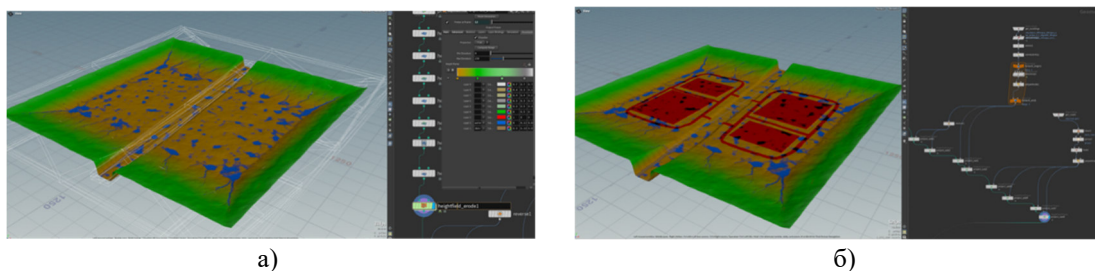


Рисунок 6. Пример разработки ландшафта в программном обеспечении The Side FX Houdini:  
 а – промежуточный вариант разработки ландшафта при помощи симуляции эрозии земли;  
 б – итоговой вариант разработки ландшафта с доработкой выравниванием

Процедурная генерация дорог осуществлена при помощи обработки кривых Безье, расставленных в игровом движке, которые впоследствии импортируются в программное обеспечение The Side FX Houdini. Далее кривые Безье подвергаются уменьшению количества точек на кривых до минимального расстояния между точками в 250 сантиметров. При помощи ландшафта, сгенерированного ранее, точкам присваивается тип наземной или надземной (рис. 7).

Разделение кривой производится по вышеуказанным типам. Из наземного типа создается геометрия наземных дорог с постобработкой в виде поиска пересечений и генерации перекрестков. В местах пересечений на геометрию дорог накладывается атрибут маски для разделения на геометрию, которая не подвергается обработке, и геометрию, трансформирующуюся в перекресток при помощи обработки вершин, граничащих с двумя пересекающимися дорогами. Перекресток преобразуется в сплошную плоскость для последующего объединения с оставшейся геометрией дорог.

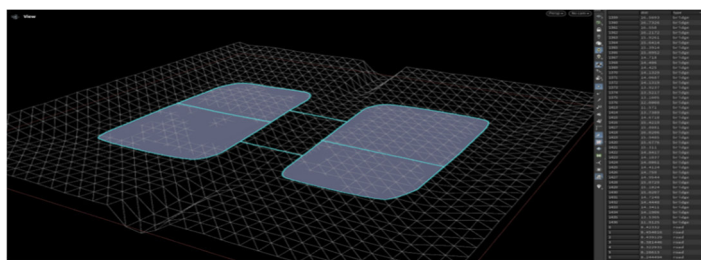


Рисунок 7. Результат присвоения точкам типа дороги с использованием геометрии ландшафта

Из надземного типа генерируется геометрия надземных дорог, где в зависимости от высоты над ландшафтом геометрия представлена сплошной заливкой фундамента при высоте от 500 до 750 сантиметров или мостом при высоте выше 750 сантиметров. Генерация дороги со сплошной заливкой фундамента, работающая на расстоянии от геометрии дороги до геометрии ландшафта, представлена на рисунке 8а. Генерация моста работает на основе генерации кривых из геометрии дороги и формируется при помощи параллельных и поперечных кривых структуры моста. Дополнительно на местах пересечения параллельных и поперечных кривых генерируется геометрия колон, заполняющая пространство от геометрии структуры моста до ландшафта. Результат построения представлен на рисунке 8б.

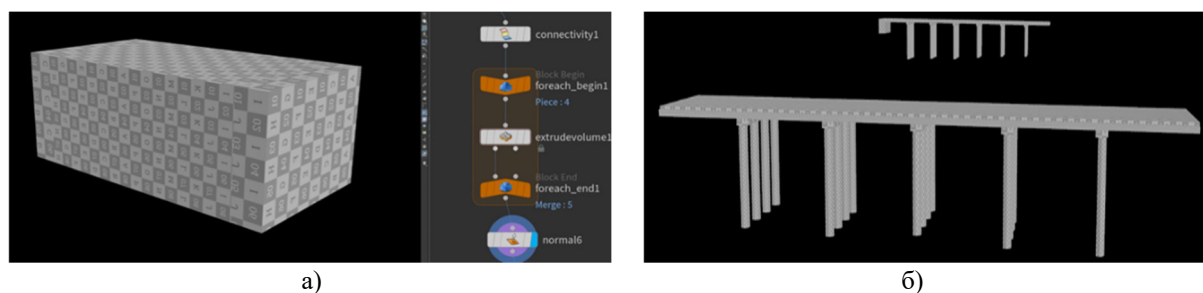


Рисунок 8. Пример сгенерированной геометрии надземных дорог:  
а – пример генерации дороги; б – пример генерации моста

Конечным этапом работы является объединение всех типов дорог в одну геометрию и последующий экспорт в игровой движок.

Для генерации зданий вся геометрия импортирована в проект, в котором она применяется, и распределена по папкам в соответствии с генератором зданий. Такая же система использована с распределением материалов и текстурных карт по проекту.

До генерации расставлены объекты примитивной геометрии ландшафта, зданий и кривых дорог и труб завода (рис. 9).

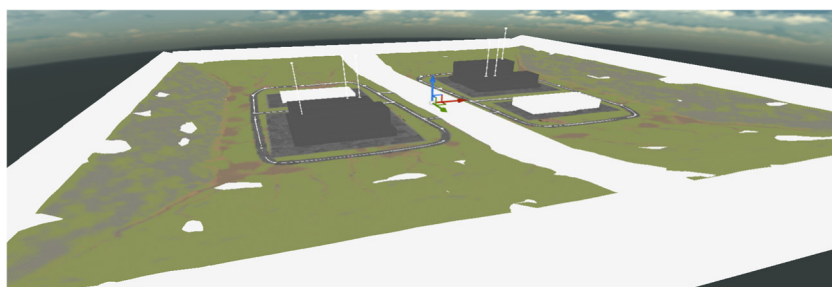


Рисунок 9. Расстановка примитивной геометрии в игровом движке

### Результаты экспериментов

Использование разработанного инструментария генерации позволило создать здания, ландшафт и дорогу локации промзоны. Настройки процедурной генерации доступны в специальной вкладке Houdini Engine в окне настроек и параметров объекта. Готовая локация представлена на рисунке 10а в режиме отображения только цвета материала на геометрии, а на рисунке 10б представлено наличие освещения текстурных карт шероховатости и металличности.

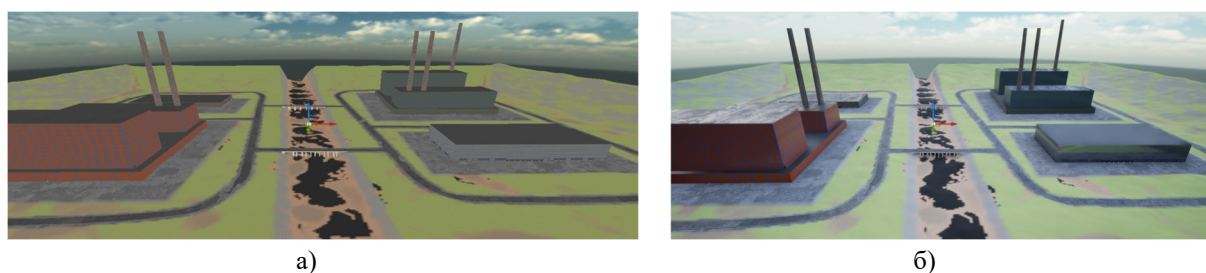


Рисунок 10. Разработанная локация в игровом движке в режиме полной визуализации: *а* – только различные цвета материалов; *б* – с учетом освещения текстурных карт шероховатости и металличности

### Обсуждение результатов

Для определения эффективности разработки продукта за основу взяты два способа создания трехмерной игровой локации с применением инструмента процедурной генерации и без него. Очевидно, что первый вариант является предпочтительным, так как здесь наблюдается возможность экономии временных и финансовых ресурсов.

При проведении сравнительного анализа для определения затрат на разработку трехмерной модели игровой локации из готовых элементов в жанре «стратегия» с процедурной составляющей и без нее каждому разработанному элементу присвоено оригинальное название:

- склад – здание склада размерами 50 метров длиной и 50 метров шириной с одним основным и тремя дополнительными этажами;
- логистический центр – здание логистического центра размерами 100 метров длиной и 50 метров шириной с одним основным и пятью дополнительными этажами;
- завод – здание завода с основной частью в 100 метров длиной и 100 метров шириной, верхней частью в 80 метров длиной и 40 метров шириной, двумя трубами высотой 25 метров;
- ландшафт – ландшафт размером 512 на 512 метров;
- дорога – дорога длиной 100 метров с перекрестком и мостом.

Для подсчета стоимости времени разработки, была взята средняя зарплата художника по окружению в Москве, составляющая в среднем 90000 рублей, и разделена на 168 часов трудового месяца для получения почасовой стоимости работника, равной 535,7 рубля в час [19].

### Сравнительная эффективность способов разработки локации

Элемент	Критерии			
	Время разработки без процедурного инструментария, ч.	Стоимость времени, потраченного на разработку, руб.	Время разработки с процедурным инструментарием, ч.	Стоимость времени, потраченного на разработку, руб.
Склад	0,187	100,18	0,086	46,07
Логистический центр	0,378	202,49	0,107	57,32
Завод	0,470	251,78	0,225	120,53
Ландшафт	0,398	213,21	0,145	77,68
Дорога	0,185	99,10	0,074	39,64
Сумма	1,618	867,76	0,637	341,24

Время разработки здания склада при использовании разработанного процедурного инструментария сократилось на 54 % по сравнению с разработкой зданий вручную из модульных элементов геометрии, здания логистического центра – на 71,7 %, здания завода – на 52,1 %. Время разработки ландшафта при помощи процедурного инструментария, использующего примитивы для формирования ландшафта, сократилось на 63,6 % по сравнению с ручным формированием ландшафта при помощи рисования через графический планшет. Время разработки дороги при помощи разработанной процедурной генерации через кривые Безье сократилось на 60 % по сравнению с ручным моделированием геометрии дороги и мостов. В итоге время на разработку локации с помощью процедурной генерации сократилось на 60,28 %.

На практике доказано, что использование разработанных процедурных инструментов приводит к принципиальному сокращению временных ресурсов для производства трёхмерной локации. При этом появляется возможность масштабирования объектов и беспрепятственного внесения изменений в готовый продукт.

## Выводы

Достигнутые результаты исследования позволяют утверждать, что использование процедурной генерации с помощью геометрических нод в Blender3D для создания цифровых локаций является актуальным перспективным направлением в игровой индустрии. Разработанные методы и модели демонстрируют, как современные технологии способствуют оптимизации процесса создания игровых миров, делая его более быстрым, экономичным и гибким. Разработка локаций при помощи процедурного инструментария позволяет художнику по окружению быстрее выполнять поставленные задачи, сокращая финансовые ресурсы. Отметим, что вопросы, касающиеся внесения изменений в готовый продукт решаются существенно проще и быстрее по сравнению с повторной разработкой геометрии или перемещением модульных элементов вручную.

## Список литературы

1. Харченко Т. В., Зорин П. С., Мысева Е. Р. Трудности и перспективы развития мирового рынка видеоигр. 2023.
2. Седых И. А. Индустрия компьютерных игр / Национальный исследовательский университет «Высшая школа экономики». 2020.
3. Тренды развития российского геймдева в 2025 году // Простые технологии. Режим доступа: <https://blog.rt.ru/b2c/trendy-razvitiya-rossiiskogo-geimdeva-v-2025-godu.htm/> (дата обращения: 7.11.2024).
4. Игровые локации в современных играх: как сюжет раскрывается через окружение // Voki Games. Режим доступа: <https://vokigames.com/ru/igrovye-lokaczii-v-sovremennyh-igrakh-kak-syuzhet-raskryvaetsya-cherez-okruzhenie/> (дата обращения: 16.12.2024).
5. Степанова А. А., Киргизов Ю. В. Способы вовлечения пользователей в игровой процесс компьютерных игр // Графический дизайн: традиции и инновации. 2021. С. 195–205.
6. McMillan J. Games, strategies, and managers. Oxford University Press, 1992.
7. Разработка локаций // info@rating-gamedev.ru. Режим доступа <https://rating-gamedev.ru/blog/razrabotka-lokaczii?ysclid=md4v04uhnk837210249/> (дата обращения: 24.05.2024).
8. McMillan J. Games, strategies, and managers. Oxford University Press, 1992.
9. From game design elements to gamefulness: defining "gamification" / Deterding S., et al. // Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments. 2011. С. 9–15.
10. Три типа игровых миров // DTF. Режим доступа: <https://dtf.ru/u/198189-naidzhel-blait/1209578-tri-tipa-igrovyyh-mirov> (дата обращения: 17.12.2024).
11. Киселева Л. Ю. Видеоигры как объекты права интеллектуальной собственности // Образование и право. 2022. №. 3. С. 93-99.
12. Гражданский кодекс Российской Федерации (часть четвертая) от 18.12.2006 № 230-ФЗ (ред. от 22.07.2024). Ст. 1240.123
13. Федеральный закон Российской Федерации от 29 декабря 2010 г. № 436-ФЗ «О защите детей от информации, причиняющей вред их здоровью и развитию».
14. Федеральный закон Российской Федерации от 25 июля 2002 г. № 114-ФЗ «О противодействии экстремистской деятельности».
15. Медведев М. Р., Фатуллаев Э. С. Правовое регулирование видеоигр по российскому законодательству // Инновации. Наука. Образование. 2020. № 14. С. 312-322.
16. Трусов А. И., Ефимов А. А. Понятие государственной тайны и перечень сведений, составляющих государственную тайну // Гуманитарно-правовые аспекты развития российского общества. 2017. С. 474-478.
17. Щербов З. О. Технологический процесс создания контента игровых локаций. 2011.
18. Левел-дизайн 101: язык разработки локаций // DTF. Режим доступа: <https://dtf.ru/gamedev/1175122-level-dizain-101-yazyk-razrabotki-lokaczii> (дата обращения: 19.12.2024).
19. Сколько зарабатывает 3d artist unreal в Москве // Фильтр Работ. Режим доступа <https://moskva.jobfilter.ru/career/3d-artist-unreal> (дата обращения: 08.05.2025).

УДК 004.93

DOI: 10.25686/978-5-8158-2474-4-2025-954-958

## Геометрическое моделирование солнцезащитных устройств

**А.Т. Дворецкий**

Севастопольский государственный университет, Севастополь, Россия

**Аннотация.** Увеличение доли солнечной энергии в энергоснабжении здания за счет использования пассивных и активных солнечных коллекторов в отопительный период и использования солнцезащиты в период охлаждения является основной задачей при строительстве и эксплуатации энергоэффективных зданий. Солнечная геометрия лежит в основе многих способов повышения эффективности солнечного отопления и охлаждения здания. В основе всех способов формообразования солнцезащитных устройств лежит геометрия видимого движения Солнца по небосводу, а именно – геометрическая модель процесса инсоляции точки на поверхности земли в течение суток. Эта модель представляет собой однопараметрическое множество солнечных лучей, приходящих в одну точку на земной поверхности в течение суток, и является суточным конусом солнечных лучей (СКСЛ). Алгоритмы проектирования солнцезащитных устройств в виде жалюзи и коробов основаны на модели СКСЛ. Предложенная модель солнцезащитных устройств в виде коробов экранирует светопрозрачные конструкции от прямой солнечной радиации в период охлаждения здания, а в отопительный период помещение максимально инсолируется.

**Ключевые слова:** суточный конус солнечных лучей, солнцезащитные устройства, климатизация помещения, энергетическая эффективность

## Geometric modeling of sun protection devices

**A.T. Dvoretzky**

Sevastopol State University, Sevastopol, Russia

**Abstract.** Increasing the share of solar energy in the energy supply of a building by using passive and active solar collectors during the heating period and using sun protection during the cooling period is the main task in the construction and operation of energy-efficient buildings. Solar geometry underlies many methods for increasing the efficiency of solar heating and cooling of a building. All methods for shaping sun protection devices are based on the geometry of the visible movement of the Sun across the sky, namely, a geometric model of the insolation process of a point on the earth's surface during the day. This model is a one-parameter set of solar rays arriving at one point on the earth's surface during the day, and is a diurnal cone of solar rays (DCSR). Algorithms for designing sun protection devices in the form of blinds and boxes are based on the DCSR model. The proposed model of sun protection devices in the form of boxes screens translucent structures from direct solar radiation during the cooling period of the building, and during the heating period the room is maximally insolated.

**Keywords:** daily cone of solar rays, sun protection devices, indoor climate control, energy efficiency

### Введение

Пассивные солнечные дома создаются в результате комплексного проектирования, которое использует местные источники энергии и материалы, климатизацию внутреннего пространства в большей мере архитектурными средствами, чем инженерными [1, 2].

Обеспечение комфортных параметров микроклимата в энергоэффективных зданиях является одним из основных требований к современному строительству [3]. Параметры микроклимата в помещениях нормируются в соответствии с ГОСТ 30494-2011.

В южных странах Европы (Португалия, Испания, Италия, Греция и другие) средняя зимняя температура в январе +10 °С, средняя летняя в июле +25 °С. Эти страны имеют летний доминирующий климат.

В Крыму средняя температура в январе -0,3 °С с холодной пятидневкой -18 °С, средняя температура в июле 22 °С с жаркой пятидневкой 25 °С. Зимой достаточно холодно, а летом жарко. Такой климат может быть назван смешанным.

Необходимо обеспечивать комфортные условия в помещениях как в холодный, так и в перегретый период года. Для снижения теплоступлений в помещения в жаркий период года наиболее эффективным способом считается использование солнцезащитных устройств [4].



### Постановка задачи

По данным Европейской организации по солнечной защите (ESSO – European Solar Shading Organization, 2012), использование внешних солнцезащитных устройств существенно сокращает потребности в энергии для отопления, кондиционирования и освещения.

На рисунке 1 показаны потребности в энергии для отопления, кондиционирования и освещения для офиса площадью 20 м<sup>2</sup> в шести столицах Европы. Офисное помещение оборудовано прозрачным двойным остеклением и находится на южной стороне. Светопрозрачная поверхность представляет собой 80 % фасада. Солнцезащитное устройство (СЗУ) установлено снаружи [5].

Согласно гистограмме потребления энергии на климатизацию помещения южной ориентации (см. рис. 1), самое существенное снижение потребления энергии в течение года наблюдается в Риме – на 75 %, в Мадриде – на 62 %, в Берлине – на 25 %.

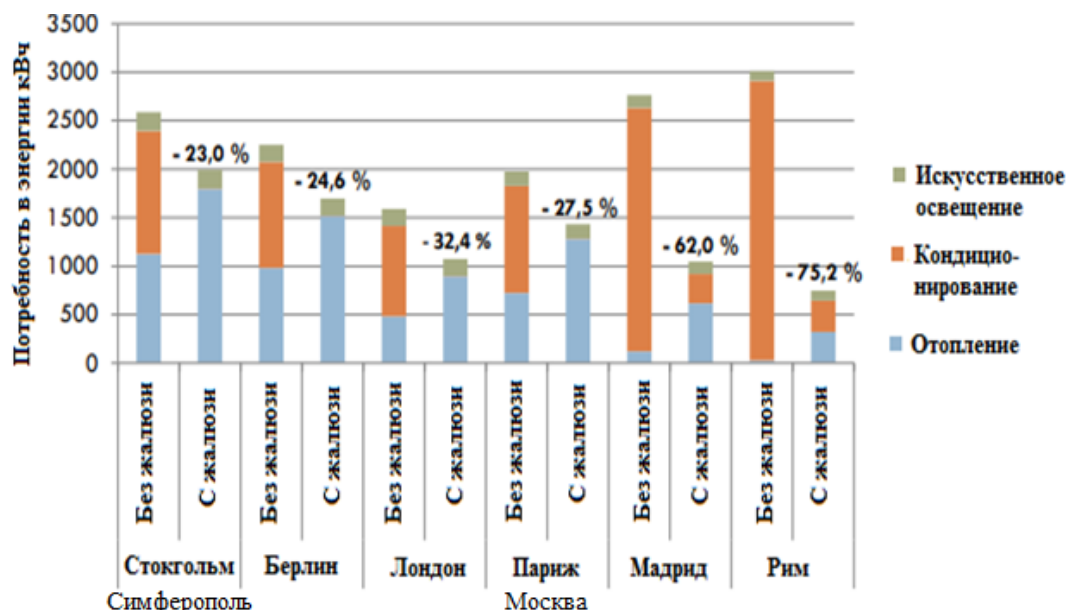


Рисунок 1. Энергопотребление в помещении южного фасада: без солнцезащиты и с наружной солнцезащитой

Косвенно, с достаточной долей вероятности, оценить энергетическую эффективность СЗУ в Симферополе и Москве можно следующим образом:

Климатические условия Москвы очень схожи с климатическими условиями Стокгольма. По уровню солнечной радиации Москва и Стокгольм находятся в одинаковых условиях: 1400–1700 солнечных часов в году. Что касается среднемесячной и среднегодовой температур этих городов, среднегодовые значения температур отличаются на 0,8 °C [6].

Климатические условия Симферополя и Парижа мало отличаются по температуре. Среднегодовые значения температур отличаются на 1,9 °C. Однако количество солнечных дней в году в Симферополе значительно больше, чем в Париже: 2000 и 1700 соответственно [6].

Энергетическая эффективность солнцезащитных устройств в Стокгольме доказана тем, что в течение года потребление энергии на климатизацию помещения снижается на 23 % при использовании жалюзи (рис. 1). Так как климатические условия Москвы очень схожи с климатическими условиями Стокгольма, то можно сделать вывод, что снижение потребления энергии в Москве будет схожим со Стокгольмом.

В течение года потребление энергии на климатизацию помещения в Париже снижается на 27,5 % при использовании жалюзи. Учитывая то, что солнечной радиации в течение года в Симферополе значительно больше, чем в Париже, и среднегодовая температура близка по значениям, снижение потребления энергии на климатизацию помещения будет больше в Симферополе (рис. 1).

Эти примеры доказывают энергетическую эффективность использования солнцезащитных устройств при климатизации зданий.

### Формообразование солнцезащитных устройств в виде коробов

Одним из основных факторов формирования микроклимата помещений является инсоляция, контроль за которой можно осуществлять с помощью солнцезащитных устройств (СЗУ). Задачей повышения энергоэффективности здания при проектировании СЗУ является определение формы устройства, при которой в период охлаждения здания солнечная радиация не проходит в помещение, а в отопительный период помещение максимально инсолируется [7].

Отличия в положениях Солнца для различных ориентаций и времени года определяются по суточному конусу солнечных лучей (рис. 2).

В основе всех методов формирования стационарных солнцезащитных устройств лежит геометрия видимого движения Солнца по небосводу, а именно геометрическая модель процесса инсоляции в точке земной поверхности в течение суток. Эта модель представляет собой однопараметрическую совокупность солнечных лучей, поступающих в одну точку земной поверхности в течение суток, и суточный конус солнечных лучей (СКСЛ).

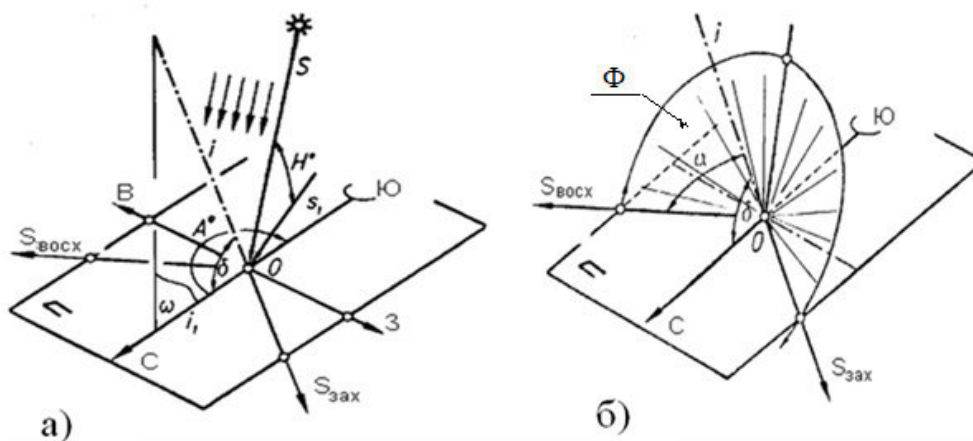


Рисунок 2. Геометрическая модель суточного конуса солнечных лучей:  $A$  – инсолируемая точка;  $\Phi$  – суточный конус солнечных лучей;  $\alpha$  – угол между образующей суточного конуса солнечных лучей (СКСЛ) и его осью;  $P$  – горизонтальная плоскость (поверхность Земли в инсолируемой точке);  $\delta$  – широта местности и угол наклона оси конуса к плоскости горизонта;  $i$  – ось СКСЛ параллельна оси вращения Земли;  $S_{\text{вост}}$  – направление на восход Солнца;  $S_{\text{зах}}$  – направление на заход Солнца; азимут восхода Солнца;  $H^0$  – угловая высота Солнца

Используя СКСЛ, можно определять следующие параметры:

- угловую высоту Солнца в полдень  $H^0_{12}$  для выбранного дня года (применяется в расчетах параметров положения солнечных коллекторов и фотоэлектрических панелей);
- азимуты восхода  $A^0_{\text{вост}}$  и захода  $A^0_{\text{зах}}$  Солнца для выбранной даты (применяется при определении продолжительности инсоляции);
- время восхода  $\tau_{\text{вост}}$  и захода  $\tau_{\text{зах}}$  Солнца.

### Результаты моделирования

Спроектировать СЗУ на южном фасаде в виде цилиндра с горизонтальной осью, перпендикулярной фасаду здания, можно по следующим алгоритму:

1. Задаётся светопроём на фасаде соответствующей ориентации. В данном примере фасад южный;
2. Задаётся поверхность СЗУ. В качестве поверхности СЗУ выбран цилиндр, ось  $i$  которого горизонтальна и перпендикулярна плоскости фасада и который легко может быть реализован в виде маркизы (рис. 3);

3. Рассчитывается половина угла при вершине суточного конуса солнечных лучей (СКСЛ). Косинус половины угла при вершине суточного конуса солнечных  $\alpha$  лучей:

$$\cos \alpha = 0,3979 \cdot \cos \gamma. \quad (1)$$

Определяется переменный угол  $\gamma$ , который равен углу поворота Земли вокруг Солнца, отсчитываемый от точки летнего противостояния:



$$\gamma = \frac{360^\circ \cdot N}{365} = 0,9863^\circ \cdot N, \quad (2)$$

где  $N$  – число суток, которые отсчитываются от 22 июня до заданного дня года;

4. Ориентируется СКСЛ так, чтобы его вершина совпадала с расчетной точкой Р.Т., а его ось принадлежала вертикальной плоскости, расположенной в направлении север-юг и была наклонена к плоскости горизонта под углом  $\delta$  (широта местности);

5. Строится контур СЗУ (рис. 3) как линия пересечения поверхности цилиндра и суточного конуса солнечных лучей.

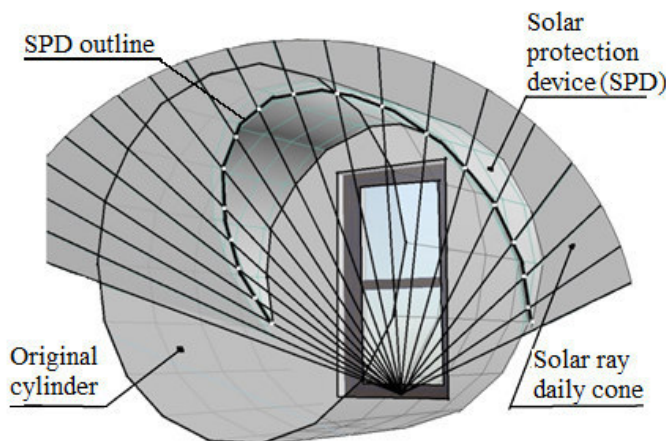


Рисунок 3. Солнцезащитное устройство в виде цилиндра

### Обсуждение результатов

При проектировании солнцезащитных устройств следует учитывать основные требования к их влиянию на микроклимат помещений и энергосбережение:

- снижение теплопоступлений в помещения в теплый период года (пассивное охлаждение);
- снижение теплопотерь и максимальные теплопоступления в помещения в холодный период года (пассивный солнечный обогрев);
- повышение зрительного комфорта, в том числе устранение слепящей яркости в производственных и общественных зданиях и сохранение визуального контакта с внешней средой в течение всего года.

Этим требованиям удовлетворяют два проекта южных фасадов. На рисунке 4 изображён фасад здания с солнцезащитными устройствами в виде цилиндрических поверхностей. На рисунке 5 изображён фасад здания с солнцезащитными устройствами в виде конуса.



Рисунок 4. Солнцезащитные устройства в виде цилиндров

## Выводы

Для проектирования СЗУ в виде коробов и инсоляционных расчетов предпочтение надо отдавать способу, основанному на использовании суточного конуса солнечных лучей, из-за его высокой информативности и универсальности.

Модель суточного конуса солнечных лучей позволяет определять значения азимутов восхода и захода Солнца, а также угловые высоты Солнца для любого дня года, что необходимо для оптимальной ориентации здания по сторонам света и выбора солнцезащитных устройств при проектировании энергоэффективных зданий.

Неважно, для какого климата спроектировано здание, – оно должно быть ориентировано на солнце, иметь естественное отопление и охлаждение, естественное освещение, естественную вентиляцию и выполнено из натуральных, возобновляемых материалов с учетом переработки.

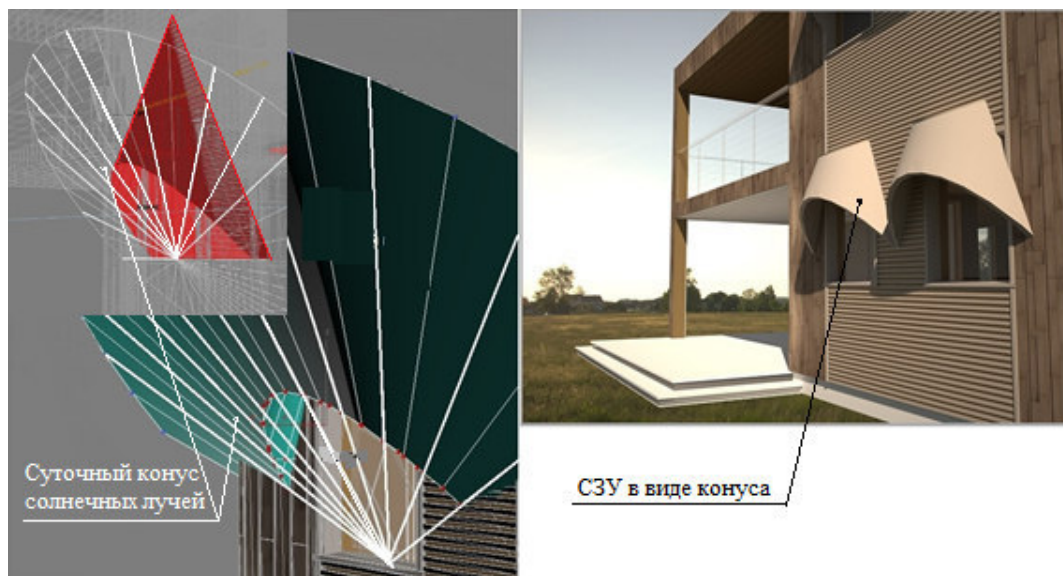


Рисунок 5. Солнцезащитное устройство в виде конуса

## Список литературы

1. Bainbridge David A., Haggard Ken. Passive Solar Architecture. Heating, Cooling, Ventilation and Daylighting Using Natural Flows // Chelsea Green Publishing. Vermont, 2011. 300 p.
2. The Business Case for Passive House. Synergy Sustainability Institute. Victoria, Canada, 2015. 70 p.
3. Дворецкий А.Т. Энергетическая эффективность солнцезащитных устройств // Строительство и техногенная безопасность. Симферополь, 2023. № 31. С. 119-127.
4. Faist V. Summary of designing passive houses // Publishing Association Building Universities. Moscow, 2008. 144 p.
5. Solar shading for low energy building / European Solar Shading Organization. 2012. 48 p.
6. Accounting of Climatic Features in Designing Solar Shading Devices / Alexander T. Dvoretzky, Alexander V. Spiridonov, Igor L. Shubin, Ksenia N. Klevets // Light&Engineering. 2018. Vol. 26, no. 2. P. 162–166.
7. Методы проектирования стационарных солнцезащитных устройств / Дворецкий А.Т., Моргунова М.А., Сергейчук О. В., Спиридонов А.В. // Светотехника. 2016. № 6. С. 43-48.



Секция 8

***ГРАФИЧЕСКИЕ ИНФОРМАЦИОННЫЕ  
СИСТЕМЫ И ИММЕРСИВНЫЕ  
ТЕХНОЛОГИИ НА СТАДИЯХ  
ЖИЗНЕННОГО ЦИКЛА ПРОДУКТА***

---

**GRAPHIC INFORMATION SYSTEMS  
AND IMMERSIVE TECHNOLOGIES  
IN THE PRODUCT LIFE CYCLE STAGES**



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия

## Формализация коэффициента совокупных потерь при каскадной трансляции геометрических моделей

М. А. Серова

Нижегородский государственный технический университет им. Р. Е. Алексеева,  
Нижний Новгород, Россия

**Аннотация.** В работе рассматривается задача корректной передачи геометрических моделей между CAD-системами в контексте обеспечения непрерывности жизненного цикла изделия. Основное внимание уделено формализации коэффициента совокупных потерь при трансляции через промежуточные среды. Для учёта дублирования ошибок на последовательных этапах введён параметр  $\Delta$ , характеризующий долю пересечения потерь. Теоретический анализ показал, что  $\Delta$  ограничен интервалом между полной независимостью и полным дублированием ошибок.

Экспериментальная часть включала 110 опытов трансляции деталей и сборок с использованием нейтральных и специализированных форматов. Результаты демонстрируют высокие значения  $\Delta$ , что указывает на значительное пересечение потерь при последовательных трансляциях. Отмечено, что величина  $\Delta$  зависит от выбора формата, используемого геометрического ядра и сложности исходной модели. Полученные выводы подтверждают системный характер ошибок передачи и могут использоваться для прогнозирования совокупных потерь при практическом обмене данными между CAD-системами.

**Ключевые слова:** геометрическая модель; трансляция геометрической информации; CAD-системы; коэффициент потерь; пересечение потерь; нейтральные форматы.

## Formalization of the Cumulative Loss Coefficient in Multi-Step CAD Model Translation

M. A. Serova

Nizhny Novgorod State Technical University n.a. R. E. Alekseev, Nizhny Novgorod, Russia

**Abstract.** This study addresses the challenge of reliable transfer of geometric models between different CAD systems, which is essential for maintaining product lifecycle continuity. The research focuses on formalizing the cumulative loss coefficient when models are translated through intermediate environments. To capture the overlap of errors occurring at successive stages, a parameter  $\Delta$  is introduced. Theoretical analysis shows that  $\Delta$  is bounded between two extremes: complete duplication of losses and their full independence.

The experimental part included 110 translation trials of parts and assemblies across CAD chains using both neutral (STEP, IGES) and proprietary formats. The results demonstrate consistently high values of  $\Delta$ , indicating a significant overlap of losses across translation steps. Moreover, the findings suggest that  $\Delta$  is influenced by the choice of format, the underlying geometric kernel, and the structural complexity of the model. These insights highlight the systemic nature of translation errors and provide a basis for predicting cumulative losses in practical scenarios of CAD data exchange.

**Keywords:** geometric model; geometric data translation; CAD systems; loss coefficient; loss overlap; neutral formats

### Введение

Проблема корректной трансляции геометрических моделей сохраняет высокую актуальность, поскольку для обеспечения полного жизненного цикла объектов необходимо передавать цифровые представления между различными программными средами. Особое значение данный процесс приобретает в ситуации роста актуальности построения полноценных цифровых двойников [1], что влечет за собой интеграции различных CAD-систем. На практике моделирование редко выполняется «с нуля»: зачастую требуется использование или доработка уже существующей геометрии с последующей интеграцией в более современную или специализированную среду.

Основные проблемы трансляции заключаются в несовпадении форматов обмена данными и различиях в способах их интерпретации экспортными узлами различных систем. В результате файлы одного и того же формата, созданные в различных средах, могут содержать неодинаковые структуры и приводить к потерям или искажениям данных [2]. Задачу трансляции геометрической модели из одной среды  $S_1$  в среду  $S_2$  целесообразно рассматривать как задачу поиска оптимального пути трансляции через промежуточную среду  $S_3$ , обеспечивающую минимальные потери информации при трансляции. Особое внимание в работе уделяется количественной оценке этих потерь, что позволяет сравнивать различные маршруты трансляции и выявлять оптимальные варианты передачи геометрической информации.

### Постановка задачи

Для подтверждения гипотезы необходимо сопоставить величину потерь при прямой трансляции из S1 в S2 с потерями, возникающими при передаче через промежуточную среду S3. Прямые трансляции могут быть оценены эмпирически или спрогнозированы на основе экспериментальных данных [3]. Однако при построении математической модели, ориентированной на использование множества программных сред и форматов, требуется формализованная методология, описывающая механизм вычисления совокупных потерь на маршруте трансляции через промежуточные среды.

Для оценки качества трансляции необходимо прийти к математическому сравнению двух путей трансляции и выбора из них трансляции с наименьшими потерями.

$$q_{S1-S3-S2} < q_{S1-S2}.$$

Сложность оценки заключается в том, что совокупные потери трансляции через промежуточную среду  $q_{S1-S3-S2}$  не могут быть сведены к сумме потерь  $q_{S1-S3} + q_{S3-S2}$ , если не введены определённые допущения относительно характера накопления ошибок. В частности, требуется зафиксировать модель взаимодействия между потерями на каждом этапе трансляции – аддитивную, мультипликативную либо взвешенную в зависимости от семантической значимости передаваемых параметров.

### Теория

Рассмотрим частный случай независимости потерь, при котором ошибки, возникающие на участке трансляции S3-S2 не зависят от потерь на предыдущем участке и оцениваются как абсолютные. Для такой ситуации применим некоторые элементы мультипликативного способа оценки качества, а именно способ агрегирования коэффициентов качества [4]. Коэффициент качества передачи K в общем случае указывает на долю корректно переданной информации. Тогда при независимых трансляциях общий коэффициент качества вычисляется как произведение коэффициентов отдельных этапов:

$$k_{S1-S3-S2} = k_{S1-S3} * k_{S3-S2} = (1 - q_{S1-S3}) * (1 - q_{S3-S2}).$$

Соответствующий коэффициент совокупных потерь принимает вид

$$q_{S1-S3-S2} = q_{S1-S3} + q_{S3-S2} - q_{S1-S3} * q_{S3-S2}.$$

Таким образом, при условии независимости потерь совокупная величина трансляционных ошибок оказывается меньше их простой суммы, что отражает эффект компенсации при мультипликативной оценке качества.

Приведённая выше формула справедлива лишь в условиях независимых и равномерно распределённых потерь при трансляции модели. Однако подобная ситуация представляет собой частный и сравнительно редкий случай. На практике потери, как правило, являются зависимыми: они включают накопление ошибок на последовательных этапах, внутренние погрешности алгоритмов экспорта/импорта, а также постепенную деградацию данных. Указанные факторы существенно усложняют процесс формализации и должны быть учтены в общей модели расчёта совокупных потерь при трансляции через промежуточные среды.

Для перехода от частного случая независимых потерь к более общей модели введём величину  $\Delta$  *доля пересечения потерь*. Эта величина условно отражает количество информации, утраченное на первом этапе трансляции и повторно учитываемое как утраченное на втором этапе. С его учётом коэффициент совокупных потерь при передаче через промежуточную среду S3 может быть записан в виде

$$q_{S1-S3-S2} = q_{S1-S3} + q_{S3-S2} - \Delta.$$

Таким образом, задача оценки суммарных потерь передачи геометрии через промежуточную среду сводится к предсказанию величины  $\Delta$ . Определим пределы изменения этой величины, для чего рассмотрим пограничные ситуации.

Возможна ситуация, при которой трансляция S3 → S2 полностью дублирует ошибки, возникшие на этапе трансляции S1 → S3. В таком случае дополнительные потери отсутствуют, а величина  $\Delta$ , отражающая долю пересечения потерь, принимает максимально возможное значение. Это означает, что суммарные потери совпадают с наибольшей из двух частных величин, поскольку все искажения второго этапа уже были учтены при первой трансляции.

$$q_{S1-S3-S2} = \max\{q_{S1-S3}; q_{S3-S2}\}, \\ \Delta = \min\{q_{S1-S3}; q_{S3-S2}\}.$$



Возможна и противоположная ситуация, когда потери на каждом из этапов трансляции относятся к различным параметрам геометрической модели. В этом случае пересечения отсутствуют,  $\Delta=0$ , и совокупный коэффициент потерь определяется как аддитивная сумма частных коэффициентов

$$q_{S1-S3-S2} = q_{S1-S3} + q_{S3-S2}, \\ \Delta = 0.$$

Такой случай отражает максимально возможное дублирование ошибок при последовательной передаче, когда каждый этап трансляции вносит уникальные искажения в структуру модели.

Однако при приравнивании коэффициента совокупных потерь к сумме коэффициентов потерь могут возникнуть некорректные случаи, где совокупные потери превысят 1, что противоречит как математической корректности, так и смыслу задачи.

Физический смысл коэффициента потерь заключается в количественной оценке утраченной информации при трансляции геометрической модели. Значение данного коэффициента отражает долю параметров, которые не удалось корректно сохранить при передаче из одной среды в другую. При этом  $q=0$  соответствует идеальной передаче без потерь, а  $q=1$  означает полную утрату информации, когда ни один из параметров исходной модели не был транслирован корректно. Исходя из физического смысла, коэффициент потерь всегда изменяется в пределах от 0 до 1, что гарантирует его интерпретацию как вероятностной или относительной меры.

$$q_{S1-S3-S2} = \max\{q_{S1-S3} + q_{S3-S2}; 1\}.$$

Совокупный коэффициент потерь при последовательной трансляции геометрической модели также должен быть ограничен единицей, поскольку утеря не может превышать весь объём передаваемой информации: модель либо сохраняет часть параметров, либо теряет их полностью при любом количестве трансляций.

Таким образом, введение величины  $\Delta$  также должно балансировать модель между двумя крайними сценариями: полной независимостью ошибок и полным их дублированием. Диапазон значений  $\Delta$  задаёт допустимые границы модели, исключая появление абсурдных результатов и обеспечивающие её применимость в практических задачах передачи геометрической информации.

$$\Delta_{min} = \max\{0; q_{S1-S3} + q_{S3-S2} - 1\}; \\ \Delta_{max} = \min\{q_{S1-S3}; q_{S3-S2}\}.$$

При любых значениях коэффициентов потерь, характеризующих прямые трансляции между системами, *величина пересечения потерь* при передаче через промежуточную среду всегда ограничена строгими пределами. Данное ограничение отражает фундаментальный принцип: повторные трансляции не способны уменьшить уже произошедшие потери, но также не могут привести к утрате информации в объёме, превышающем суммарное действие отдельных этапов передачи.

### Накопление потерь

Основная задача при формализации коэффициента совокупных потерь заключается в предсказании величины параметра  $\Delta$ . Очевидно, что определить его абсолютное значение невозможно, поскольку накопление и дублирование потерь зависят от множества скрытых факторов: особенностей геометрического ядра, версии формата и вариативности экспортно-импортных узлов в средах.

Статистический подход предполагает оценку  $\Delta$  на основе большого массива экспериментальных данных, однако объём требуемых опытов делает такой метод практически нереализуемым. В связи с этим более перспективным является категориальный подход, при котором пересечение потерь прогнозируется исходя из анализа параметров, утраченных при трансляции. Такой способ позволяет аппроксимировать  $\Delta$ , опираясь на описание геометрической модели, специфику формата передачи, а также на характеристики сред, участвующих в трансляции.

В рамках анализа программных сред сложно сделать устойчивый вывод о влиянии на величину  $\Delta$  совпадения или различия геометрического ядра сред, участвующих в трансляции. При использовании одинакового ядра CAD среды реализуют схожие процедуры экспорта и импорта, что позволяет выдвинуть гипотезу о том, что значительная доля информации, утратившаяся на первом шаге, фиксируется как аналогичные потери на втором, что увеличивает пересечение потерь.

Использование сред с различными геометрическими ядрами, напротив, должно влечь за собой различия в интерпретации геометрических моделей. Подобная закономерность отмечается в

исследованиях по совместимости CAD-систем, демонстрирующих, что унификация геометрического ядра снижает вероятность случайных ошибок, но усиливает эффект дублирования систематических потерь при каскадной трансляции [5]. Таким образом, совпадение ядра обеспечивает высокую предсказуемость потерь, но одновременно увеличивает величину пересечения  $\Delta$ , тогда как различие ядер делает процесс менее предсказуемым, но снижает эффект дублирования потерь.

Дублирование потерь может быть связано с использованием идентичных форматов на различных этапах трансляции модели. Связано это в первую очередь с ограниченными возможностями формата хранить конкретную информацию. Однако в работах последних лет отмечается, что повторное использование одного и того же формата в трансляционных цепочках сопровождается накоплением систематических ошибок и затрудняет последующую реконструкцию модели [6].

Использование для трансляции различных форматов уменьшает пересечение потерь, однако увеличивает вероятность возникновения новых, уникальных ошибок. На эти ошибки влияют не только формат и его характеристики, но и специфика формата в конкретной среде. Последовательная трансляция в различных форматах сопровождается накоплением систематических ошибок и затрудняет последующую реконструкцию модели.

В оценке качества передачи и количества потерь при трансляции принимается исходное допущение, что геометрическая модель в среде S1 при передаче  $S1 \rightarrow S2$  и в среде S2 при передаче  $S2 \rightarrow S3$  содержит полный набор параметров, включающий до 30 различных характеристик геометрии и метаданных, принятых за параметры модели и описанных в [7]. Именно полнота модели используется как эталон для сопоставления и анализа качества передачи. Однако на практике трансляция часто осуществляется не для иерархических геометрических структур с развернутым набором атрибутов, а для сравнительно упрощенных моделей, представляющих собой лишь геометрическое описание с ограниченным числом параметров. В таких случаях расчёт потерь относительно «идеальной» геометрической модели приводит к заниженной оценке, так как часть параметров изначально отсутствует и не может быть утрачена. Следовательно, чем выше структурная сложность исходной модели, тем больше вероятность наложения и дублирования потерь при последовательных трансляциях, что усиливает расхождение между теоретической и практической величиной коэффициента потерь. Исходя из этого рассуждения можно утверждать, что существует зависимость величины  $\Delta$  от величины коэффициента потерь на каждом из шагов.

### Результаты экспериментов

Теоретический анализ совокупных потерь и обоснование введения величины  $\Delta$  объясняют механизмы накопления или дублирования ошибок при трансляции геометрических моделей. Однако верификация выведенных ограничений может быть продемонстрирована только на реальных данных, полученных в ходе экспериментальной трансляции моделей между различными системами. Для подтверждения предложенной модели и выявленных закономерностей необходимо провести ряд опытов по последовательной трансляции геометрической модели для оценки величины совокупных потерь.

Для проверки выдвинутых теоретических положений проведён комплекс экспериментов, направленных на выявление закономерностей дублирования и накопления ошибок при трансляции геометрических моделей. На данном этапе задачей являлось сопоставление поведения трансляций внутри семейств CAD-систем, основанных на единых геометрических ядрах. Такой подход позволяет зафиксировать влияние ядра как фактора, определяющего структуру и характер потерь.

В первую серию опытов включена цепочка КОМПАС-3D  $\rightarrow$  Inventor  $\rightarrow$  Fusion360, представляющая семейство, использующее геометрическое ядро ASM (Autodesk ShapeManager/ACIS). Вторая серия опытов построена по аналогичной схеме, но на базе ядра Parasolid и включает цепочку КОМПАС-3D  $\rightarrow$  SolidWorks  $\rightarrow$  Solid Edge. Трансляция геометрических моделей осуществлялась с использованием различных форматов представления данных. В исследовании были задействованы как универсальные нейтральные форматы (STEP, IGES), традиционно применяемые для обеспечения межсистемной совместимости, так и специализированные форматы ограниченного распространения, включая .x\_t и .iam, позволяющие зафиксировать особенности отдельных CAD-систем. Такой выбор форматов обеспечил более широкий охват сценариев трансляции и позволил учесть как стандартные, так и редкие практические случаи передачи моделей.



Эксперименты проведены на двух типах тестовых моделей, отличающихся по геометрической сложности: твердотельной детали и сборке. Деталь включает фаски, резьбы, сквозные и глухие отверстия, выточки, построенные с использованием встроенных инструментов CAD-системы. Часть элементов имеет параметрические зависимости, что позволяет оценить сохранение семантической информации при передаче. Помимо указанных выше элементов, сборка содержит сопряжения, массивы и стандартные соединения, что существенно усложняет её структуру и позволяет анализировать не только сохранение геометрии, но и корректность передачи связей между компонентами.

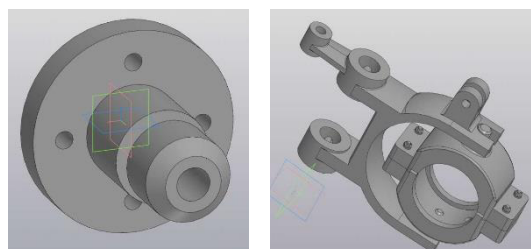


Рисунок 1. Геометрические модели для эксперимента: деталь (слева) и сборка (справа)

Общая методика оценки качества трансляции строится следующим образом. Геометрическая модель описывается набором из девятнадцати параметров: точность узлов, ребер, поверхностей, объёмов; сохранение типа моделирования, точность фасок, геометрическая и иерархическая параметризация, эскизы и чертежи, массивы, материалы; точность массы, плотности, площади, объема; центр масс, возможность использования кириллицы и размер файла. Каждому параметру присваивается весовой коэффициент, определяющий его значимость в структуре модели. Данный коэффициент формируется с учетом экспертной оценки, основанной на предположении о трудоёмкости восстановления соответствующего параметра в случае его утраты при трансляции.

При трансляции геометрической модели из одной среды в другую передача каждого параметра оценивается обособленно. Весовой коэффициент параметра домножается на показатель его сохранности при передаче. При этом значение показателя принимается равным 0, если параметр не удалось передать, 0,5 – в случае передачи с ошибками или неточностями, и 1 – если параметр был воспроизведён корректно и полностью. Сумма утраченной информации о параметрах представляет собой коэффициент потерь при данной трансляции. Совокупные потери оцениваются аналогичным способом для геометрической модели спустя две трансляции.

В рамках экспериментальной части исследования было проведено 110 опытов трансляции геометрических моделей через промежуточную систему. Коэффициенты потерь прямых трансляций, а также коэффициент совокупных потерь получены эмпирическим путем по описанной выше технологии. На их основе была рассчитана доля пересечения потерь  $\Delta$ . Результаты первых 10 опытов отражены в таблице.

### Результаты эксперимента

i	Формат S1 - S3	Формат S3 - S2	Потери S1 → S3	Потери S3 → S2	Совокупные потери	Доля пересечения	$\Delta_{min}$	$\Delta_{max}$	$\Delta_{норм}$
1	F1	F6	0,3015	0,3285	0,3350	0,2950	0	0,3015	0,9784
2	F1	F7	0,3015	0,3445	0,3510	0,2950	0	0,3015	0,9784
3	F1	F1	0,3015	0,4140	0,4140	0,3015	0	0,3015	1
4	F1	F2	0,3015	0,6950	0,6950	0,3015	0	0,3015	1
5	F1	F3	0,3015	0,4265	0,4330	0,2950	0	0,3015	0,9784
6	F1	F4	0,3015	0,4330	0,4330	0,3015	0	0,3015	1
7	F2	F6	0,6610	0,3285	0,6950	0,2945	0	0,3285	0,8965
8	F2	F7	0,6610	0,3445	0,6950	0,3105	0,0055	0,3445	0,8997
9	F2	F1	0,6610	0,4140	0,6950	0,3800	0,0750	0,4140	0,8997
10	F2	F2	0,6610	0,6950	0,6950	0,6610	0,356	0,6610	1

В экспериментальной части используется зависимость между долей пересечения потерь и коэффициентом совокупных потерь. Основным интересом заключается в возможности экстраполяции величины  $\Delta$  на иные сценарии трансляции, но вычисление  $\Delta$  без проведения эмпирических испытаний затруднительно.

Однако границы изменения величины  $\Delta$  могут быть рассчитаны, они зависят исключительно от коэффициентов потерь, зафиксированных при прямых трансляциях, входящих в рассматриваемую цепочку. Таким образом, для каждого конкретного случая трансляции возможно определить диапазон допустимых значений  $\Delta$  без проведения эксперимента. В теоретической части работы были выведены соответствующие зависимости, что позволяет использовать их для расчёта индивидуальных границ  $\Delta$ : максимальное значение, соответствующее полному дублированию потерь, и минимальное значение, отражающее их полное различие.

Для сопоставимости полученных результатов целесообразно использовать не абсолютные значения доли пересечения потерь  $\Delta$ , а их нормализованные величины. Такая нормализация осуществляется относительно рассчитанного диапазона изменения  $\Delta$ , определяемого на основе коэффициентов прямых трансляций. Применение нормализованной  $\Delta^{\text{норм}}$  позволяет свести все наблюдения к единой шкале и устранить зависимость оцениваемой доли от конкретных коэффициентов потерь. Преобразование значения  $\Delta$  в его нормализованную форму  $\Delta^{\text{норм}}$  осуществляется по формуле

$$\Delta^{\text{норм}} = \frac{\Delta - \Delta_{\min}}{\Delta_{\max} - \Delta_{\min}}.$$

Полученная нормализованная доля, а также рассчитанные на основе коэффициентов потерь прямых трансляций пограничные значения доли пересекающихся потерь также занесены в таблицу 1. Для удобства распределение нормализованных значений представлено в виде гистограмм на рисунке 2.

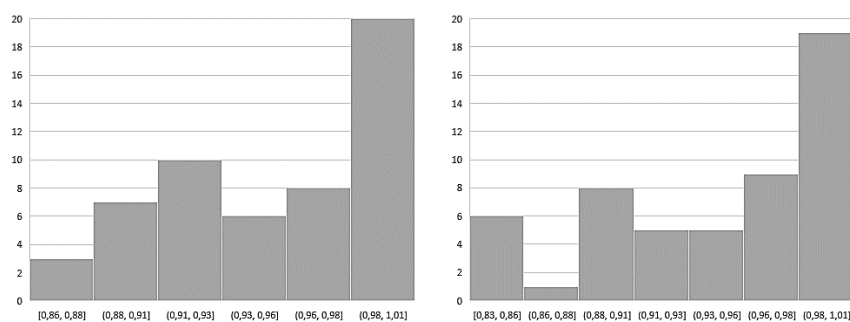


Рисунок 2. Распределение  $\Delta^{\text{норм}}$ : при трансляции детали (слева) и сборки (справа)

Таким образом, на основе проведённых расчётов и эмпирических наблюдений были получены значения  $\Delta$ , а также их нормализованные распределения. Построенные гистограммы позволяют выявить характерные диапазоны изменения величины и очертить область наиболее вероятных значений. Полученные данные представляют собой основу для дальнейшего анализа и требуют интерпретации с точки зрения закономерностей трансляции геометрической информации между системами.

### Обсуждение результатов

Объём экспериментальных данных, включающий 110 опытов по трансляции деталей и сборок, не является достаточным для полной аппроксимации результатов. Ограниченность выборки не позволяет в полной мере зафиксировать редкие сценарии трансляции с минимальной долей пересечения потерь. Тем не менее даже при таком числе наблюдений выявляются устойчивые закономерности.

Распределение значений нормализованной дельты демонстрирует ненормальный характер, что подтверждает системный, а не случайный характер потерь. Диапазон  $\Delta^{\text{норм}}$  находится в пределах 0,8317–1, при этом значения ниже 0,83 отсутствуют. Это свидетельствует о том, что даже в условиях высоких коэффициентов потерь доля пересечения потерь остаётся высокой.

Наибольшая плотность значений в обеих группах опытов фиксируется в интервале 0,89–0,99, что отражает высокий уровень дублирования ошибок. При этом для трансляций во второй группе опытов

диапазон значений несколько уже, нижняя граница смещена вверх, а разброс меньше по сравнению с первой группой. Таким образом, трансляция сборок характеризуется большей стабильностью и меньшей вариативностью совокупных потерь. Однако между опытами с формально идентичными условиями по трансляции деталей и сборок устойчивая корреляция не наблюдается: пересечение потерь зависит от конкретной комбинации систем, форматов и геометрии моделей.

Подтвердить влияние геометрического ядра на основании текущего объема экспериментов невозможно. Однако отмечается связь между использованием специализированных форматов и принадлежностью систем к одному и тому же ядру. Это частично объясняет распределение  $\Delta^{\text{норм}}$ , а именно высокие значения нормализованной дельты и насыщение вблизи верхней границы. Таким образом, зафиксированное дублирование потерь является проявлением системной зависимости, а не случайного фактора.

Дополнительный интерес представляет влияние формата передачи. В случаях, когда в цепочке S1–S3 и S3–S2 использовался один и тот же формат,  $\Delta^{\text{норм}}$  достигала значения 1, что зафиксировано в 20 опытах. Следовательно, при одинаковых форматах вероятность дублирования потерь максимальна. При этом величины коэффициентов потерь при прямых трансляциях не оказывают взаимного влияния: максимально возможное значение  $\Delta$  не зависит от совпадения этих коэффициентов, а определяется условиями трансляции.

Некоторые опыты с одинаковыми парами форматов показали расхождения как в значениях  $\Delta^{\text{норм}}$ , так и в совокупных потерях. Аналогичные различия проявляются при работе с разными геометрическими моделями и программами, что указывает на локальную зависимость параметра от конкретных условий трансляции.

В целом эксперименты демонстрируют, что совокупные потери при двукратной трансляции геометрической информации остаются ограниченными и близкими к минимальным значениям.

Следует отметить, что *полученные в эксперименте значения доли пересечения потерь  $\Delta$  во всех случаях находились внутри теоретически рассчитанных границ*. Однако ни в одном опыте  $\Delta$  не достигла минимального значения. Это указывает на то, что даже при неблагоприятных сценариях трансляции потери имеют значительное пересечение. При этом можно предположить, что усреднённое значение  $\Delta$  располагается ближе к верхней границе диапазона, однако данное утверждение требует дополнительной проверки на расширенной выборке. Таким образом, теоретически рассчитанные границы доли  $\Delta$  сохраняют практическую значимость и могут использоваться в прикладных задачах оценки качества передачи геометрической информации, задавая допустимый интервал значений для анализа и прогнозирования.

## Выводы

Результаты экспериментов показывают, что существующая методика оценивания качества передачи имеет ограниченную применимость. Большинство трансляций не сохранили параметризацию и зависимости, полностью утратили информацию о массивах, а сведения о материалах деталей оказались искажёнными. Для практических задач целесообразно оценивать сохранность отдельных параметров. Повторение экспериментов на иной системе параметров может позволить получить более нормированное распределение и выявить дополнительные закономерности.

Анализ полученных результатов позволяет поставить вопрос о возможности экстраполяции разработанных формул на случай последовательной трансляции через несколько промежуточных систем. Теоретически такая экстраполяция возможна, однако с увеличением числа шагов накапливаются как ошибки экспорта, так и ошибки импорта, вследствие чего коэффициент совокупных потерь перестаёт иметь однозначную интерпретацию. Дополнительно следует учитывать и практический аспект: увеличение количества промежуточных трансляций ведёт к росту трудоёмкости процесса передачи моделей, что делает подобные маршруты нецелесообразными.

Результаты экспериментов также позволяют выдвинуть гипотезу о зависимости величины пересечения потерь  $\Delta$  не только от используемых форматов передачи и геометрического ядра, но и от характеристик самой модели. В частности, на примере опытов с более сложными объектами (сборочные модели) зафиксировано незначительное повышение коэффициента совокупных потерь по

сравнению с аналогичными опытами, проведёнными на менее сложных моделях (отдельные детали) при использовании тех же программных сред и форматов.

### Список литературы

1. Data Exchange with Support for the Neutral Processing of Formats in Computer-Aided Design/Computer-Aided Manufacturing Systems / P. Kuryło, P. Frankovský, M. Malinowski, et al. // Applied Sciences. 2023. Vol. 13, iss. 17. P. 9811. DOI: 10.3390/app13179811
2. Problems of data exchange between BIM and facility management systems, standardization and ensuring data continuity in BIM throughout the lifecycle / R. B. Otranto, et al. // ITcon. 2025. URL: <http://www.itcon.org> (дата обращения: 19.08.2025).
3. Филинских А. Д., Мерзляков И. Н., Курушин Е. А. Оценка потери данных геометрической модели в условиях импортозамещения // Труды Международной конференции по компьютерной графике и зрению «Графикон». 2022. № 32. С. 917–924. DOI: 10.20948/graphicon-2022-917-924; EDN: GHRKCP
4. Multiplicative method of multi-criteria analysis based on expected criteria values / M. M. Žižović, M. R. Žižović, N. Ž. Damljanić, K. Ž. Pavlović // Reports in Mechanical Engineering. 2023. Vol. 4, iss. 1. Pp. 224–234. DOI: 10.31181/rme230402224z
5. Schnitger C. New Math. The Hidden Costs of Kernel Change. Schnitger Corporation, 2020. 24 p. URL: <https://schnitgercorp.com/wp-content/uploads/2020/12/SchnitgerCorp-Hidden-Cost-of-Kernel-Change-2020.pdf> (дата обращения: 19.08.2025).
6. Evaluation of CAD Data Exchange Using Neutral File Formats / P. Kuryło, A. Przesmycki, A. Wróbel, S. Wierzbicki // Applied Sciences. 2023. Vol. 13, iss. 17. P. 9811. DOI: 10.3390/app13179811
7. Boytyakov A., Filinskikh A. Data conversion model using the principles of geometric model structure proximity comparison // GraphiCon 2021: Proceedings of the 31st International Conference on Computer Graphics and Vision, Nizhny Novgorod, September 27–30, 2021. Nizhny Novgorod, 2021. Vol. 3027. Pp. 1064–1072. EDN: PABKMS

## **Влияние степени интерактивности приложения на эффективность экспозиционной терапии в виртуальной реальности**

**Е. Д. Шанин, А. А. Спирин, А. А. Шутов**

Нижегородский государственный технический университет им. Р. Е. Алексеева,  
Нижний Новгород, Россия

**Аннотация.** Фобии остаются одной из актуальных проблем психического здоровья, а экспозиционная терапия признана одним из наиболее эффективных методов их лечения. Особый интерес представляет терапия в виртуальной реальности. Проведенный анализ подтвердил, что экспозиционная терапия в виртуальной реальности действительно демонстрирует высокую результативность в лечении фобий. Однако большинство существующих решений ограничены отсутствием взаимодействия пользователя с виртуальной средой. Экспериментальные данные указывают на то, что внедрение интерактивных элементов способно повысить эффективность терапии. Тем не менее для окончательного подтверждения этой гипотезы требуются дополнительные исследования.

**Ключевые слова:** фобии, экспозиционная терапия, виртуальная реальность, интерактивность.

## **The impact of the degree of interactivity of an application on the effectiveness of exposure therapy in virtual reality**

**E. D. Shanin, A. A. Spirin, A. A. Shutov**

Nizhny Novgorod State Technical University n.a. R. E. Alekseev, Nizhny Novgorod, Russia

**Abstract.** Phobias remain one of the pressing issues in mental health, and exposure therapy is recognized as one of the most effective treatment methods. Of particular interest is therapy in virtual reality (VR). The conducted analysis confirms that virtual reality exposure therapy (VRET) indeed demonstrates high efficacy in treating phobias. However, most existing solutions are limited by the lack of user interaction with the virtual environment. Experimental data suggest that the introduction of interactive elements can enhance the effectiveness of therapy. Nevertheless, further research is required to conclusively confirm this hypothesis.

**Keywords:** phobias, exposure therapy, virtual reality, interactivity.

### **Введение**

Фобии являются серьезной проблемой в современном обществе. Есть множество способов борьбы с этими расстройствами. Наиболее эффективным является экспозиционная терапия [1]. Подвидом данного типа терапии является терапия в виртуальной реальности. Однако приложения для экспозиционной терапии в виртуальной реальности не используют потенциал технологий виртуальной реальности в высокой степени интерактивности с окружением. В статье рассматривается вопрос, может ли более высокая степень интерактивности сделать терапию фобий в виртуальной реальности более эффективной.

### **Постановка задачи**

Основной задачей статьи является выявление связи между степенью интерактивности приложения виртуальной реальности для терапии фобии и эффективностью лечения этой фобии.

### **Теория**

Фобия – психическое расстройство, при котором человек испытывает иррациональный страх перед конкретным событием или предметом [1]. Фобии могут значительно мешать жизни и рабочей деятельности человека. У человека с аэрофобией (боязнь летать) может возникать сильная тревога, панические атаки, учащённое сердцебиение и потливость при одной лишь мысли о полёте на самолёте. Это может привести к избеганию командировок, отказу от путешествий или отказу от выгодных предложений о работе, требующих перемещения.

Другой пример – социофобия, при которой человек боится социальных ситуаций, где возможна оценка со стороны других, что может мешать участию в публичных выступлениях, собеседованиях или даже обычном общении с коллегами. Такие ограничения снижают качество жизни и могут приводить к социальной изоляции, депрессии и снижению самооценки.

Согласно множеству исследований, пожизненно конкретными фобиями страдают от 3–15 % [2] людей в экстремальной степени, что мешает этим людям выполнять ежедневные задачи. Распространенность и серьезность расстройств приводит к выводу, что необходимо эффективное лечение. Есть ряд терапий, которые уже помогают людям с фобиями:

- когнитивно-поведенческая терапия;
- медикаментозное лечение;
- экспозиционная терапия.

Когнитивно-поведенческая терапия включает в себя выявление проблемных или иррациональных моделей мышления и помощь в формировании новых, более адаптивных способов мышления о сложных ситуациях. Этот вид терапии может включать в себя и экспозиционную терапию.

Экспозиционная терапия помогает людям преодолевать страхи, постепенно приближаясь к тем вещам, которые их пугают. Терапия безопасно подвергает пациента его фобии в контролируемой обстановке. Целью экспозиции является десенсибилизация – привыкание человека к объекту его страха или к ситуации, которая вызывает страх, и постепенное снижение его тревожности.

Медикаментозное лечение фобий включает использование различных препаратов. Наиболее часто назначаемыми лекарствами являются бензодиазепины (например, ксанакс, ативан, клонопин) [3].

Однако медикаментозное лечение работает только на короткий срок. Препараты необходимо принимать непосредственно прямо перед столкновением со страхом, и только тогда будет эффект. Этот способ лечения не несет особой пользы сам по себе [3]. Для преодоления фобий нужен другой подход.

Самым эффективным способом лечения является когнитивно-поведенческая терапия в связи с экспозиционной терапией. Данный способ успешно значительно уменьшает уровень тревоги уже после 10 сеансов [3]. Однако у экспозиционной терапии есть серьезный недостаток – недоступность. Для экспозиции необходимо столкнуть человека с его страхом напрямую. Если речь идет о страхе темноты, то найти среду для экспозиции не составит труда. Но если фобия – это боязнь высоты, полетов или глубин, то проведение терапии становится крайне недоступным и оттолкнет большинство пациентов.

Это проблему решают приложения виртуальной реальности. Пациенту предлагается пройти курс лечения, используя очки виртуальной реальности под наблюдением психотерапевта. Этот метод позволяет создавать опыт с высокой степенью погружения, моделируя реалистичные и контролируемые среды, в которых пациент может столкнуться со своим страхом в безопасной обстановке.

Согласно многочисленным исследованиям, экспозиционная терапия фобий в виртуальной реальности действительно является эффективным способом лечения фобий [4]. Некоторые клиники (в том числе в России) уже активно применяют данную технологию, например, клиника «Сова» [5].

Клиника предоставляет лечение боязней высоты, полетов, насекомых, замкнутых пространств и еще ряда фобий. Также «Сова» уверяет своих пациентов в эффективности терапии в виртуальной реальности. Более того, клиника утверждает, что подобный вид терапии даже эффективнее традиционной экспозиционной терапии на рисунке 1 и своих высказываниях на сайте. Согласно исследованиям, описанным в медицинской зарубежной литературе, эффективность VR-психотерапии выше, чем при обычной терапии стандартными методами лечения [5].

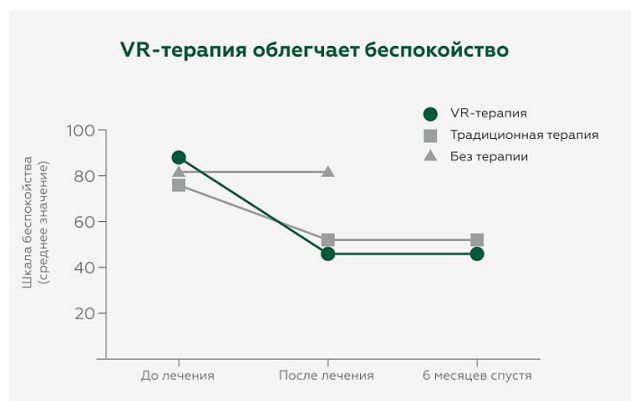


Рисунок 1. График с сайта клиника «Сова»

В клинике применяется собственное приложение, в котором пользователю предстоит оказаться в различных местностях и под контролем психотерапевта перемещаться по ним. На рисунке 2 представлен скриншот из приложения.

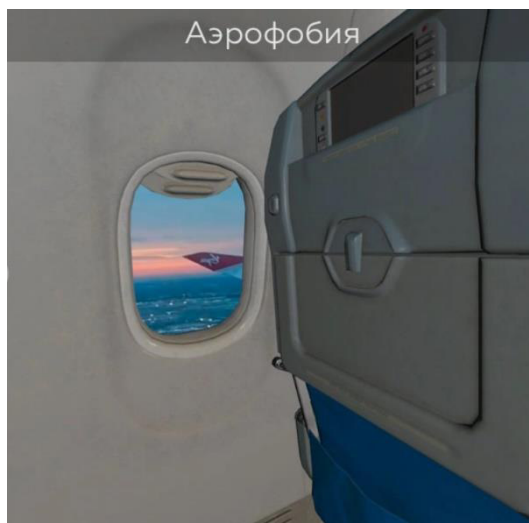


Рисунок 2. Скриншот из приложения клиники «Сова»

Но лечение фобий практикуется не только в этой клинике. На рынке есть еще ряд приложений, которые с помощью терапевта эффективно помогают пациентам с различными фобиями. Рассмотрим Phobia Exposure VR и Virtually Better.

Phobia exposure VR постепенно погружает пользователя в среду виртуальной реальности, связанную со страхом пользователя [7]. Это приложение значительно менее продвинутое, в отличие от приложения из клиники «Сова». Здесь пользователь все время находится в одной комнате. В эту комнату по настройке самого пользователя помещаются объекты фобии. Это могут пауки, змеи и насекомые. Пользователь может настроить их количество, а также начать просто с картинок. Скриншот из приложения представлен на рисунке 3.



Рисунок 3. Скриншот из Phobia Exposure VR

Следующим программным продуктом является Virtually Better [6]. Это решение уже больше напоминает приложение из клиники «Сова». Здесь пользователя помещают в различные локации и все также предлагают постепенно столкнуться со своим страхом. Среди фобий есть боязнь полетов, высот, погодных условий и темноты. Скриншот из Virtually Better представлен на рисунке 4.

На сайте Virtually Better можно найти видеоролик, демонстрирующий само приложение [6]. В ролике видно, что пациенту, как и в клинике «Сова», дается возможность перемещаться по локации.

Изучая вышеперечисленные приложения, можно сделать вывод, что они сталкивают пользователя с его страхом путем исследования уровня, на котором присутствуют объекты страха. Однако технологии виртуальной реальности позволяют сделать куда больше. Контроллеры виртуальной



реальности позволяют взаимодействовать с различными объектами. С помощью контроллеров можно перемещать объекты, держаться за объекты, нажимать кнопки и т.д. В технологии виртуальной реальности лежит потенциал, который упомянутые выше приложения не пытаются использовать.



Рисунок 4. Скриншот из Virtually Better

Однако есть приложение, которое использует этот потенциал, – Nope Challenge VR. Приложение представляет собой игру в виртуальной реальности, которая сталкивает пользователя с целым рядом фобий (и не только фобий). В приложении имеется возможность брать объекты, перемещать их, взаимодействовать с персонажами, перемещаться по лестнице и еще ряд действий. Скриншот из Nope Challenge VR представлен на рисунке 5.

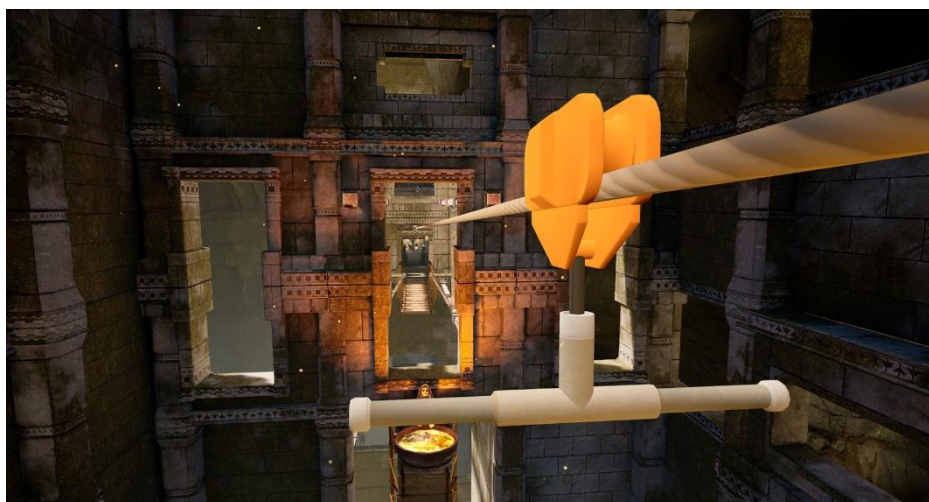


Рисунок 5. Скриншот из Nope Challenge VR

При этом у Nope Challenge VR есть сильный недостаток – это не приложение для лечения фобий. Nope Challenge VR разрабатывалось как развлекательное приложение, не нацеленное на медицинские цели. Также нет никаких доказательств, что это приложение действительно может помочь кому-то в преодолении фобии.

Это оставляет открытым вопрос: «Может ли повышение степени интерактивности в приложении увеличить эффективность терапии?». Для ответа на этот вопрос было разработано приложение виртуальной реальности для помощи в терапии акрофобии (боязни высоты) с интерактивными элементами. Также было проведено тестирование на фокус-группе. Пользователям из фокус-группы предстояло пройти три испытания в приложении. Все испытания связаны с боязнью высоты и постепенно увеличиваются в сложности (как и в вышеперечисленных примерах).

Начальная секция не включает в себя интерактивных элементов. Пользователям необходимо пройти по мосту со стеклянными бортиками на высоте 35 этажа. В первом испытании встречаются интерактивные элементы в виде горизонтальных и вертикальных лестниц. Скриншот первого испытания представлен на рисунке 6.

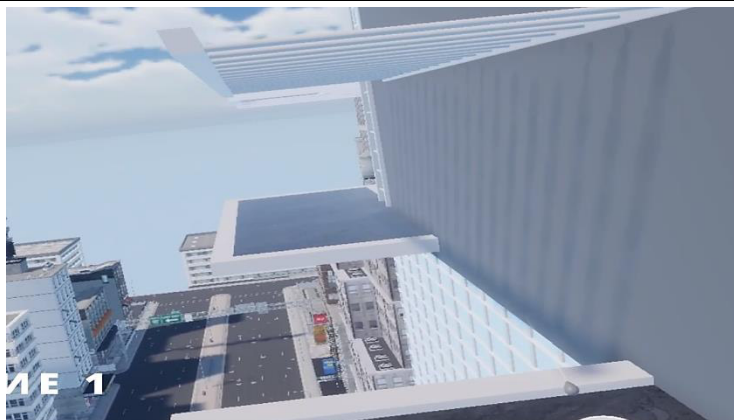


Рисунок 6. Скриншот первого испытания

Во втором испытании участникам было необходимо передвигаться по мостам из деревянных досок, убирая коробки на пути, что также заставляет посмотреть вниз с высоты 35 этажа. Это добавляет уровень взаимодействия, где пользователю необходимо не только перемещать своего персонажа, но и взаимодействовать со средой. Скриншот второго испытания представлен на рисунке 7.



Рисунок 7. Скриншот второго испытания

В третьем и финальном испытании пользователи должны были спуститься по лестнице на стену небоскреба, после чего взяться за доску с ее помощью переместиться на следующую лестницу. Скриншот третьего испытания представлен на рисунке 8.



Рисунок 8. Скриншот третьего испытания

Для разработки приложения необходимо было создать 3D-модели небоскреба, его интерьера, а также доработать модель города. Все модели были созданы с помощью программы Blender и экспортированы в формате FBX вместе с текстурами в разрешении 1024x1024 пикселей. Это помогло создать реалистичную картинку в конечном продукте. Скриншот модели интерьера из Blender представлен на рисунке 9.

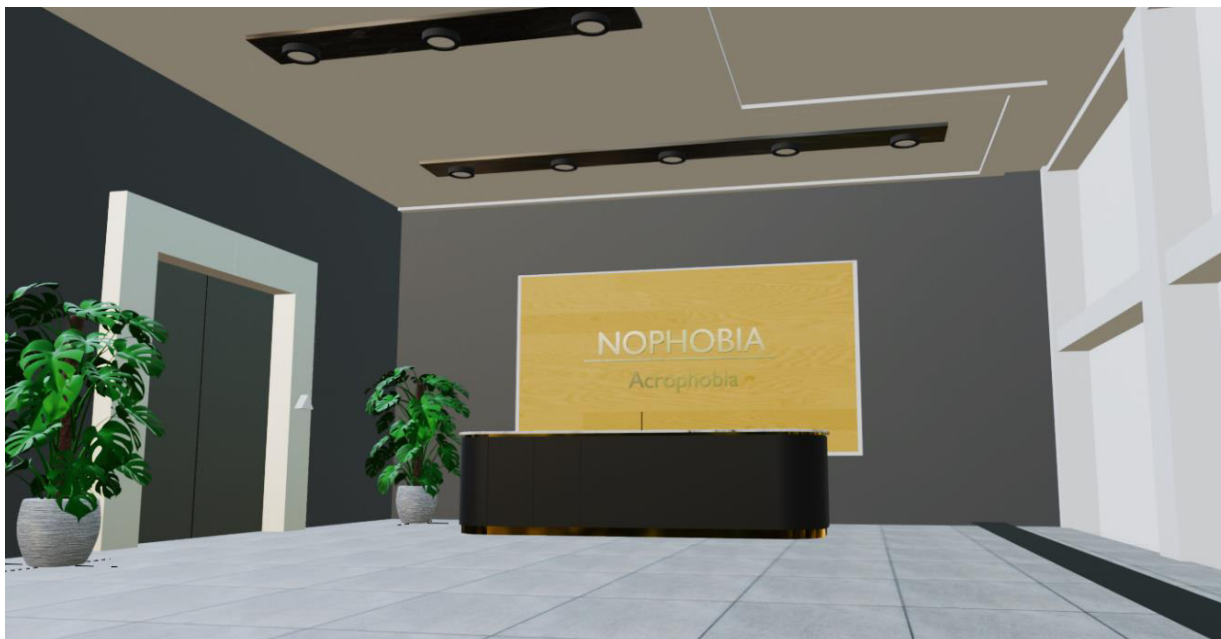


Рисунок 9. Скриншот модели интерьера

После разработки все модели были экспортированы в Unity, где уже была собрана сцена с использованием High Defenition Rendering Pipeline вместе с Reflection Probe, что помогло сделать реалистичное освещение. Скриншот с Reflection Probe (слева) и без (справа) в Unity представлен на рисунке 10.

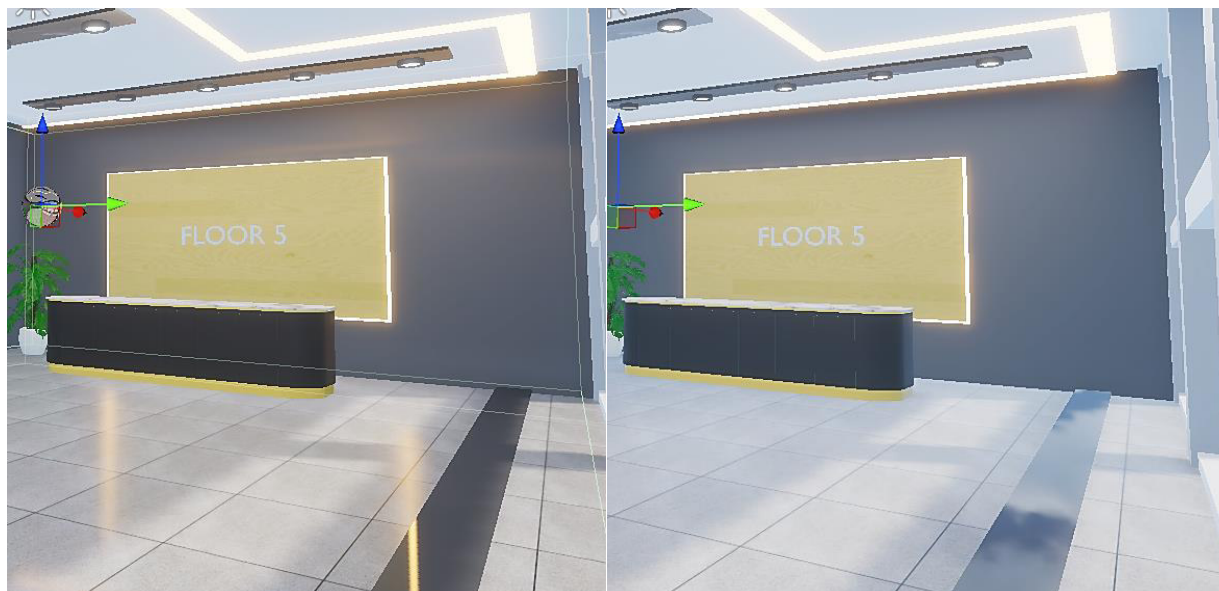


Рисунок 10. Reflection probe

Все эксперименты проводили, используя очки виртуальной реальности HTC Vive Pro. Благодаря встроенным наушникам в шлеме удалось добиться большей степени погружения с помощью звука. В разработке используются звуки ветра, громкость которых динамически регулируется в зависимости от положения пользователя. Шлем использовался вместе с контроллерами и датчиками движения для



отслеживания перемещения участника. Это позволило сделать взаимодействие с приложением виртуальной реальности удобным и понятным.

Помимо этого, в продукт был добавлен пользовательский интерфейс для настройки уровней, а также подсказки, чтобы человеку даже без опыта в виртуальной реальности было несложно освоиться в программе. Скриншот интерфейса представлен на рисунке 11.



Рисунок 11. Скриншот пользовательского интерфейса

### Результаты экспериментов

Всего проведено две сессии тестирования с шестью участниками. Вторая сессия была проведена через неделю после первой. Также после прохождения участникам было предложено ответить на ряд вопросов. В течение сессий у всех постоянно измерялся пульс. Результаты измерений максимального пульса участников на обоих сеансах показан на рисунке 12.

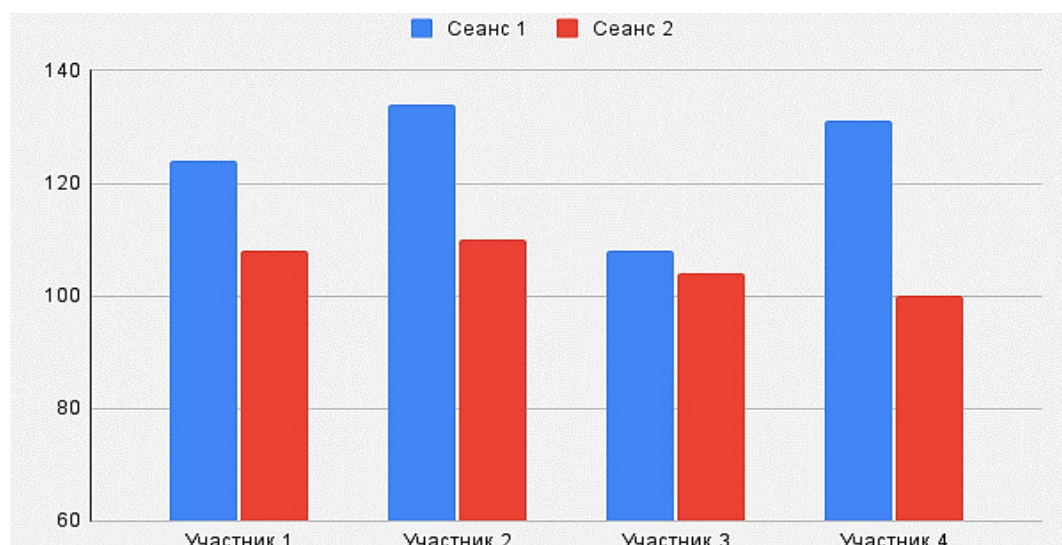


Рисунок 12. Результаты измерений пульса участников

После прохождения испытаний участникам был задан вопрос: «Хотели бы вы пройти подобный сеанс снова?». На рисунке 13 представлена статистика ответов на этот вопрос.



Рисунок 13. Результаты ответов на вопрос о наличии желания вновь пройти тест

### Обсуждение результатов

Согласно данным, представленным на рисунке 12, можно сделать вывод об успешном эффекте десенсибилизации (уменьшения тревожности), то есть об эффективности приложения как помощника в терапии акрофобии. А результаты на рисунке 13 говорят о высоком желании участников попробовать терапию снова. Именно большой процент участников с таким желанием может повысить эффективность терапии. У пациентов будет большее желание вернуться и побороть фобию путем экспозиционной терапии в виртуальной реальности.

Стоит учесть, что вывод о том, что более высокая степень интерактивности действительно увеличит эффективность терапии, делать рано, так как выборка довольно маленькая. Однако результаты говорят о том, что дальнейшие исследования в этой сфере не будут бесполезны и есть высокая вероятность доказать выдвинутую гипотезу.

### Выводы

В ходе изучения различных способов терапии фобий было выявлено, что экспозиционная терапия в виртуальной реальности действительно может помочь в лечении фобий. Анализ результатов экспериментов свидетельствует о том, что наличие интерактивных элементов может повысить степень эффективности терапии фобий в виртуальной реальности, однако для полноценного доказательства гипотезы необходимы дальнейшие исследования в данном направлении.

### Список литературы

1. Фобии: симптомы, лечение, диагностика. URL: [https://comfort.sovamed.ru/zabolevaniya/fobii-simptomy-lechenie-diagnostics/?utm\\_referrer=](https://comfort.sovamed.ru/zabolevaniya/fobii-simptomy-lechenie-diagnostics/?utm_referrer=) (дата обращения: 14.08.2025).
2. William W.E., Bienvenu O. J., Meloyan B. Specific phobias // The Lancet Psychiatry. 2018. No 5-8. Pp. 678–686.
3. Medication for Phobias. URL: <https://cogbtherapy.com/medication-for-phobias> (дата обращения: 14.08.2025).
4. VR в медицине: лечение фобий и реабилитация с помощью виртуальной реальности. URL: <https://medgorod-clinic.ru/stati/vr-v-meditsine--lechenie-fobiy-i-reabilitatsiya-s-pomoshchyu-virtualnoy-realnosti/> (дата обращения: 14.08.2025).
5. VR-психотерапия. URL: <https://voronezh.sovamed.ru/vr-psihoterapiya> (дата обращения: 14.08.2025).
6. Phobias Suite. URL: <https://virtuallybetter.com/Phobias/> (дата обращения: 14.08.2025).
7. Phobia Exposure VR. URL: [https://store.steampowered.com/app/2509570/Phobia\\_Exposure\\_VR/?l=russian](https://store.steampowered.com/app/2509570/Phobia_Exposure_VR/?l=russian) (дата обращения: 14.08.2025).

## Алгоритм адаптивного разбиения облака точек с использованием иерархии ограничивающих объемов

Е. С. Глумова

Нижегородский государственный технический университет им. Р. Е. Алексеева,  
Нижний Новгород, Россия

**Аннотация.** В исследовании представлены результаты разработки алгоритма разбиения пространства на основе структуры иерархии ограничивающих объемов BVH (Bounding Volume Hierarchy) для задачи трансляции пространственных данных. Цель исследования – разработать алгоритм разбиения облака точек, адаптирующийся к локальной плотности распределения точек. В данной работе предлагается следующий подход разбиения облаков точек на основе BVH-структуры с адаптивным критерием стоимости разбиения point-SAH с поправками на контраст плотностей и высот, статистической проверкой на наличие явной кластерной структуры данных на основе алгоритма g-means, динамическими порогами на квантованной решетке. Комбинирование этих механизмов позволяет ускорить процессы импорта облаков точек за счет более качественного распределения точек по блокам, минимизировать пустые блоки и облегчить обработку на аппаратных ресурсах, не обладающих высокой вычислительными мощностями. Представлено подробное описание алгоритма, его математической модели, а также эксперименты, демонстрирующие влияние параметров на качество разбиения. Обсуждаются ограничения текущего исследования и направления дальнейшей работы.

**Ключевые слова:** облако точек, трехмерное моделирование, иерархические структуры, BVH, SAH, g-means, CloudCompare.

## Algorithm Adaptive Point Cloud Partitioning based on Bounding Volume Hierarchy

E. S. Glumova

Nizhny Novgorod State Technical University named after R.E. Alekseev, Nizhny Novgorod, Russia

**Abstract.** The research presents a novel space partitioning algorithm based on Bounding Volume Hierarchies (BVH) designed for the efficient translation of spatial data. The research objective is to develop a point cloud partitioning algorithm that adapts to local variations in point density. To this end, this paper proposes a BVH construction approach featuring several key components: an adaptive point-SAH splitting criterion with cost adjustments for density and height contrast; a statistical test based on the g-means algorithm to detect distinct cluster structures; and dynamic termination thresholds applied on a quantized grid. The combination of these mechanisms accelerates point cloud import and processing by ensuring a more balanced distribution of points across blocks, minimizing empty partitions, and reducing the computational load on low-power hardware. The paper provides a detailed description of the algorithm, its formal mathematical model, and a series of experiments demonstrating the impact of its parameters on the partitioning quality. Finally, the limitations of the current research and potential avenues for future work are discussed.

**Keywords:** point cloud, 3D modelling, Hierarchical Structures, BVH, SAH, g-means, CloudCompare.

### Введение

Трехмерное моделирование играет все более важную роль в управлении жизненным циклом промышленных и гражданских объектов. В основе этого процесса лежат технологии наземного лазерного сканирования (НЛС), обеспечивающие получение высокоточных цифровых моделей. Одной из наиболее перспективных методологий является Scan-to-BIM, позволяющая интегрировать данные НЛС в цифровые информационные модели (ЦИМ) для решения задач на всех этапах жизненного цикла объекта – от проектирования до эксплуатации и демонтажа [1].

Результатом НЛС является облако точек – неструктурированный набор трехмерных координат. Работа с ним сопряжена с проблемой большого объема данных, который накладывает серьезные ограничения на методы их хранения, обработки и трансляции между программными средами. Для решения этой проблемы традиционно применяются иерархические методы пространственного разбиения. Принято различать структуры, управляемые пространством, и структуры, управляемые данными. Несмотря на преимущества структур, управляемых данными, стандартные алгоритмы их построения не в полной мере адаптируются к локальным статистическим особенностям распределения точек. Гипотеза данной работы заключается в том, что интеграция статистического анализа плотности непосредственно в критерии построения иерархии ограничивающих объемов BVH позволит

значительно повысить качество разбиения, формируя блоки, которые более точно соответствуют семантическим кластерам в облаке точек.

Целью данного исследования является разработка алгоритма разбиения облака точек на основе BVH, адаптирующегося к локальной плотности. Для проверки гипотезы предложен алгоритм построения BVH-дерева, включающий следующие ключевые элементы:

- адаптивный критерий стоимости разбиения point- SAH с поправками на контраст плотностей и высот;
- статистическая проверка на наличие явной кластерной структуры данных на основе алгоритма g-means;
- динамический порог для листовых узлов, зависящий от их физического объема, что предотвращает избыточное дробление в разреженных областях;
- механизм устойчивого разделения на основе квантованной решетки и строгих полуинтервалов для исключения потерь или дублирования точек.

Предложенная комбинация эвристик и механизмов направлена на создание разбиения, которое не только ускоряет процессы импорта и обработки данных за счет сбалансированного распределения нагрузки, но и формирует блоки, более соответствующие естественной структуре сканируемого объекта.

### Обзор существующих подходов

Иерархические структуры являются фундаментальным инструментом для организации пространственных данных. Классические подходы, такие как k-d деревья и октодеревья, имеют известные ограничения при работе с неоднородными облаками точек. K-d дерево может становиться чрезмерно глубоким и создавать геометрически несбалансированные ячейки, тогда как октодерево страдает от генерации множества пустых узлов в разреженных областях и недостаточной детализации в плотных. Для решения этих проблем были предложены гибридные структуры, например, сочетающие октодерево на верхних уровнях с локальными R\*-деревьями в листьях [2]. Однако построение таких сложных индексов для сверхбольших облаков точек сопряжено со значительными вычислительными затратами.

С ростом объемов данных ключевыми направлениями исследований стали обработка данных, превышающих объем оперативной памяти, и параллелизм. Значимой работой является метод, который позволяет строить октодерево для миллиардов точек за линейное время путем потоковой обработки данных с диска [3]. Этот подход, реализованный в системе Potree [4], стал стандартом для веб-визуализации. Для ускорения построения на параллельных архитектурах, таких как GPU, широко используются методы, основанные на кодах Мортон [5], которые сводят задачу построения иерархии к высокоэффективной сортировке. Однако качество таких деревьев, как правило, уступает структурам, построенным с использованием более сложных эвристик, поскольку коды Мортон эквивалентны равномерному разбиению пространства и не учитывают реальную форму облака точек.

Наиболее гибкой структурой, управляемой данными, является иерархия ограничивающих объемов BVH. В отличие от октодеревьев, BVH группирует геометрию, а не делит пространство, что делает структуру изначально адаптивной. Для построения качественных BVH широко применяется эвристика площади поверхности SAH [6], которая оценивает «стоимость» разделения узла на основе площадей ограничивающих объемов и количества примитивов в них. Критерий SAH нацелен на минимизацию вероятности пересечения условного запроса (например, луча при трассировке) с дочерними узлами и стал стандартом для построения высококачественных деревьев. Однако классический SAH предполагает равномерное распределение запросов по объему узла, что не вполне релевантно для задачи разбиения облака точек, где целью является семантическая и плотностная однородность блоков.

Развитие подходов к обработке «вне ядра» также повлияло на форматы хранения данных. Если ранние системы, такие как EPT (Entwine Point Tiles) [7], генерировали огромное число мелких файлов, то современные форматы, такие как COPC (Cloud-Optimized Point Cloud) [8], интегрируют иерархический индекс непосредственно в единый laz-файл. Это подтверждает тенденцию на использование встроенных иерархических структур для обеспечения быстрого потокового доступа к данным. Такие библиотеки, как PDAL [9] и LAStools [10], предоставляют конвейеры для обработки и преобразования данных в эти форматы. Тем не менее, они в первую очередь ориентированы на хранение и трансляцию,



а не на семантически осмысленное разбиение сцены. Таким образом, остается актуальной задача совершенствования самих критериев построения иерархий, чтобы они лучше соответствовали внутреннему устройству и плотностным характеристикам облаков точек.

### Теория

Полный цикл работы алгоритма разбиения облака точек на основе BVH-структуры, адаптирующегося к локальной плотности, отражает рисунок 1.

Процесс состоит из трех основных этапов:

- квантизация пространства;
- рекурсивное деление пространства на основе статистической или плотностной проверок;
- постобработка и слияние узлов.

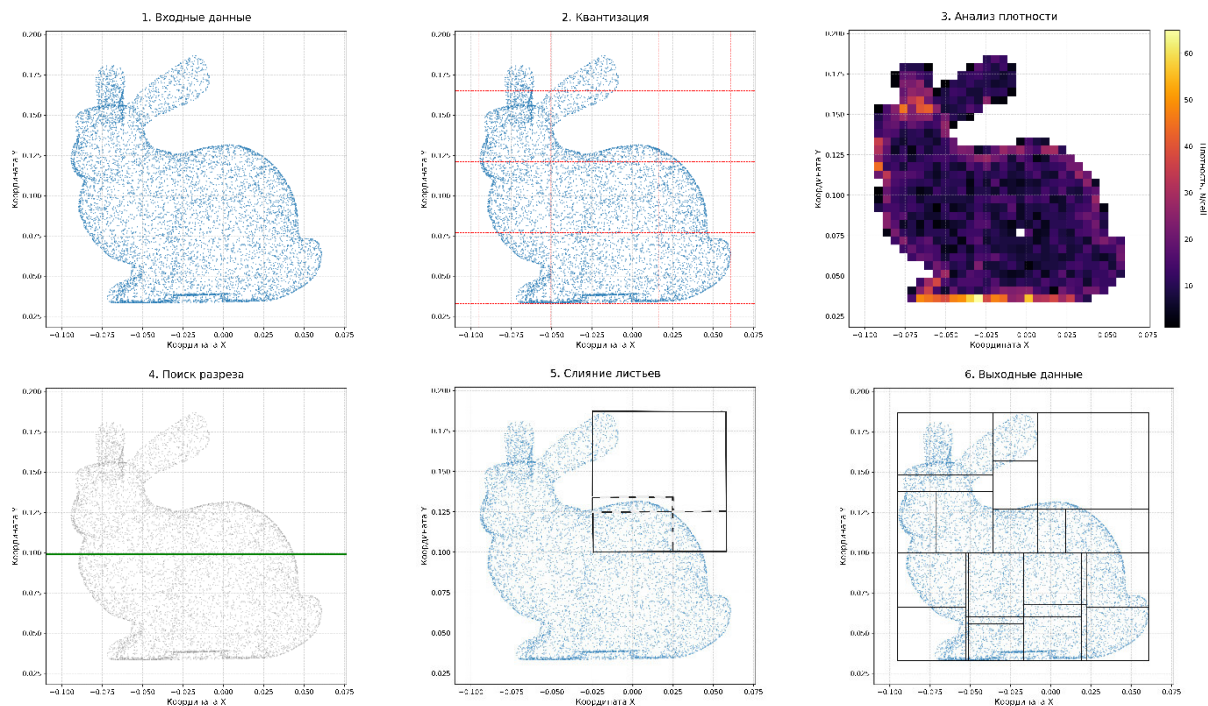


Рисунок 1. Полный цикл работы алгоритма разбиения облака точек на основе BVH-структуры (2D-проекция)

Исходные координаты точек в метрической системе преобразуются в дискретную целочисленную сетку (рис. 1, 1, 2), затем происходит анализ их плотности (рис. 1, 3). Выполняется рекурсивное деление пространства «сверху-вниз». Для каждого узла на основе геометрических и статистических проверок ищется оптимальная плоскость разреза (рис. 1, 4). После завершения основного этапа построения выполняется проход по дереву «снизу-вверх», в ходе которого избыточно мелкие соседние листья могут быть объединены (рис. 1, 5) для получения итогового сбалансированного набора блоков (рис. 1, 6).

Формальная постановка задачи выглядит следующим образом. Дано конечное множество точек  $X = \{x_1, x_2, \dots, x_n\}$ , являющееся подмножеством пространства  $R^3$ . Требуется найти иерархическую структуру  $T$ , которая определяет партицию множества  $X$  на  $m$  непересекающихся подмножеств  $L_1, L_2, \dots, L_m$ . Эти подмножества  $L_i$  соответствуют листовым узлам дерева  $T$  и удовлетворяют следующим свойствам разбиения:

- объединение всех подмножеств равно исходному множеству:  $\bigcup_{i=1}^m L_i = X$ ;
- подмножества попарно не пересекаются:  $L_i \cap L_j = \emptyset$  при  $i \neq j$ ;
- каждое подмножество непустое:  $L_i \neq \emptyset$ .

Искомое разбиение  $P^* = \{L_1, \dots, L_m\}$  и соответствующая ему иерархия  $T^*$  должны быть оптимальными с точки зрения комплексной целевой функции  $C(T)$ , которая минимизирует геометрические и максимизирует статистические свойства однородности элементов разбиения. Задача

нахождения оптимальной иерархии  $T^*$  решается путем локальной оптимизации на каждом шаге рекурсивного разбиения с использованием жадного подхода, логика которого представлена на рисунке 2.

Для обеспечения вычислительной устойчивости все координаты исходного облака точек преобразуются в целочисленную сетку. Процесс построения дерева  $T$  эквивалентен рекурсивному разбиению множеств. Корневому узлу дерева соответствует исходное множество  $X$ . Каждый внутренний узел соответствует некоторому подмножеству  $S \subseteq X$ .

Классическая эвристика SAH для трассировки лучей оценивает стоимость разделения узла. Данный критерий модифицируется для задачи разбиения облака точек. Во-первых, опускается константа стоимости обхода узла ( $C_T$ ), так как целью является не минимизация глубины дерева, а качество итоговых листовых узлов.

Во-вторых, вводятся дополнительные члены за разделение по границам с высоким контрастом плотности и высот.

В результате модифицированная стоимость point-SAH ( $C'_{split}$ ) для разреза по оси  $j$  и порогу  $t$  определяется следующим образом (1):

$$C'_{split}(S, j, t) = (SA(AABB(S_L)) \cdot |S_L| + SA(AABB(S_R)) \cdot |S_R|) - \lambda_D \cdot \Delta_{dens} - \lambda_H \cdot \Delta_{height} \quad (1)$$

где  $S_L = \{x \in S \mid x_j \leq t\}$  и  $S_R = \{x \in S \mid x_j > t\}$  – дочерние множества точек;

$|S_L|, |S_R|$  – количество точек (мощность множеств) в левом и правом подузлах;

$SA(AABB(S_L)), SA(AABB(S_R))$  – площади поверхностей  $AABB$  дочерних узлов;

$\lambda_D$  и  $\lambda_H$  – это неотрицательные весовые коэффициенты, управляющие влиянием членов за плотность и высоту соответственно;

$\Delta_{dens}$  – нормированный контраст плотности между дочерними множествами  $S_L$  и  $S_R$ ;

$\Delta_{height}$  – нормированный контраст высот между дочерними множествами  $S_L$  и  $S_R$ .

Контраст плотности ( $\Delta_{dens}$ ) определяется как нормированная разность оценок плотности в дочерних узлах (2).

$$\Delta_{dens} = \frac{|\rho(S_L) - \rho(S_R)|}{\rho_*} \quad (2)$$

Здесь  $\rho(S_i) = \frac{|S_i|}{\mu(AABB(S_i))}$  – оценка плотности точек в узле  $S_i$ , где  $\mu(\cdot)$  – это пространственная мера (объем  $AABB$  для 3D-анализа или площадь его проекции на плоскость  $XY$  для 2.5D-анализа);

$\rho_*$  – это целевая или средняя плотность в родительском узле, используемая для нормировки.

Контраст по высоте ( $\Delta_{height}$ ) определяется как разность средних  $z$ -координат (высот) дочерних узлов, нормированная на общий диапазон высот в родительском узле (3).

$$\Delta_{height} = \frac{|Z_{avg}(S_L) - Z_{avg}(S_R)|}{Z_{range}(S)}, \quad (3)$$

где  $Z_{avg}(S_i) = \left(\frac{1}{|S_i|}\right) \cdot \sum_{x \in S_i} x^3$  – функция, вычисляющая среднее значение координаты  $Z$  (ось  $j = 3$ ) для множества  $S_i$ ;

$Z_{range}(S) = \max_{x \in S}(x^3) - \min_{x \in S}(x^3)$  – полный диапазон высот в родительском множестве  $S$ , используемый для нормировки.

Минимизация модифицированной целевой функции  $C_{split}$  поощряет такие разрезы, которые не только создают геометрически компактные дочерние узлы, но и проходят по границам областей с различной плотностью или высотой. Таким образом, алгоритм стремится находить не просто геометрически, а семантически и статистически более осмысленные границы для разделения данных.

Для предотвращения статистически необоснованных («псевдо») делений вводится дополнительное ограничение. Разделение множества  $S$  по оси  $j$  рассматривается как допустимый кандидат только в том случае, если распределение точек вдоль этой оси является неоднородным.

Пусть  $\Pi_j(S) = \{x_j \mid x \in S\}$  – множество проекций точек  $S$  на ось  $j$ .

Пусть  $\mathcal{H}_0$  – нулевая гипотеза о том, что выборка  $\Pi_j(S)$  порождена нормальным распределением.

Процесс принятия решения о разделении узла имеет иерархическую структуру. Приоритет отдается статистически значимому разрезу, предложенному проверкой  $g$ -means. В его отсутствие

рассматривается лучший кандидат с точки зрения минимизации  $C_{split}$ . Разрез по SAN принимается только в том случае, если он обеспечивает достаточное снижение стоимости. Если ни один из критериев не дает убедительного кандидата, узел становится листом.

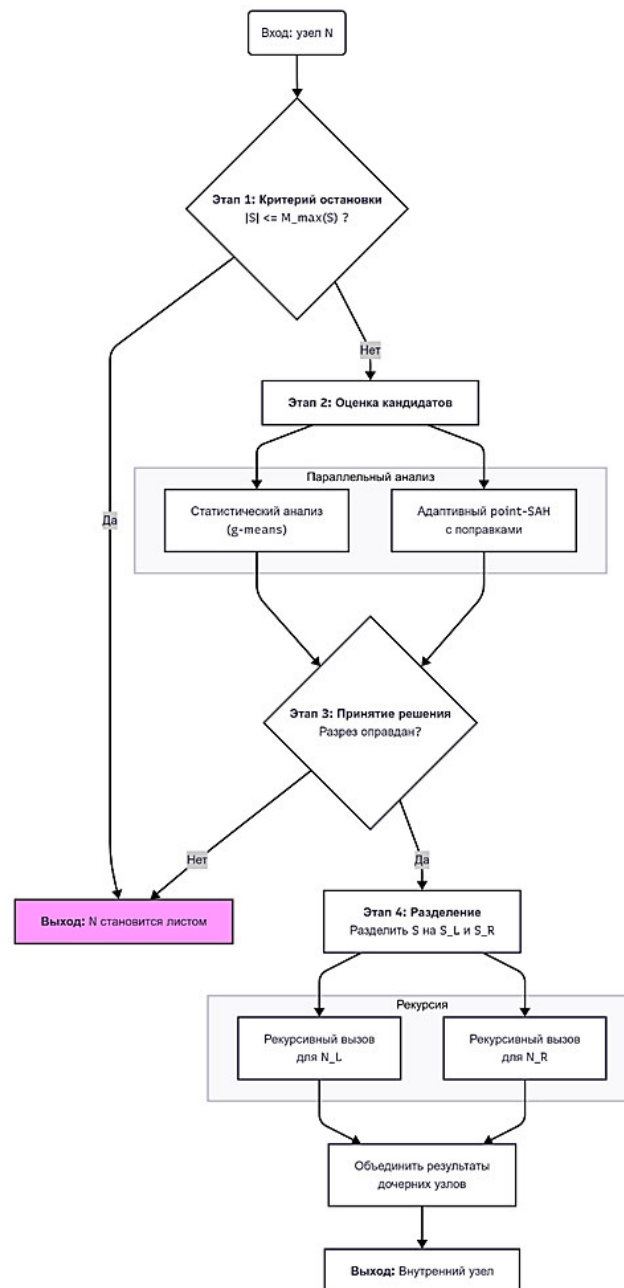


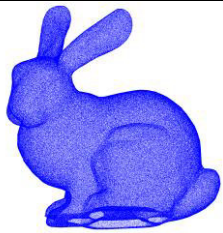
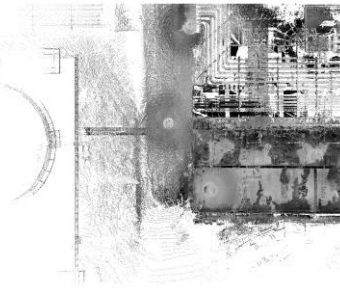
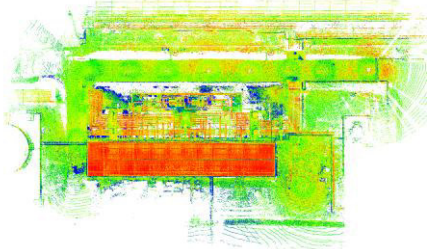
Рисунок 2. Блок-схема процесса деления для одного узла N

Жадный характер рекурсивного построения «сверху-вниз» может приводить к избыточной сегментации. Для исправления локально неоптимальных разрезов применяется этап постобработки. Алгоритм выполняет обход построенного дерева «снизу-вверх». На каждом шаге он проверяет пары соседних листовых узлов. Если суммарное количество точек в двух таких узлах не превышает динамического порога  $M_{max}(S)$  для их общего родительского узла  $S$ , они сливаются, а родительский узел становится листовым. Этот этап позволяет улучшить баланс итогового разбиения, укрупняя излишне детализированные блоки.

### Результаты экспериментов

Для оценки работоспособности алгоритма была проведена серия экспериментов на трех наборах данных, различных по масштабу, плотности и геометрической сложности (табл. 1).

Таблица 1. Характеристика исходных данных

№	Визуализация	Название	Число точек	Особенности
1		Stanford_Bunny	999 966 точек	Характеризуется однородной плотностью и сложной, гладкой геометрией
2		6_data	15 754 340 точек	Характеризуются наличием как крупных плоских поверхностей, так и мелких конструкций (трубопроводы), а также присутствием шума и выбросов
3		Alldata	48 389 508 точек	Данные отличаются сильно выраженной неоднородностью плотности и сложной пространственной организацией

Все эксперименты проводились на персональной вычислительной системе с CPU Intel Core i7-14700K, 64 ГБ ОЗУ, NVIDIA GeForce RTX 4060 Ti).

Поскольку главной целью работы алгоритма является подготовка блоков для дальнейшего использования в средах геометрического моделирования, то ключевым параметром можно считать максимальное число точек в блоке – размера узла. Число точек, как и занимаемый ими битовый объем, ограничивает возможность работы с облаком.

Для оценки влияния параметра размера узла на характеристики построенного BVH-дерева исследовались:

- скорость построения структуры;
- число и распределение листьев;
- устойчивость работы критерия SAH в условиях различных типов облаков точек.

В таблице 2 представлены усредненные показатели производительности алгоритма для трех тестовых наборов данных.

Таблица 2. Усредненные показатели работы алгоритма по наборам данных

№	Набор данных	Число точек	Время построения дерева, с	Скорость построения дерева, млн точек/ с	Число листьев	Число узлов
1	Stanford_Bunny	999 966 точек	5.011945	0.405027	24.375	47.75
2	6_data	15 754 340 точек	14.2330961	1.407944	10.45	19.9
3	Alldata	48 389 508 точек	33.14676842	1.640115	13.89474	26.78947

Необходимо отметить, что для набора Stanford\_Bunny наблюдается низкая скорость построения дерева. Этот набор данных, несмотря на свой малый объем, был включен в экспериментальную часть для демонстрации работы на равноплотных облаках. Однако при таком количестве точек в облаке существенную долю времени занимают накладные расходы на инициализацию структур, выделение

памяти и управление потоками. В результате измеряемая производительность оказывалась заниженной по сравнению с крупными наборами данных.

Для оценки качества построения деревьев была проведена серия экспериментов на трех наборах данных, в которой изменялось только значение порога максимального числа точек в листе при фиксированных прочих настройках алгоритма. На рисунке 3 представлено распределение размеров итоговых листовых узлов при различных значениях порога. График демонстрирует возможность управлять формированием блоков путем изменения числа точек в узле. Использование модифицированного критерия SAN обеспечивает контролируемое распределение размеров листьев при изменении максимального числа точек в листе. При увеличении заданного порога медианный размер получаемых листьев закономерно возрастает. Форма кривых распределения указывает на степень однородности итоговых блоков. При более низких значениях порога кривая имеет более крутой наклон – размеры листьев плотно сконцентрированы в узком диапазоне. С увеличением порога кривая становится более полой, что указывает на увеличение разброса в размерах итоговых блоков.

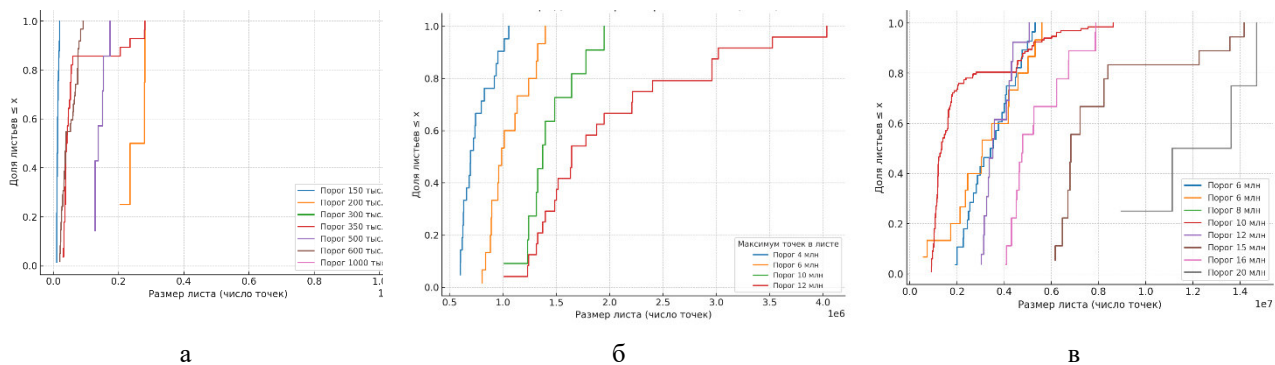


Рисунок 3. Распределение размеров листьев при различных порогах максимального числа точек в листе:  
а – Stanford\_Bunny; б – 6\_data; в – Alldata

Также была исследована зависимость скорости построения дерева от размера листа (рис. 4). График иллюстрирует, что при достижении определенного порога дополнительные разбиения перестают быть оправданными.

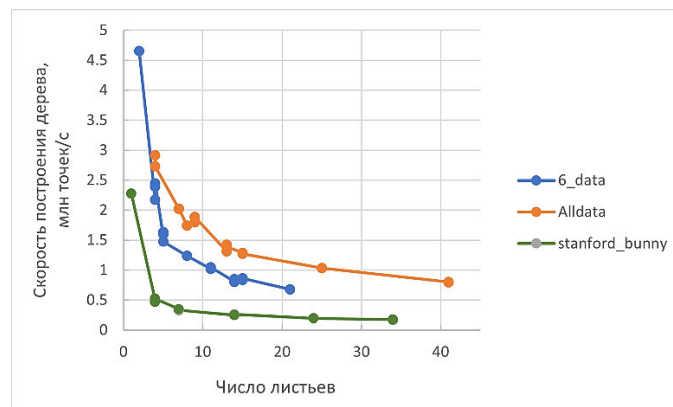


Рисунок 4. Зависимость скорости построения дерева от размера листа

Пример полученной BVH-структуры приведен на рисунке 5. Видно, что дерево является несбалансированным, разные ветви уходят на разную глубину, что порождает листья неоднородного размера. Рисунок показывает, что дерево не делится равномерно, например, значительная часть модели формирует крупный блок с пустотами, а голова оказывается в маленьком блоке (рис. 5, а). Возможна и обратная ситуация (рис. 5, б), когда дерево получается слишком детализированным, появляются мелкие блоки, хотя крупные блоки за счет малого количества точек внутри остаются большими.

Разрез для среднеразмерного набора 6\_data приведен на рисунке 6. Как видим, итоговые ограничивающие объемы формируются с учетом всех точек, включая отдаленные артефакты, что может приводить к созданию избыточно пустых пространств внутри блоков.



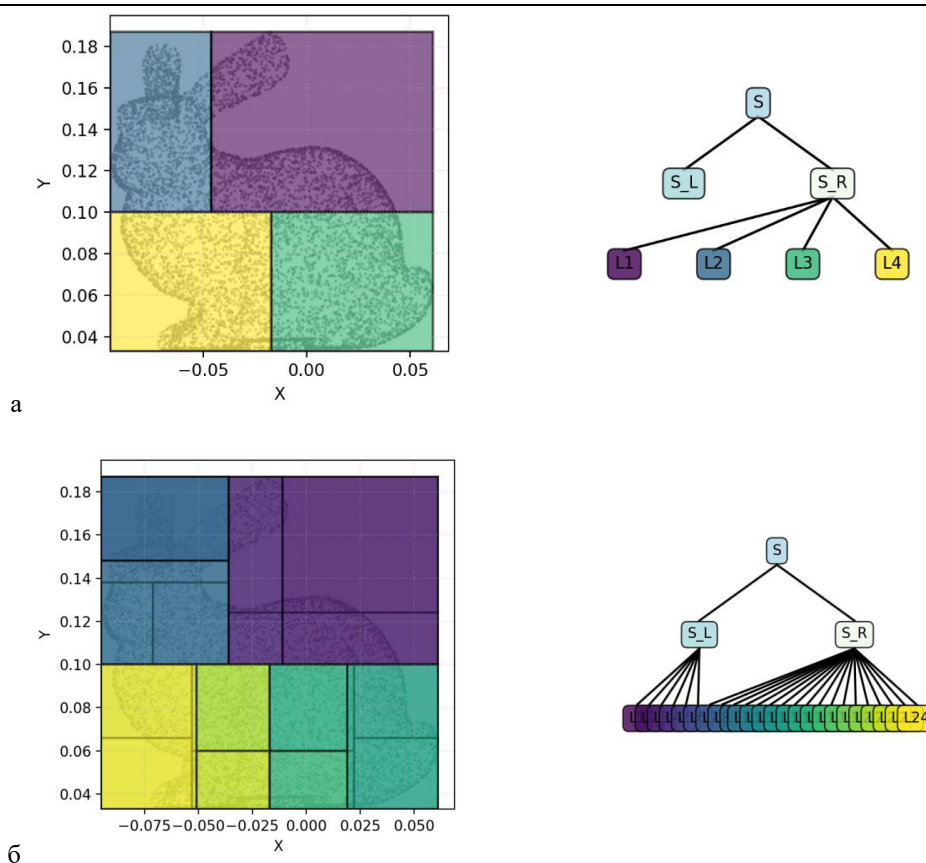


Рисунок 5. Пример несбалансированной BVH-структуры при малом числе блоков для Stanford\_Bunny:  
 а – при малом числе блоко; б – при большом

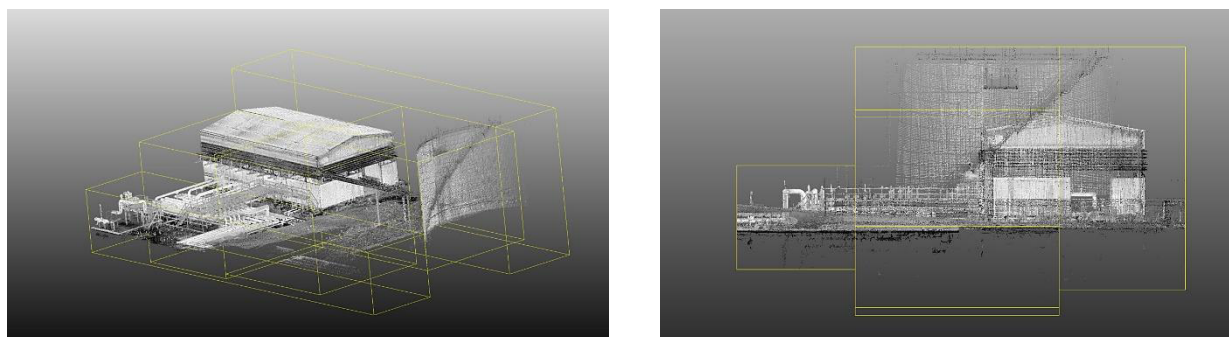


Рисунок 6. Пример разбиения на блоки для набора данных 6\_data

### Обсуждение результатов

Экспериментальное исследование показало, что предложенный алгоритм построения BVH-структуры решает задачу разбиения облака точек. Базовый критерий SAH обеспечивает геометрически «выгодные» разрезы, однако без дополнительных поправок он склонен формировать несбалансированные деревья. Модифицированный критерий с учетом плотностных и высотных поправок, а также с использованием g-means-теста, напротив, приводит к более равномерному распределению размеров листьев, что подтверждается результатами на рисунке 3.

Введение коэффициентов для плотности и высоты делает алгоритм чувствительным к пространственным неоднородностям. Это особенно важно для промышленных объектов и крупномасштабных сканов, где присутствуют как плотные фрагменты (конструкции, оборудование), так и разреженные (открытые площадки, шум, выбросы). Поправки предотвращают чрезмерное укрупнение или, наоборот, чрезмерное измельчение листьев, но требуют тонкой настройки.

Ограничение на число узлов может определять размер блока для процессов экспорта или импорта в среды геометрического моделирования. При слишком малом значении формируется большое число

мелких блоков и наблюдается резкое замедление построения и рост числа операций ввода/вывода. И наоборот, задав слишком большой порог, можно получить неудобные для обработки блоки.

### **Заключение**

В данной работе представлен алгоритм разбиения облака точек на основе иерархии ограничивающих объемов структуры с критерием стоимости разбиения point-SAH, поправками на контраст плотностей и высот, статистической проверкой на наличие явной кластерной структуры данных на основе алгоритма g-means, динамическими порогами на квантованной решетке. Алгоритм демонстрирует стабильность и предсказуемость результатов даже при вариации параметров, а также предоставляет интуитивно понятные средства управления разбиением. Визуальный и статистический анализ подтвердил, что полученные блоки семантически и геометрически соответствуют объектам. По сравнению с октодеревом или равномерной сеткой BVH создает меньше пустых регионов и умеет останавливать деление в разреженных зонах, что приводит к меньшему числу файлов без потери детализации в плотных областях.

В дальнейшем необходимо исследовать влияние плотностных и высотных надбавок на качество и время разбиения, определить оптимальные значения в зависимости от свойств облака точек. Перспективным направлением является разработка методов динамического выбора коэффициентов на основе локальных статистик (средняя плотность, вариативность высот, наличие выбросов). Такой подход позволит повысить универсальность алгоритма без ручной настройки.

### **Список литературы**

1. Глумова Е. С., Дудаков Н. Ю., Филинских А. Д. Обработка результатов наземного лазерного сканирования для создания цифровых моделей // Научно-технический вестник Поволжья. 2024. № 12. С. 119–124. EDN: JMRRPA
2. A hybrid spatial indexing structure of massive point cloud based on octree and 3D R\*-tree / W. Wang et al. // Applied Sciences. 2021. Vol. 11, iss. 20. P. 9581.
3. Schütz M., Ohrhallinger S., Wimmer M. Fast out-of-core octree generation for massive point clouds // Computer Graphics Forum. 2020. Vol. 39, iss. 7. Pp. 155–167.
4. PotreeConverter. Release 2.0 // GitHub. URL: <https://github.com/potree/PotreeConverter/releases/tag/2.0> (дата обращения: 10.08.2025).
5. Pantaleoni J., Luebke D. HLBVH: Hierarchical LBVH construction for real-time ray tracing of dynamic geometry // Proceedings of the Conference on High Performance Graphics. 2010. С. 87–95.
6. Bounding Volume Hierarchies // Physically Based Rendering: From Theory to Implementation. URL: [https://pbr-book.org/3ed-2018/Primitives\\_and\\_Intersection\\_Acceleration/Bounding\\_Volume\\_Hierarchies](https://pbr-book.org/3ed-2018/Primitives_and_Intersection_Acceleration/Bounding_Volume_Hierarchies) (дата обращения: 10.08.2025).
7. Entwine Point Tiles Specification. Hobu Inc., 2021. URL: <https://entwine.io> (дата обращения: 09.08.2025).
8. Cloud Optimized Point Cloud Specification 1.0. Hobu Inc., 2021. URL: <https://copc.io> (дата обращения: 09.08.2025).
9. Point Data Abstraction Library: Documentation. PDAL Project, 2023. URL: <https://pdal.io> (дата обращения: 11.08.2025).
10. LAStools: Efficient Lidar Processing Software. Rapidlasso GmbH, 2022. URL: <https://rapidlasso.de> (дата обращения: 11.08.2025).



## Анализ функционально ориентированной оценки передачи и восстановления геометрической модели

Е. А. Курушин

Нижегородский государственный технический университет им. Р. Е. Алексеева,  
Нижний Новгород, Россия

**Аннотация.** В работе рассматривается применение функционально ориентированной методики оценки передачи и восстановления геометрических моделей при взаимодействии систем 3D-моделирования. На примере передачи сцены из Autodesk 3ds Max в Blender проведена количественная оценка сохранения параметров модели при использовании форматов \*.abc, \*.obj и \*.stl. Установлено, что формат \*.abc обеспечивает наибольшую полноту передачи геометрии и анимации, тогда как \*.obj и \*.stl ограничиваются лишь сохранением базовой геометрии. Полученные результаты подтверждают применимость функционально ориентированного подхода, однако выявляют его ограничения: отсутствие учёта обязательных параметров (например, истории построений в CAD-моделях) и специфик реализации форматов. Это обуславливает необходимость дальнейшего развития методики с включением классификации параметров по критичности и анализа особенностей форматов передачи данных.

**Ключевые слова:** геометрическая модель, оценка потерь, функционально ориентированная оценка, передача данных, CAD, DCC

## Analysis of function-oriented evaluation of 3d geometric model transfer and reconstruction

E. A. Kurushin

Nizhny Novgorod State Technical University. R.E. Alekseev, Nizhny Novgorod, Russia

**Abstract.** The paper presents the application of a function oriented methodology for evaluating the transfer and reconstruction of 3D geometric models between modeling systems. A case study of transferring a scene from Autodesk 3ds Max to Blender was carried out, with quantitative assessment of model parameters preserved through \*.abc, \*.obj, and \*.stl file formats. The results show that the \*.abc format provides the most complete transfer of geometry and animation, while \*.obj and \*.stl are limited to basic geometry preservation. The study confirms the applicability of the function oriented approach, but also reveals its limitations, namely the lack of consideration of mandatory parameters (such as construction history in CAD models) and format-specific implementation issues. This highlights the need to further develop the methodology by introducing parameter criticality classification and incorporating analysis of format-specific features.

**Keywords:** geometric model, loss estimation, data transmission, function oriented evaluation, CAD, DCC.

### Введение

Геометрическая модель изделия является ключевым элементом информационной поддержки жизненного цикла изделия. Она выступает основой для автоматизированного формирования конструкторской документации, проведения инженерных расчётов (включая анализ прочности, динамики, теплопередачи и др.) и пр. Таким образом, корректность и полнота описания геометрической модели напрямую определяют эффективность последующих этапов жизненного цикла изделия.

На практике разработка геометрических моделей может выполняться в различных программных продуктах разных вендоров по определенным причинам, наряду с чем может возникнуть необходимость перехода на другое программное обеспечение в рамках модернизации или интеграции производственного процесса [6, 9, 13]. Однако прямой перенос модели из одной системы в другую, как правило, невозможен. Это связано с использованием проприетарных форматов хранения данных, спецификации которых закрыты и не предназначены для межсистемного обмена.

Для обеспечения совместимости применяются нейтральные форматы передачи данных, такие как IGES, STEP и др., специально разработанные для межплатформенного взаимодействия. Их особенностью является сохранение базовой геометрической информации при частичной или полной утрате ряда дополнительных параметров модели (например, зависимостей построения, атрибутов материалов, метаданных) [4, 10, 12–15]. Таким образом, при трансляции данных между системами

неизбежно возникают информационные потери, влияющие на полноту и корректность модели в системе-приёмнике [1].

В ряде исследований использовалась качественная оценка передачи моделей, фиксирующая факт сохранения или утраты отдельных параметров [6, 8, 9, 11, 13, 14]. Такие оценки не позволяют учесть степень значимости утраченной информации для последующего использования модели.

Функционально ориентированная методика [2], предложенная в современных исследованиях, обеспечивает возможность количественной оценки передачи и восстановления геометрических моделей. Она учитывает как фактическое сохранение параметров, так и их значимость в контексте применения модели. Такой подход позволяет формировать количественную оценку передачи и восстановления данных.

### Постановка задачи

Каждый формат описывается спецификацией или стандартом, определяющим его назначение, параметры и правила описания модели. Так, например, файл формата VRML считается валидным, если он начинается с заголовка `#VRML V2.0 utf8`, синтаксически соответствует грамматике спецификации и использует только допустимые узлы и поля либо их корректные расширения. К обязательным элементам относятся наличие заголовка и хотя бы одного корневого узла в сцене; к рекомендуемым – использование узла `WorldInfo` для описания сцены, а также явное указание `Viewpoint` и `NavigationInfo`, в противном случае применяются значения по умолчанию. Все остальные элементы, такие как `Background`, `Fog`, `Sound`, `Script` и другие, являются опциональными, при этом их порядок и количество не ограничены [3]. Это означает, что при реализации подпрограмм чтения и записи файлов внешних непроприетарных форматов разработчик может не учесть ряд параметров, характеризующих геометрическую модель.

Для корректной оценки передачи геометрической модели недостаточно учитывать только параметры исходного объекта и результаты импорта в систему-приёмник. Необходимо также рассматривать характеристики самого формата данных, используемого при экспорте. Таким образом, постановка задачи сводится к проверке применимости существующей функционально ориентированной методики для оценки корректности передачи геометрических моделей между системами.

### Теория

Алгоритм расчета функционально-информационной метрики представлен на рисунке 1.

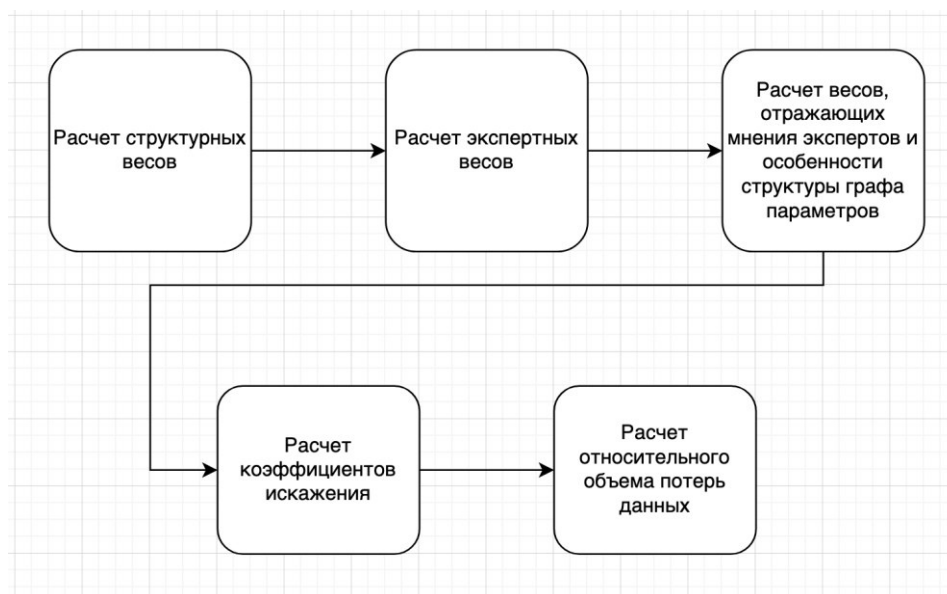


Рисунок 1. Алгоритм расчета функционально-информационной метрики

**Расчет структурных весов** проводится на основании параметров модели, образованных по группам этих параметров с помощью процесса интеграции и дифференциации, основанных на анализе свойств и различий, присущем процедуре классификации.

Процесс построения графа [2] делится на этапы:

1. Обозначается глобальный признак;
2. Определяются первичные критерии оценки (параметры ГМ по которым идет оценка);
3. Формируются локальные признаки для построения дерева.

Глобальным признаком определена сложность восстановления геометрической модели. В качестве локальных свойств первого яруса выделяются геометрические свойства модели, атрибутивная информация о модели и свойства передающего файла, где геометрические свойства модели делятся на свойства сборочной единицы и её отдельных деталей, которые являются локальными признаками второго яруса. Аналогично проводится последующее объединение и распределение параметров по схожим признакам.

Описанная последовательность построения графа параметров геометрической модели характеризует отличительные особенности CAD-систем, тогда как каждый частный случай применения методики может характеризоваться другими описательными характеристиками описательного объекта. Так, по результатам расчета структурных весов параметров, граф может иметь вид, представленный на рисунке 2.

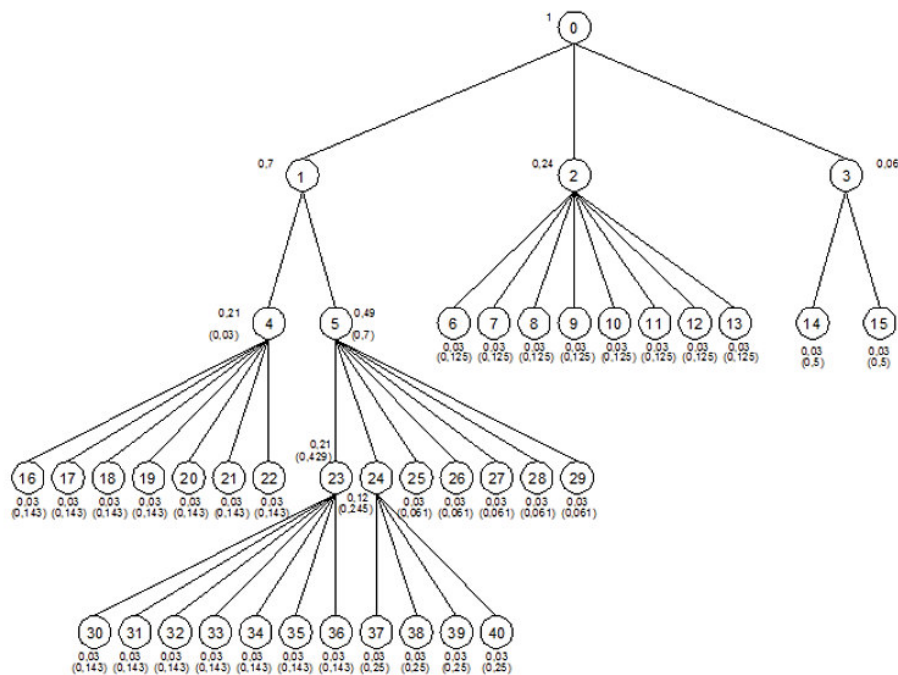


Рисунок 2. Граф параметров геометрической модели с рассчитанными структурными весами

**Расчет экспертных весов** проводится по результатам опроса специалистов, работающих в ППС различных вендоров, на предмет сложности восстановления того или иного параметра ГМ при передаче её из одной системы в другую с помощью метода экспертных оценок. Средняя оценка каждого параметра рассчитывается по формуле

$$w_i^{\exists} = \frac{\sum_{j=1}^m w_{ij}}{\sum_{j=1}^m \sum_{i=1}^m w_{ij}}, \quad (1)$$

где  $w_i^{\exists}$  – вес  $i$ -го параметра, подсчитанный по оценкам всех экспертов;

$w_{ij}$  – вес  $i$ -го параметра, данный  $j$ -им экспертом;

$n$  – число параметров;

$m$  – число экспертов.

$$w_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, \quad (2)$$

$w_{ij}$  – вес  $i$ -го параметра, данный  $j$ -им экспертом;

$x_{ij}$  – оценка  $i$ -го параметра, данная  $j$ -им экспертом;

$n$  – число параметров.

**Расчет весов, отражающих мнения экспертов и особенности структуры графа параметров, проводится по формуле**

$$w_i^{\text{ЭС}} = \frac{w_i^{\text{Э}}}{w_i^{\text{С}} \sum_{i=1}^n \frac{w_i^{\text{Э}}}{w_w^{\text{С}}}}. \quad (3)$$

Для **расчета коэффициентов искажения** передачи геометрическая модель импортируется в систему-приемник и анализируется на предмет наличия параметров. Так, проставляются коэффициент передачи  $i$ -го параметра  $k_i$ , принимающий значения: 0 – параметр передать не удалось; 0,5 – параметр удалось передать не полностью; 1 – параметр передан полностью. Это позволяет рассчитать коэффициент искажения параметров по формуле

$$p_i = 1 - k_i. \quad (4)$$

Для оценки потерь данных при передаче модели между системами рассчитывается **относительный объем потерь данных** как сумма потерь по каждому параметру:

$$K_Z = \sum_{i=1}^n w_i^{\text{ЭС}} p_i. \quad (5)$$

### Результаты экспериментов

Для оценки передачи геометрической модели был проведен эксперимент с использованием предварительно подготовленной модели в программе 3ds Max для оценки потерь данных при передаче в Blender в условиях импортозамещения [5]. Подготовлена модель контейнера и помещена на сцену, в которой имеется ландшафт, представляющий из себя объект plane с травой, выполненной с помощью модификатора Scatter. Также в сцене присутствует камера и один источник освещения (рис. 3). Некоторые элементы имеют анимацию. Так, камера совершает облет вокруг контейнера по кривой Helix, контейнер за это время поднимается на гидравлическом приводе.

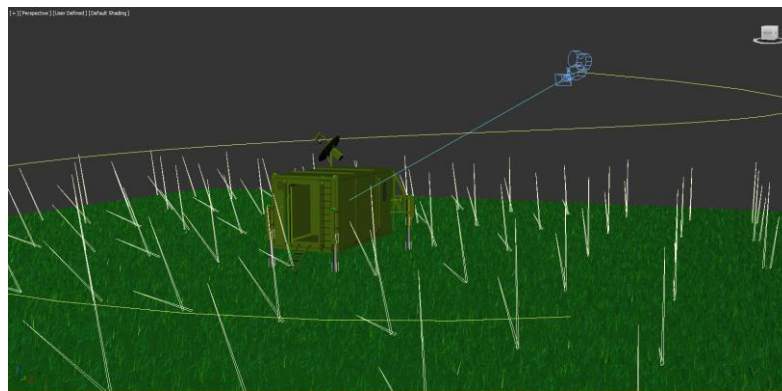


Рисунок 3. Передаваемая сцена

Выделены параметры, по которым строился граф и производилась оценка потерь при передаче геометрической модели:

1. Геометрии
  - 1.1. Группировка элементов
  - 1.2. Объект
    - 1.2.1. Положение
    - 1.2.2. Габариты
  - 1.3. Окружение
    - 1.3.1. Положение
    - 1.3.2. Габариты
  - 1.4. Материалы
    - 1.4.1. Базовый цвет
    - 1.4.2. Шероховатость

- 1.4.3. Отражение
- 1.4.4. Прозрачность
- 1.4.5. Эмиссия
- 1.5. Отбрасывание теней
- 2. Камера
  - 2.1. Положение
  - 2.2. Направление камеры
  - 2.3. Параметры камеры
    - 2.3.1. Объектив
    - 2.3.2. Тип
    - 2.3.3. focal depth
  - 2.4. Привязка камеры к кривой
- 3. Освещение
  - 3.1. Параметры источника освещения
    - 3.1.1. Тип
    - 3.1.2. Множитель
    - 3.1.3. Цвет
    - 3.1.4. Создание теней
  - 3.2. Положение
- 4. Анимация
  - 4.1. Ключи анимации камеры
  - 4.2. Ключи анимации модели

Следовательно, граф с рассчитанными иерархическими весами будет выглядеть так, как показано на рисунке 4 (справа обозначены уровни (ярусы) графа).

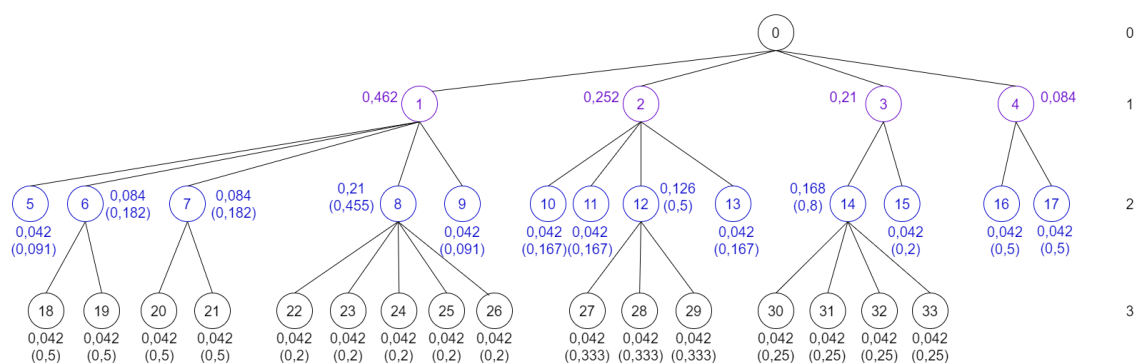


Рисунок 4. Граф параметров геометрической модели

Перечень параметров ГМ, подлежащих оценке:

1. Группировка элементов (геометрия);
2. Положение (геометрия, объект);
3. Габариты (геометрия, объект);
4. Положение (геометрия, окружение);
5. Габариты (геометрия, окружение);
6. Базовый цвет (геометрия, материалы);
7. Шероховатость (геометрия, материалы);
8. Отражение (геометрия, материалы);
9. Прозрачность (геометрия, материалы);
10. Эмиссия (геометрия, материалы);
11. Отбрасывание теней (геометрия);
12. Положение (камера);
13. Направление камеры (камера, параметры);
14. Объектив (камера, параметры);

15. Тип (камера, параметры);
16. focal depth (камера, параметры);
17. Привязка камеры к кривой;
18. Тип (освещение, параметры);
19. Множитель (освещение, параметры);
20. Цвет (освещение, параметры);
21. Создание теней (освещение, параметры);
22. Положение (совещение);
23. Ключи анимации камеры;
24. Ключи анимации модели.

С помощью метода экспертных оценок необходимо получить данные от специалистов, работающих в различных профессиональных программных средах геометрического моделирования, о сложности восстановления того или иного параметра ГМ, при передаче ее из одной программной среды в другую. Оценка проводилась методом непосредственных оценок по 10-балльной шкале, где 1 – параметр легко восстановить, 10 – параметр очень сложно восстановить. Результаты опроса представлены в таблице 1.

Таблица 1. Оценки экспертов

$i/j$	1	2	3	4	5	6	7	8	9	10
1	7	4	9	5	9	6	3	8	3	4
2	6	8	9	5	7	5	6	8	3	9
3	7	9	9	2	7	9	5	7	6	5
4	10	9	9	4	8	6	6	7	4	8
5	9	10	10	5	8	6	8	9	5	8
6	9	6	8	3	8	7	8	9	2	6
7	7	7	9	6	8	8	4	8	3	8
8	7	3	10	5	8	8	5	8	7	5
9	9	4	8	4	8	8	2	8	7	6
10	8	10	8	5	9	4	8	8	2	2
11	6	10	9	3	9	4	4	8	6	8
12	6	6	10	5	8	2	8	7	2	4
13	7	7	9	3	8	9	4	9	3	8
14	8	4	10	8	8	4	7	9	8	7
15	6	2	8	4	10	4	4	10	5	7
16	9	5	9	7	8	8	5	8	7	4
17	9	2	9	5	7	3	3	9	3	7
18	8	4	9	5	8	5	6	7	7	8
19	6	2	10	7	8	8	5	9	6	7
20	9	3	10	5	9	4	8	8	8	7
21	10	3	9	4	9	6	7	8	3	5
22	10	5	9	3	8	3	7	10	4	3
23	6	4	9	4	8	8	5	9	2	6
24	6	6	9	4	8	4	6	9	3	4

В таблице 1  $j$  – порядковый номер эксперта,  $i$  – порядковый номер параметра, определенного для оценки. Весовой коэффициент, отражающий только мнения экспертов, рассчитывается по формуле (1). Итоги расчета экспертных весов представлены в таблице 2.

На основе полученных результатов рассчитывается весовой коэффициент, отражающий мнения экспертов о сложности восстановления  $i$ -го параметра ГМ и особенности структуры графа параметров геометрических моделей по формуле (3). Результаты расчетов приведены в таблице 2.

Таблица 2. Рассчитанные весовые коэффициенты

№ эксперта	$w_i^C$	$w_i^Э$	$w_i^{ЭC}$
1	0,091	0,036	0,09
2	0,5	0,042	0,019

Окончание таблицы 2

№ эксперта	$w_i^c$	$w_i^э$	$w_i^{эс}$
3	0,5	0,043	0,02
4	0,5	0,045	0,021
5	0,5	0,05	0,023
6	0,2	0,041	0,047
7	0,2	0,044	0,05
8	0,2	0,043	0,049
9	0,2	0,041	0,047
10	0,2	0,041	0,047
11	0,091	0,043	0,108
12	0,167	0,036	0,049
13	0,167	0,042	0,057
14	0,333	0,048	0,033
15	0,333	0,037	0,025
16	0,333	0,046	0,032
17	0,167	0,035	0,048
18	0,25	0,044	0,04
19	0,25	0,044	0,04
20	0,25	0,046	0,042
21	0,25	0,040	0,036
22	0,2	0,038	0,043
23	0,5	0,038	0,017
24	0,5	0,037	0,017

Модель была экспортирована во все форматы, представленные на экспорт, таким образом, необходимо проанализировать коэффициент передачи для 15 файлов:

- \*.3ds
- \*.abc
- \*.ai
- \*.ase
- \*.dae
- \*.dwf
- \*.dwg
- \*.dxf
- \*.fbx
- \*.flt
- \*.igs
- \*.obj
- \*.pxproj
- \*.stl
- \*.wrl

Необходимо последовательно передать геометрическую модель из системы S1 (Autodesk 3ds Max 2019) в систему S2 (Blender 2.93.6) и оценить коэффициент передачи по каждому параметру для всех форматов файлов для расчета коэффициента искажения по формуле (4). Каждый экспортированный файл открывается в Blender. При импорте моделей выяснилось, что некоторые из них не поддерживаются, следовательно, коэффициент передачи всех параметров равен 0.

При экспорте модели в формате .fbx возникла ошибка «Version 6900 unsupported, must be 7100 or later», следовательно, можно считать, что данные не передались. Следует также сделать вывод о неудачной передаче сцены посредством формата .dae ввиду того, что его размер составляет 0 КБ.

Значит, коэффициент искажения для данных форматов будет следующим:

- $K_{3ds} = 1$
- $K_{ai} = 1$



- $K_{ase} = 1$
- $K_{dae} = 1$
- $K_{dwf} = 1$
- $K_{dwg} = 1$
- $K_{dx} = 1$
- $K_{fbx} = 1$
- $K_{flt} = 1$
- $K_{igs} = 1$
- $K_{pxproj} = 1$
- $K_{wrl} = 1$

Файл .abc открылся в Blender. На изображении ниже представлен вид с камеры. Сразу стоит отметить, что передача материалов не удалась совсем. Размеры и положение геометрических примитивов соответствуют исходным. Ключи анимации передались практически все: камера движется по траектории, заданной в 3ds Max, контейнер поднимается, однако не опускается лестница. Камера определяется как перспективная (изначально target). Вероятно, дело в другом подходе к организации принципа камеры в продукте Blender. Следует отметить, что камера на всем протяжении анимации сфокусирована на контейнере. Объектив и сопутствующие параметры передались успешно. Что касается источника освещения, то удалось передать только его координаты.

Модель в формате .stl успешно импортировать в Blender, однако она импортировалась как единый объект. Отсутствует возможность проверки положения камеры и освещения для выставления оценок, однако делая выводы из габаритов результирующей полученной модели, можно сделать вывод о том, что источник освещения был проигнорирован при экспорте. Возможность отбрасывания теней отсутствует. Для проверки были добавлены камера и освещение.

Файл .obj успешно импортировался в Blender. Как и ожидалось, сцена импортировалась повернутой на -90 градусов относительно оси x. Камера, освещение и анимация отсутствуют. Выполненный рендер показал, что геометрия не отбрасывает тени. Была также попытка передать материалы. Базовые цвета изменились совершенно и распределились по непонятному признаку. На импортированных объектах имеется ряд переданных параметров, однако приходит мысль, что Blender попросту установил дефолтные значения.

Таблица 3. Коэффициенты передачи модели

Параметр ( $i$ )	$k_i$ .abc	$k_i$ .stl	$k_i$ .obj
<b>Геометрии</b>			
Группировка элементов	1	0	0
Положение (объект)	1	1	1
Габариты (объект)	1	1	1
Положение (окружение)	1	1	1
Габариты (окружение)	1	1	1
Базовый цвет (материалы)	0	0	0,5
Шероховатость (материалы)	0	0	0,5
Отражение (материалы)	0	0	0
Прозрачность (материалы)	0	0	0
Эмиссия (материалы)	0	0	0,5
Отбрасывание теней	0,5	0	0
<b>Камера</b>			
Положение	1	0	0
Направление камеры	1	0	0
Объектив	1	0	0
Тип	0,5	0	0
focal depth	0	0	0
Привязка камеры к кривой	1	0	0

Окончание таблицы 3

Параметр ( <i>i</i> )	$k_i .abc$	$k_i .stl$	$k_i .obj$
Освещение			
Тип	0	0	0
Множитель	0	0	0
Цвет	0	0	0
Создание теней	0	0	0
Положение	1	0	0
Анимация			
Ключи анимации камеры	1	0	0
Ключи анимации модели	0,5	0	0

Таким образом на основе полученных данных относительный объем потерь данных рассчитывается по формуле (5) и составляет  $K_{abc} = 0,505$ ,  $K_{stl} = 0,917$ ,  $K_{obj} = 0,885$ .

### Обсуждение результатов

Подводя итоги проделанной работы можно сделать вывод, что для передачи сцены из Autodesk 3ds Max 2019 в Blender 2.93.6 возможно использовать форматы .abc, .stl, .obj ввиду того, что остальные форматы, доступные в первой системе для экспорта, вторая система не поддерживает. Вопрос о возможности использования формата .fbx остается открытым ввиду того, что Blender указанной версии не поддерживает версию файла, которую установила первая система при экспорте, из-за чего требуется отдельная конвертация. В целом, относительно успешно с поставленной задачей справился только формат abc ввиду успешной передачи камеры и анимации, однако материалы им проигнорированы. Форматом .obj удалось меньше число параметров, в то время как посредством формата .stl удалось передать лишь геометрию. Корректность передачи зависит как от выбранного формата, так и от особенностей реализации импорта/экспорта в конкретных программных продуктах, что явно демонстрируется на примерах форматов .fbx и .dae.

Проведенное исследование показало, что форматы, посредством которых возможен перенос из системы один в систему два, заведомо не имеют возможности описания определенных параметров. Из этого следует необходимость учета спецификации формата. С учетом этого данные заведомо могут потеряться в процессе переноса из-за различий реализации программного метода чтения и записи формата разработчиком программного продукта. Возникает необходимость комплексного анализа обозначенных подпрограмм на предмет реализации заявленных в спецификации поддерживаемых параметров.

Наряду с этим необходимо учитывать параметры геометрической модели, без успешного переноса которых дальнейшая поддержка моделей будет сопоставима с повторной разработкой модели в системе два. Так, в DCC средах основой является полигональная модель, обладающая материалами, нормальными, скелетом анимации и прочими параметрами, в то время как в CAD моделях важной составляющей является история построения модели, отсутствие которой значительно усложняет процесс внесения изменений в такую модель, что в перспективе может означать существенные трудовые и финансовые затраты на стадиях информационной поддержки жизненного цикла изделий. Это обуславливает необходимость в введении дополнительных параметров оценки передачи геометрической модели.

### Выводы

Функционально-ориентированная оценка передачи и восстановления геометрических моделей является научно-обоснованным показателем передачи и восстановления геометрических моделей в профессиональных программных средах. По результатам проделанной работы предлагается введение дополнительных параметров расчета, отражающих особенности «нейтральных форматов» передачи данных. Основанные на спецификациях и стандартах форматы нередко имеют проблему «диалектов». Проблема здесь – фактическая реализация этих стандартов в конкретном программном обеспечении.

Выбор оптимального пути переноса геометрических данных должен учитывать особенности каждого конкретного формата и его возможностей, а также ключевых параметров, при отсутствии которых рассмотрение определенного формата для передачи геометрической модели между системами нецелесообразно. Так, каждый частный случай расчета оценки передачи и восстановления ставит во главу определенный ряд параметров, без успешного переноса которых рассматривать определенный путь передачи является нецелесообразным.

### Список литературы

1. Об эффективности обмена информацией между САПР / Л. И. Райкин, И. Л. Райкин, И. Н. Мерзляков и др. // *Universum: технические науки*. 2014. №2 (3). Ст. 4.
2. Филинских А. Д., Райкин Л. И. Функционально-ориентированная оценка передачи и восстановления геометрических моделей // *Вестник БГТУ им. В. Г. Шухова*. 2014. №4. С. 176–179.
3. The Virtual Reality Modeling Language Specification // Stanford Computer Graphics Laboratory. URL: [http://www.graphics.stanford.edu/courses/cs248-98-fall/Assignments/Assignment3/VRML2\\_Specification/spec/index.html](http://www.graphics.stanford.edu/courses/cs248-98-fall/Assignments/Assignment3/VRML2_Specification/spec/index.html)
4. Методика переноса 3D-моделей между системами CAD и CAE / И. А. Севцов, А. А. Юдин, К. В. Жихарев, И. И. Кочегаров // *Труды Международного симпозиума «Надежность и качество»*. Т. 1. Пенза: Пензенский государственный университет, 2016. С. 288–293.
5. Поставщик ПО для строительства Autodesk приостановил работу в России. // РБК.ru. 04.03.2022. URL: <https://www.rbc.ru/politics/04/03/2022/6221b6209a7947f98925c8d7>
6. Большаков В. П., Бочков А. Л., Лячек Ю. Т. Проблемы обмена графическими данными между CAD-системами // *Компьютерные инструменты в образовании*. 2013. №2. С. 37–46.
7. Филинских А. Д., Соснина О. А., Бойтяков А. А. Иерархическое пространство параметров геометрической модели // *Вестник БГТУ им. В. Г. Шухова*. 2015. №2. С. 131–134.
8. Филинских А. Д., Бяшеров А. Х. Анализ передачи параметрической и графической информации на основе экспериментальных данных // *Вестник БГТУ им. В. Г. Шухова*. 2012. №2. С. 164–166.
9. Галанина М. Н. Анализ перспектив импортозамещения программного обеспечения в области BIM-систем // *КОГРАФ-2022: сборник материалов 32-й Всероссийской научно-практической конференции по графическим информационным технологиям и системам*, Нижний Новгород, 18–21 апреля 2022 года. Нижний Новгород: НГТУ им. Р.Е. Алексеева, 2022. С. 95–102.
10. Журавлева Е. С., Томчинская Т. Н. Программное взаимодействие компонентов BIM на примере информационной системы НГТУ им. Р.Е. Алексеева // *КОГРАФ-2022: сборник материалов 32-й Всероссийской научно-практической конференции по графическим информационным технологиям и системам*, Нижний Новгород, 18–21 апреля 2022 года. Нижний Новгород: НГТУ им. Р.Е. Алексеева, 2022. С. 115–121.
11. Лахов А. Я., Лахов К. А. Конвертация библиотечных объектов двухконтурных геодезических оболочек // *КОГРАФ-2020: сборник материалов 30-й Всероссийской научно-практической конференции по графическим информационным технологиям и системам*, Нижний Новгород, 13–16 апреля 2020 года. Нижний Новгород: НГТУ им. Р.Е. Алексеева, 2020. С. 60–65.
12. Лахов А. Я. Трансляция геометрических моделей геодезических оболочек в нейтральный sat формат // *Инженерный вестник Дона*. 2018. №3 (50). Ст. 139.
13. Проблемы трансляции графических данных CAD-систем / С. Л. Кенин, В. Ф. Барабанов, А. М. Нужный, Н. И. Гребенникова // *Вестник Воронежского государственного технического университета*. 2013. №3-1. С. 4–8.
14. Романова Е. Б., Гагаринов П. А. Конвертация данных из eCAD Altium Designer в MCAD Solid Edge // *Известия высших учебных заведений. Приборостроение*. 2019. Т. 62, № 2. С. 178–184.
15. Романова Е. Б., Трифонова Т. А. Трансляция данных между системами EDA и MCAD // *Известия высших учебных заведений. Приборостроение*. 2016. Т. 59, № 4. С. 288–293.

## Информационная система инвентаризации вуза с применением BIM-технологий

А. Е. Николаев, Е. В. Сахно, А. Д. Филинских, Т. Н. Томчинская  
Нижегородский государственный технический университет им. Р.Е. Алексеева,  
Нижний Новгород, Россия

**Аннотация.** В целях упрощения процессов учета, оптимизации эксплуатации и повышения уровня автоматизации управления материально-технической базой университета исследованы вопросы повышения эффективности процессов инвентаризации оборудования, минимизации человеческого фактора и улучшения качества учета данных. Выполнен небольшой обзор разрабатываемых вузами BIM-систем, решающих задачу оптимизации эксплуатации зданий и оборудования – готовых и находящихся в разработке. Представлена информационная система инвентаризации оборудования НГТУ им. Р. Е. Алексеева. Реализованный функционал включает в себя возможность управления данными об оборудовании, его ремонте и заявках на ремонт, а также управления данными о зданиях и помещениях университета, управления пользователями, их правами и предоставления инструментов для планирования ремонта и эксплуатации оборудования.

**Ключевые слова:** BIM-технологии, инвентаризация, вуз, учёт оборудования, цифровой двойник, информационная система.

## Information system of inventory of the University using BIM technologies

A. E. Nikolaev, E. V. Sakhno, A. D. Filinskikh, T. N. Tomchinskaya  
Nizhny Novgorod State Technical University, n.a. R.E. Alekseev, Nizhny Novgorod, Russia

**Abstract.** In order to simplify accounting processes, optimize operation and increase the level of automation of the management of the material and technical base of the university, the issues of increasing the efficiency of equipment inventory processes, minimizing the human factor and improving the quality of data accounting were investigated. A short review of BIM systems developed by universities that solve the problem of optimizing the operation of buildings and equipment – ready and under development – has been performed. The information system of equipment inventory of the R.E. Alekseev NSTU is presented. The implemented functionality includes the ability to manage data about equipment, its repairs and repair requests. It also manages data on university buildings and premises, manages users, their rights, and provides tools for planning equipment repairs and operation.

**Keywords:** BIM technologies, inventory, university, equipment accounting, building management services, information system.

### Введение

В современных условиях цифровизации все большее значение приобретает использование информационных систем для автоматизации процессов управления и инвентаризации [1]. Одной из ключевых задач таких систем является эффективное управление оборудованием, включая его учет, мониторинг состояния и планирование эксплуатации. Это особенно актуально для крупных предприятий и образовательных учреждений, которые, как правило, имеют разветвленную сеть корпусов и помещений [2, 3], большое количество разнообразного учебного оборудования.

Для решения этих задач предназначены BIM-технологии. Под BIM (Building Information Modeling) понимают технологию, представляющую собой процесс коллективного создания и использования данных об объекте на всех этапах его жизненного цикла – от планирования до строительства, эксплуатации и сноса [4]. Большинство общих исследований указывают на BIM как на технологию, которая охватывает все дисциплины и связанную информацию, относящуюся как к разработке проекта, так и к требуемым этапам выполнения, а затем и эксплуатации.

Таким образом, BIM-модель можно представить в качестве центрального хранилища информации о реальной инфраструктуре [5–10]. Применение BIM особенно заметно в объектах строительства. Однако современные требования общества, такие как устойчивое развитие, создали новую тенденцию к внедрению BIM для существующих зданий [5].

Исследования показывают, что многие вузы активно включают в свои образовательные программы изучение BIM-технологий и принимают участие в разработке региональных BIM-проектов. Например, на кафедре информационных технологий Санкт-Петербургского государственного университета

архитектуры и строительства (СПбГАСУ) студенты и преподаватели создают объектно-ориентированные цифровые модели зданий и строительных процессов, включая архитектуру, конструкции и инженерные системы. Результаты включают чертежи, спецификации и анализы коллизий на этапах проектирования и строительства. Разработаны BIM-модели корпусов кампуса [11].

Московский государственный строительный университет (МГСУ) запустил программу «Информационное моделирование в строительстве», где студенты учатся разрабатывать и управлять BIM-моделями объектов и реализовывать проекты информационного моделирования всего жизненного цикла строительного объекта [12].

Студенты Санкт-Петербургского политехнического университета Петра Великого (СПбПУ) и Технического университета Граца приняли участие в совместном проекте и создали 4D-модель здания на основе BIM-модели, выполненной в Revit [13].

В рамках магистратуры «BIM-технологии в проектировании и строительстве» студенты МИСИС участвуют в реальных проектах: например, моделируют тоннели Московского метрополитена, осуществляют 3D-мониторинг систем, создают информационные платформы проекта совместно с промышленными партнёрами [14].

Студенты Инженерной академии Российского университета дружбы народов создают банк данных объектов культурного наследия на базе BIM-технологий [15].

Многие вузы разрабатывают BIM-модели своих учебных заведений и решают конкретные задачи, связанные с управлением учебным процессом и эксплуатацией здания. Например, в рамках выпускных квалификационных работ студентами Уральского федерального университета имени первого Президента России Б.Н. Ельцина (УрФУ) выполнены обмерные работы главного корпуса, построена 3D-модель здания в Autodesk Revit и при помощи технологии 3D-печати создан объемный макет главного корпуса УрФУ [16].

Преподавателями и студентами Инженерной академии Российского университета дружбы народов (Сибстрин) разработана информационная модель объектов капитального строительства: главного корпуса НГАСУ и спортивно-оздоровительного корпуса [17].

Ряд российских и зарубежных вузов создает цифровые двойники зданий, использует датчики и интегрирует BIM-модели с FM-системами (Facility Management) и IoT-системами (Internet of Things) для обслуживания инфраструктуры и проведения инвентаризации оборудования.

Так, Дальневосточный федеральный университет (ДВФУ) активно внедряет цифровые технологии: запущена платформа «Умный кампус», включающая цифровой двойник имущественного комплекса – микрорайона площадью более 850 000 м<sup>2</sup>, с управлением инженерными системами, аналитикой использования и обслуживания зданий. Цифровой двойник помогает отслеживать историю обслуживания, смету и состояние оборудования [18].

В статье [19] авторы из Казанского государственного архитектурно-строительного университета (КГАСУ) отмечают, что информационное моделирование активно применяется на стадии проектирования и строительства, генерируя большой объем информации об объекте. На этапе эксплуатации эта информация почти не используется, хотя эксплуатационным службам она необходима для принятия эффективных решений.

В работе рассматривается создание цифрового двойника на примере зданий кампуса КГАСУ и предложен метод мониторинга микроклимата в учебных аудиториях.

Студент из Московского государственного политехнического университета (МГПУ) разработал BIM-решение Inworks для эксплуатации зданий, включая учёт и мониторинг состояния оборудования. Проект уже презентовали представителям Минстроя России и мэрии Москвы. Проект является частью Комплексной системы управления техническим обслуживанием (КСУТО), разрабатываемой партнером проекта – компанией КСУТО. Совместно они стремятся полностью оцифровать весь цикл эксплуатации зданий МГПУ – от документооборота до создания цифровых двойников [20].

Студенты магистратуры Новгородского государственного университета (НовГУ) разрабатывают BIM-модели объектов университетского кампуса. Цифровые двойники есть у главного корпуса, университетского общежития. Ведётся кропотливая работа над каждым отдельным помещением с коммуникациями, мебелью и внутренней отделкой. Готовые модели планируется использовать для эксплуатации кампуса [21].

Из зарубежного опыта хочется отметить интересную работу авторов из Политехнического университета Мадрида (Испания) [5]. Разработана система управления на основе информационного моделирования зданий (BIM) для факультета гражданского строительства (ETSICCP) Мадридского политехнического университета с долгосрочной перспективой развития. Создана информационно-управленческая система на основе BIM-модели университетского кампуса. Этот проект разработан для всех участников его функционирования: студентов, преподавателей и руководителей. Система BIM-FM позволяет управлять классными комнатами, спортивными сооружениями, распределением шкафчиков и любой другой информацией о местоположении. Управление включает в себя возможность бронирования мест, составления учебного расписания, считывания температуры и влажности, определения вместимости учебных мест и отображения информации о техническом обслуживании оборудования. Высокая степень детализации модели, включая характеристики распределительной коробки, пропускной способности, кондиционирования воздуха или проектора, подробно описаны для каждого помещения, а официальные документы, прикрепленные к веб-сайту, позволяют преподавателям выбирать оптимальное помещение с учетом значительно большего количества переменных, чем при использовании традиционных методов.

Университет Райерсона и Йоркский университет Канады применяют технологии информационного моделирования для обеспечения безопасной эксплуатации зданий. Метод заключается в интеграции данных из сети датчиков Интернета вещей (IoT) здания с FM-BIM. Такой подход позволяет конечному пользователю выбирать желаемый временной интервал для визуализации и легко просматривать пространственно-временные данные об эксплуатационных характеристиках здания [22].

В технологическом институте Unites (Новая Зеландия) разработана информационная система, которая интегрирует BIM с FM [23]. Совместное использование этой информации и внедрение системы и ее приложений привело к улучшению рабочего процесса, более эффективному доведению информации об объекте до конечных пользователей, а также значительной экономии средств в нескольких областях. Unites – это высшее учебное заведение с тремя кампусами в Окленде, где обучается более 23000 студентов каждый год и работает около 800 сотрудников.

Несмотря на то что большое количество предприятий и учебных заведений работает над использованием BIM-систем на этапе эксплуатации, многие авторы отмечают, что, в то время как BIM-модели уже активно применяются на этапах проектирования и строительства, их использование на этапе эксплуатации находится лишь в зачаточном состоянии [5, 23, 24]. К тому же сотрудники отделов эксплуатации, как правило, не владеют навыками работы с BIM-системами.

Для решения задач автоматизации процесса инвентаризации среди отечественных компаний-разработчиков информационных систем наибольшей популярностью пользуются программные приложения на базе платформы 1С отечественного разработчика «Фирма 1С». Продукты данной компании охватывают практически все сферы бизнеса, подлежащие автоматизации. Но высокая стоимость, находящаяся в прямой зависимости от количества пользователей не позволяет предоставить доступ к системе (например, для подачи заявок на ремонт) всем заинтересованным сотрудникам и ограничивает область её внедрения во все административно-хозяйственные подразделения университета. Также программный продукт предъявляет высокие требования к аппаратному обеспечению [25, 26].

Возникает необходимость разработки приложения, позволяющего просматривать, редактировать и дополнять информацию, содержащуюся в BIM-моделях, сотрудникам служб эксплуатации. Приложение должно быть ориентировано на упрощение процессов учета, оптимизации эксплуатации и повышение уровня автоматизации управления материально-технической базой университета.

Потенциальными пользователями системы могут быть как сотрудники административно-хозяйственного отдела университета, так другие категории сотрудников, которым будет предоставлен функционал просмотра информации об оборудовании их подразделений и подачи заявок на его ремонт.

### **Постановка задачи**

Основная задача инвентаризации в вузах заключается не только учете мебели и компьютеров. Это часть управления активами, эксплуатацией и развитием материально-технической базы [16]. Ключевыми

задачами являются ведение актуального реестра оборудования, контроль физического состояния (исправность); фактическое местоположение; постановка на баланс и списание; техническое обслуживание и формирование заявок на ремонт; контроль сроков поверки. Инвентаризация в BIM – это не просто учет оборудования или помещений по факту. Это ведение точной 3D-модели, где каждый элемент – не просто геометрия, а база данных (атрибуты, серийные номера, паспорт, статус).

Идеальная инвентаризация характеризуется:

- скоростью (минимальное время проведения процедуры, варьирующееся от нескольких часов до нескольких дней в зависимости от количества объектов);
- экономичностью (отсутствие необходимости дополнительных финансовых затрат);
- качеством (получение достоверных и точных данных);
- точностью (соответствие фактических остатков данным учетной системы).

Этот идеальный стандарт является ориентиром для всех организаций, стремящихся к эффективному управлению своими ресурсами. Для решения выявленных проблем требуется разработка информационной системы, способной выполнять следующие функции:

- централизованное хранение и управление данными об оборудовании;
- интеграцию атрибутивных и пространственных данных помещений и оборудования;
- возможность гибкой конфигурации. Система должна легко адаптироваться к возможным изменениям объектов реального мира: строительству новых зданий и модернизации существующих, перепланировке помещений, изменениям структуры подразделений университета, добавлению новых типов зданий, помещений и оборудования;
- возможность добавления собственных наборов параметров у разных типов оборудования, что необходимо для адаптивности системы к оборудованию разных типов;
- возможность генерации уникальных идентификаторов в виде QR-кодов для упрощения идентификации и ускорения доступа к данным;
- генерацию и доставку уведомлений о плановых ремонтах и сроках вывода оборудования из эксплуатации;
- возможность назначения ответственных сотрудников за оборудование для повышения контроля.

### **Приложение инвентаризации «ИНВЕНТ НГТУ»**

Силами студентов и преподавателей Нижегородского государственного технического университета им. Р.Е. Алексеева (НГТУ) разрабатывается BIM-модель шестого учебного корпуса университета. Корпус состоит из десяти блоков. Построенная цифровая модель представлена в виде архитектурно-строительной части: перегородки, полы потолки, внешние фасады здания, лестничные пролеты, окна и двери. Модель разработана в инструментальной среде Revit и включает инженерные коммуникации, элементы сантехнических приборов, вентиляции, электроосвещения, такие как несколько типов потолочных светильников, вентиляционные решетки, розетки, шкафы для шланга пожаротушения, радиаторы, унитазы, раковины и другие.

В связи с санкционной политикой модель корпуса экспортирована в российский программный продукт Renga (рис. 1).

Разработка трёхмерной BIM-модели корпуса берёт начало в более масштабном проекте BIM-GIS интегрированной системы. Целью проекта BIM-GIS-системы с единой общей базой данных является создание трёхмерной модели всех корпусов университета, которая позволит аккумулировать всю информацию, необходимую руководству университета для эксплуатации зданий или в образовательных целях. Эта BIM-модель поможет руководству университета улучшить академическую жизнь студентов, облегчит учебный процесс преподавателей и интегрирует работу обслуживающего и административного персонала [27].

Для достижения основных целей исследования рассмотрены два основных этапа. Первый связан с моделированием инфраструктуры и разработкой трёхмерной цифровой модели корпуса. Второй задачей было внедрение адекватных инструментов для управления инфраструктурой с использованием BIM-модели здания.

С учетом возможности использования модели сотрудниками университета для проведения процесса инвентаризации оборудования было разработано приложение «ИНВЕНТ НГТУ».



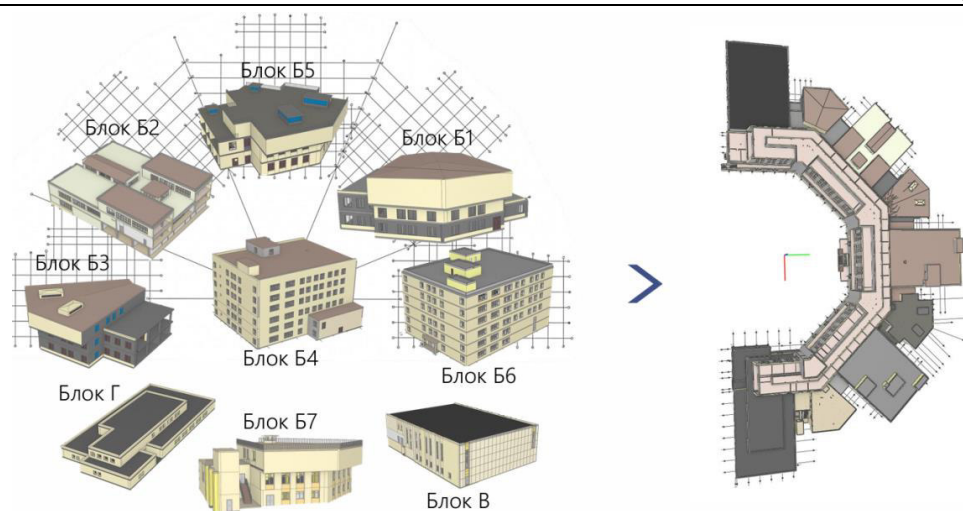


Рисунок 1. BIM-модель учебного корпуса НГТУ им. Р.Е. Алексеева

В приложение инвентаризации модель загружена через универсальный строительный формат IFC. В качестве архитектуры системы выбрана клиент-серверная архитектура, так как она лежит в основе большинства современных веб-приложений, включая информационные системы инвентаризации. Этот подход делит приложение на две части: клиентскую (frontend), отвечающую за взаимодействие с пользователем, и серверную (backend), занимающуюся обработкой запросов, управлением бизнес-логикой и взаимодействием с базой данных. Такое разделение обеспечивает высокую гибкость, масштабируемость и удобство сопровождения системы [28].

В контексте системы инвентаризации оборудования клиент-серверная архитектура имеет дополнительные преимущества. Она поддерживает многопользовательскую работу, обеспечивая одновременный доступ сотрудников и администраторов к централизованной базе данных. Гибкость интерфейсов позволяет пользователям работать как через веб-приложение, так и через мобильные платформы. Централизованное хранение данных на сервере упрощает управление большим объемом информации об оборудовании, его параметрах и статусах. Более того, серверная часть может интегрироваться с внешними системами, например бухгалтерскими или аналитическими платформами.

В представленном проекте применяется архитектурный шаблон проектирования Model–View–Controller (MVC), или Модель–Представление–Контроллер, – подход, который разделяет приложение на три основных, независимых друг от друга компонента: Модель (Model), Представление (View) и Контроллер (Controller). Такой подход обеспечивает четкое разделение бизнес-логики, пользовательского интерфейса и обработки данных, что делает разработку, поддержку и масштабирование приложения более удобными [8].

Четкое разделение ответственности – одно из главных преимуществ MVC. Компонент Model отвечает за взаимодействие с данными: их получение, сохранение и обработку. Например, в системе инвентаризации это работа с таблицами оборудования и связями между ними. Компонент View фокусируется на отображении данных пользователю, таких как списки оборудования или формы для редактирования. Controller, в свою очередь, управляет пользовательскими запросами, взаимодействует с Model и возвращает обработанные данные для отображения через View.

Серверная часть системы инвентаризации представляет собой ядро приложения, где сосредоточена вся бизнес-логика, обработка запросов пользователей, поступающих от клиентской части и взаимодействие с базой данных. Она реализована на языке PHP с использованием фреймворка Laravel, что позволило обеспечить надежность, безопасность и удобство разработки. Серверная часть взаимодействует с базой данных PostgreSQL для хранения и извлечения данных, а также предоставляет API-интерфейсы для работы клиентских приложений.

Клиентская часть системы инвентаризации играет ключевую роль в обеспечении удобного взаимодействия пользователей с функционалом приложения. Она отвечает за отображение данных,

управление интерфейсом и обработку пользовательских действий. Для реализации клиентской части выбран подход серверного рендеринга (SSR) на базе шаблонов Blade, встроенных в Laravel, что позволило обеспечить простоту, производительность и удобство работы с интерфейсом.

Серверный рендеринг (SSR) обеспечивает генерацию HTML-страниц на сервере перед их отправкой в браузер пользователя. Такой подход ускоряет время загрузки страницы, так как пользователь сразу получает полностью сформированный HTML-документ.

Стилизация интерфейса выполнена с использованием Tailwind CSS – CSS-фреймворка, который предлагает утилитарный подход к дизайну. Вместо написания кастомных классов или стилей Tailwind предоставляет готовые утилиты, такие как flex, grid, text-center, bg-gray-500, и многие другие, которые задают конкретные стили. Это значительно ускоряет разработку и упрощает поддержку.

Интерфейс системы состоит из нескольких основных элементов, включая страницы управления оборудованием, зданий, помещений и других связанных объектов. Шаблоны Blade используются для построения каждого из этих компонентов, обеспечивая единый стиль и структуру. Компоненты интерфейса разделены на логические блоки, что облегчает их повторное использование и поддержку.

Для создания динамичной и интерактивной пользовательской среды в системе инвентаризации оборудования применяется сочетание технологий серверного рендеринга и JavaScript-компонентов. В данной работе используется как нативный JavaScript, так и JavaScript-фреймворк Vue.js, что дает возможность оптимально распределить функциональные задачи между этими инструментами в зависимости от их сложности и назначения.

При разработке интерактивной 3D-модели здания в системе инвентаризации оборудования использован инструмент – библиотека Three.js, построенная на базе технологии WebGL. Этот выбор обусловлен необходимостью создания визуально комфортного и функционального интерфейса для отображения пространственных данных, таких как структура зданий, расположение помещений и размещение оборудования.

Данный Vue-компонент отвечает за рендеринг и интерактивное управление 3D-моделью здания, включая помещения и оборудование. Компонент состоит из нескольких функциональных блоков, обеспечивающих удобный интерфейс и работу с трехмерными объектами.

Компонент поддерживает следующие интерактивные функции:

- выбор режима отображения. Пользователь может переключаться между режимами просмотра (здание, помещения, оборудование) с помощью панели управления. Методы `updateLayers` и `setVisibility` отвечают за изменение видимости объектов в зависимости от выбранного режима;
- выбор этажа. Панель управления позволяет фильтровать объекты по этажам. Это делает отображение более удобным для анализа;
- выбор объектов. Пользователь может кликнуть на объект в модели, чтобы получить подробную информацию. Метод `onMouseClicked` определяет, на какой объект кликнул пользователь, и открывает модальное окно с данными.

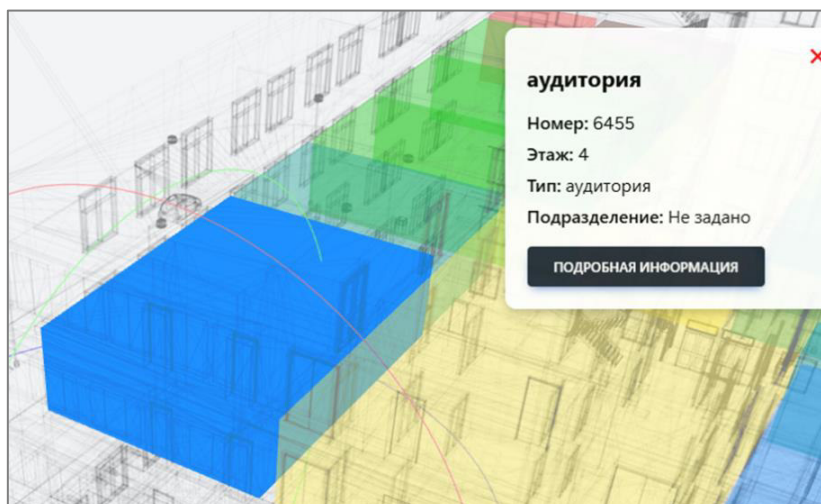


Рисунок 2. Модальное окно с подробной информацией о помещении

Предусмотрено отдельное модальное окно (рис. 2), появляющееся при выборе объекта. Оно отображает структурированную информацию, включая название объекта, его характеристики и ссылку на страницу с подробной информацией. При закрытии окна выделение объекта снимается.

Общий вид разработанного интерфейса в режиме отображения помещений всех этажей представлен на рисунке 3. Различными цветами выделены помещения разных типов.

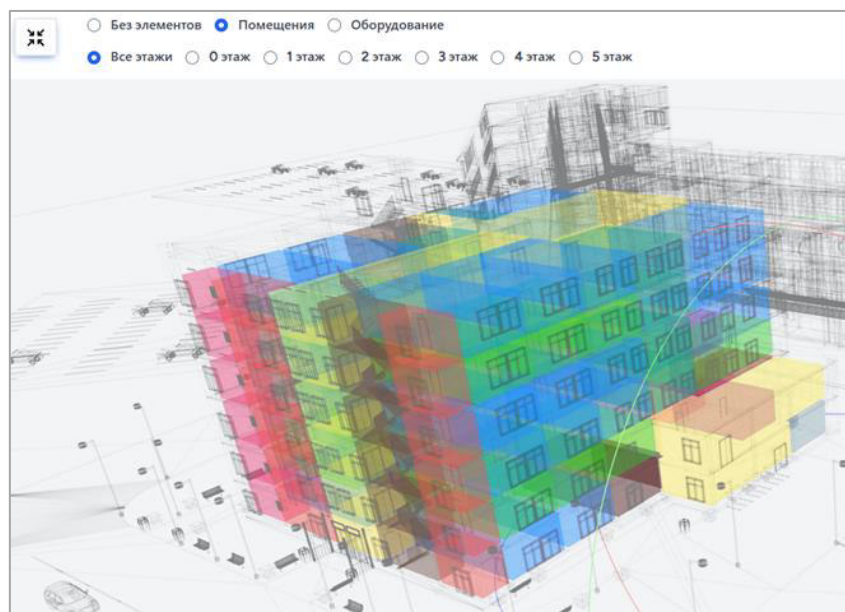


Рисунок 3. 3D-модель блока учебного корпуса в режиме отображения всех помещений

Каждый элемент (помещение, оборудование) на 3D-модели интерактивен, по клику на него открывается модальное окно с краткой информацией о нём и кнопкой «Подробнее», которая ведёт на страницу этого элемента с подробной информацией.

Список всех помещений НГТУ доступен на странице «Помещения». Предусмотрена фильтрация по типу помещений, зданию, этажу, а также поиск по номеру, наименованию и подразделению. На странице помещения доступна общая информация о помещении, об оборудовании в помещении, а также фотографии этого помещения в обычном формате и в формате 360, т. е. панорамные фотографии с углом обзора 360 (рис. 4) Также доступна кнопка «Просмотр 3D-модели».



Рисунок 4. Панорамный снимок учебной аудитории

На странице оборудования (рис. 5) также доступна кнопка «Печать QR-кода», при нажатии на которую QR-код со ссылкой на оборудование добавляется в очередь на печать. Данный документ (рис. 6) можно распечатать и наклеить на соответствующее оборудование. Каждый элемент содержит инвентарный номер и QR-код со ссылкой на страницу оборудования в информационной системе.



Таким образом, при наличии цифрового двойника корпусов и помещений вуза появляется возможность автоматизации размещения оборудования по аудиториям. В модели видно, где что стоит: не нужно бегать с бумажками по аудиториям.

Оборудование: инв. № 190080011693, компьютер

ПРОСМОТР 3D МОДЕЛИ ПЕЧАТЬ QR-КОДА РЕДАКТИРОВАТЬ УДАЛИТЬ

Основные данные	
Инвентарный номер:	190080011693
Тип оборудования:	компьютер
Наименование:	компьютер
Описание:	
Месторасположение (здание):	Учебный корпус №6
Этаж:	4
Месторасположение (помещение):	6449 (компьютерный класс)

Рисунок 5. Страница оборудования (фрагмент)



Рисунок 6. PDF-документ с QR-кодами оборудования

На странице «Оборудование» доступен список ремонтов оборудования НГТУ. Просматривать список всех ремонтов могут лишь только пользователи системы с ролью «Администратор». Пользователям системы с ролью «Пользователь» доступна информация только о ремонтах оборудования их подразделения и дочерних. Предусмотрена фильтрация по типу ремонта, статусу ремонта, а также поиск по описанию ремонта, инвентарному номеру или наименованию оборудования.

Список заявок на ремонт оборудования доступен на странице «Заявки на ремонт оборудования». Просматривать список всех заявок могут только администраторы и системы. Пользователям с ролью «Пользователь» доступна информация только об их заявках. Предусмотрена фильтрация по статусу заявки, а также поиск по номеру или описанию заявки, инвентарному номеру или наименованию оборудования.

Пользователи системы могут создать новую заявку на ремонт оборудования как при помощи кнопки «Создать новую заявку» на странице списка заявок, так и непосредственно на странице неисправного оборудования при помощи кнопки «Создать заявку на ремонт». Во втором случае инвентарный номер оборудования в форме создания заявки заполнится автоматически.

В форме создания заявки пользователи могут выбрать неисправное оборудование, указать краткое и расширенное описание возникшей проблемы. После создания заявки к ней также можно дополнительно приложить фотографии оборудования.

В приложении «ИНВЕНТ НГТУ» предусмотрены две основные роли: администратор и пользователь. У каждой из них есть свои права и ограничения, которые определяют дальнейший доступ к функциональности всей системы.

Роль пользователя позволяет сотрудникам различных подразделений просматривать информацию о самом здании, его помещениях, оборудовании и всех его характеристиках, а также выполнять поиск оборудования по выбранным критериям, печатать QR-коды на него и создавать заявки на ремонт, отслеживать статус их выполнения. По просмотру информации о зданиях и их помещениях у роли «Пользователь» никаких ограничений нет, а вот просмотр оборудования из подразделения, к которому пользователь не принадлежит, будет ему запрещен.

Роль «Администратор» же позволяет сотрудникам, отвечающим за управление системой, управлять всеми данными о зданиях, помещениях и оборудовании. Пользователь с ролью «Администратор» обладает расширенными полномочиями. Ему доступны такие функции, как настройка типов

оборудования, назначение ответственных сотрудников, возможность просмотра 3D-моделей зданий, а также их загрузки. Также таким пользователям открыты возможности добавления новых графических элементов (оборудования и помещений). Администраторы обладают возможностью управления учетными записями всех пользователей. Эта возможность включает в том числе назначение ролей и подтверждение регистрации для аккаунтов сотрудников. Сотрудникам, отвечающим за управление системой, доступны и другие возможности. Например, отслеживание состояния оборудования, в том числе вывод из эксплуатации и плановый ремонт. Также администраторам доступны управление заявками на ремонт, внесение изменений в статус таких заявок и возможность добавлять комментарии.

Кроме того, в системе присутствует роль «Неизвестный», которая присваивается пользователю сразу после процесса регистрации, при котором указывается электронная почта сотрудника. С данной ролью у пользователей нет доступа к системе, за исключением страниц регистрации и авторизации. Для того чтобы у «Неизвестного» появился полноценный доступ, необходимо, чтобы администратор системы изменил роль нового пользователя, подтвердив, что «Неизвестный» на самом деле является сотрудником университета и ему разрешен доступ к информационной системе инвентаризации НГТУ им. Р.Е. Алексеева «ИНВЕНТ НГТУ».

Особое внимание уделялось автоматизации процессов разработки и развертывания. Инфраструктура системы организована с использованием Docker и Docker Compose, что позволило обеспечить идентичность окружений разработки и эксплуатации. Настроенный CI/CD процесс на базе GitLab обеспечил стабильность и оперативность развертывания обновлений.

### Заключение

Разработана информационная система инвентаризации оборудования НГТУ им. Р.Е. Алексеева. Реализованный функционал включает в себя возможность управления данными о зданиях и помещениях университета, данными об оборудовании, его ремонте и заявках на ремонт. Возможность управления пользователями и их правами, а также предоставления инструментов для анализа и планирования ремонта и эксплуатации оборудования.

Одной из ключевых особенностей системы является интеграция данных о помещениях и оборудовании с 3D-моделями зданий. С использованием библиотеки Three.js создано визуальное представление блока учебного корпуса университета, в котором интегрирована атрибутивная и пространственная информация об оборудовании и помещениях. Данный подход позволяет не только эффективно управлять оборудованием, но и визуализировать его распределение в пространстве, что особенно полезно для крупных учреждений с разветвленной инфраструктурой, таких как НГТУ им. Р.Е. Алексеева.

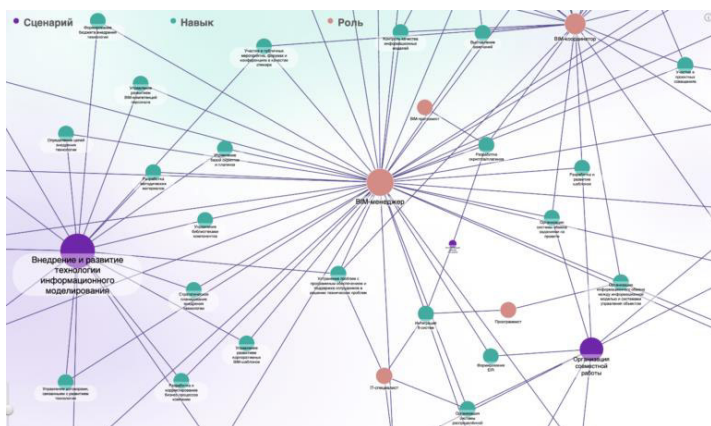
В серверной части системы использовался подход MVC, что позволило реализовать удобную расширяемую архитектуру приложения. Дизайн интерфейса системы создан с учетом адаптации под мобильные устройства, что обеспечило удобство в использовании системы на различных платформах.

Внедрение данной системы в эксплуатацию не только упрощает работу административных и хозяйственных служб, но и способствует более рациональному использованию ресурсов университета. В будущем возможна доработка и расширение системы, включая интеграцию с внешними сервисами и аналитическими платформами. Такой подход позволит университету оставаться на передовых позициях в области использования информационных технологий для управления своей инфраструктурой.

### Список литературы

1. Рассел Д. Инвентаризация. VSD, 2012. 662 с.
2. Ирагелова У. А. Инвентаризация основных средств как метод фактического контроля // Инновационная наука. 2020. № 3. URL: <https://cyberleninka.ru/article/n/inventarizatsiya-osnovnyh-sredstv-kak-metod-fakticheskogo-kontrolya-1> (дата обращения: 16.07.2025).
3. Вяткин В. В. Повышение качества отслеживания оборудования путем разработки и внедрения информационной системы инвентаризации СГУВТ // Научно-образовательный журнал для студентов и преподавателей «StudNet». 2022. №6. С. 6538–6548.
4. BIM – Building Information Modeling. Информационное моделирование зданий и сооружений // Электронный журнал «Tadviser». Государство. Бизнес. Технологии. 10.05.2010. URL: <https://www.tadviser.ru/a/53184> (дата обращения: 25.07.2025).

5. Pavón R. M., Alvarez A. A., Alberti M.G. BIM-Based Educational and Facility Management of Large University Venues // *Applied Sciences*. 2020. Vol. 10, iss. 22. Art. 7976. URL: <https://doi.org/10.3390/app10227976>
6. Wang X. The BIM Handbook: A Guide to Building Information Modeling for owners, managers, designers, engineers, and contractors // *Australasian Journal of Construction Economics and Building*. 2012. Vol 12, iss. 3. Pp. 101–102.
7. Watson A. Digital buildings – Challenges and opportunities // *Advanced Engineering Informatics*. 2011. Vol. 25, iss. 4. Pp. 573–581.
8. Exploring how information exchanges can be enhanced through Cloud BIM / A. Redmond, A. Hore, M. Alshawi, R. West // *Automation in Construction*. 2012. Vol. 24. Pp. 175–183.
9. Donath D. Bauaufnahme und Planung im Bestand. Vieweg+Teubner: Wiesbaden, Germany, 2009.
10. Building Information Modeling (BIM) for transportation infrastructure – Literature review, applications, challenges, and recommendations / A. Costin, A. Adibfar, H. Hu, S. S. Chen // *Automation in Construction*. 2018. Vol. 94. Pp. 257–281.
11. BIM-технологии моделирования зданий. URL: <https://www.spbgasu.ru/university/faculties-and-departments/bim-tehnologii-modelirovaniya-zdaniy/> (дата обращения: 26.07.2025).
12. Информационное моделирование в строительстве. URL: <https://mgsu.ru/postupayushchim/Magistratura/Perechen-realizuemyykh-programm/construction/informatsionnoe-modelirovanie-v-stroitelstve> (дата обращения: 26.07.2025).
13. Опыт создания цифровой модели здания в рамках международной образовательной программы на платформе BIM 360 / М. А. Романович, В. Г. Всеволожская, А. Р. Кузьминых и др. // Информационное моделирование в задачах строительства и архитектуры (BIMAC 2024): материалы VII Международной научно-практической конференции, Санкт-Петербург, 15–17 мая 2024 года. Санкт-Петербург: СПбГАСУ, 2024. С. 237–245. URL: [https://bim.conf.spbgasu.ru/wp-content/uploads/2024/05/BIMAC\\_2024.pdf](https://bim.conf.spbgasu.ru/wp-content/uploads/2024/05/BIMAC_2024.pdf) (дата обращения: 25.06.2025).
14. BIM-технологии в проектировании и строительстве. URL: <https://misis.ru/applicants/admission/magistracy/faculties/informatikaivtmag/bim-teh/> (дата обращения: 18.07.2025).
15. Халаби С. М., Савельева Л. В., Плотникова О. Г. Внедрение технологий информационного моделирования в инженерно-архитектурное образование // *Architecture and Modern Information Technologies*. 2017. №3(40). С. 322–331. URL: [http://marhi.ru/AMIT/2017/3kvart17/23\\_khalabi-savelievplotnikova/index.php](http://marhi.ru/AMIT/2017/3kvart17/23_khalabi-savelievplotnikova/index.php) (дата обращения: 11.07.2025).
16. Машкин О. В. История успеха внедрения BIM в Институте строительства и архитектуры УрФУ. URL: <https://bim.vc/news/istoriya-uspekha-urfu-platforma-bim-vc-reshaet-problemy-vuza> (дата обращения: 20.07.2025).
17. Карелин Д. В., Чмир Ю. Э. ТИМ как инструмент комплексного развития университета // Информационное моделирование в задачах строительства и архитектуры (BIMAC 2024): материалы VII Международной научно-практической конференции, Санкт-Петербург, 15–17 мая 2024 года. Санкт-Петербург: СПбГАСУ, 2024. С. 260–265. URL: [https://bim.conf.spbgasu.ru/wp-content/uploads/2024/05/BIMAC\\_2024.pdf](https://bim.conf.spbgasu.ru/wp-content/uploads/2024/05/BIMAC_2024.pdf) (дата обращения: 25.06.2025).
18. Цифровой полигон «Умный кампус». URL: <https://2030.dvfu.ru/czifrovoj-poligon-umnyj-kampus> (дата обращения: 25.06.2025).
19. Степанов И. О., Крайнов Д. В. Применение цифрового двойника на стадии эксплуатации здания // Информационное моделирование в задачах строительства и архитектуры (BIMAC 2024): материалы VII Международной научно-практической конференции, Санкт-Петербург, 15–17 мая 2024 года. Санкт-Петербург: СПбГАСУ, 2024. С. 237–245. URL: [https://bim.conf.spbgasu.ru/wp-content/uploads/2024/05/BIMAC\\_2024.pdf](https://bim.conf.spbgasu.ru/wp-content/uploads/2024/05/BIMAC_2024.pdf) (дата обращения: 25.06.2025).
20. Студент Московского Политеха представил инновационное BIM-решение для эксплуатации зданий. URL: <https://mospolytech.ru/news/student-moskovskogo-politekha-predstavil-innovatsionnoe-bim-reshenie-dlya-ekspluatatsii-zdaniy> (дата обращения: 25.06.2025).
21. Цифровой город будущего: уникальную методику эксплуатации зданий и сооружений создают в Новгородском университете. URL: <https://gazon.media/nauka/tekhnicheskie-nauki/tsifrovoy-gorod-buduschego-unikalnuyu-metodiku-ek/> (дата обращения: 25.06.2025).
22. Building Automation System Data Integration with BIM: Data Structure and Supporting Case Study / C. Quinn, A. Z. Shabestari, T. Mistic, et al. // *Automation in Construction*. 2020. Vol. 118. Art. 103257. URL: <https://arxiv.org/abs/2205.05518> (дата обращения: 25.06.2025).
23. Parsanezhad P., Dimyadi J. Effective facility management and operations via a BIM-based integrated information system // *Proceedings of CIB W070, W111 & W118 International Conference*. 05 2014. URL: [https://www.researchgate.net/publication/262362615\\_Effective\\_Facility\\_Management\\_and\\_Operations\\_via\\_a\\_BIM-based\\_Integrated\\_Information\\_System](https://www.researchgate.net/publication/262362615_Effective_Facility_Management_and_Operations_via_a_BIM-based_Integrated_Information_System) (дата обращения: 27.07.2025).
24. Sodislab. URL: <https://www.sodislab.com/ru> (дата обращения: 28.02.2024).
25. Функциональные возможности программного комплекса 1С: BIM 6D // Отраслевые и специализированные решения 1С:Предприятие. URL: <https://solutions.1c.ru/catalog/bim6d/features> (дата обращения: 27.07.2025).
26. 1С:Предприятие 8. Инвентаризация и управление имуществом // Отраслевые и специализированные решения 1С:Предприятие. URL: <https://solutions.1c.ru/catalog/asm/features>
27. Tomchinskaya T., Galanina M. University Management System Based On BIM-GISTechnologies // *Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (GraphiCon 2020)*. Part 2. 2020-12-17 / Journal article. DOI: 10.51130/graphicon-2020-2-4-26
28. Шевченко К. К., Бондаренко Т. В. Клиент-серверная архитектура // Образование. Наука. Производство: материалы XII Международного молодежного форума, Белгород, 01–20 октября 2020 года. Белгород: БГТУ им. В.Г. Шухова, 2020. С. 2046–2049.



Секция 9

***ПОДГОТОВКА КАДРОВ ВЫСШЕЙ  
КВАЛИФИКАЦИИ В ОБЛАСТИ  
ИНЖЕНЕРНОЙ ГЕОМЕТРИИ,  
КОМПЬЮТЕРНОЙ ГРАФИКИ  
И ЦИФРОВОЙ ПОДДЕРЖКИ  
ЖИЗНЕННОГО ЦИКЛА ИЗДЕЛИЙ***

---

**TRAINING OF HIGHLY QUALIFIED PERSONNEL  
IN THE FIELD OF ENGINEERING GEOMETRY,  
COMPUTER GRAPHICS AND DIGITAL SUPPORT  
OF THE PRODUCT LIFE CYCLE**



**GraphiCon 2025**

30 сентября – 2 октября 2025  
Йошкар-Ола, Россия



## Геометрические модели и алгоритмы проектирования отражающих экранов в акустике

Е. А. Замятина

Ростовский государственный университет путей сообщения, Ростов-на-Дону, Россия

**Аннотация.** При проектировании зданий и сооружений общественного назначения важно учитывать акустические параметры помещения. Данные параметры оказывают значительное влияние на комфортность эксплуатации помещений. Одним из способов улучшения акустических параметров является применение отражающих экранов. В работе рассмотрены вопросы разработки таких экранов на основе методов геометрической акустики.

**Ключевые слова:** геометрическая модель, акустические параметры, отражающий экран, каркасная поверхность, сплайн

## Geometric models and algorithms for designing reflective screens in acoustics

E. A. Zamyatina

Rostov State Transport University, Rostov-on-Don, Russia

**Abstract.** When designing buildings and public facilities, it is important to take into account the acoustic parameters of the room. These parameters have a strong influence on the comfort of the premises operation. One of the ways to improve acoustic parameters is to use reflective screens. The paper considers the issues of developing such screens based on geometric acoustics methods.

**Keywords:** geometric model, acoustic parameters, reflective screen, frame surface, spline

### Введение

При проектировании поверхностей внутренних помещений объектов общественного назначения необходимо учитывать акустические свойства помещений. Эти характеристики оказывают сильное влияние на комфортность эксплуатации объектов.

В настоящее время для расчета акустических параметров помещений применяются волновая, статистическая и геометрическая модели.

Так как звук имеет волновую природу, то *волновая модель* наиболее полно отражает явления, связанные с его распространением в помещениях. Но данная модель требует большого объема вычислений. В связи с этим волновая модель применяется для помещений несложной геометрической формы [1].

*Статистическая модель* [2] предполагает, что звуковое поле в помещении изотропное и диффузное. Для расчетов используются усредненные параметры звукового излучения. Эта модель не учитывает геометрическую форму помещения, но эффективна для расчета времени реверберации и других акустических параметров.

*Геометрическая модель* не учитывает волновую природу излучения, поэтому имеет ряд ограничений [3]. При соблюдении условий применимости геометрическая модель дает возможность получить результаты расчета акустических параметров при использовании поверхностей сложной геометрической формы. В расчетах акустических параметров помещений геометрическая модель применяется достаточно давно, но ее применение было ограничено трудоемкими процессами графических построений. Поэтому, как правило, строились плоские модели (разрезы, планы) распространения звука в помещениях. Применение геометрической модели в системах компьютерной графики значительно расширяет ее возможности, позволяя реализовать геометрическое моделирование акустических процессов в пространстве для поверхностей, имеющих сложную геометрическую форму.

### Постановка задачи

В статье рассмотрены вопросы совершенствования существующих и разработка новых геометрических моделей и алгоритмов создания отражающих поверхностей на основе сплайновых методов с учетом акустических свойств. Для этих целей необходимо:

- 1) разработать геометрическую модель создания поверхностей экранов, содержащих заданные точки и отражающих звук точечного источника на заданную поверхность;
- 2) разработать, на основе предложенной геометрической модели, геометрические и вычислительные алгоритмы создания поверхностей отражающих экранов;
- 3) разработать методику оптимизации поверхности отражающих экранов для повышения их эффективности;
- 4) провести вычислительные эксперименты по расчету акустических параметров помещений при применении отражающих экранов, созданных на основе предложенной геометрической модели создания поверхностей экранов.

### Теория

Среди основных методов образования поверхностей наибольшее распространение получили *каркасный* и *каркасно-кинематический*. Для решения поставленных задач наиболее подходит каркасный метод, так как при создании поверхности с помощью данного метода [4] есть возможность наиболее полно учитывать заданные условия в определенных точках поверхности.

Каркасные поверхности определяются непрерывными каркасами, т. е. однопараметрическими наборами линий, принадлежащими поверхности, взятыми с шагом параметра, меньшим по модулю любой наперед заданной положительной величины.

Геометрической частью определителя каркасной поверхности является ее дискретный каркас, представляющий собой дискретный набор линий, принадлежащих определяемой поверхности. Алгоритмическая часть определителя такой поверхности представляет собой описание процесса определения точек и линий, принадлежащих поверхности, при непрерывном изменении параметров.

Для задания геометрической части определяемых поверхностей, т. е. дискретного линейного каркаса, в работе использовались кубические сплайны [5]. Как известно, сплайны являются мощным современным средством геометрического моделирования. Многие современные графические системы (*AutoCAD*, *3DMAX* и др.) поддерживают моделирование с помощью сплайнов. Кубические сплайны не требуют больших затрат вычислительных ресурсов, вместе с этим позволяют строить плавные линии в пространстве, проходящие через заданные точки, имеющие в каждой точке непрерывные производные первого и второго порядков.

Рассмотрим алгоритмическую часть определителя каркасных поверхностей, т. е. набор действий, которые необходимо выполнить для определения различных параметров поверхности. Приведены два варианта: аппроксимация участка поверхности и аппроксимация всего заданного отсека поверхности.

Рассмотрим *аппроксимацию отсека поверхности*. Пусть поверхность задана сетью. Возьмем в пространстве некоторую линию  $m$ . Определим первые и вторые производные в ее точках при переходе к новому параметру. Пусть задана линия

$$\vec{r} = \vec{r}(s). \quad (1)$$

Перейдем к новому параметру  $u$ . Будем считать, что параметрам точек  $M_i - s_i$  соответствуют новые параметры  $u_i$ . Запишем зависимость параметра  $s$  от  $u$  в следующем виде:

$$s = f(u). \quad (2)$$

Тогда с учетом (1) и (2) уравнение линии имеет такой вид:

$$\vec{r} = \vec{r}(f(u)). \quad (3)$$

Определим зависимость (3). Выполним аппроксимацию функции (3) плоским сплайном, который обозначим через  $n$  (рис. 1). Имея такую зависимость, можно определить параметр точки на линии –  $s^T$  при заданном параметре  $u^T$  (см. рис. 1) и из уравнения (3) найти координаты точки.

Найдем первую и вторую производные по новому параметру  $u$ . Продифференцировав (9) по параметру  $u$ , получим

$$\frac{d\vec{r}}{du} = \frac{d\vec{r}(f(u))}{ds} \frac{df(u)}{du}. \quad (4)$$

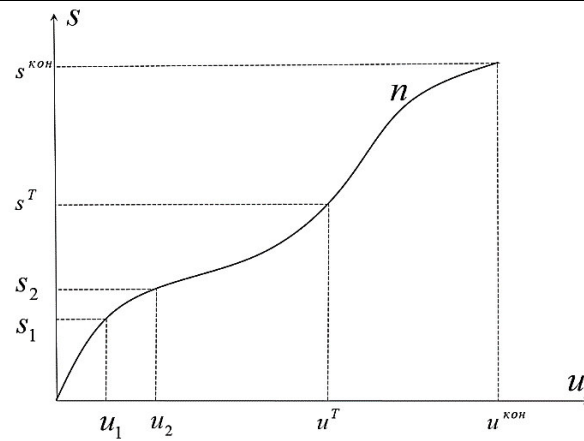


Рисунок 1. Зависимость  $s = f(u)$

Положение точки на сплайне  $n$  (см. рис. 1) зависит от параметра  $\sigma$ , поэтому

$$\frac{df}{du} = \frac{ds}{d\sigma} / \frac{du}{d\sigma}. \quad (5)$$

Подставив (5) в (4), для первых производных получим

$$\frac{d\vec{r}}{du} = \frac{d\vec{r}}{ds} \frac{ds}{d\sigma} / \frac{du}{d\sigma}. \quad (6)$$

Для определения вторых производных продифференцируем (5) по  $u$ :

$$\frac{d^2\vec{r}}{du^2} = \frac{d^2\vec{r}}{ds^2} \left( \frac{df}{du} \right)^2 + \frac{d\vec{r}}{ds} \frac{d^2f}{du^2}. \quad (7)$$

Найдем вторую производную функции  $f$  по параметру  $u$ :

$$\frac{d^2f}{du^2} = \left( \frac{du}{d\sigma} \cdot \frac{d^2f}{d\sigma^2} - \frac{d^2u}{d\sigma^2} \cdot \frac{df}{d\sigma} \right) / \left( \frac{du}{d\sigma} \right)^2. \quad (8)$$

Подставив (5) и (8) в (7), получим значение искомой второй производной.

Пусть заданы параметры некоторой точки поверхности  $K$  –  $u_k$  и  $v_k$ . Найдем координаты этой и производные в ней. Определим на каждом сплайне  $k_i$  точку  $N_i$ , соответствующую параметру  $v = v_k$  (3). В каждой точке  $N_i$  определим первые производные  $\frac{d\vec{r}_i^k(f_i(v_k))}{dv}$  (6) и вторые производные

$\frac{d^2\vec{r}_i^k(f_i(v_k))}{dv^2}$  (7). Через найденные точки построим сплайн  $l_k$  (рис. 2). Точка  $K$  будет принадлежать этому сплайну. Построим для сплайна  $l_k$  зависимость параметра  $s_k$  от  $u$  (2) (см. рис. 2). По формуле (3) определим координаты точки  $K$ , по (7), (8) – производные высших порядков.

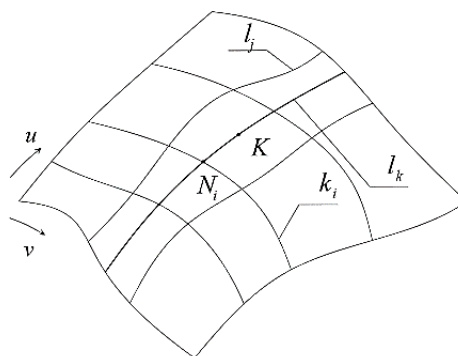


Рисунок 2. Построение сплайна  $l_k$

Рассмотрим конструирование каркасных поверхностей, соответствующих заданным требованиям. В качестве требований, предъявляемых к поверхности, взяты условия оптимального отражения излучения от конструируемой поверхности на заданную поверхность и принадлежности конструируемой поверхности заданных точек. Такие условия часто используются для конструирования отражающих экранов при решении задач в геометрической оптике и акустике.

Рассмотрим формирование отражающей поверхности. Пусть заданы  $I$  – точечный источник излучения, массив точек  $A_{ij}$ , где  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$  и отсек поверхности  $\Omega$ . Используя сплайновый метод, построим отсек поверхности  $\Sigma$ , содержащий точки массива  $A_{ij}$  и равномерно отражающий излучение источника  $I$  на отсек  $\Omega$ . Массив и отсек поверхности согласованы. Под согласованием подразумевается установление взаимно однозначного соответствия между точками массива и точками отсека поверхности таким образом, чтобы сумма длин отрезков, соединяющих соответственные точки, была минимальной.

Определим в точках  $A_{ij}$  векторы нормалей  $\vec{N}_{ij}$  поверхности  $\Sigma$  (рис. 3)

$$\vec{N}_{ij} = \frac{\vec{B}_{ij} - \vec{A}_{ij}}{|\vec{B}_{ij} - \vec{A}_{ij}|} + \frac{\vec{I} - \vec{A}_{ij}}{|\vec{I} - \vec{A}_{ij}|}, \quad (9)$$

где  $\vec{I}$ ,  $\vec{A}_{ij}$ ,  $\vec{B}_{ij}$  – радиусы-векторы соответствующих точек  $I$ ,  $A_{ij}$ ,  $B_{ij}$ .

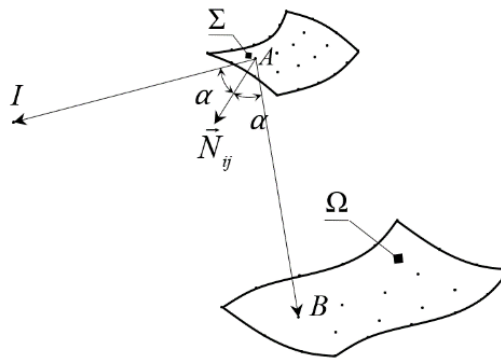


Рисунок 3. Определение вектора нормали

Возьмем ряд точек массива  $\vec{A}_{ij}$  при фиксированном значении  $j$ . Построим сплайн, проходящий через эти точки таким образом, чтобы касательные в них были перпендикулярны вектору нормали (9) и имели наименьшие углы  $\varphi$  с векторами  $\vec{a}_{ij} = \vec{A}_{i+1j} - \vec{A}_{ij}$ . Вектор касательной  $\vec{t}_{ij}^v$  найдем как направляющий вектор линии пересечения  $t_{ij}^v$  плоскостей (6), содержащей векторы  $\vec{a}_{ij}$  и  $\vec{N}_{ij}$ , и  $\beta_{ij}$ , проходящей через точку  $A_{ij}$  и перпендикулярной  $\vec{N}_{ij}$

$$t_{ij}^v = \alpha_{ij} \cap \beta_{ij}, \quad (10)$$

где  $\vec{a}_{ij} \in \alpha_{ij}$ ;  $\vec{N}_{ij} \in \alpha_{ij}$ ;  $A_{ij} \in \beta_{ij}$ ;  $\vec{N}_{ij} \perp \beta_{ij}$ .

При построении сплайна явно задать касательные можно только в начальной и конечной точках. Для того чтобы сплайн имел заданные касательные во всех точках, введем вблизи  $A_{ij}$  дополнительные точки, которые будут определять касательные (10).

Таким образом, построив сплайны, проходящие через ряды точек массива  $D_{ij}$  с фиксированным  $j$ , получим линии, проходящие через точки  $A_{ij}$  и имеющие в них касательные, перпендикулярные нормалям  $\vec{N}_{ij}$ . Обозначим построенные сплайны через  $n_j$ . Их уравнения

$$\vec{r}(u) = r_{n_j}(u, v^j). \quad (11)$$

Построим в каждой точке  $D_{ij}$  векторы нормалей  $\vec{N}_{ij}$ , перпендикулярные векторам  $\vec{t}_{ij}^v$

$$\vec{t}_{ij}^v = \frac{dr_{n_j}(v_j^{i'})}{dv}, \quad (12)$$

где  $v_j^{i'}$  – параметр, соответствующий точке  $D_{ij}$  на сплайне (11).

В точках массива  $D_{ij}$ , соответствующих точкам заданного массива  $A_{ij}$ , векторы  $\vec{N}_{ij}$  совпадают с векторами нормалей в этих точках, определенных по соотношениям (10). В ближайших точках (соответствующих дополнительным точкам  $C_{ijk}^1$  и  $C_{ijk}^2$ ) построим  $\vec{N}_{ij}$  таким образом, чтобы отраженные в них лучи имели минимальное отклонение от  $B_{ij}$ . Определим вектор  $\vec{N}_{ij}'$  по формуле

$$\vec{N}_{ij}' = \frac{\vec{B}_{ij} - \vec{D}_{ij}}{|\vec{B}_{ij} - \vec{D}_{ij}|} + \frac{\vec{I} - \vec{D}_{ij}}{|\vec{I} - \vec{D}_{ij}|}, \quad (13)$$

где  $i'' = \text{Round}\left(\frac{i' + 2k}{2k + 1}\right)$ ;  $\text{Round}$  – функция, округляющая значение аргумента до ближайшего целого.

Если бы вектор  $\vec{N}_{ij}'$  был нормалью в точке  $D_{ij}$ , то луч, отраженный в ней, точно попадал бы в  $B_{ij}$ . Но это возможно только для точек, соответствующих  $A_{ij}$ , так как во всех остальных точках  $\vec{N}_{ij}'$  не перпендикулярен вектору касательной  $\vec{t}_{ij}^v$  (12). Поэтому построим его так, чтобы угол  $\psi$  между ним и  $\vec{N}_{ij}'$  был минимальным. Вектор  $\vec{N}_{ij}'$  определяется как направляющий вектор прямой  $n_{ij}$ , которая является линией пересечения плоскостей  $\gamma_{ij}$  и  $\delta_{ij}$ . Плоскость  $\gamma_{ij}$  содержит векторы  $\vec{t}_{ij}^v$  (12) и  $\vec{N}_{ij}'$  (13). Плоскость  $\delta_{ij}$  содержит точку  $D_{ij}$  и перпендикулярна вектору  $\vec{t}_{ij}^v$ .

Построим векторы касательных в точках  $D_{ij}$ , перпендикулярные  $\vec{N}_{ij}'$  и составляющие с векторами  $\vec{b}_{ij}$  минимальный угол:

$$\vec{b}_{ij} = \begin{cases} \vec{D}_{ij+1} - \vec{D}_{ij}, & \text{если } j = 1; \\ \vec{D}_{ij} - \vec{D}_{ij-1}, & \text{если } j \neq 1. \end{cases}$$

где  $\vec{t}_{ij}^u$  определим аналогично тому, как определили  $\vec{t}_{ij}^v$  (10). Найдем линию пересечения плоскостей  $\alpha'_{ij}$  и  $\beta'_{ij} - t_{ij}^u$

$$t_{ij}^u = \alpha'_{ij} \cap \beta'_{ij}, \quad (14)$$

где  $\vec{b}_{ij} \in \alpha'_{ij}$ ;  $\vec{N}_{ij} \in \alpha'_{ij}$ ;  $D_{ij} \in \beta'_{ij}$ ;  $\vec{N}_{ij} \perp \beta'_{ij}$ .

Направляющий векторы линий  $t_{ij}^u$  (14) –  $\vec{t}_{ij}^u$  является искомым вектором.

Построим сплайны, проходящие через точки  $D_{ij}$ , при фиксированном значении  $i'$  и имеющие в них касательные (14). Введем вблизи  $D_{ij}$  дополнительные точки  $E_{i'jk}^1$  и  $E_{i'jk}^2$ ,  $k = 1, 2, \dots, M$ , радиусы-векторы которых определяются по соотношениям

$$\vec{E}_{i'jk}^1 = \vec{D}_{ij} - \frac{\vec{t}_{ij}^u}{|\vec{t}_{ij}^u|} d_1 k, \quad j = 1, 2, \dots, m-1; \quad \vec{E}_{i'jk}^2 = \vec{D}_{ij} + \frac{\vec{t}_{ij}^u}{|\vec{t}_{ij}^u|} d_2 k, \quad j = 2, 3, \dots, m,$$

где  $d_1 = \frac{|\vec{b}_{ij}|}{N}$ ;  $d_2 = \frac{|\vec{b}_{ij+1}|}{N}$ .

После ввода дополнительных точек имеем массив  $G_{i'j'}$ , где  $j' = 1, 2, \dots, 2M(m-1) + m$ . Точкам массива  $A_{ij}$  соответствуют точки  $G_{2M(i-1)+i, 2M(j-1)+j}$ . Через точечные ряды  $G_{i'j'}$ , при фиксированном  $i'$ , строим сплайны  $k_{i'}$ , имеющие в этих точках заданные касательные (14).

Через точки массива  $G_{i'j'}$ , при фиксированном  $j'$ , строим сплайны  $l_{j'}$ . Сплайновым методом строим поверхность, используем в качестве исходных данных сеть  $k_{i'}$ ,  $l_{j'}$  (рис. 4).

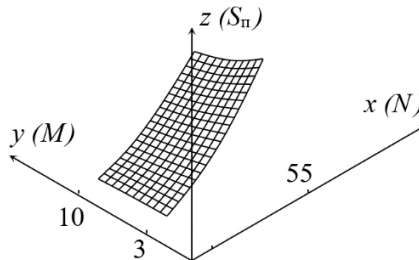


Рисунок 4. Полученная поверхность экрана

Для определения оптимальных параметров создается сплайновая поверхность в трехмерном пространстве описанным выше способом. Координате  $x$  соответствует параметр  $N$ ,  $y$  – параметр  $M$ ,  $z$  – полезная площадь  $S_n$ . Затем определяется локальный максимум координаты  $z$  (т. е. полезной площади  $S_n$ ). Рассматриваются значения  $z$  на границах области и в точках экстремума в заданной области. Для определения точек экстремума определим частные производные  $z$  по  $x$  и  $y$ . Известно

$$\begin{cases} z_u = z_x x_u + z_y y_u; \\ z_v = z_x x_v + z_y y_v, \end{cases} \quad (15)$$

где для краткости обозначим  $\frac{\partial z}{\partial u} = z_u$ ,  $\frac{\partial z}{\partial v} = z_v$ ,  $\frac{\partial x}{\partial u} = x_u$ ,  $\frac{\partial x}{\partial v} = x_v$ ,  $\frac{\partial y}{\partial u} = y_u$ ,  $\frac{\partial y}{\partial v} = y_v$ .

Из системы (15) имеем

$$\begin{cases} z_x = \frac{z_v x_u - z_u x_v}{x_u y_v - x_v y_u}; \\ z_y = \frac{z_u y_v - z_v y_u}{x_u y_v - x_v y_u}. \end{cases} \quad (16)$$

Для определения необходимого условия экстремума в точках области приравняем уравнения системы (16) к 0:

$$\begin{cases} z_v x_u - z_u x_v = 0; \\ z_u y_v - z_v y_u = 0. \end{cases} \quad (17)$$

Решив систему (17), получим точки, в которых выполняется необходимое условие экстремума. Из найденных точек определяем ту, которая имеет максимальное значение координаты  $z$ . На рисунке 5 приведен пример, где эта поверхность не имеет локальных максимумов, поэтому максимальное значение принимает на границе области, которую легко определить визуально.

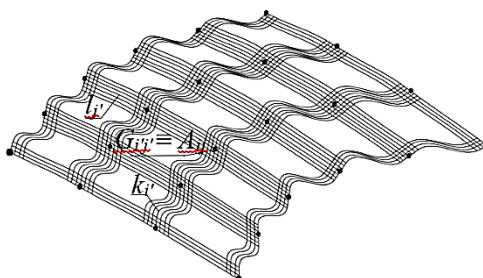


Рисунок 5. Оптимизация поверхности

## Результаты экспериментов

Проведен численный эксперимент расчета акустических параметров проектируемого помещения с использованием отражающих экранов, созданных на основе предложенной геометрической модели. Для расчета акустических параметров помещений применялось специализированное программное обеспечение моделирования и анализа акустики помещений *Catt-acoustic*. Результаты расчетов приведены на рисунке 6 в виде цветowych диаграмм. Средние значения звукового давления равны без отражающего экрана 78,283714 дБ, с экраном – 82,206857 дБ.

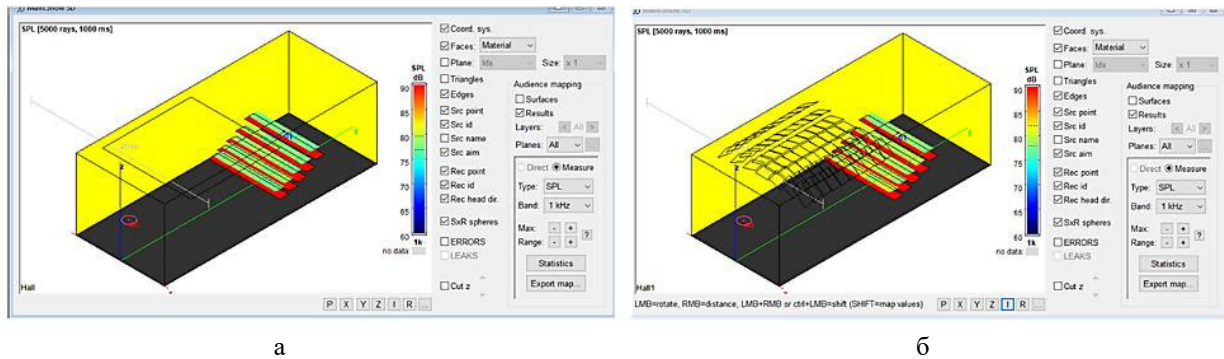


Рисунок 6. Цветовые диаграммы распределения звукового давления: а – без отражающего экрана; б – с экраном

Эхограммы первых отражений в одной из точек поверхности зрителей приведены на рисунке 7. На рисунке 7а показана эхограмма без применения отражающего экрана, а на рисунке 7б – с отражающим экраном.

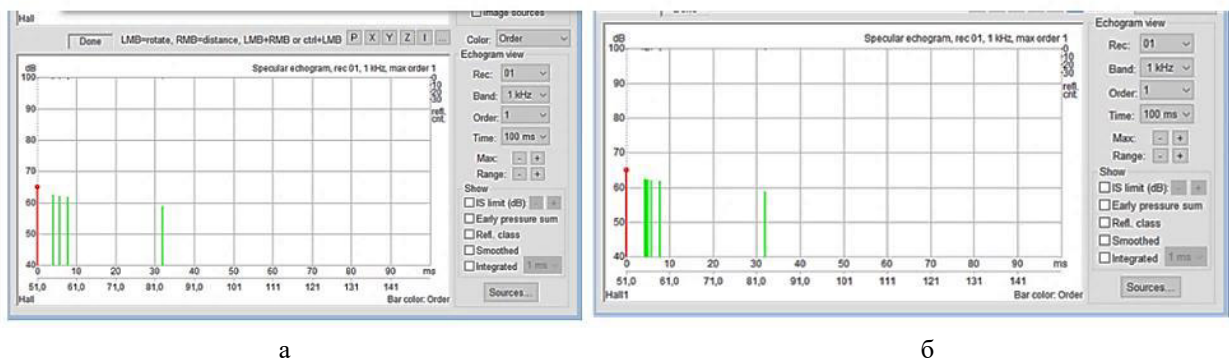


Рисунок 7. Эхограммы первых отражений ( | – прямые лучи, | – отраженные лучи)

## Обсуждение результатов

Проведенные исследования показывают, что введение отражающего экрана увеличивает среднее звуковое давление на зону поверхности зрителей за счет первых мало запаздывающих отражений на 5 % (см. рис. 6б).

## Выводы и заключение

По проделанной работе можно сделать следующие выводы:

1. Разработана геометрическая модель создания поверхностей экранов, отражающих звук точечного источника на заданную поверхность. Применение разработанной модели позволит значительно ускорить проектирование помещений с улучшенными акустическими параметрами на этапе эскизного проектирования. Разработанные модели могут быть использованы как для создания отражающих экранов, так и для разработки поверхностей ограждающих конструкций помещений.
2. Разработан пакет программ для автоматизированного проектирования отражающих экранов. Программы, входящие в разработанный пакет, могут быть использованы как самостоятельно при проектировании помещений, так и в составе автоматизированных систем архитектурно-строительного проектирования.



3. Разработана методика оптимизации поверхности отражающих экранов для повышения их эффективности при изменении задаваемых основных геометрических параметров, основанная на расчёте эффективной площади отражающих экранов. Эта методика позволяет оценить основные акустические параметры помещения при различных вариантах архитектурных решений и выбрать наилучший из них. Исследованы виды поверхностей экранов, их эффективность, при изменении задаваемых основных геометрических параметров. Результаты данного исследования позволяют проектировщику более быстро определить параметры создаваемой поверхности экрана, для получения большей эффективности.

4. Исследована эффективность предложенной модели при проектировании помещений с учетом акустических свойств. Исследования производились путем проведения численного эксперимента с использованием специализированного программного обеспечения *Catt-acoustic* для моделирования и анализа акустики помещений, которые показали увеличение величины среднего звукового давления на поверхность зрителей на 5 и более процентов.

**Научный руководитель** Приходько Виктор Маркович, заведующий кафедрой «Начертательная геометрия и графика», Ростовский государственный университет путей сообщения.

#### Список литературы

1. Помпеи А., Сумбатьян М. А., Тодоров Н. Ф. Компьютерные модели в акустике помещений // Акустический журнал. 2009. Т. 55, №6. С. 760–771.
2. Абзалилова С. А., Дмитриев В. Л. История развития акустики // Современные научные исследования и инновации. 2019. № 1. URL: <https://web.snauka.ru/issues/2019/01/88494> (дата обращения: 08.12.2024).
3. Лейзер И. Г. О применении методов геометрической акустики для расчета отражений звука от плоских поверхностей // Акустический журнал. 1966. Т. XII, вып. 2. С. 206–212.
4. Котов И. И. Алгоритмы конструирования каркасных поверхностей. М.: МАИ, 1975. 63 с.
5. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения. М: Мир, 1972. 319 с.

## Разработка программного обеспечения для построения октаэдрических сеток

Е. Е. Заяц, А. В. Февральских

Московский авиационный институт (национальный исследовательский университет),  
Москва, Россия

**Аннотация.** Статья посвящена вопросам разработки программного обеспечения, предназначенного для построения контрольнообъемных сеточных моделей на основе архимедовых тел. Изложены основы используемого геометрического метода, приведена общая архитектура программы и основные положения алгоритма ее работы.

**Ключевые слова:** разработка программного обеспечения, полиэдрические сеточные модели, численное моделирование, механика сплошных сред, метод конечных объемов.

## Development of software for constructing octahedral meshes

E. E. Zayats, A. V. Fevralskikh

Moscow Aviation Institute, Moscow, Russia

**Abstract.** The work is devoted to the issues of developing software intended for constructing control volume mesh models based on Archimedean solids. The basics of the geometric method used are presented, the general architecture of the program and the main provisions of the algorithm of its operation are given.

**Keywords:** software development, polyhedral mesh models, numerical modeling, continuum mechanics, finite volume method.

### Введение

Численное моделирование – ключевой инструмент, который широко применяется для решения задач механики сплошных сред. Он основан на применении методов, многие из которых используют сеточные модели. Среди них можно выделить такие, как метод контрольных объемов (FVM), метод конечных элементов (FEM) и метод конечных разностей (FDM). Данные методы зарекомендовали себя как универсальные и надежные для широкого круга задач. Однако на точность и устойчивость получаемых численных решений напрямую влияют тип и качество используемой сеточной модели. Сеточные модели с различной морфологией элементов по-разному влияют на объем требуемых для моделирования вычислительных ресурсов, на скорость сходимости решения и на его гладкость. В частности, сеточные модели на основе тетраэдров, как правило, характеризуются большими затратами вычислительных ресурсов ввиду необходимости их большей концентрации для получения гладких физических полей, чем, например, сеточные модели на основе полигексов. С другой стороны, сеточные модели на основе полигексов могут требовать большей концентрации элементов для расчета градиентов физических величин. Кроме того, затраты памяти и количества операций на обработку одного полигекса превышают аналогичные затраты на обработку гексаэдров и тетраэдров в ходе численного интегрирования уравнений динамики. Эти обстоятельства подталкивают к разработке геометрических методов построения компромиссных сеточных моделей и соответственно методов проектирования программного обеспечения для их генерации.

На сегодняшний день существуют различные программные решения, которые позволяют строить разнообразные сеточные модели, применяемые в ряде задач, а именно: поверхностные, объемные гекса-сетки, тетра-сетки, смешанные или полиэдрические (на основе полигексов). Примерами таких решений являются программные пакеты Ansys Fluent Meshing [1–3], OpenFOAM, FlowVision и многие другие. Помимо того, что выбор морфологий сеточных элементов в этих программах является весьма ограниченным, архитектура этих программ и интерфейс зачастую не позволяют выработать эффективные методики работы с ними для решения задач цифровой поддержки жизненного цикла промышленных изделий. Это создает барьер для эффективного использования необходимых вычислительных методов инженерами, особенно в случаях, когда важна скорость подготовки модели.

Целью данной работы является разработка геометрического метода построения сеточных моделей на основе архимедовых тел и методика проектирования программного обеспечения для их создания

в ходе решения задач проектирования изделий. В рамках исследования для достижения цели решаются следующие задачи:

- разработка геометрического метода построения сеточных моделей на основе архимедовых тел,
- разработка общей архитектуры программного обеспечения,
- реализация и тестирование модуля построения сеточной модели.

В рамках данного доклада рассматривается процесс реализации вычислительного модуля, а также приводятся описания общей архитектуры разрабатываемого приложения и выбранного метода построения сеточной модели.

Разрабатываемое приложение сможет стать связующим звеном между этапами геометрического проектирования и численного моделирования в составе SCADA системы (Supervisory Control and Data Acquisition) и позволит сократить время подготовки расчетных данных и снизить требования к экспертизе пользователя.

### Постановка задачи

Наиболее значимыми характеристиками расчетных сеток являются скорость генерации и качество итоговой модели. Классический подход к построению полиэдрических сеточных моделей, реализованный в Ansys Fluen Meshing, основан на предварительной генерации поверхностной сетки с последующей генерацией объемных элементов (рис. 1а). Несмотря на высокую точность аппроксимации границ, данный подход требует тщательной подготовки STL геометрии и заметно уступает в скорости альтернативному подходу, который основан на обрезке заранее сгенерированной объемной сетки (рис. 1б). Основным преимуществом альтернативного подхода является высокая степень его автоматизации, обусловленная отсутствием необходимости ручного исправления геометрических дефектов модели и трудоемкой настройки параметров генерации. Таким образом, в качестве основного алгоритма построения полиэдрической сеточной модели, закладываемого в разрабатываемое приложение, предлагается выбрать метод на основе обрезки модели по заданной STL геометрии.



Рисунок 1. Сравнение блок-схем алгоритмов построения сеточных моделей

## Теория

В основе вычислительного модуля разрабатываемой программы, предназначенной для построения полиэдрических сеточных моделей, лежит геометрический метод, использующий архимедовы тела. Данный метод предполагает последовательное заполнение пространства многогранниками с учетом величины шага сетки – параметра, задаваемого пользователем. Генерация начинается с определения геометрических координат центров многогранников, составляющих модель. Координаты вершин многогранников находятся перестановкой значений, рассчитываемых на основе величины шага сетки и выбранного типа многогранника. Пример получаемой данным методом сеточной модели показан на рисунке 2, где бирюзовым изображены кубookтаэдры, а лиловым – октаэдры.

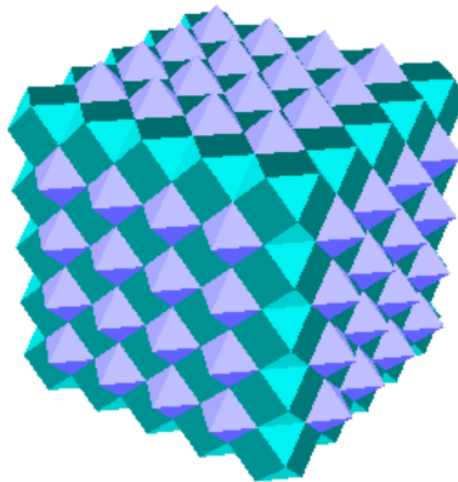


Рисунок 2. Сеточная модель на основе октаэдров и кубookтаэдров для куба

В результате сравнительного анализа различных сеточных моделей, генерируемых данным методом, было выявлено, что сеточная модель на основе октаэдров и кубookтаэдров является наиболее предпочтительной для вычислений, поскольку она обладает наибольшей информативностью в сравнении с другими рассмотренными моделями. Результаты данного анализа в том числе продемонстрированы на графике зависимости количества ячеек сетки  $N$  как меры ее информативности от шага  $a$  в логарифмическом масштабе (рис. 3), где под методом 1 понимается метод построения сеточной модели на основе усеченных октаэдров, 2 – октаэдров и кубookтаэдров, 3 – кубookтаэдров и ромбокубookтаэдров, 4 – усеченных октаэдров и усеченных кубookтаэдров.

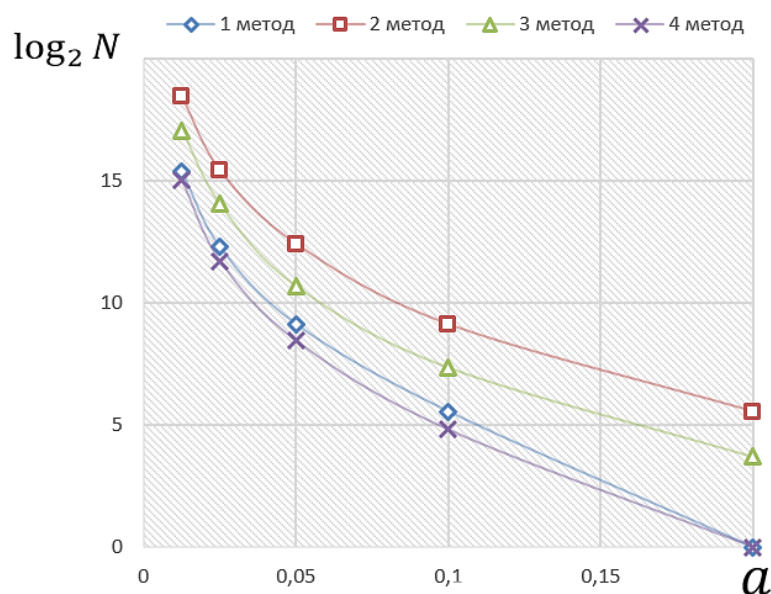


Рисунок 3. Зависимость количества ячеек сетки  $N$  от шага  $a$  для выбранного метода в логарифмическом масштабе

### Разработка модуля построения сеточной модели

Разработка модуля ведется с применением объектно-ориентированного подхода, в рамках которого были введены следующие сущности:

- вершина многогранника – точка в трехмерном пространстве, описываемая тремя значениями координат ( $x$ ,  $y$ ,  $z$ ), для хранения которых используется тип данных с плавающей точкой, при этом точность вычислений определяется пользователем и ограничивается вычислительной архитектурой, на которой производится расчет;
- ребро многогранника – описывается двумя вершинами многогранника;
- грань многогранника – участок плоскости, описываемый вектором вершин, которые образуют замкнутый контур, и вектором нормали, который описывается тремя значениями и определяет ориентацию грани в пространстве;
- геометрическое граничное условие – описывается вектором граней;
- многогранник – ячейка сетки, которая описывается координатами центральной точки, значением длины ребра и видом многогранника (гексаэдр, октаэдр, кубооктаэдр и другие); в процессе обрезки многогранника геометрическим граничным условием вычисляются и преобразуются векторы его вершин и граней, обновленные значения запоминаются;
- сеточная модель – описывается вектором многогранников.

Алгоритм генерации сеточной модели модулем ранее был продемонстрирован в виде блок-схемы на рисунке 1б. Поясняя данную схему, отметим, что в качестве параметров сеточной модели программа принимает величину шага сетки и ее вид. Предполагается, что сгенерированная модель описывается векторами усеченных и обычных многогранников, причем для описания усеченных дополнительно вычисляются и хранятся значения их граней.

### Общая архитектура программы

На рисунке 4 показана общая архитектура разрабатываемой программы, где стрелками показаны два возможных режима ее работы: сплошной линией – режим построения сеточной модели, пунктирной линией – режим получения модели внешней системой.

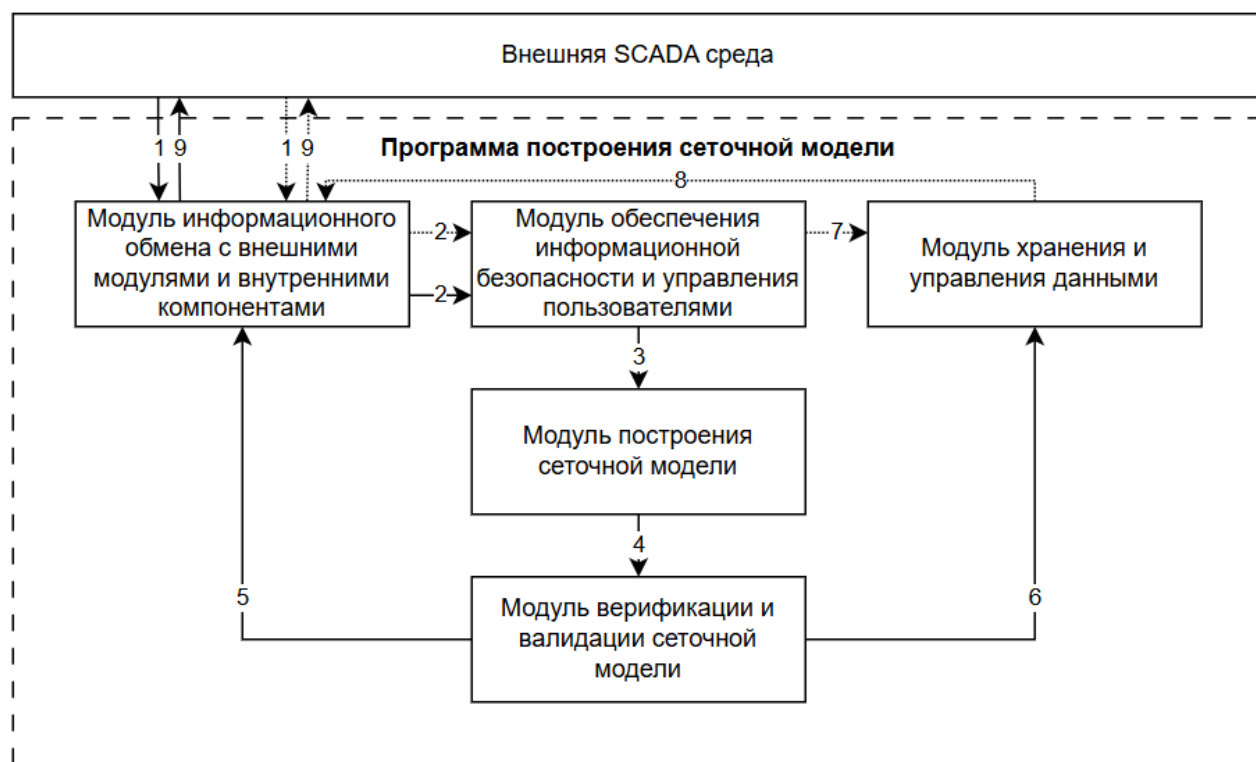


Рисунок 4. Общая архитектура разрабатываемой программы для построения октаэдрических сеточных моделей

Цифрами на стрелках отмечены следующие функциональные связи между модулями программы:

- 1) запуск программы,
- 2) авторизация пользователя и предоставление ему прав на основе ролевой модели,
- 3) настройка параметров и запуск расчета сеточных моделей,
- 4) запуск верификации и валидации сеточных моделей,
- 5) возврат статуса процесса построения модели внешней системе,
- 6) сохранение результатов моделирования совместно с параметрами расчетов в библиотеку результатов,
- 7) параметрический запрос результатов расчета внешней системой из библиотеки,
- 8) передача результатов расчета внешней системе,
- 9) выход из программы.

Задачи модулей программы приведены в таблице.

#### Задачи модулей программы построения октаэдрических сеточных моделей

№	Название модуля	Описание задачи
1	Модуль информационного обмена с внешними модулями и внутренними компонентами	Реализация API для внешних компонентов
		Реализация конвертеров форматов
		Поддержка версионности
2	Модуль обеспечения информационной безопасности и управления пользователями	Управление пользователями на основе ролевой модели
		Авторизация пользователей
		Аутентификация пользователей
		Журнализация событий
		Антивирусная защита
3	Модуль хранения и управления данными	Обеспечение механизмов ввода и вывода данных по запросу внешних и внутренних компонентов приложения или пользователя
		Обеспечение целостности данных
		Механизмы резервного копирования
4	Модуль построения сеточной модели	Подготовка полиэдрической сеточной модели в рамках заданных граничных условий
5	Модуль верификации и валидации сеточной модели	Проверка сгенерированной модели на соответствие заданным условиям

Приведем краткое описание сценариев функционирования разрабатываемой программы. Режим построения сеточной модели можно описать последовательностью следующих шагов:

- 1) авторизация пользователя, получение прав на основе ролевой модели,
- 2) настройка параметров расчета сеточных моделей,
- 3) построение сеточной модели на основе введенных параметров,
- 4) верификация и валидация сеточной модели,
- 5) сохранение результатов расчета совместно с его параметрами в библиотеку результатов,
- 6) завершение сеанса работы.

В режиме получения расчетной модели последовательно выполняются следующие шаги:

- 1) авторизация пользователя, получение прав на основе ролевой модели,
- 2) параметрический запрос результатов расчета внешней системой из библиотеки,
- 3) передача результатов расчета внешней системе,
- 4) завершение сеанса работы.

#### Выводы

Разработан геометрический метод построения сеточных моделей на основе архимедовых тел. Разработана общая архитектура программного обеспечения, предназначенного для построения сеточных моделей, на основе данного метода. Сформулированы основные принципы разработки вычислительного модуля построения сеточных моделей.

**Список литературы**

1. Arocena V. M., Danao L. A. M. Improving the Modeling of Pressure Pulsation and Cavitation Prediction in a Double-Volute Double-Suction Pump Using Mosaic Meshing Technology // Processes. 2023. Vol. 11, iss. 3. P. 660. DOI: 10.3390/pr11030660; EDN: SUEJNA
2. Численное моделирование гидродинамических процессов в тракте теплоносителя заборного теплообменного аппарата / А. М. Белов, С. С. Гоняева, В. В. Кожемякин, Н. А. Морозов // Морские интеллектуальные технологии. 2022. № 4-2(58). С. 87–91. DOI: 10.37220/MIT.2022.58A.010; EDN: CUXPXM
3. Февральских А. В. Проектирование аэродинамической компоновки скоростного амфибийного судна с использованием технологий цифрового двойника // Морской вестник. 2021. № 3(79). С. 14–19. EDN: PPIAOV



## Цифровые двойники в задачах валидации экспериментальных исследований и установок

О. И. Ведяйкина, П. А. Хазов

Нижегородский государственный архитектурно-строительный университет,  
Нижний Новгород, Россия

**Аннотация.** Статья посвящена изучению проблемы повышения эффективности физических исследований путем создания цифрового двойника сложной технической установки. Актуальность темы обусловлена необходимостью валидации экспериментальных исследований, вследствие часто сложности визуализации и верификации их результатов. Предложен алгоритм создания цифрового двойника экспериментальной установки на примере малогабаритной дозвуковой аэродинамической трубы, который в дальнейшем может быть использован для проведения виртуальных испытаний, с целью повышения восприятия физики процесса, а также создания методики учета погрешностей реального эксперимента и интеграции цифровых технологий в жизненный цикл установок.

**Ключевые слова:** цифровой двойник, компьютерное моделирование, геометрическая оптимизация, валидация исследования, валидация экспериментальной установки

## Digital twins in the validation of experimental studies and installations

O. I. Vediaikina, P. A. Khazov

Nizhny Novgorod State University of Architecture and Civil Engineering, Nizhny Novgorod, Russia

**Abstract.** The article studies the problem of increasing the efficiency of physical research by creating a digital twin of a complex technical installation. The relevance of this topic is due to the need to validate experimental research, which is often difficult to visualize and verify. The article proposes an algorithm for creating a digital twin of an experimental installation, using the example of a small-sized subsonic wind tunnel, which can be used for conducting virtual tests to improve the understanding of the physics of the process, as well as for creating a methodology for accounting for the errors of real experiments and integrating digital technologies into the life cycle of installations.

**Keywords:** digital twin, computer modeling, geometric optimization, research validation, experimental setup validation.

### Введение

На сегодняшний день одной из наиболее актуальных задач в области инженерии и технологий является разработка эффективных методов, обеспечивающих цифровую поддержку для разнообразных технических объектов, систем, в том числе процессов исследования, практически во всех областях технических наук и промышленности [1–3]. Особое внимание стоит уделить сложным, уникальным и высокотехнологичным экспериментальным установкам, где реализация физических экспериментов часто сталкивается с трудностями, связанными с визуализацией получаемых результатов [4].

Уровень возможностей, которого в настоящее время достигло компьютерное моделирование, позволяет создавать надежные методики и алгоритмы цифровой поддержки экспериментальных исследований и установок, применимых в дальнейшем в инженерной практике [5–8].

Одним из направлений в разработке таких алгоритмов является применение концепций цифровых двойников и цифровых (виртуальных) испытаний [9–11]. Цифровой двойник (ЦД) представляет собой виртуальную копию реального объекта, которая отражает полностью как его геометрию, так и все физические данные, сопровождающие процессы, происходящие в установке. Использование ЦД значительно упрощает отслеживание измеряемых и изучаемых параметров, визуализацию сложных экспериментальных процессов и лежит в основе проведения цифровых испытаний, направленных на определение качественных и количественных характеристик свойств объекта, что, в свою очередь, позволит существенно повысить эффективность физических экспериментов. Предлагаемый алгоритм валидации экспериментальных установок представлен на рисунке 1.

Таким образом, развитие технологий ЦД и методов компьютерного моделирования не только улучшает качество испытаний, содействует поддержанию жизненного цикла экспериментальных установок, но и способствует более глубокому пониманию происходящих физических процессов, что открывает новые возможности для оптимизации процессов в различных отраслях науки.

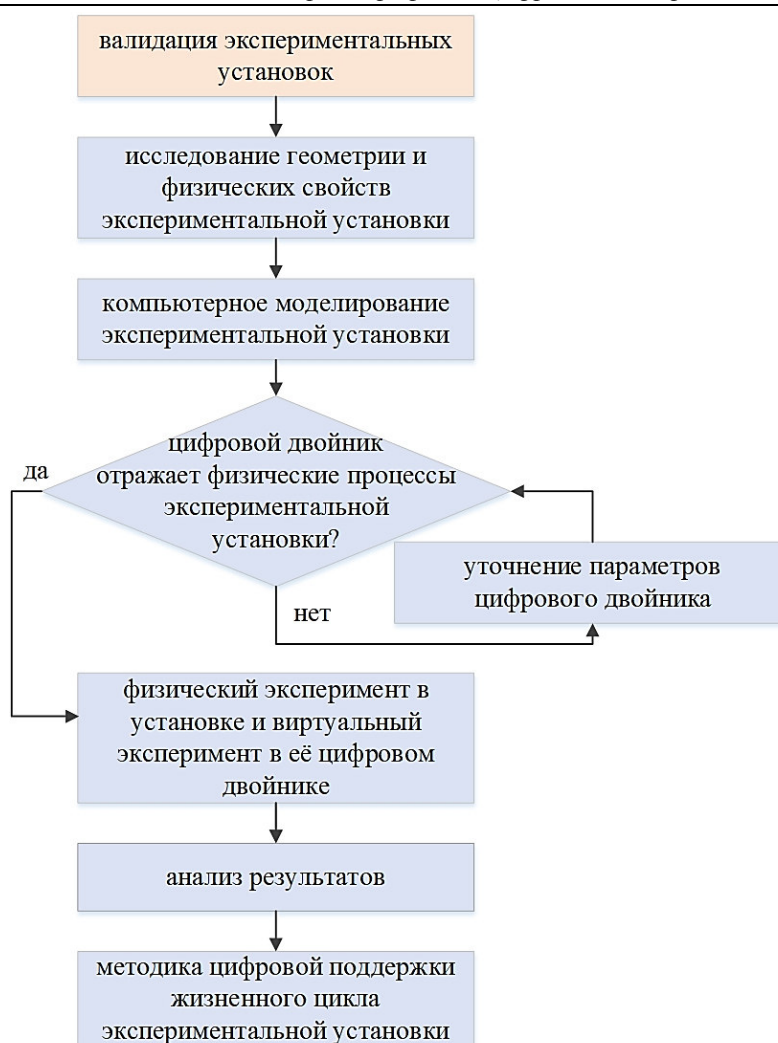


Рисунок 1. Алгоритм цифровой поддержки в разработке методики валидации экспериментальной установки

### Постановка задачи

Создание ЦД технических объектов может служить для валидации процессов, происходящих в них, с целью продления их жизненного цикла. Примером такой экспериментальной установки является дозвуковая малогабаритная аэродинамическая труба.

Проектирование современных зданий и сооружений, особенно тех, что обладают сложной, нестандартной геометрической формой, ставит перед инженерами сложную задачу: правильно оценить аэродинамическую нагрузку на их конструкции. Для получения максимально точных данных традиционно используются модельные эксперименты в крупногабаритных трубах. Результаты экспериментов, проведенных в таких установках, отличаются высокой достоверностью, подтвержденной многолетней практикой [12]. Однако это крайне сложные инженерные сооружения, требующие больших финансовых и трудовых затрат. В качестве более экономичной альтернативы могут рассматриваться аэродинамические трубы малого диаметра, но их ограниченные размеры и мощности существенно влияют на качество получаемых результатов: возникают искажения картины обтекания моделей и, соответственно, погрешности в измерениях, что делает их применение в серьезных исследованиях практически невозможным [13, 14]. В контексте всего вышесказанного создание ЦД для такой установки является крайне актуальной задачей, так как цифровые испытания в ней позволят не только выявить недостатки и несовершенства в её работе, но и детально визуализировать их.

ЦД способствует проведению валидации (проверки на соответствие) результатов физических исследований, которые позволяют значительно расширить возможности использования малогабаритных установок, сделав их результаты сравнимыми по точности с результатами крупногабаритных аналогов, что впоследствии увеличит их жизненный цикл.

## Теория

Валидация экспериментальных исследований осуществляется для вертикальной малогабаритной дозвуковой аэродинамической трубы лаборатории кафедры отопления и вентиляции ННГАСУ с открытой рабочей областью (рис. 2, 3).



Рисунок 2. Аэродинамическая труба – прототип цифрового двойника

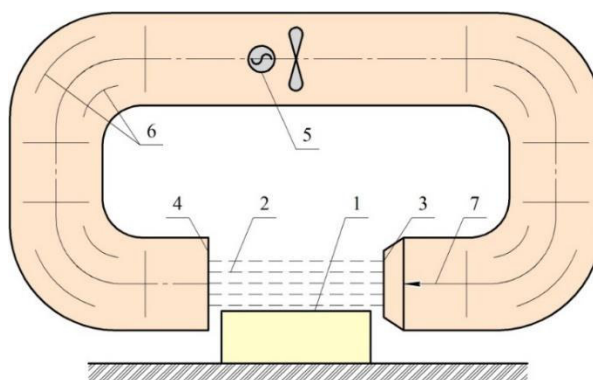


Рисунок 3. Схема аэродинамической трубы: 1 – подмакетник; 2 – рабочая область; 3 – выходное сечение; 4 – входное сечение; 5 – аэродинамическая труба с осевым вентилятором; 6 – направляющие ребра; 7 – направление ветрового потока

Для создания ЦД аэродинамической трубы в ней проводился эксперимент без установки макета в рабочей области с целью изучения картин распределения воздушных потоков. Измерения производились в плоскости выходного и входного сечений аэродинамической трубы, а также в различных точках рабочей области электронным термоанемометром Testo 405v1.

Виртуальные испытания проводились в программном комплексе ANSYS с использованием модуля вычислительной гидрогазодинамики ANSYS CFX с заданием следующих условий:

- размеры расчетной области ограничены реальными размерами помещения, где расположена установка, обладающая свойствами несжимаемой газовой среды;
- все твердые поверхности (труба, подмакетник, пол) имеют условие, обеспечивающее отсутствие проникания воздушных потоков сквозь них;
- вход воздушных потоков осуществляется через сечение выходной части трубы с заданной скоростью, достигающей 12 м/с в выходном сечении рабочей области;
- выход воздушных потоков происходит через сечение во входной части трубы, а также по области вокруг неё;
- расчетная сетка формируется адаптивной, то есть со сгущением к рабочей области и укрупнением на внешних границах;
- остальные параметры выбирались в процессе исследования (рис. 4); для достижения наиболее точного схождения результатов физического и виртуального экспериментов будут описаны ниже.

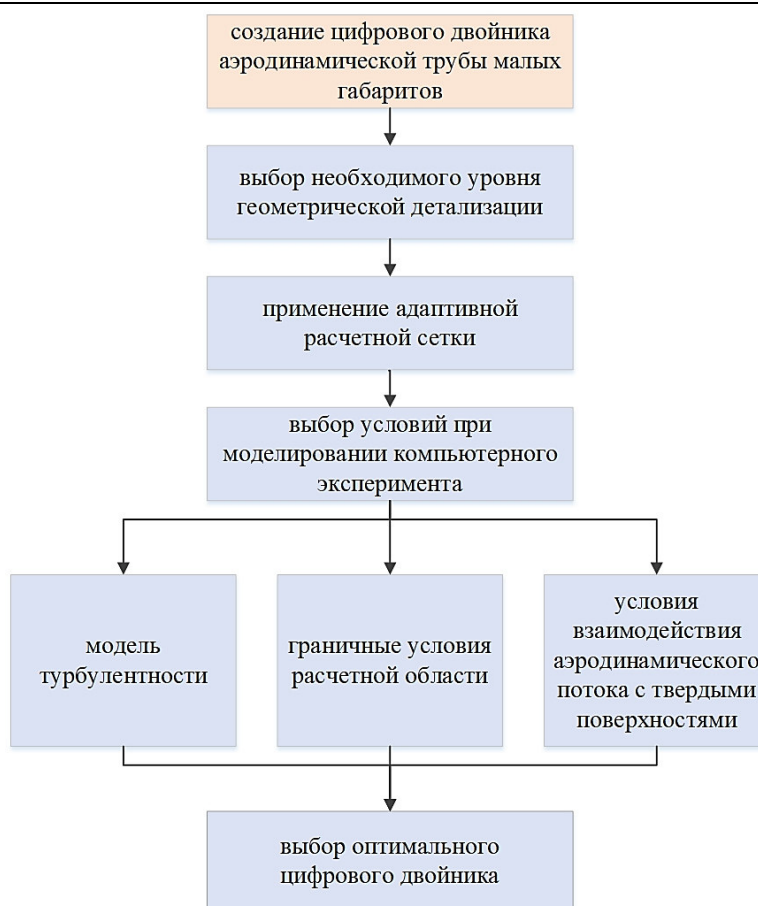


Рисунок 4. Алгоритм создания цифрового двойника на примере малогабаритной аэродинамической трубы

### Результаты экспериментов

Первым этапом создания ЦД экспериментальной установки является создание её геометрической модели и установление необходимого уровня геометрической детализации [15]. Уровни отличаются степенью имитации – от коротких выпусков до полной модели (рис. 5), при этом в последнем варианте имеется разрыв в месте установки вентилятора для конструктивной возможности задания ветрового потока при выполнении компьютерного исследования.

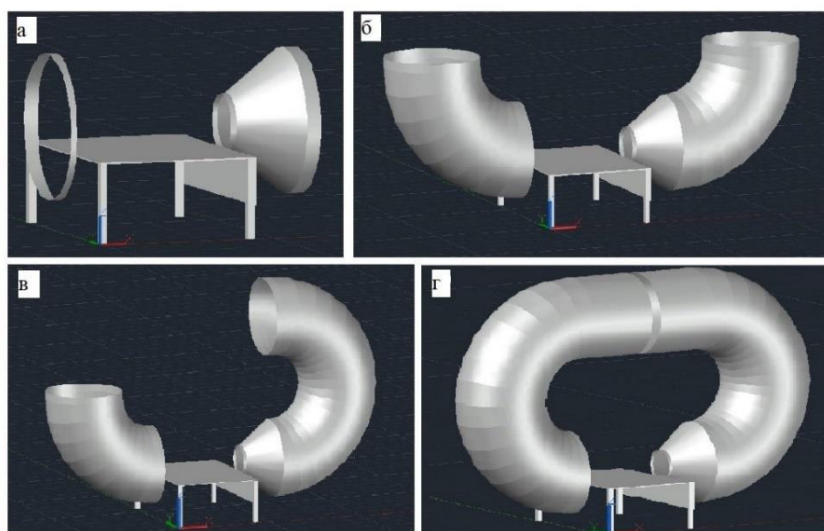


Рисунок 5. Уровни геометрической детализации моделей аэродинамической трубы: а) первый вид – короткие выпуски трубы; б) второй вид – длинные выпуски трубы; в) третий вид – часть трубы с выходным сечением, практически полностью повторяющим конструкцию установки; г) четвертый вид – целая труба



Предложенное альтернативное решение геометрических моделей продиктовано стремлением создать наиболее эффективный ЦД установки, способный достоверно воспроизводить её особенности и протекающие в ней процессы, одновременно минимизируя затраты цифровых ресурсов.

Визуальное распределение ветровых потоков над поверхностью подмакетника достаточно близкое: максимальные показатели модулей скоростей и завихрения возникают в одних областях (рис. 6).

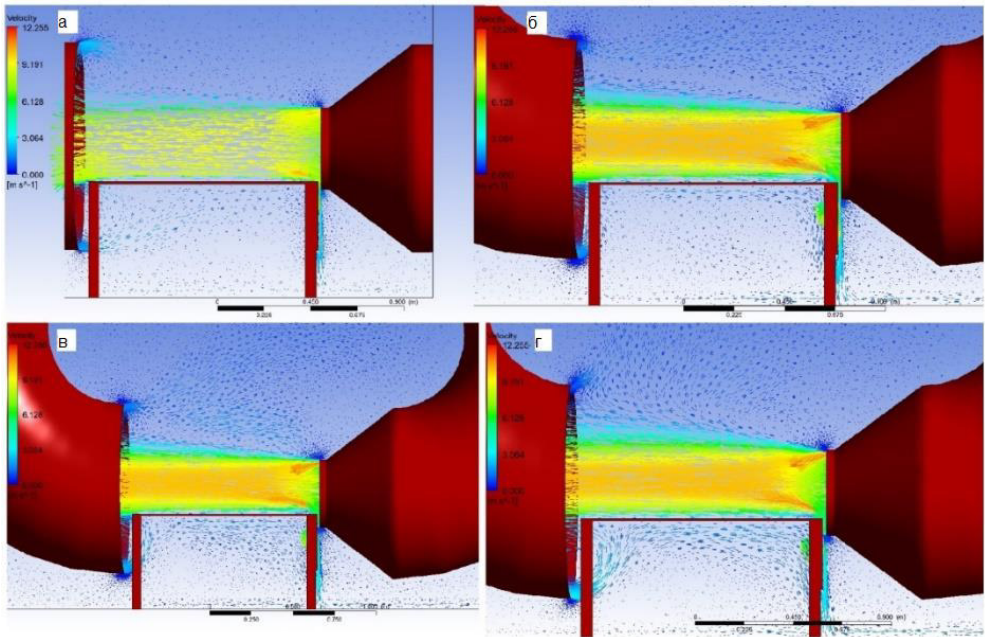


Рисунок 6. Распределение воздушных потоков в рабочей области ЦД аэродинамической трубы при различных уровнях геометрической детализации: а) первый вид – короткие выпуски трубы; б) второй вид – длинные выпуски трубы; в) третий вид – часть трубы с выходным сечением, практически полностью повторяющим конструкцию установки; г) четвертый вид – целая труба

Более детальный анализ картины распределения ветровых потоков выявил значительное влияние уровня геометрической детализации (табл. 1). Наиболее отличаются значения в первом виде геометрической модели, при этом расхождения в точечных значениях скоростей достигают почти 50 %. Недостаточная согласованность результатов не даёт возможности исключения из предложенных геометрий, так как даже самый упрощенный и отличающийся в показателях вариант убирать из исследования не представляется целесообразным ввиду важности учёта ресурсоёмкости расчетов каждого варианта.

Таблица 1. Сравнение значений модулей скоростей четырех видов геометрических моделей цифровых двойников

Высота точки от уровня пола, м	Первый вид	Второй вид	Третий вид	Четвертый вид	Максимальная разница, %
Значения скоростей на выходном сечении, м/с					
0,56	2,66	4,69	4,85	5,06	47,43
0,60	9,32	9,51	9,75	9,97	6,52
0,70	8,54	9,71	9,68	9,64	12,06
0,80	8,62	9,73	9,71	9,70	11,42
Значения скоростей в середине рабочей области, м/с					
0,56	7,92	5,06	5,52	6,21	36,14
0,60	8,56	7,43	7,67	7,99	13,24
0,70	9,09	10,14	10,11	10,07	10,39
0,80	9,03	10,01	10,0	9,92	9,85
Значения скоростей на входном сечении, м/с					
0,56	7,07	4,98	5,61	6,77	29,56
0,60	8,20	7,52	7,81	8,21	8,40
0,70	8,30	10,10	10,07	10,0	17,80
0,80	8,14	10,05	10,03	9,79	19,03
Значения максимальных скоростей, м/с					
	11,41	12,09	12,26	12,03	6,93

Следующим критически важным аспектом при создании ЦД является выбор оптимальной модели турбулентности как неотъемлемой части ветрового потока, оказывающего значительное влияние на распределение аэродинамических сил. В CFD-моделировании (Computational Fluid Dynamics – вычислительная гидрогазодинамика), подразделе механики сплошных сред, предназначенном для вычисления характеристик потоковых процессов, существует множество моделей турбулентности, различающихся по сложности и точности.

Для исследования влияния модели турбулентности были выбраны три самые распространенные модели, возможные для выбора в вычислительном модуле гидрогазодинамики ANSYS CFX:

- k-Epsilon (k-ε), основанная на решении уравнений энергии турбулентности (k) и скорости её диссипации (ε), широко используется для свободных течений, но плохо описывает пристеночные;
- Shear Stress Transport (SST) сочетает в себе преимущества двух моделей: нивелирует недостатки вышерассмотренной модели в областях близких к стенкам, используя модель k-ω;
- Baseline Reynolds Stress (BSL RS) используется для более точного описания турбулентных потоков в сравнении с предыдущими моделями, подходит для высокоточных аэродинамических расчетов, но требует больше вычислительных ресурсов и времени.

Компьютерное моделирование продемонстрировало, что различные модели турбулентности обеспечивают схожие результаты генерируемых воздушных потоков в рабочей области установки (табл. 2). Тем не менее модель SST имеет весомые преимущества: расчёт был осуществлен минимум в два раза быстрее, следовательно, требуется гораздо меньше вычислительной мощности. Кроме того, модель отличается высокой точностью вычислений и демонстрирует наилучшую сходимост погрешностей решений. Иными словами, данная модель позволяет достичь аналогичных результатов, как и более сложные, однако делает это намного эффективнее с точки зрения временных и аппаратных затрат.

Таблица 2. Сравнение значений модулей скоростей при разных моделях турбулентности

Высота точки от уровня пола, м	Модель турбулентности			Максимальная разница, %
	SST	k-ε	BSL RS	
Значения скоростей на выходном сечении, м/с				
0,56	5,06	4,86	4,95	<b>3,97</b>
0,60	9,97	9,69	9,62	3,46
0,70	9,64	9,70	9,68	0,62
0,80	9,70	9,73	9,76	0,60
Значения скоростей в середине рабочей области, м/с				
0,56	6,21	5,58	5,71	<b>10,14</b>
0,60	7,99	7,52	7,56	5,87
0,70	10,07	10,13	11,12	9,43
0,80	9,92	9,96	9,99	0,79
Значения скоростей на входном сечении, м/с				
0,56	6,77	6,04	5,97	<b>11,81</b>
0,60	8,21	7,81	7,86	4,90
0,70	10,0	10,10	10,07	0,90
0,80	9,79	9,91	10,01	2,12
Значения максимальных скоростей, м/с				
	12,025	11,918	12,082	1,36

Существенное влияние на движение воздушных потоков оказывает условие его взаимодействия с твердыми поверхностями (свойства поверхностей с условием Wall). В ПК ANSYS по умолчанию поверхности с условием Wall имеют свойство No Slip («нет скольжения»), подразумевающее, что скорость потока на границе совпадает со скоростью самой стенки, то есть равна нулю. При этом данным свойством можно пренебречь, если происходит моделирование движения воздушных масс при малых скоростях, поскольку вязкость потока в таком случае пренебрежительно мала, что отражено в свойстве Free Slip («свободное скольжение»), означающее скольжение потока вдоль границы без взаимодействия.

Результаты компьютерного моделирования распределения ветровых потоков наглядно показывают существенное влияние свойств поверхностей (рис. 7). В случае No Slip, предполагающего полное прилипание воздуха к поверхности, наблюдается практически полная остановка потока непосредственно у поверхности подмакетника. Заметно, что после этого начального «замедления» поток остается относительно равномерным по высоте до уровня выходного сечения, демонстрируя минимальное затухание вплоть до входного сечения трубы. Это контрастирует с Free Slip, характеризующимся отсутствием прилипания воздуха к поверхности, где поток демонстрирует существенное затухание с высотой. При этом наблюдается волновой характер течения, более адекватно отражающий реальную картину, так как столкновение потока с поверхностью подмакетника неизбежно приводит к возникновению возмущений и небольшим скачкам скорости. В данном случае скорость потока на поверхности подмакетника не стремится к нулю, поток постепенно затухает к верхней границе рабочей области и ослабевает ко входному сечению трубы.

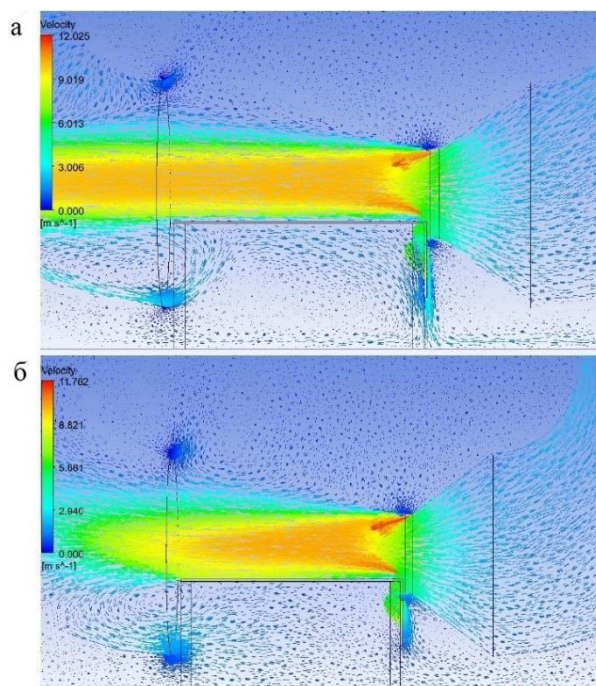


Рисунок 7. Распределение воздушных потоков в рабочей области ЦД аэродинамической трубы при разных свойствах поверхности: а – No Slip; б – Free Slip

Завершающим сравниваемым параметром являются условия на границах исследуемой расчетной области:

- Symmetry, используемое для моделирования симметричных потоков или симметричной геометрии, часто применяется при воспроизведении натурных аэродинамических экспериментов в строительстве;
- Outlet используется для моделирования свободного выхода потока из расчетной области, при этом давление «выхода» является задаваемым параметром;
- Opening используется также для моделирования свободного выхода потока, но позволяет учитывать также и возможность входа (например, в результате разницы давлений).

Результаты компьютерного эксперимента показали незначительные расхождения в картинах распределения воздушных потоков и возможность применимости всех видов условий в зависимости от решаемой задачи. В конкретной задаче поиска ЦД для дозвуковой малогабаритной аэродинамической трубы наиболее отражающим реальную ситуацию на границах расчетной области можно принять условие Outlet.

### Обсуждение результатов

Визуальное распределение ветровых потоков в рабочей области экспериментальной установки (малогабаритная аэродинамическая труба) и её ЦД (четвертый вид геометрической детализации),



полученного в результате выполнения действий алгоритма (рис. 4), достаточно близки, что отражается в схожем градиенте изменения скоростей по вертикальному направлению в рабочей области, а также практически идентичному затуханию к входному сечению (рис. 8). Различия наблюдаются вблизи выходного сечения, что может быть получено вследствие сложности физического замера показаний, связанного с неравномерностью создаваемого ветрового потока в малогабаритной трубе, являющейся её функциональной особенностью. При этом вышеуказанное отличие несущественно, так как установка исследуемого в рамках проведения аэродинамического эксперимента макета на подмакетник осуществляется в центральной части рабочей области.

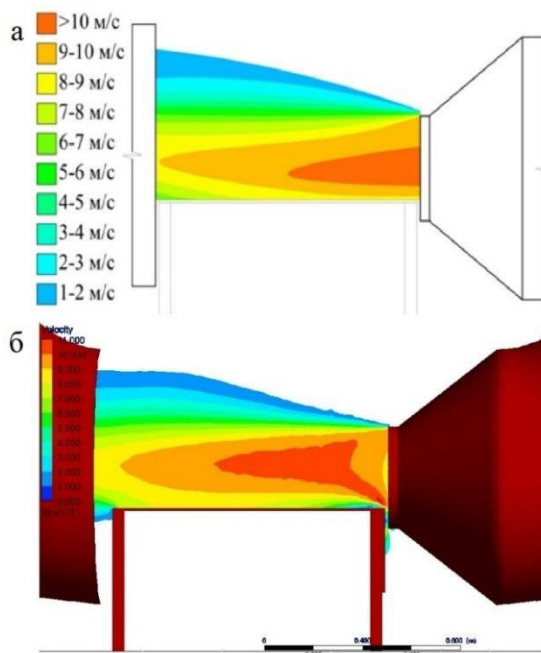


Рисунок 8. Распределение воздушных потоков в рабочей области реальной аэродинамической трубы (а) и её цифрового двойника (б)

Отличия в отдельных точках в реальном потоке и его компьютерном варианте остаются в пределах 15 % (табл. 3), при этом для оценки разницы в усредненных показателях сравниваются значения средних скоростей в каждом из трёх ключевых областей рабочей области и максимальные отклонения в этом случае не превышают 5 % – что, безусловно, является показателем достаточно точным для цифрового двойника экспериментальной установки.

Таблица 3. Сравнение значений модулей скоростей реальной аэродинамической трубы и её ЦД

Высота точки от уровня подмакетника (пола), м	Скорости на выходном сечении, м/с			Скорости в середине рабочей области, м/с			Скорости на входном сечении, м/с		
	Реальная труба	ЦД	Разница, %	Реальная труба	ЦД	Разница, %	Реальная труба	ЦД	Разница, %
0,1 (0,57)	8,58	9,57	13,77	8,4	8,37	0,35	6,3	5,5	12,78
0,08 (0,64)	10,7	10,21	4,6						
0,10 (0,66)				10,0	9,88	1,19	8,61	8,46	1,71
0,16 (0,72)	10,5	9,65	8,1						
0,20 (0,76)				9,5	10,06	5,56	9,0	8,36	7,1
0,24 (0,80)	9,6	9,82	2,27						
0,30 (0,86)				7,0	7,33	4,5	6,9	6,32	8,42
0,32 (0,88)	9,0	7,76	13,78						
0,40 (0,96)				2,9	2,83	2,59	3,25	3,55	8,48

Высота точки от уровня подмакетника (пола), м	Скорости на выходном сечении, м/с			Скорости в середине рабочей области, м/с			Скорости на входном сечении, м/с		
	Реальная труба	ЦД	Разница, %	Реальная труба	ЦД	Разница, %	Реальная труба	ЦД	Разница, %
0,50 (1,06)				0,8	0,83	3,15	1,3	1,35	3,7
0,60 (1,16)				0,5	0,5	0,8	0,91	0,97	6,31
<b>Среднее значение скорости</b>	9,676	9,4	<b>2,83</b>	5,585	5,68	<b>1,73</b>	4,602	4,41	<b>4,18</b>

## Заключение

В ходе исследования был изучен вопрос создания ЦД экспериментальной установки на примере малогабаритной дозвуковой аэродинамической трубы. Этот процесс подразумевает создание точной компьютерной модели физической системы, способной имитировать поведение реальной установки во всех режимах эксплуатации.

Следующим этапом развития данного направления (рис. 1) будет являться использование полученного ЦД для проведения цифровых испытаний и исследований с целью выявления потенциальных ошибок и погрешностей реальных экспериментов, оценка степени их влияния на конечные результаты и разработка методики их учета и коррекции. Это, в свою очередь, даст возможность интегрировать цифровые технологии поддержки полного жизненного цикла сложных исследовательских установок, что обеспечит повышение точности проводимых экспериментальных исследований, сокращение их сроков и оптимизацию затрат на научные изыскания.

## Список литературы

1. Февральских А. В. Цифровая поддержка проектирования аэрогидродинамической компоновки скоростного амфибийного судна // Вестник компьютерных и информационных технологий. 2025. Т. 22, № 1 (247). С. 35–41. DOI: 10.14489/vkit.2025.01.pp.035-041; EDN: LLPGGC
2. Рутман Ю. Л., Мелешко В. А. Применение пакета ANSYS CFX для исследования аэродинамической неустойчивости сооружения // Вестник гражданских инженеров. 2011. № 1 (26). С. 58–61.
3. Singh J., Roy A. K. CFD simulation of the wind field around pyramidal roofed single-story buildings // SN Applied Sciences. 2019. Vol. 1, iss. 11. Art. 1425 (). DOI: 10.1007/s42452-019-1476-2
4. Ведяйкина О. И., Хазов П. А., Шилов С. С. Алгоритм цифровой поддержки работоспособности малогабаритных аэродинамических экспериментальных установок // Вестник Южно-Уральского государственного университета. Серия: Строительство и архитектура. 2025. Т. 25, № 2. С. 62-71. DOI: 10.14529/build250208; EDN: RZJJND
5. Идрисова Ю. В., Типеев А. Н., Перевертайло Д. А. Методика разработки цифровой модели мехатронной системы мотор-шпиндель для исследования его виброустойчивости // Materials. Technologies. Design. 2020. Т. 2, № 1(2). С. 59–68. EDN: OMWQMJ
6. Февральских А. В. Компьютерное моделирование аэродинамики пропеллера в российском пакете программ Flowvision // Русский инженер. 2024. № 4(85). С. 32–35. EDN: IPFHNT
7. Ветрянщиков И. А. К вопросу о компьютерном моделировании и анализе аэродинамики группы зданий городского района // Известия ТулГУ. Технические науки. 2021. № 12. С. 253–255.
8. Deineko A. V., Shamshurin A. I., Kazaryan N. A. Selection of the computational model of wind flow in the problems of computational architectural and civil engineering aerodynamics in accordance with regulatory and technical documents // International journal for computational civil and structural engineering. 2019. Vol. 15, iss. 1. Pp. 14–28.
9. Швачко Е. В., Васильев А. Н., Шишляев В. В. Цифровой двойник скважин – инструмент оптимизации процесса проектирования и строительства // Газовая промышленность. 2024. № 1 (860). С. 36–41. EDN: TWLHBD
10. Сапелкин Р. И., Матренинский С. И., Мищенко В. Я. Цифровой двойник процесса вулканизации материала защитного покрытия из каучукового бетона на поверхности трубного металла // Научный журнал строительства и архитектуры. 2025. № 1(77). С. 102–110. DOI: 10.36622/2541-7592.2025.77.1.010; EDN: JQJBDI

- 
11. Кабалдин Ю. Г., Шатагин Д. А., Аносов М. С., Кузьмишина А. М. Цифровой двойник процесса стружкообразования // Вестник машиностроения. 2020. № 11. С. 81–86. DOI: 10.36652/0042-4633-2020-11-81-8; EDN: XKIABS
12. Белостоцкий А. М., Дубинский С. И., Афанасьева И. Н. Численное моделирование задач строительной аэродинамики. Разработка методик расчета ветровых воздействий и исследование реальных объектов // Вестник МГСУ. 2010. № 4-5. С. 182–185.
13. Экспериментальное исследование ветрового воздействия на строительные конструкции в аэродинамической трубе с малой рабочей частью / О. И. Поддаева, А. Н. Федосова, С. А. Саврасов, А. Т. Каграманян // Научно-технический вестник Поволжья. 2018. № 5. С. 140–143.
14. Экспериментальное исследование распределения аэродинамических коэффициентов на высотное здание / А. А. Сатанов, М. Л. Поздеев, А. В. Симонов, А. П. Помазов, П. А. Хазов // Приволжский научный журнал. 2022. № 3. С. 43–51.
15. Ведяйкина О. И. Валидация экспериментальных методов исследований в малогабаритных аэродинамических трубах // Актуальные проблемы компьютерного моделирования конструкций и сооружений: тезисы докладов IX Международного симпозиума / ФГБОУ ВО «Пензенский государственный университет архитектуры и строительства». Пенза: Изд-во ПГУ, 2025. С. 50–51.

## Алгоритмы автоматизации определения аэродинамических показателей поверхностей отрицательной гауссовой кривизны

С. С. Шилов

Нижегородский государственный архитектурно-строительный университет,  
Нижний Новгород, Россия

**Аннотация.** В статье описываются процессы цифровой поддержки поверхностной с отрицательной гауссовой кривизной и сопутствующие процессы, касающиеся геометрического и математического моделирования, а также параметризации форм. Приведено геометрическое описание тороидной и глобoidной поверхностей наложением необходимых ограничений для создания конечной конфигурации. Представлены результаты компьютерного моделирования ветровых потоков при их взаимодействии с исследуемыми поверхностями. При систематизации полученных результатов разработаны алгоритмы автоматизации определения аэродинамических характеристик, которые вошли в основу инженерных расчетов, позволяющих получать необходимые данные без проведения дополнительных вычислительных операций.

**Ключевые слова:** поверхность с отрицательной гауссовой кривизной, цифровая поддержка проектирования, аэродинамические воздействия

## Geometric methods and algorithms for automating the modeling of aerodynamics of surfaces with negative Gaussian curvature

S. S. Shilov

Nizhny Novgorod State University of Architecture and Civil Engineering,  
Nizhny Novgorod, Russia

**Abstract.** The article describes the processes of digital support for a surface with negative Gaussian curvature and related processes related to geometric and mathematical modeling, as well as shape parameterization. A geometric description of the toroidal and globoid surfaces is given by imposing the necessary constraints to create the final configuration. The results of computer simulation of wind flows in their interaction with the studied surfaces are presented. When systematizing the results obtained, algorithms for automating the determination of aerodynamic characteristics have been developed, which form the basis of engineering methods that allow obtaining the necessary data without additional computational operations.

**Keywords:** negative Gaussian curvature surface, digital design support, aerodynamic effects.

### Введение

Геометрические формы играют ключевую роль в инженерном деле, определяя функциональность, надежность и эффективность применяемых элементов различных технических систем [1, 2]. Разнообразие форм способствует возникновению определенных сложностей у инженеров и проектировщиков при решении поставленных перед ними задач. Принцип наличия согласованности между применяемой формой и требуемым функционалом позволяет оптимизировать как отдельные конструктивные элементы, так и объекты проектирования в целом.

Поверхности с отрицательной гауссовой кривизной характеризуются наличием седловидных форм, в которых одна главная кривизна положительна, а другая – отрицательна. Такие поверхности обладают уникальными свойствами, такими как гибкость, устойчивость к деформациям и способность адаптироваться к различным нагрузкам. Геометрические описания таких поверхностей рассмотрены в работах [3, 4].

Воздушные потоки играют важную роль в различных областях науки и техники, включая аэродинамику, климатологию и инженерию. Понимание взаимодействия воздушных потоков с различными геометрическими формами помогает оптимизировать конструкции самолетов, зданий и других объектов, обеспечивая эффективность и безопасность эксплуатации. Вопрос взаимодействия потоков воздуха с различными геометрическими формами рассматривался многими учеными в научных работах [5].

Взаимодействие криволинейных поверхностей с потоками воздушных масс, процесс оптимизации и автоматизации моделирования аэродинамических процессов легли в основу настоящего исследования.

### Постановка задачи

Целью исследования является разработка алгоритмов автоматизации моделирования аэродинамики поверхностей отрицательной гауссовой кривизны. Для достижения поставленной цели решался ряд задач:

- геометрическое описание рассматриваемых тороидных и глобоидных поверхностей;
- систематизация результатов компьютерного моделирования;
- разработка алгоритмов автоматизированного расчета аэродинамических параметров на поверхности с отрицательной гауссовой кривизной.

### Теория

Объектом исследования являются поверхности отрицательной гауссовой кривизны. Одна поверхность является результатом вырезания внутренней части тороидной поверхности из цилиндрического объема (рис. 1). Вторая поверхность в начертательной геометрии носит название глобоида и образуется путем вращения части дуги вокруг неподвижной оси (рис. 2) [6, 7].

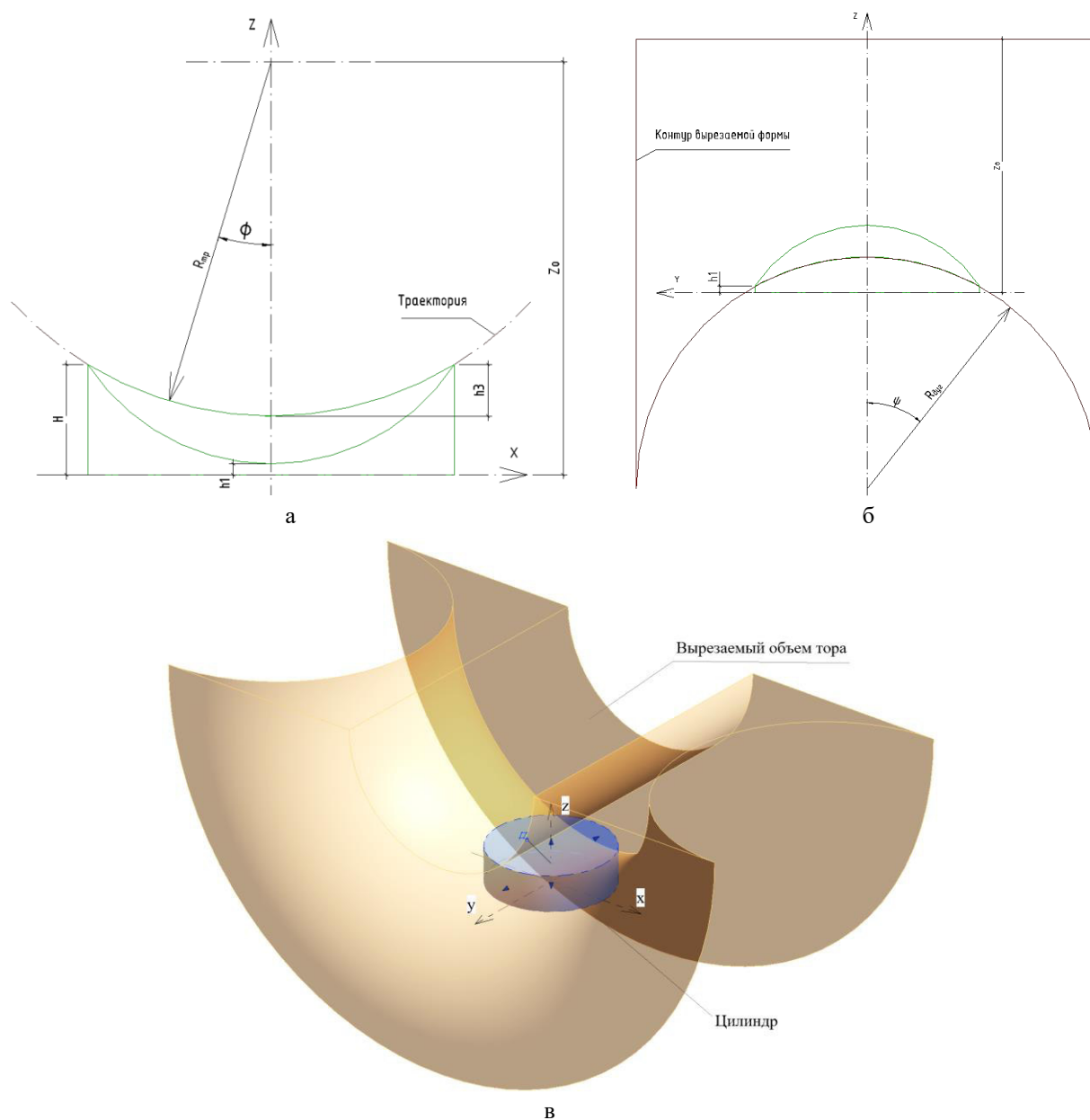


Рисунок 1. Геометрическая схема тороидной поверхности: а – вид спереди; б – вид сбоку; в – общий вид

Поверхность тора описывается в параметрическом виде (1). В рассматриваемой задаче интерес представляет только внутренняя часть тороидной поверхности, которая обеспечивается наложением

ограничений (2). Итоговая поверхность ограничивается цилиндром, соответственно, координаты точек поверхности должны удовлетворять равенству (1) и удовлетворять условиям (3).

$$\begin{cases} x = ((R_{mp} + R_{dyz}) + R_{dyz} \cos \psi) \cdot \cos \varphi; \\ y = R_{dyz} \sin \psi; \\ z = ((R_{mp} + R_{dyz}) + R_{dyz} \cos \psi) \cdot \sin \varphi + (H + R_{mp} - h_3). \end{cases} \quad (1)$$

$$\begin{cases} \psi \in \left[-\frac{\pi}{2}; -\pi\right] \cup \left[\frac{\pi}{2}; \pi\right]; \\ \varphi \in [\pi; 2\pi]. \end{cases} \quad (2)$$

$$\begin{cases} -D/2 \leq x \leq D/2; \\ -D/2 \leq y \leq D/2; \\ 0 < z \leq H. \end{cases} \quad (3)$$

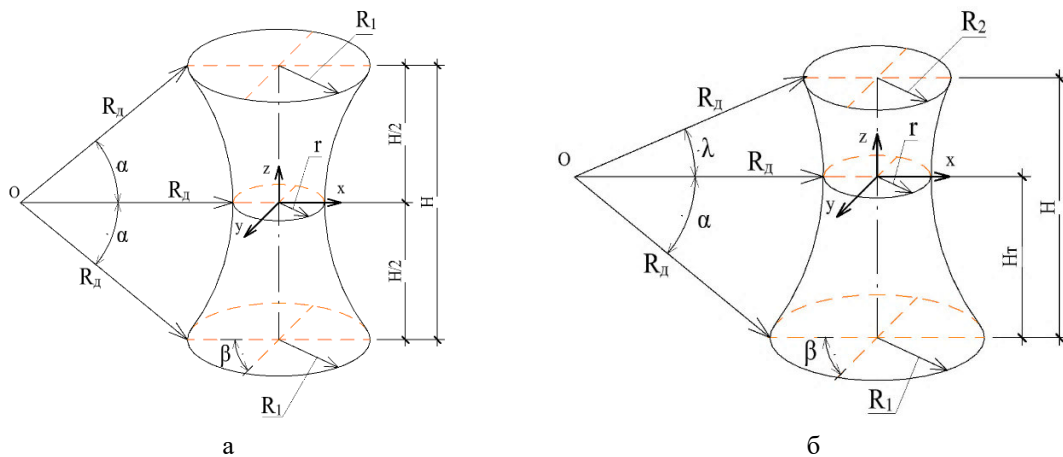


Рисунок 2. Геометрическая схема глобоидной поверхности:  
при равных (а) и различных (б) размерах верхнего и нижнего оснований

Поверхность глобоида можно выразить в виде параметрической зависимости (4), записанной для тороидной поверхности с некоторыми ограничениями:

$$\begin{cases} x(z) = ((R_o + r) - R_o \cos \alpha) \cdot \cos \beta; \\ y(z) = ((R_o + r) - R_o \cos \alpha) \cdot \sin \beta; \\ z \in \left[-\frac{H}{2}; \frac{H}{2}\right]. \end{cases} \quad (4)$$

В основе компьютерного моделирования ветровых потоков лежат уравнения Навье–Стокса, которые описывают движение жидкости и газа [8] (5), а также уравнение неразрывности (6).

$$\frac{\partial \vec{v}}{\partial t} = -(\vec{v} \cdot \nabla) \vec{v} + \nu \Delta \vec{v} - \frac{1}{\rho} \nabla p + \vec{f}; \quad (5)$$

$$\frac{\partial p}{\partial t} + \nabla \cdot (p \vec{v}) = 0, \quad (6)$$

где  $\nabla$  – оператор набла;  $\Delta$  – векторный оператор Лапласа;  $t$  – время;  $\nu$  – коэффициент кинематической вязкости;  $\rho$  – плотность;  $p$  – давление;  $\vec{v}$  – векторное поле скоростей;  $\vec{f}$  – векторное поле силовых масс.

Приведенные в данном разделе геометрические методы построения поверхности отрицательной кривизны легли в основу построения параметрических моделей в среде информационного моделирования, что позволяет создавать множество форм, экономя трудовые и временные ресурсы [9]. Уравнение движения газов использовалось при разработке алгоритмов автоматизации определения аэродинамических характеристик.

### Результаты экспериментов

По результатам моделирования торонидной поверхности были выделены характерные аэродинамические зоны (рис. 3) и систематизированы картины распределения аэродинамических характеристик в зависимости от изменения коэффициента формы  $k$  (рис. 4), который определяется по следующему выражению:

$$k = \frac{H - h_2 - h_3}{D},$$

где  $h_2$  – высота сдвиговой дуги;  $h_3$  – высота дуги-траектории;  $H$  – общая высота;  $D$  – диаметр.

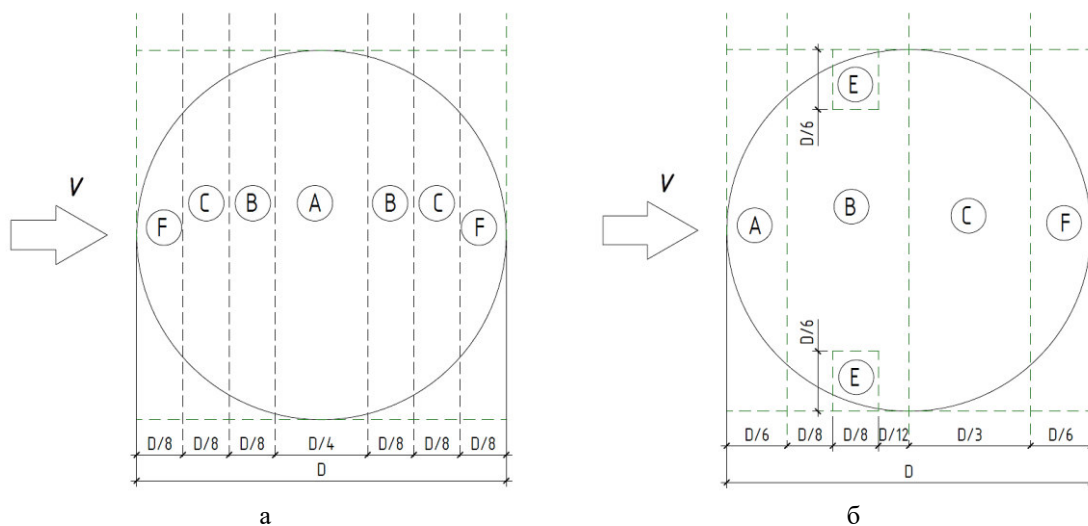


Рисунок 3. Размеры аэродинамических зон: а – «направление 0»; б – «направление 90»

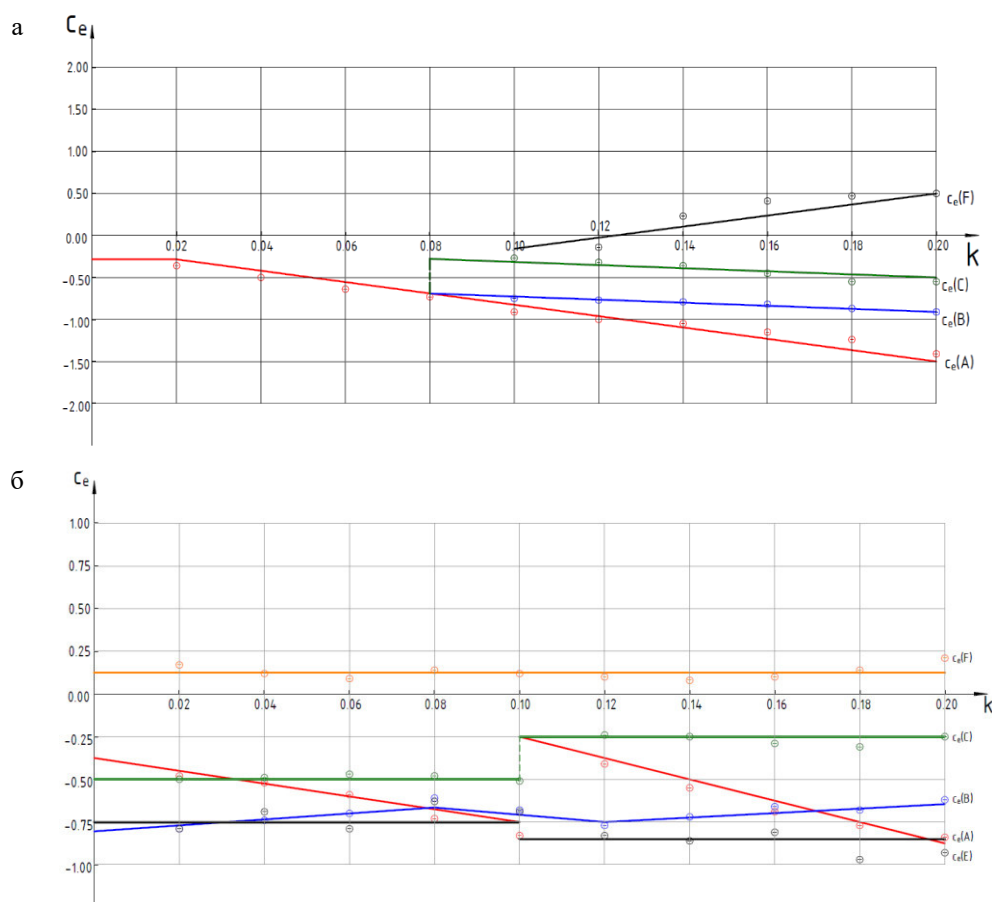


Рисунок 4. Картины изменения аэродинамических коэффициентов при направлении потоков: а – «направление 0»; б – «направление 90»



Для разработки алгоритмов автоматизации моделирования аэродинамики поверхностей отрицательной гауссовой кривизны определялись характерные картины для каждой поверхности. В ходе анализа результатов многократно проведенных компьютерных экспериментов тороидной поверхности были получены три характерные картины распределения аэродинамических коэффициентов по поверхности по одному направлению движения воздушных масс и одна картина по перпендикулярному направлению. Следующим этапом являлась систематизация геометрических фигур, характеризующих изменения аэродинамических коэффициентов, и установление принадлежности искомой точки той или иной аэродинамической зоне. Данный алгоритм был опубликован в научной статье [10].

При анализе результатов компьютерного моделирования аэродинамических особенностей поверхности глобоида было установлено, что имеются три характерные картины распределения исследуемых параметров (рис. 5–7). Каждая картина была разделена на области, для которых были получены осредненные значения аэродинамических коэффициентов. Для возможности выделения момента, когда меняется характерная картина, вводится коэффициент формы, который в общем случае определяется по формуле

$$k = \frac{R_1 + R_2 - 2r}{2H}.$$

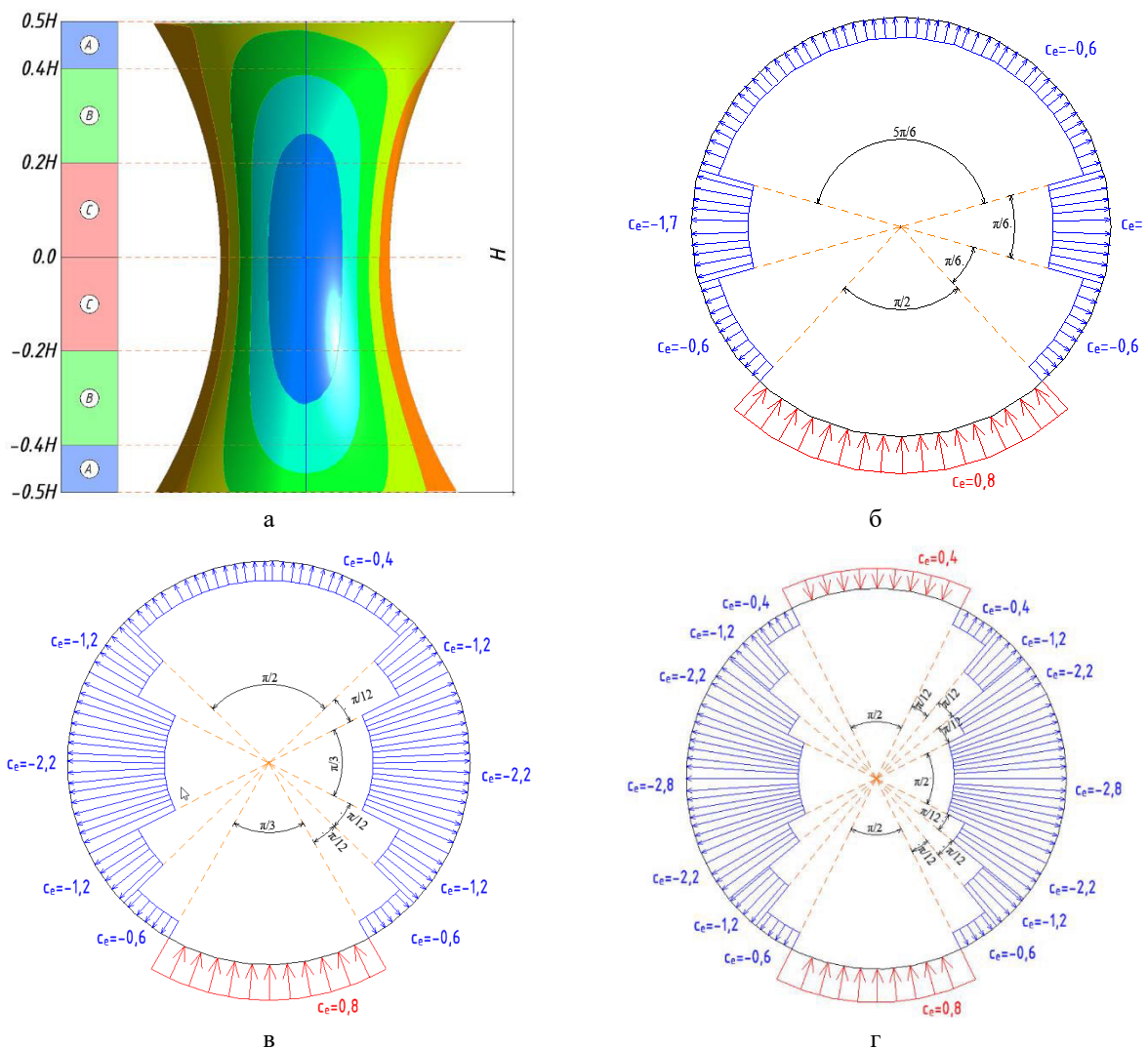


Рисунок 5. Картины распределения аэродинамических коэффициентов:  
а – общий вид первой картины при  $R_1 = R_2$ ; б – для зоны "А"; в – для зоны "В"; г – для зоны "С"

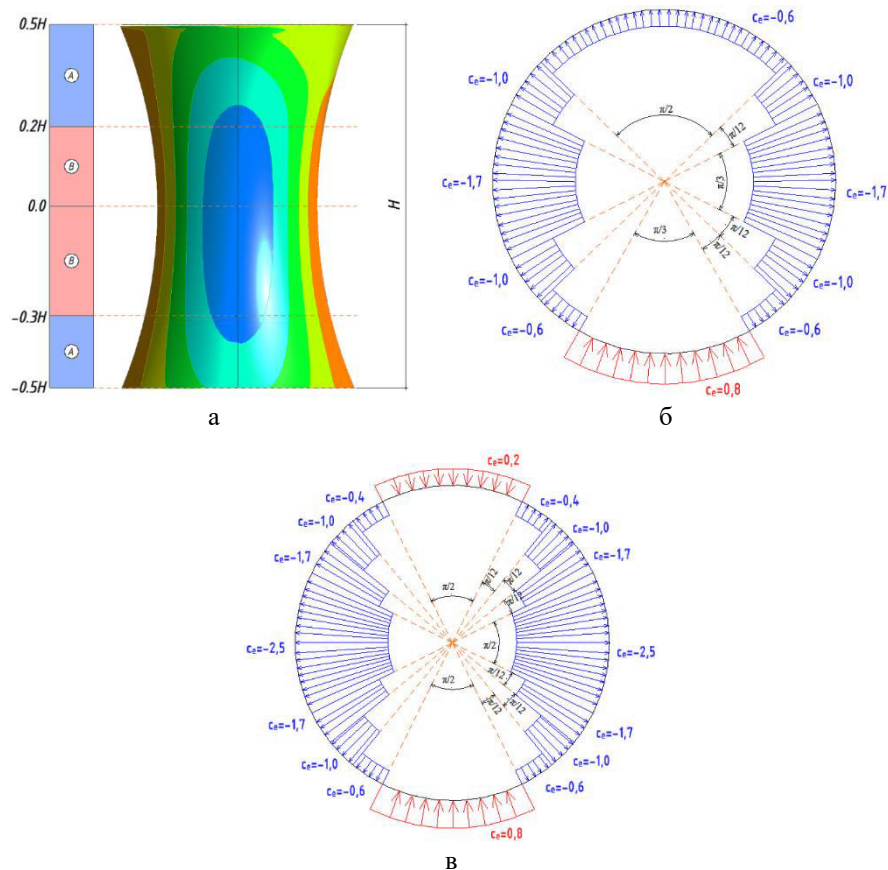


Рисунок 6. Картины распределения аэродинамических коэффициентов:  
 а – общий вид второй картины при  $R_1 = R_2$ ; б – для зоны "А"; в – для зоны "В"

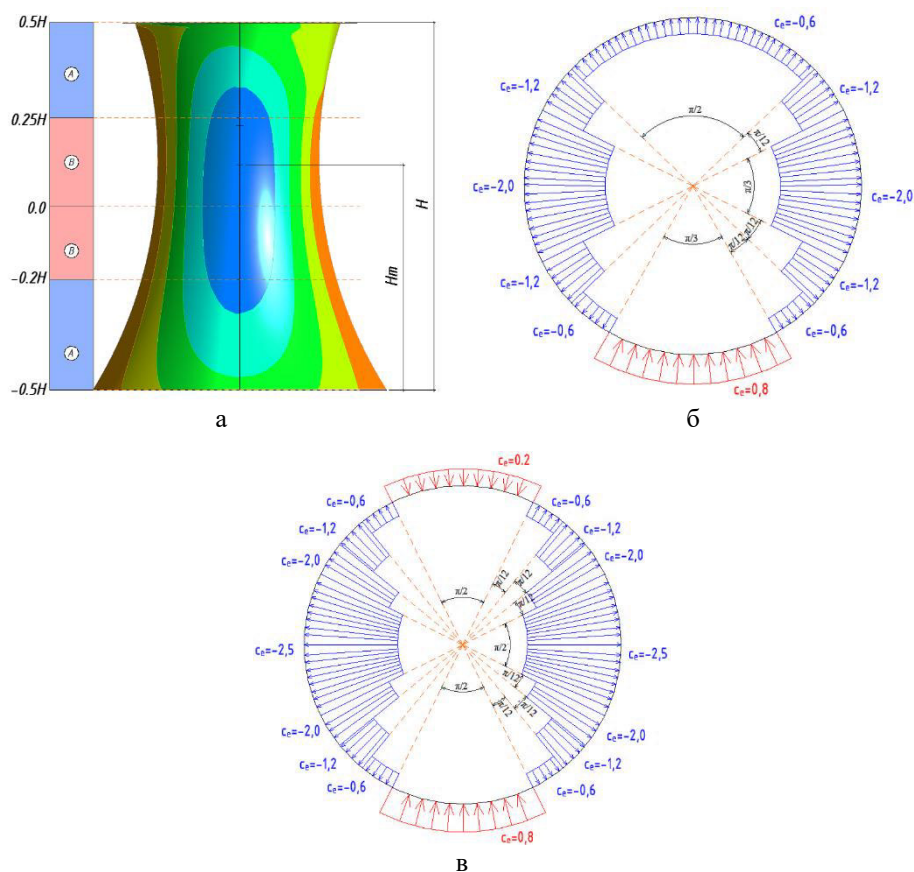


Рисунок 7. Картины распределения аэродинамических коэффициентов:  
 а – общий вид картины при  $R_1 \neq R_2$ ; б – для зоны "А"; в – для зоны "В"

### Обсуждение результатов

Разработанные способы определения аэродинамических характеристик исследуемых поверхностей расширяют существующие положения нормативных баз, в частности в строительных нормах, в которых заложены формы, являющиеся крайним случаем исследуемых поверхностей: сводчатая поверхность – крайняя формы тороидной поверхности (рис. 8), цилиндр – глобоида (рис. 9). Достоверность проверенных исследований достигается строгостью математического аппарата, большого количества вычислительных экспериментов и общей согласованностью с действующими нормативными документами.

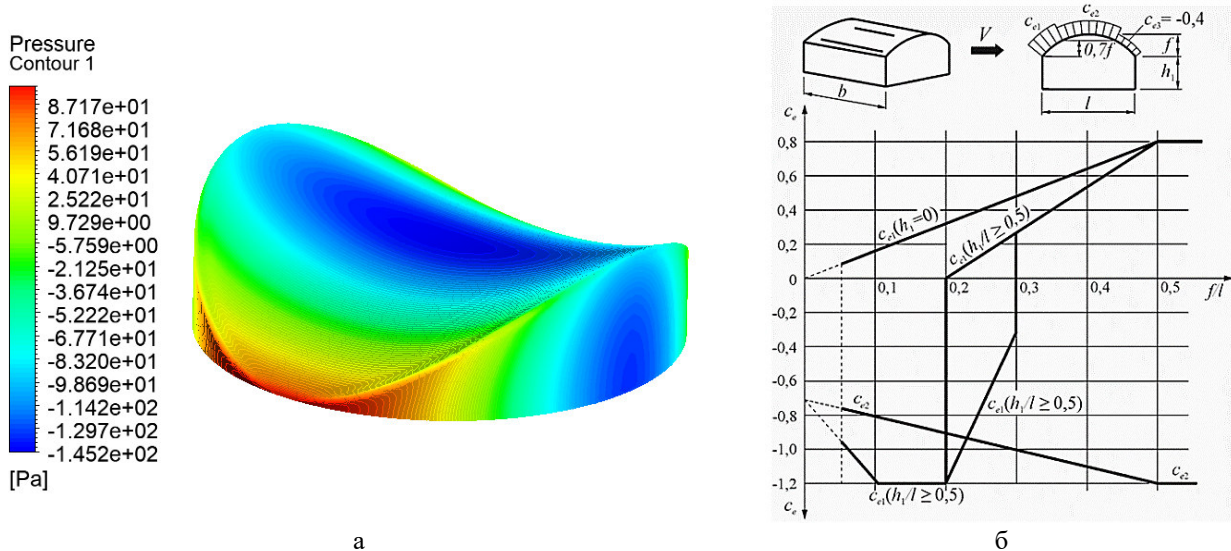


Рисунок 8. Картины распределения аэродинамических коэффициентов: *а* – компьютерное моделирование для тороидной поверхности; *б* – фрагмент действующего нормативного документа

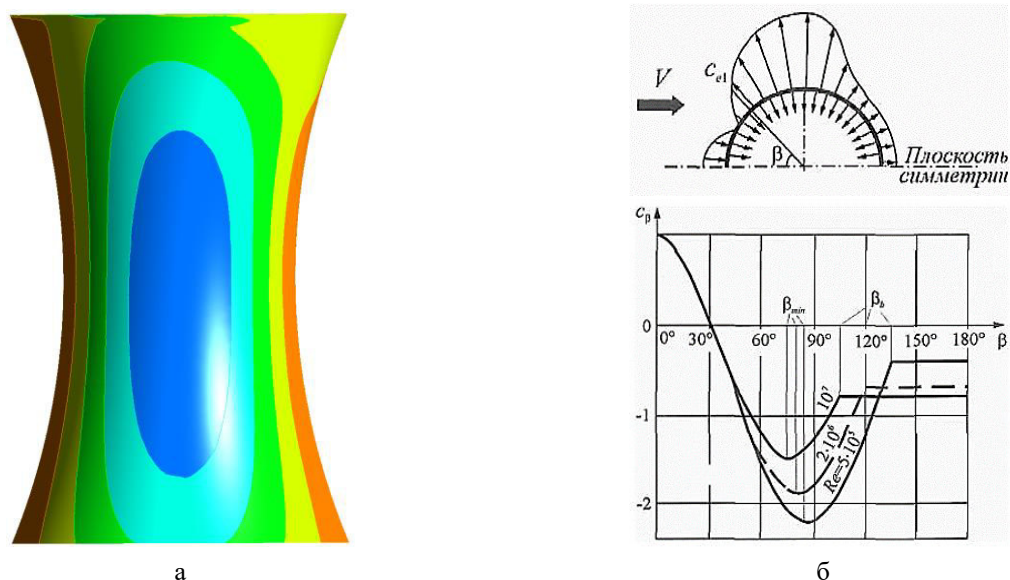


Рисунок 9. Картины распределения аэродинамических коэффициентов: *а* – компьютерное моделирование для глобоидной поверхности; *б* – фрагмент действующего нормативного документа (прил. В.1.12 СП 20.13330.2016)

С использованием представленных в статье цифровых моделей поверхностей с отрицательной гауссовой кривизной были разработаны алгоритмы автоматизированного определения аэродинамических характеристик в конкретно заданной точке на тороидных и глобоидных поверхностях (рис. 10, 11), которые могут быть внедрены в проектные процедуры на различных этапах жизненного цикла зданий и сооружений.

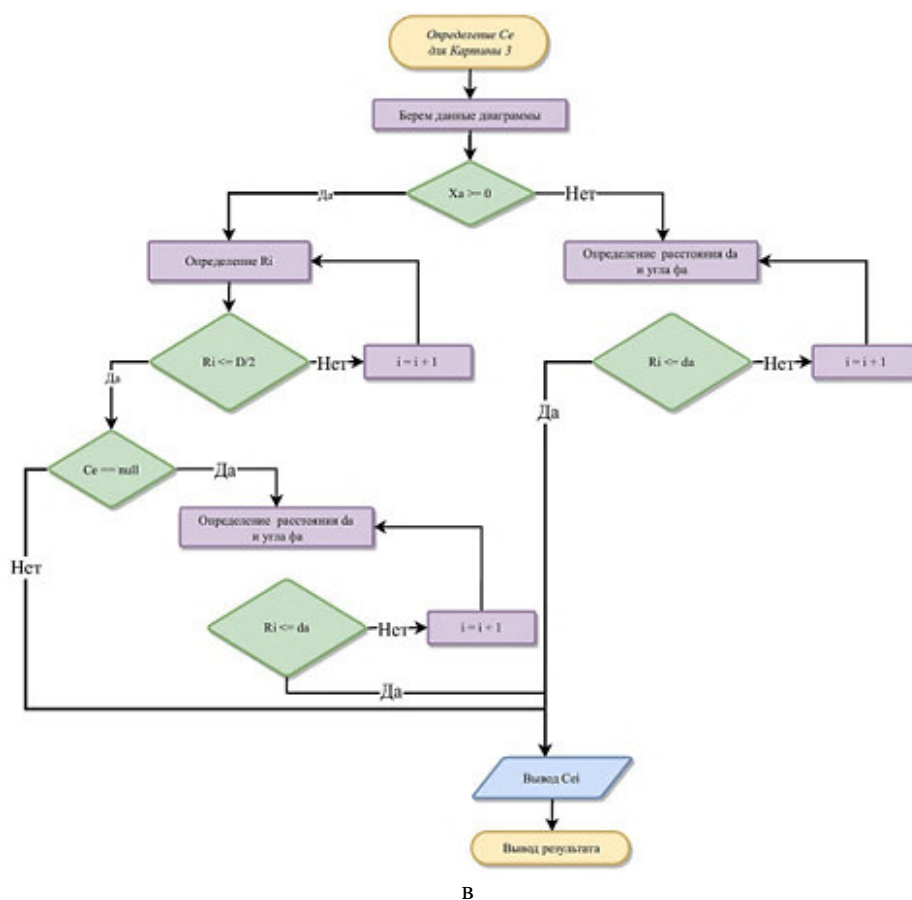
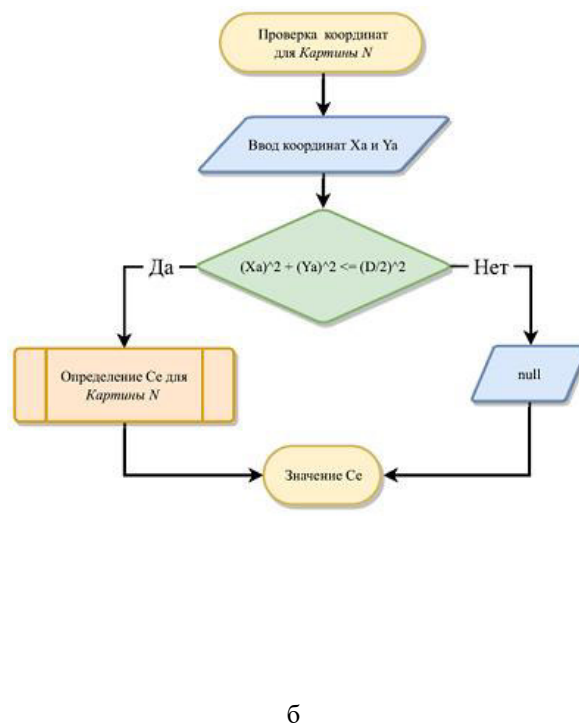
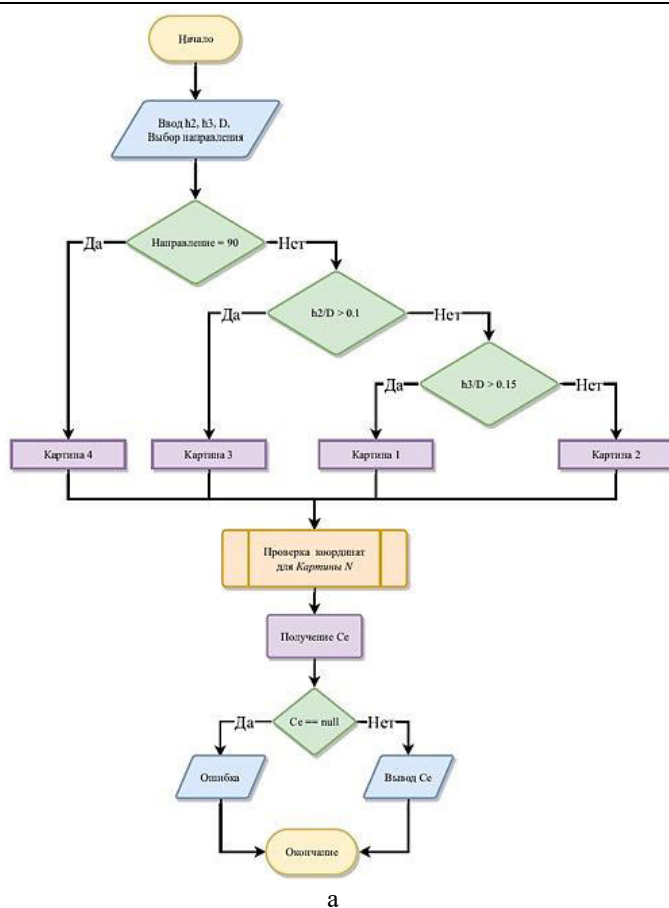


Рисунок 10. Алгоритм определения аэродинамических коэффициентов на тороидной поверхности:  
а – блок-схема основной программы; б, в – блок-схема подпрограмм



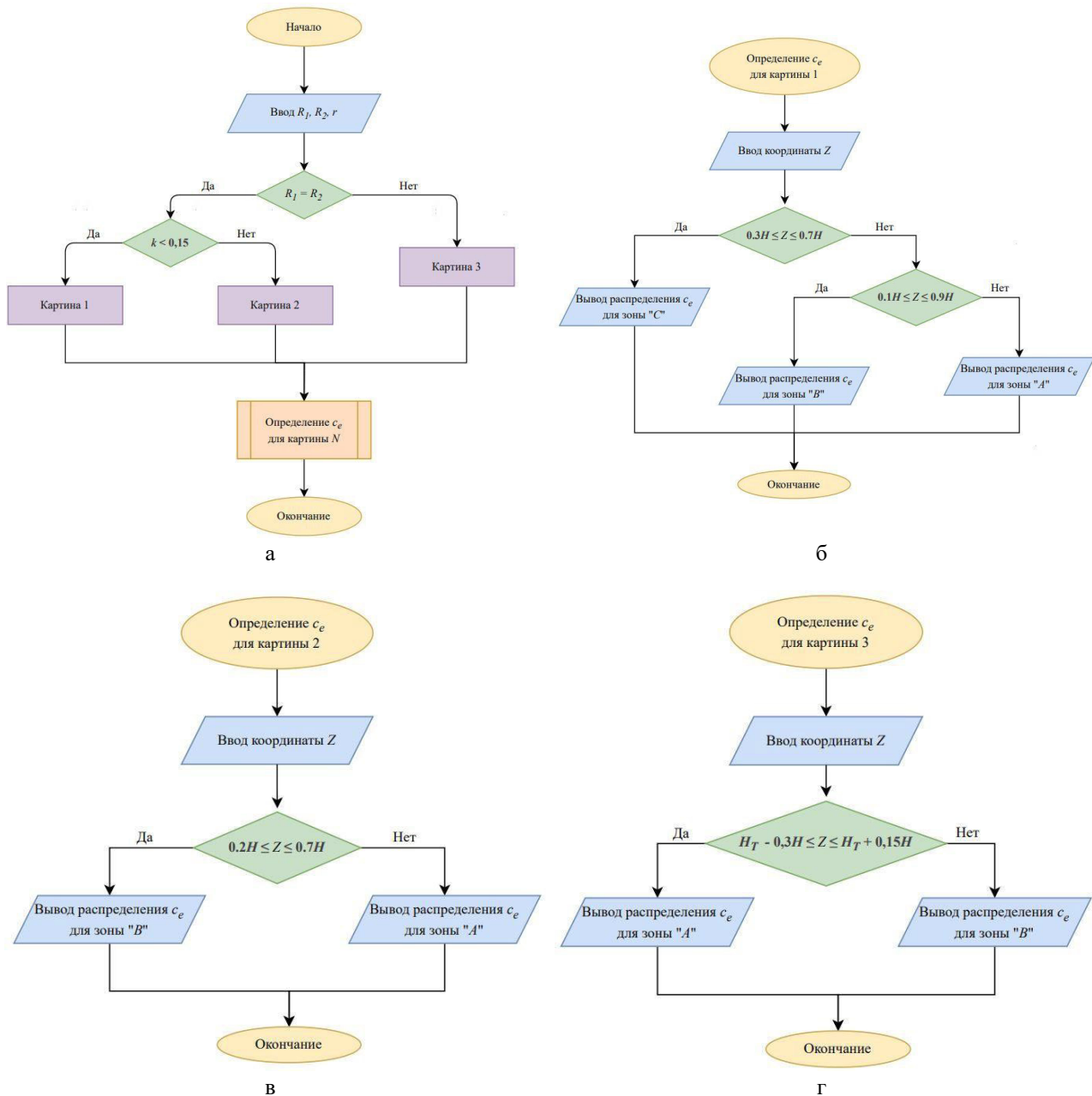


Рисунок 11. Алгоритм определения аэродинамических коэффициентов на тороидной поверхности:  
а – блок-схема основной программы; б-г – блок-схема подпрограмм

## Выводы

1. Решены задачи параметрического моделирования семейств геометрических поверхностей тороидального и глобоидного типов, пригодных для интеграции в современные системы автоматизированного проектирования (САПР) и стандарты проектирования различных технических объектов, таких как строительные конструкции, автомобили и самолеты.

2. На основании систематизации результатов многократно проведенных вычислительных экспериментов и геометрического анализа изополей распределения аэродинамических коэффициентов разработаны инженерные расчеты аэродинамических параметров поверхностей отрицательной гауссовой кривизны.

3. Разработаны алгоритмы автоматизированного расчета аэродинамических параметров поверхностей отрицательной гауссовой кривизны, ставшие основой для реализованного программного обеспечения «Система автоматизированного расчета аэродинамических коэффициентов на поверхность двоякой кривизны» (свидетельство о государственной регистрации № 2025666216 от 24.06.2025 г.).

**Научный руководитель** Хазов Павел Алексеевич, Нижегородский государственный архитектурно-строительный университет.

### Список литературы

1. Нартова Л. Г., Бодрышев В. В. Геометрическое моделирование технических поверхностей избранных классов // Вестник БГТУ. 2018. № 8 (69). С. 4–13.
2. Иванов В. Н. Формообразование поверхностей на заданных криволинейных планах // Строительная механика инженерных конструкций и сооружений. 2017. № 13. С. 3–8.
3. Моделирование кривых 2-го порядка и поверхностей оболочек инженерных сооружений на их основе / Е. В. Конопацкий, О. С. Воронова, С. И. Ротков, М. В. Лагунова, А. А. Бездичный // Строительство и техногенная безопасность. 2021. № 22 (74). С. 101–110.
4. Крысько А. А. Геометрическое и компьютерное моделирование криволинейных поверхностей мембранных покрытий на прямоугольном плане // Строительство и техногенная безопасность. 2020. № 18 (70). С. 98–106.
5. Поддаева О. И., Дуничкин И. В. Архитектурно-строительная аэродинамика // Вестник МГСУ. 2017. Т. 12, № 6. С. 602–609. DOI: 10.22227/1997-0935.2017.6.602-609
6. Sobolak M., Jagielowicz P. E. The methods of globoid surface modeling in CAD // Archives of Materials Science and Engineering. 2016. No. 81. Pp. 76–84.
7. Golebsky R., Boral P. Globoid surface shaped with turning and envelope method // MATEC Web of Conferences, 2019. No. 254. Pp. 2–12.
8. Темам Р. Уравнения Навье–Стокса. Теория и численный анализ. 2-е изд. М.: Мир, 1981. 408 с.
9. Шилов С. С., Федоров А. В., Молева А. А. Параметрическое моделирование как способ увеличения жизненного цикла зданий и сооружений // Приволжский научный журнал. 2024. № 3. С. 231–239.
10. Шилов С. С., Хазов П. А., Ведяйкина О. И. Методика цифровой поддержки определения аэродинамических показателей на поверхности двойкой кривизны // Вестник компьютерных и информационных технологий. 2025. № 7. С. 23–31.



Секция 10

# ***ЦИФРОВАЯ ТРАНСФОРМАЦИЯ ОБРАЗОВАНИЯ***

---

## **DIGITAL TRANSFORMATION OF EDUCATION**



**GraphiCon 2025**  
30 сентября – 2 октября 2025  
Йошкар-Ола, Россия



## Возможности систем искусственного интеллекта в решении тестовых заданий по начертательной геометрии и инженерной графике

О. А. Моисеева

Поволжский государственный технологический университет, Йошкар-Ола, Россия

**Аннотация.** В статье проведен комплексный анализ возможностей современных систем искусственного интеллекта (ИИ) при выполнении тестовых заданий в рамках онлайн-курса «Инженерная графика» с акцентом на выявление их сильных и слабых сторон в работе с текстовой и, что наиболее важно, графической информацией. Представлен обзор практик и инструментов ИИ, активно используемых студентами для решения учебных задач. Выделены наиболее популярные и удобные для данных целей нейросетевые платформы. Основу исследования составило экспериментальное тестирование возможностей трех ведущих мультимодальных нейросетей – Google Gemini, DeepSeek и GigaChat – на реальных тестовых материалах курса. Исследование выявило значительный разрыв в успешности выполнения нейросетями разных типов заданий. Показана относительно высокая эффективность ИИ при работе с текстовыми вопросами. Однако при решении графических заданий все протестированные модели продемонстрировали существенные ограничения.

**Ключевые слова:** искусственный интеллект, нейросети, инженерная графика, начертательная геометрия, онлайн-обучение, академическая честность, тестирование, Google Gemini, DeepSeek, GigaChat, мультимодальные модели, пространственное мышление.

## The capabilities of artificial intelligence systems in solving test tasks in descriptive geometry and engineering graphics

O. A. Moiseeva

Volga State University of Technology, Yoshkar-Ola, Russia

**Abstract.** A comprehensive analysis of the capabilities of modern artificial intelligence (AI) systems is carried out when performing test tasks within the online course «Engineering Graphics», with an emphasis on identifying their strengths and weaknesses in working with text and, most importantly, graphic information. The article provides an overview of AI practices and tools actively used by students to solve educational problems. The most popular and convenient neural network platforms for these purposes are highlighted. The study was based on experimental testing of the capabilities of three leading multimodal neural networks – Google Gemini, DeepSeek and GigaChat – on real test course materials. The study revealed a significant gap in the success of neural networks in completing different types of tasks. Relatively high efficiency of AI was demonstrated when working with text questions. However, when solving graphic tasks, all tested models demonstrated significant limitations.

**Keywords:** artificial intelligence, neural networks, engineering graphics, descriptive geometry, online learning, academic integrity, testing, Google Gemini, DeepSeek, GigaChat, multimodal models, spatial thinking.

### Введение

Основной идеей цифровизации образования является эффективное и адаптивное использование современных технологий с целью сделать образовательный процесс результативным. Обязательное внедрение электронной образовательной среды, создание и использование онлайн-курсов, разработка электронных ресурсов, заданий с автоматизированным контролем – это всё стало неотъемлемой частью образовательного процесса в современном вузе. Преподавателями кафедры начертательной геометрии и графики Поволжского государственного технологического университета при ресурсной поддержке Поволжского регионального центра компетенций в области онлайн-обучения были разработаны онлайн-курсы по дисциплинам кафедры. Курсы прошли экспертизу, размещены на портале открытого образования Волгатеха (<https://moored.net/>), находятся в открытом доступе.

Онлайн-курсы используются для организации самостоятельной работы студентов как очной, так и заочной форм обучения. Автоматизированный контроль уровня освоения теоретических материалов на онлайн-курсах осуществляется через решение тестовых заданий. А так как результаты самостоятельной работы учитываются в системе оценки знаний студентов по изучаемым дисциплинам, то вопрос о достоверности этих результатов всегда открыт. Желание списать, подсмотреть и сэкономить время – явление не новое, но с появлением и развитием искусственного интеллекта (ИИ) это сделать становится легче, а проконтролировать и уличить в академической нечестности – сложнее.

Вопросы востребованности и использования инструментов ИИ студентами активно обсуждаются исследователями [1–6]. Исследования показывают, что студенты высоко оценивают потенциал нейросетей для решения учебных и творческих задач. Количество обращений к помощи ИИ-инструментов постоянно увеличивается. Учитывая то, что тема ИИ остается популярной, а технологии продолжают развиваться, можно прогнозировать дальнейшее увеличение уровня вовлеченности студентов [5]. Преподаватели озабочены тем, что обучающиеся слишком полагаются на искусственный интеллект, излишне ему доверяют и не понимают, что при этом теряется смысл образования, а все труды преподавателей по разработке заданий оказываются бесполезными [1]. Запретить использование искусственного интеллекта не представляется возможным и является бессмысленным, поэтому педагогическим сообществом предлагается обучить студентов принципам использования, предоставить необходимую методологическую базу для работы с ИИ и воспитать критическое отношение к результатам генерации [1].

В данной работе представлены исследования по работе инструментов искусственного интеллекта с тестовыми заданиями онлайн-курса «Инженерная графика», размещенного на портале открытого образования Волгатека (<https://mooped.net/>). Онлайн-курс «Инженерная графика» состоит из трех разделов: 1) Начертательная геометрия; 2) Проекционное черчение; 3) Машиностроительное черчение. В каждом разделе представлены темы согласно рабочей программе. Каждая тема имеет блоки: теоретический, практический и контроль (тест). В конце каждого раздела приведен итоговый тест. В конце курса также запланировано итоговое тестирование по всему курсу. Какие из инструментов ИИ являются наиболее популярными среди студентов? Возможно ли преодолеть порог прохождения тестов онлайн-курса, используя только инструменты ИИ? Как усовершенствовать задания теста, чтобы исключить при их решении использование ИИ? Какие негативные последствия возникают при использовании ИИ для решения тестов? Ответы на данные вопросы позволят усовершенствовать онлайн-курсы по дисциплинам кафедры и эффективнее использовать их в дальнейшем.

### Постановка задачи

Целью работы является анализ возможностей систем искусственного интеллекта выполнения тестовых заданий онлайн-курса «Инженерная графика». Для достижения поставленной цели требуется решить следующие задачи:

- определить наиболее популярные и доступные системы искусственного интеллекта, выявить наиболее подходящие для решения заданий по начертательной геометрии и инженерной графике;
- выявить типы заданий тестов, наиболее сложные для решения системами ИИ;
- проанализировать ответы ИИ на тестовые задания разных типов;
- выявить уровень способности ИИ в прохождении итогового тестирования по курсу «Инженерная графика». При необходимости скорректировать тестовые задания онлайн-курса.

### Теория

#### 1. Системы ИИ для решения педагогических тестов

Впервые термин «искусственный интеллект» был использован в 50-е годы, но до сих пор общепризнанного определения ИИ не существует, так как происходит постоянное его развитие и совершенствование. Согласно Национальной стратегии развития ИИ на период до 2030 года в Российской Федерации, искусственный интеллект – это комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений [7]. Под термином «искусственный интеллект» понимают целый ряд технологий, методов и инструментов, которые являются взаимосвязанными и накладываются друг на друга.

Актуализация проблемы использования студентами возможностей ИИ в процессе обучения фиксируется исследователями началом 2023 года. За достаточно короткое время интерес к исследованию

влияния ИИ на высшее образование возрос, определилась тематика основных направлений и особенностей изучения применения инструментов ИИ в образовании. Авторами выделяются как позитивные стороны использования ИИ в образовании (безграничные возможности и перспективы ИИ в образовании), так и нарастающие проблемы и риски для организации обучения (оценка качества образования в ходе цифровой трансформации высшей школы) [8]. Ряд авторов [1, 5, 8] представляют результаты анализа целей использования инструментов ИИ. Среди целей использования инструментов ИИ наиболее популярными являются:

- 1) работа с текстом (подготовка курсовых и дипломных работ, эссе, докладов, сообщений);
- 2) генерация и обработка визуального контента (изображений и видео);
- 3) перевод с иностранных языков;
- 4) создание презентаций;
- 5) программирование (написание и оптимизация программного кода, исправление ошибок);
- 6) поиск и анализ информации (выполнение учебных заданий, решение тестовых заданий).

Написание студенческих работ, создание презентаций и решение учебных задач доверяют искусственному интеллекту 60 % обучающихся в возрасте от 18 до 24 лет [1]. Интересующие студентов запросы способны удовлетворить несколько десятков инструментов с искусственным интеллектом, с каждым годом их становится всё больше. В основном эти инструменты опираются на нейросети. Нейронные сети используются для решения подобных задач, так как способны на распознавание образов и естественного языка, другие методы ИИ могут оказаться менее эффективными.

Для решения тестовых заданий используются инструменты для работы с текстовой и знаковой информацией. Наиболее популярными оказались ChatGPT; Google Gemini; Deepseek; Giga.chat; Kampus.ai; Study24.ai. Разработчики уверяют, что эти инструменты помогают решать тесты, предоставляя развернутые объяснения и анализ результатов, тем самым способствуют повышению уровня знаний по изучаемой теме. Вместе с тем предупреждают, что нейросеть может выдавать неверные решения, способна снизить собственные навыки критического мышления и анализа, а также может нарушать правила учебного заведения. Для использования всех вышеперечисленных инструментов требуется четко формулировать запрос (промпт), т. е. необходимо описать вопрос подробно, включая контекст и формулировки для выдачи максимально точного результата. Для получения ответа нейросеть проходит три этапа: 1) обрабатывает текст (распознает ключевые слова и их взаимосвязь в вопросе); 2) анализирует контекст (алгоритмы определяют общий смысл вопроса); 3) генерирует ответ (предлагает наиболее подходящий вариант ответа). Большинство нейросетей способны распознавать вопросы по фото и анализировать изображения при высоком их качестве [9–11]. Нейросети работают с опорой на базы данных: чем больше данных имеется для обработки, тем точнее становятся её решения.

Для того чтобы понять, как справляются на практике популярные инструменты с педагогическими тестами, рассмотрим типы тестовых заданий онлайн-курса «Инженерная графика» и сгруппируем их по форме представления информации для ИИ.

Онлайн-курс создан на платформе Moodle, которая позволяет создавать различные типы вопросов. При создании тестов курса «Инженерная графика» использовались типы:

- множественный выбор: включает в себя формулировку задания и несколько вариантов ответа, из которых один или несколько могут быть верными (задаётся настройками задания); как в формулировке задания, так и в ответах можно использовать рисунки, форматированный текст и медиафайлы;
- верно-неверно: представляет собой утверждение; студенты должны отметить его истинность или ложность;
- на соответствие: студентам предлагается к каждой позиции одного списка подобрать соответствие из другого списка; форматированный текст, графику и мультимедиа; можно использовать только в одном из двух списков;
- краткий ответ: в качестве ответа студент должен вписать слово или несколько слов, которые сравниваются с эталонными; можно задавать несколько вариантов верного ответа и использовать подстановочные знаки, например, чтобы учесть возможное склонение слова студентами;

- выбор пропущенных слов: в тексте задания есть пропущенные слова, которые заполняются с помощью выпадающих меню;

- перетаски на изображение: вопрос задается текстом, а ответ формируется студентом перетаскиванием изображения в определенную зону на фоновом изображении.

Анализ банка вопросов онлайн-курса показал, что все типы вопросов, кроме «перетаски на изображение», формулируются либо только текстом, либо текстом с изображением. При подробном анализе вопросов, содержащих текст и изображения, также выделились две группы:

1) вопрос заключен в тексте, а изображение лишь подкрепляет текст, не несет дополнительную информацию;

2) вопрос частично сформулирован в тексте, но уточняющая информация отражается на изображении.

В результате анализа банка вопросов онлайн-курса «Инженерная графика», для удобства проведения исследования, все многообразие вопросов разделено по способу представления информации на две группы: текстовые и графические задания. Под текстовыми понимаются задания, в которых и вопрос, и ответ выражаются только через текст. В графических заданиях суть вопроса и/или ответа заключена вся или частично на изображении. Было выявлено соотношение текстовых и графических заданий в банке вопросов, оно отражено на рисунке 1. В банке вопросов наблюдается преобладание графических заданий (71,5 %), текстовые задания составляют немногим больше четверти всех вопросов банка (28,5 %).

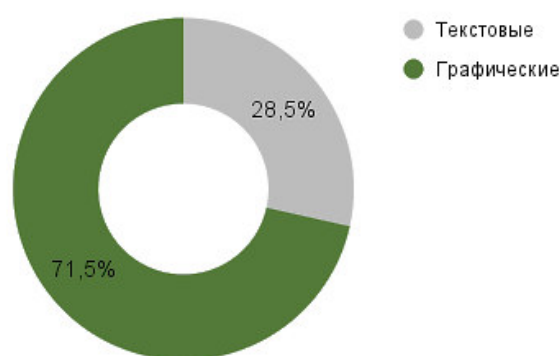


Рисунок 1. Соотношение текстовых и графических заданий теста

Сопоставляя возможности вышеперечисленных нами инструментов ИИ с выделенными группами тестовых заданий, можно сделать вывод: на все вопросы без уточняющего промта способны сгенерировать ответ (пока не рассматриваем правильность и достоверность ответа) шесть инструментов ИИ из представленных семи. ChatGPT способен автоматизировано выполнять задание только по текстовому описанию. Для его использования для решения графических заданий требуется подробное словесное описание изображений, чего студенту, погруженному в тему поверхностно, сделать будет трудно. Для оставшихся шести инструментов восприятие визуального контента тоже является сложной задачей, так как требуется не простое описание изображений, а интерпретация схем, проекций, сечений – того, что требует пространственного мышления. Также заметим, что только три нейросети из представленных являются бесплатными: Google Gemini; Deepseek; Giga.chat. Остальные имеют лишь 2–3 бесплатных запроса, что уменьшает их популярность среди студентов.

2. Процедура использования инструментов ИИ при выполнении текстовых и графических тестовых заданий онлайн-курса.

Для изучения возможностей инструментов ИИ при решении тестовых заданий онлайн-курса рассмотрим нейросети Google Gemini; DeepSeek; GigaChat, так как они способны анализировать текстовую и графическую информацию, являются бесплатными, простыми в использовании.

Google Gemini – это модель искусственного интеллекта, разработанная компанией Google DeepMind. Она доступна в трех версиях: 1) Gemini Nano встроена в смартфоны Pixel; 2) Gemini Ultra доступна в веб-версии на сайте Gemini Advanced и в одноименных приложениях на Android и iOS. (использование

платное – \$19,99 в месяц. В России Gemini Advanced недоступен без VPN); 3) Gemini Pro – бесплатное использование в чат-боте Gemini, требуется авторизация в аккаунте Google [12]. Для решения тестовых заданий использовался третий вариант, т. е. для генерации ответа на вопрос теста использовался чат-бот Gemini, в который встроен Gemini Pro. Для этого в браузере Chrome с авторизованным аккаунтом Google на смартфоне при использовании Google объектива производилась съемка тестового задания, в диалоговом окне отражали краткий запрос, получали ответ, подкрепляемый рассуждениями.

DeepSeek – это большая языковая модель с открытым исходным кодом, разработанная одноименной китайской компанией. На данный момент пользователи в России могут получить доступ к версии DeepSeek V3 через сайт. Для этого нужно перейти на сайт DeepSeek, нажать Start Now; пройти регистрацию (можно с помощью аккаунта Google); откроется диалоговое окно с чат-ботом. В нем, помимо обычного ответа, можно выбрать опцию DeepThink для рассуждений или Search для поиска данных в Интернете. Также для анализа можно прикрепить до 50 файлов разного формата размером до 100 МБ каждый [13]. Для решения тестовых заданий использовался чат-бот сайта DeepSeek на смартфоне. В диалоговом окне имеется возможность сфотографировать задание теста и написать запрос. Как и в случае с Gemini, DeepSeek выдавал ответ, подкрепленный рассуждениями.

GigaChat – это нейросеть от Сбера, созданная специально для русскоязычных пользователей. Чат-бот построен на нейросетевом ансамбле модели NeONKA (Neural Omnimodal Network with Knowledge-Awareness), включающем в себя несколько нейросетей для разных задач. Разработкой занимались команды SberDevices и Sber AI при поддержке Института искусственного интеллекта AIRI. Для получения доступа необходимо зайти на сайт <https://giga.chat/gigachat/>, авторизоваться через существующий Сбер ID или войти по номеру телефона [14]. Без регистрации доступна только «облегченная» версия чат-бота, но она обработает лишь текстовые запросы, а вот создавать картинки, обрабатывать файлы и распознавать изображения можно только в полной версии. Поэтому для тестирования использовалась полная версия с регистрацией. Как и в предыдущих случаях, со смартфона загружалось фото задания и запрос, чат-бот генерировал ответ с рассуждениями.

Все три нейросети распознавали текст по фото, распознавали изображения на фото, рассуждали и делали выводы, выдавая ответы. Правильность ответов и достоверность рассуждений рассмотрим далее.

### Результаты экспериментов

Для решения поставленных задач в проведении эксперимента выделим два основных направления:

- 1) оценивание возможностей и эффективности современных нейросетевых моделей (Google Gemini, DeepSeek, GigaChat) при решении графических тестовых заданий онлайн-курса «Инженерная графика»;
- 2) выявление способности нейросети пройти пороговый уровень итогового тестирования по курсу «Инженерная графика».

*1. Оценивание возможностей и эффективности современных нейросетевых моделей (Google Gemini, DeepSeek, GigaChat) при решении графических тестовых заданий онлайн-курса «Инженерная графика»*

При проведении эксперимента использовались одни и те же устройства при равных условиях. Тестирование проводилось последовательно на актуальных версиях моделей нейросетей Google Gemini, DeepSeek, GigaChat. Нейросетям последовательно предъявлялись фотографии реальных тестовых заданий онлайн-курса. Каждая фотография содержит:

- текстовую формулировку вопроса/задачи;
- варианты ответов (для выбора или ввода числового/текстового ответа);
- важное графическое изображение (чертеж, схема, эскиз, аксонометрия, комплексный чертеж детали и т. д.), содержащее существенную информацию, необходимую для решения.

Текст запроса к нейросети и количество уточняющих запросов зависит от опыта. Для примера рассмотрим несколько опытов с заданиями из разных разделов онлайн-курса.

#### Опыт 1. С одним текстовым запросом

Фото вопроса теста представлено на рисунке 2. Текст запроса к нейросети: «Реши задание множественного выбора по фото, возможно несколько правильных ответов».

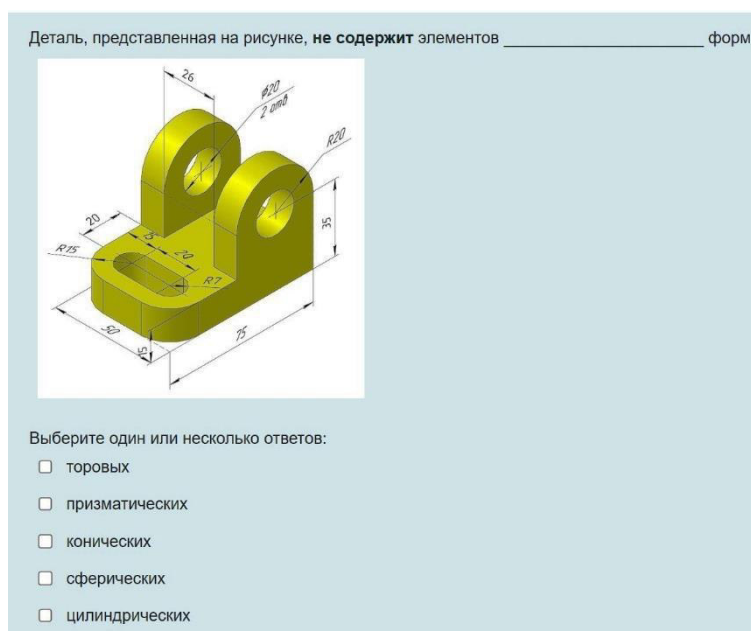


Рисунок 2. Фото тестового задания для опыта 1

Первый опыт показывает (табл. 1), что две нейросети из трех выдают правильный ответ на графическое задание, способны проанализировать форму детали по наглядному изображению. Третья ошибается на этапе интерпретации текстовой части задания, при анализе геометрической формы детали оперирует не геометрическими понятиями, ответ генерирует неверный.

**Таблица 1. Результаты решения тестового задания (опыт 1)**

Ключевые вопросы	Нейросетевые модели		
	Google Gemini	DeepSeek	GigaChat
Способность нейросети корректно интерпретировать комплексную информацию, представленную одновременно в текстовой и графической форме на изображении теста	Способна	Способна	Способна, но с ошибкой (в текстовой части вопроса проигнорировала частицу «не»)
Точность в рассуждениях ИИ при анализе геометрической формы детали	Всё верно	Всё верно	В рассуждениях используются некорректные термины
Выбор правильного ответа из предложенных вариантов	Даёт правильный ответ	Даёт правильный ответ, но с припиской «Для точного ответа требуется визуальный анализ изображения»	Даёт <b>неправильный</b> ответ

## Опыт 2. С двумя текстовыми запросами

Фото вопроса теста представлено на рисунке 3. Первый запрос: «Реши задание множественного выбора по фото, возможно несколько правильных ответов».

Второй уточняющий запрос: «Точки  $L$ ,  $M$ ,  $K$ ,  $D$  принадлежат параллелям конуса. Какие из точек являются видимыми на фронтальной проекции чертежа?»

Во втором примере предлагалось решить задание из раздела «Начертательная геометрия». Для всех трех нейросетей задание оказалось нерешаемым даже при наличии уточняющего запроса (табл. 2). При этом все три при интерпретации графической части задания определили, что дан чертеж, но не смогли правильно определить фигуру, данную на чертеже. Все три также использовали при рассуждениях верные формулировки правил начертательной геометрии при определении видимости на чертеже.



Рисунок 3. Фото тестового задания для опыта 2

Таблица 2. Результаты решения тестового задания (опыт 2)

Ключевые вопросы	Нейросетевые модели		
	Google Gemini	DeepSeek	GigaChat
Способность нейросети корректно интерпретировать комплексную информацию, представленную одновременно в текстовой и графической форме на изображении теста	Способна	Способна	Способна
Точность в рассуждениях ИИ при анализе геометрической формы по чертежу	Неверно	Неверно	Верно (определено, что на чертеже даны проекции конуса)
Использование в рассуждениях ИИ правил начертательной геометрии	Использует	Использует	Использует
Выбор правильного ответа из предложенных вариантов	Даёт неправильный ответ	Даёт неправильный ответ	Даёт неправильный ответ

**Опыт 3. С количеством уточняющих запросов, необходимых для достижения правильного ответа**

Фото вопроса теста представлено на рисунке 4. Первый запрос стандартный: «Реши задание множественного выбора по фото», формулировки последующих направляли нейросети к правильному ответу.

В третьем примере демонстрируется решение задания с одним правильным ответом. Наиболее эффективной в этом примере оказалась нейросеть DeepSeek, которая с первого запроса сгенерировала правильный ответ и объяснила его. Наибольшее количество уточняющих запросов (всего 5) для генерации правильного ответа понадобилось GigaChat (табл. 3).

Всего было подробно проанализировано 30 заданий, и на основе полученных данных заключили, что наиболее эффективно себя показала нейросеть Google Gemini, далее с небольшим отрывом от нее идет DeepSeek, замыкает рейтинг GigaChat с существенным отставанием.





Рисунок 4. Фото тестового задания для опыта 3

Таблица 3. Результаты решения тестового задания (опыт 3)

Ключевые вопросы	Нейросетевые модели		
	Google Gemini	DeepSeek	GigaChat
Количество запросов для получения верного ответа на задание теста	2	1	5

## 2. Результаты экспериментального исследования по выявлению способности нейросети пройти пороговый уровень итогового тестирования по курсу «Инженерная графика»

При проведении второй части эксперимента использованы реальные задания из итогового теста курса, в тесте 40 вопросов. Всего было проведено 5 сессий тестирования. Для прохождения порогового уровня итогового теста необходимо набрать 50 % правильных ответов. Тестирование проходила нейросеть Google Gemini, как показавшая наибольшую эффективность на предыдущем этапе эксперимента. Структура заданий итогового теста соответствует соотношению графических и текстовых заданий банка вопросов курса (таким образом настроены категории вопросов банка): 28,5 % заданий – текстовые вопросы (теория, терминология), 71,5 % заданий – задачи, требующие интерпретации визуальной информации (схемы, чертежи, эпюры, аксонометрические изображения, разрезы, сечения).

Каждое задание передавалось нейросети в виде фотографии, содержащей полный контекст: текст вопроса, варианты ответов (где применимо) и ключевое изображение. К каждой фотографии прилагался исключительно краткий запрос: «Решите задание». Никаких дополнительных пояснений, контекста курса или подсказок не предоставлялось.

Диаграмма на рисунке 5 показывает, что Google Gemini набрала 43 % правильных ответов, чего недостаточно для прохождения порогового уровня итогового теста.

Проведем детальный анализ полученных результатов.

При выполнении текстовых заданий (всего в тесте 28,5 %) продемонстрирована относительно высокая успешность, 87,7 % правильных решений (см. рис. 6а). Нейросеть корректно отвечала на вопросы, требующие знания определений, стандартов ЕСКД и теоретических основ. Типичные ошибки: неточности в формулировках или подмена терминов близкими по смыслу, но не строго корректными в контексте инженерной графики.

При выполнении графических заданий (всего в тесте 71,5 %) выявлены существенные трудности. Рисунок 6б демонстрирует диаграмму результатов, 25,2 % правильных ответов. Типичные ошибки: интерпретация чертежей (ошибки в чтении комплексных чертежей, сборочных чертежей, определении

формы детали по проекциям, нахождении линий пересечения поверхностей); распознавание элементов (проблемы с точным распознаванием и интерпретацией условных обозначений, условных изображений резьбы, штриховок, линий различного типа (осевые, выносные, размерные); пространственное мышление (ограниченная способность корректно преобразовывать информацию чертежа и соотнести ее с наглядным изображением (и наоборот) для решения задачи; работа с размерностями (частые ошибки в интерпретации или генерации числовых ответов, требующих точных расчетов на основе размеров с чертежа).

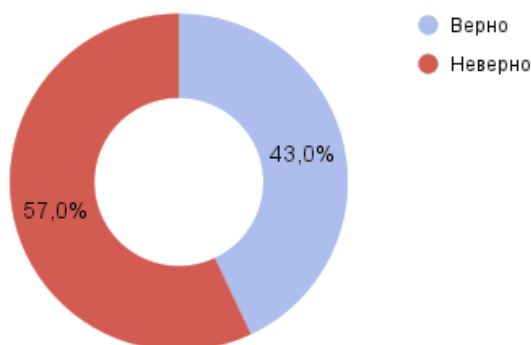


Рисунок 5. Результат решения заданий итогового теста нейросетью Google Gemini

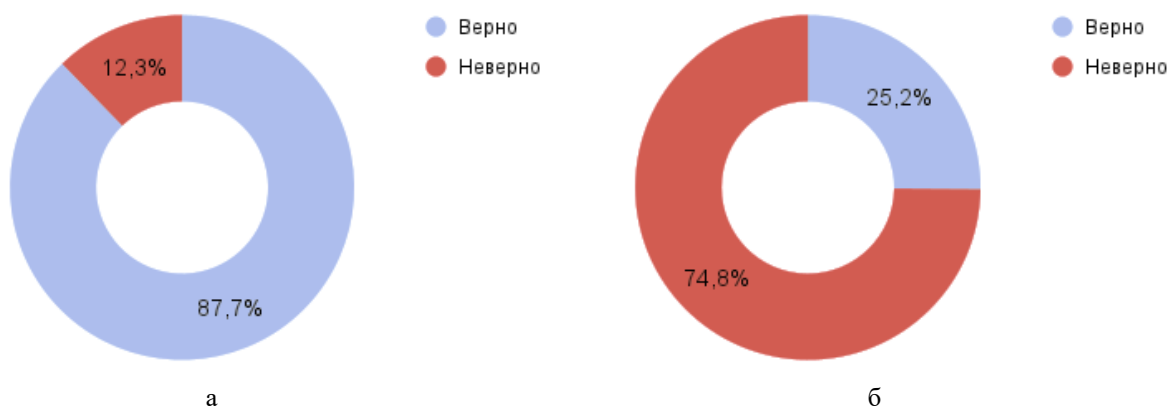


Рисунок 6. Результат решения заданий итогового теста нейросетью Google Gemini:  
а – результат решения текстовых заданий; б – результат решения графических заданий

Встречались задания с низким качеством изображений, что затрудняло корректную интерпретацию. В то же время встречались решения, в которых наблюдались ложные рассуждения, но нейросеть давала (угадывала) правильный ответ. Процент первых и вторых явлений незначительный.

*Общий вывод по возможностям Gemini:* на текущем уровне развития (Gemini 1.5 Pro) нейросеть не обладает достаточными возможностями для самостоятельного и корректного решения большинства практических заданий по инженерной графике, критически зависящих от анализа чертежей и изображений. Ее сильные стороны лежат в области обработки текста и базовой теории, но не в глубоком понимании инженерно-графической информации.

### Обсуждение результатов

Представленные выше результаты исследования говорят о значительной устойчивости тестовых заданий онлайн-курса к инструментам ИИ. Это позволит использовать имеющиеся задания в дальнейшем и обновлять их в соответствии с планом развития онлайн-курса. При разработке новых тестовых заданий следует опираться на принцип: создавай задания, требующие глубокого понимания контекста, пространственных преобразований и связей между элементами чертежа, а не просто распознавания шаблонов.

Другая важная сторона проблемы использования инструментов ИИ при решении заданий по начертательной геометрии и инженерной графике – это обучение студентов принципам использования

ИИ и воспитание критического отношения к результатам генерации. Нейросеть может дать развернутое, научнообразное объяснение своего неверного решения. Она использует правильные термины (проекция, сечение, аксонометрия), но связывает их в псевдологичную цепочку, ведущую к ошибке. Студен рискует поверить этому красивому объяснению и усвоить ложное знание. Это создаст фундаментальные пробелы и искажения в понимании, которые проявятся позже и будут мешать изучать сопромат, детали машин, ТММ и т. д.

### Заключение

Результаты проведенных эксперимента позволяют сделать следующие выводы:

1. *О текущем уровне ИИ.* Результаты подтверждают, что, несмотря на впечатляющие возможности мультимодальных моделей, точная интерпретация сложных инженерных чертежей и решение на их основе пространственных задач остаются вызовом для современных ИИ.

2. *О защите академической честности.* Исследование свидетельствует, что итоговое тестирование текущего формата (с преобладанием графических заданий) обладает высокой устойчивостью к попыткам прохождения только с помощью краткого запроса к нейросети без дополнительных пояснений. Требуемое пространственное мышление и специфические навыки чтения чертежей являются действенным барьером.

3. *Об адаптации обучения:* результаты подчеркивают непреходящую важность развития у студентов навыков пространственного мышления и работы с графической информацией. Формирование этих компетенций сложно заменить или автоматизировать на текущем уровне развития ИИ.

4. *О перспективах ИИ в образовании.* Исследование указывает на потенциальные направления для разработчиков ИИ: улучшение моделей компьютерного зрения для инженерной графики, обучение на специализированных наборах чертежей, развитие способности к пространственным рассуждениям.

Дальнейшие исследования в направлении использования инструментов ИИ видится не только в повышении устойчивости тестов к ним, но и использование нейросетей в образовательной деятельности.

### Список литературы

1. Отношение студентов и преподавателей к использованию инструментов с искусственным интеллектом в вузе / К. И. Буякова, Я. А. Дмитриев, А. С. Иванова, А. В. Фещенко, К. И. Яковлева // Образование и наука. 2024. Т. 26, № 7. С. 160–193. DOI:10.17853/1994-5639-2024-7-160-193

2. Костикова Л. П., Есенина Н. Е., Ольков А. С. Искусственный интеллект в образовательном процессе современного университета: результаты опроса студентов // Концепт. 2025. № 2. URL: <https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-obrazovatelnom-protsesse-sovremennogo-universiteta-rezultaty-oprosa-studentov> (дата обращения: 28.06.2025).

3. Искусственный интеллект для учебной аналитики и этапы педагогического проектирования: обзор решений / Е. А. Другова, И. И. Журавлева, У. С. Захарова, В. Е. Сотникова, К. И. Яковлева // Вопросы образования. 2022. № 4. URL: <https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-dlya-uchebnoy-analitiki-i-etapy-pedagogicheskogo-proektirovaniya-obzor-resheniy> (дата обращения: 25.06.2025).

4. Субботина М. В. Искусственный интеллект и высшее образование – враги или союзники // Вестник РУДН. Серия: Социология. 2024. Т. 24, № 1. С. 117–183. DOI: 10.22363/2313-2272-2024-24-1-176-183

5. Терехова Е. С., Пучкова Н. Н., Новикова Л. В. Анализ востребованности использования нейросетей для решения учебных задач // Концепт. 2024. № 8. URL: <https://cyberleninka.ru/article/n/analiz-vostrebovannosti-ispolzovaniya-neyrosetey-dlya-resheniya-uchebnyh-zadach> (дата обращения: 28.06.2025).

6. Бермус А. Г., Сизова Е. В. Этические аспекты внедрения технологий искусственного интеллекта в классическом университете: анализ отношения студенческой аудитории // Непрерывное образование: XXI век. 2025. Т. 13, № 2. С. 1–16. DOI: 10.15393/j5.art. 2025.10585

7. Национальная стратегия развития искусственного интеллекта на период до 2030 года. URL: [https://base.garant.ru/72838946/#block\\_1000](https://base.garant.ru/72838946/#block_1000) (дата обращения: 28.06.2025).

8. Российские студенты о возможностях и ограничениях использования искусственного интеллекта в обучении / И. А. Алешковский, А. Т. Гаспарисвили, Н. П. Нарбут, О. В. Крухмалева, Н. Е. Савина // Вестник РУДН. Серия: Социология. 2024. № 2. URL: <https://cyberleninka.ru/article/n/rossiyskie-studenty-o-vozmozhnostyah-i-ogranicheniyah-ispolzovaniya-iskusstvennogo-intellekta-v-obuchenii> (дата обращения: 28.06.2025).

9. GigaChat: сайт. 2023. URL: <https://giga.chat> (дата обращения: 01.06.2025).

- 
10. Лучшие нейросети для решения тестов по фото: удобные ИИ для помощи в онлайн учебе. URL: <https://dtf.ru/luchshii-rating/3206008-luchshie-neiroseti-dlya-resheniya-testov-po-foto-udobnye-ii-dlya-pomoshi-v-onlain-uchebe> (дата обращения: 18.06.2025).
11. ChatGPT и искусственный интеллект в высшем образовании. Краткое руководство. URL: [https://unesdoc.unesco.org/ark:/48223/pf0000385146\\_rus](https://unesdoc.unesco.org/ark:/48223/pf0000385146_rus) (дата обращения: 18.06.2025).
12. ИИ-модель Gemini: что это и как ей пользоваться. URL: <https://trends.rbc.ru/trends/industry/65732a7c9a79473c1aaacb35?from=copy> (дата обращения: 18.06.2025).
13. Что такое нейросеть DeepSeek: чат-бот с опцией поиска в интернете. URL: <https://trends.rbc.ru/trends/industry/67920b739a794768fab8fe92?from=copy> (дата обращения: 18.06.2025).
14. Обзор нейросети GigaChat от Сбера – что это такое, как работает и как ею пользоваться. URL: <https://www.banki.ru/news/daytheme/?id=11015390> (дата обращения: 18.06.2025).

УДК 378:514

DOI: 10.25686/978-5-8158-2474-4-2025-1052-1062

## Организация исследовательской и проектной деятельности при изучении графических дисциплин в условиях цифровизации высшего образования

С. О. Турецких

Поволжский государственный технологический университет, Йошкар-Ола, Россия

**Аннотация.** Статья посвящена актуальной проблеме организации исследовательской и проектной деятельности обучающихся при изучении графических дисциплин в условиях цифровизации высшего образования. В данной работе представлен системный анализ теоретических подходов к пониманию сущности исследовательской и проектной деятельности в психолого-педагогической науке. Детально описывается опыт Поволжского государственного технологического университета по внедрению этих видов деятельности при изучении дисциплины «Начертательная геометрия и инженерная графика». Особое внимание уделено разработке системы вовлечения студентов через олимпиады и конкурсы, использованию цифровых инструментов, интеграции с профессиональной средой через прикладные проекты. На основе проведенного опроса обучающихся и метода включенного наблюдения удалось зафиксировать основные признаки познавательной деятельности. Разработан комплекс практических рекомендаций, включающих совершенствование цифровой образовательной среды, оптимизацию системы наставничества и развитие механизмов поощрения студентов.

**Ключевые слова:** цифровизация образования, высшее профессиональное образование, графические дисциплины, исследовательская деятельность обучающихся, проектная деятельность обучающихся.

## Organization of research and project activities in the study of graphic disciplines in the context of digitalization of higher education

S. O. Turetskikh

Volga State University of Technology, Yoshkar-Ola, Russia

**Abstract.** The article is devoted to the urgent problem of organizing the research and project activities of students in the study of graphic disciplines in the context of digitalization of higher education. This paper presents a systematic analysis of theoretical approaches to understanding the essence of research and project activities in psychological and pedagogical science. The experience of the Volga State Technological University in introducing these types of activities in the study of the discipline "Descriptive Geometry and Engineering Graphics" is described in detail. Special attention is paid to the development of a student engagement system through Olympiads and competitions, the use of digital tools, and integration with the professional environment through applied projects. Based on the conducted survey of students and the method of included observation, key factors of students' motivation for research and project activities have been identified. A set of practical recommendations has been developed, including: improving the digital educational environment, optimizing the mentoring system and developing student incentive mechanisms.

**Keywords:** digitalization of education, higher professional education, graphic disciplines, research activities of students, project activities of students.

### Введение

Сегодня адаптация системы профессионального образования к запросам цифровой экономики и цифрового общества на подготовку компетентных кадров, обладающих высоким уровнем цифровой грамотности, способных реализовать свой профессиональный потенциал на основе широкого использования современных цифровых инструментов и технологий, как отмечают в своем исследовании И. Д. Столбова, Л. В. Кочурова и К. Г. Носов, является насущной необходимостью. Востребованное качество обучения будущих специалистов достигается посредством построения развивающей и технологичной образовательной среды, удовлетворяющей запросам современной экономики и высокотехнологичного производства. Однако, как справедливо подмечают авторы, «ожидаемый эффект от решения стратегических задач по улучшению качества цифрового образования невозможно получить лишь за счет перевода образовательных ресурсов и других обязательных составляющих учебно-воспитательного процесса из аналогового вида в цифровой. На новом этапе развития цифровых технологий помимо дополнения задачами, связанными с совершенствованием нынешних систем, поддерживающих образование, возникает стратегическая образовательная задача с коренной реорганизацией и трансформацией процесса обучения, а также определения концептуальных

основ для создания системы эффективного цифрового профессионального образования как фундамента подготовки кадров для инновационной экономики страны» [1, с. 53].

Подготовка квалифицированных специалистов, способных эффективно решать задачи развития не только профессиональной деятельности, но и науки, возможна лишь при тесной взаимосвязи учебного процесса с научно-исследовательской деятельностью. Нацеливание процесса кадровой подготовки в высшей школе на формирование профессиональной компетентности в системе ценностных ориентаций предполагает решительно новые подходы к научно-исследовательской деятельности студентов в вузе, в которой научно-исследовательская деятельность становится инструментом профессиональной коммуникации, повышения мотивация к изучению будущей специальности [2].

Для устранения разрыва между практикой высшего технического образования и реальной инженерной деятельностью необходимо внедрять инновационные образовательные технологии, направленные на формирование у обучающихся способности работать в команде, компетенций в области современных информационных и коммуникативных технологий, а также готовности осуществлять проектирование на основе пространственного моделирования [3].

3D-моделирование позволяет обучающимся приобретать важные навыки для укрепления своих теоретических знаний и осознания своей будущей профессии. Однако, по мнению Т. В. Машаровой, Н. А. Бушмелевой, М. С. Перевозчиковой, И. Ю. Хлобыстовой, овладение этими инструментами требует от самих обучающихся «значительных мыслительных усилий, применения воображения, инновационного мышления, способности к самостоятельной созидательной деятельности» [4, с. 429].

Выпускникам технических вузов по роду своей профессиональной деятельности необходимо понимать и самостоятельно уметь работать с чертежной документацией, создавать 3D-модели машин, механизмов, зданий и сооружений, при этом используя цифровые возможности [5]. «Начертательная геометрия и инженерная графика» как одна из первых учебных дисциплин графического цикла в вузе отвечает за формирование указанных умений и навыков.

В связи с дефицитом времени, отводимого в учебных планах на общеобразовательную предметную подготовку при принятой уровневой системе высшего образования, как отмечают И. Д. Столбова, Е. П. Александрова, Л. В. Кочурова и К. Г. Носов, становятся особо актуальными вопросы предметного обучения и встраивания его в процесс формирования требуемой компетентностной модели будущего инженера [3], а привлечение обучающихся к исследовательской и проектной деятельности, сопряженной с учебным процессом, начиная с первого года обучения, «является одним из важнейших средств повышения качества подготовки и воспитания специалистов с высшим профессиональным образованием» [6, с. 518].

В связи с вышесказанным изучение опыта и описание методики организации исследовательской и проектной деятельности при обучении графическим дисциплинам в условиях цифровизации высшего образования актуально в наши дни.

### **Постановка задачи**

В настоящей работе автор ставит перед собой цель поделиться опытом организации исследовательской и проектной деятельности обучающихся по инженерным направлениям подготовки в ФГБОУ ВО «Поволжский государственный технологический университет» при изучении дисциплины «Начертательная геометрия и инженерная графика» в условиях цифровизации образования.

Для достижения поставленной цели были выдвинуты следующие задачи:

- дать определение исследовательской и проектной деятельности обучающихся;
- охарактеризовав особенности научно-исследовательского и проектного методов, обосновать необходимость их совместного использования при реализации самостоятельной работы обучающихся при изучении графических дисциплин;
- представить методику организации исследовательской и проектной деятельности обучающихся при изучении дисциплины «Начертательная геометрия и инженерная графика» студентами 1 курса технического вуза;
- провести опрос среди обучающихся 1 курса, активно принимающих участие в исследовательской и проектной деятельности при изучении дисциплины «Начертательная геометрия и инженерная графика»;

– разработать рекомендации по повышению мотивации обучающихся к исследовательской и проектной деятельности в условиях цифровизации высшего образования на примере изучения графических дисциплин.

Научная новизна исследования заключается в разработке целостной системы организации исследовательской и проектной деятельности обучающихся при изучении дисциплины «Начертательная геометрия и инженерная графика» в техническом вузе, включающий комплексный подход, сочетающий традиционные подходы с цифровыми технологиями.

Практическая значимость подтверждается: внедрением методики в образовательный процесс Поволжского государственного технологического университета, положительной динамикой участия студентов в научно-исследовательских мероприятиях, разработанными рекомендациями по повышению мотивации обучающихся к исследовательской и проектной деятельности.

### Теория

Множество работ отечественных и зарубежных исследователей посвящено различным аспектам исследовательской и проектной деятельности обучающихся в вузе.

В работе Н. В. Бирюковой рассмотрено различие между учебно-исследовательской и научно-исследовательской работой обучающихся. По мнению автора, «данные виды деятельности являются неотъемлемой составляющей современного профессионального образования, их главная цель состоит в формировании исследовательских компетенций обучающихся. При этом учебно-исследовательская деятельность направлена преимущественно на получение образовательного результата, научно-исследовательская деятельность – на получение новых объективных научных знаний» [7, с. 994]. Следует отметить, что «именно поиск новых знаний позволяет студенту сформировать такие компетенции, как самостоятельно обрабатывать информацию, логически последовательно выстраивать свои суждения и аргументированно отстаивать свою позицию, расширять свой опыт и углубить свои знания, использовать научный стиль в своих работах» [8, с. 225].

В теории контекстного обучения учебно-исследовательская и научно-исследовательская работы обучающихся являются формами учебно-профессиональной деятельности. Так, «принимая участие в научных исследованиях, работая над курсовыми, дипломными работами, студенты остаются в позиции обучающихся и в то же время по целям, содержанию, формам, процессу и требованиям к полученным результатам – в позиции специалистов» [7, с. 994].

Как отмечает Е. А. Коган, «научно-исследовательскую работу со студентами можно рассматривать как процесс и как результат. Чаще всего научно-исследовательская деятельность представляется как результат, выраженный в количестве опубликованных статей, количестве участников и побед в олимпиадах и различных конкурсах и т. д. Тем не менее мало внимания уделяется самому процессу этой деятельности: информированию студентов, мотивации, подготовке студентов к конференциям, коммуникациям с преподавателем и т.д. Без должной организации этого процесса эффект достигнуть трудно. За результатами НИРС часто стоит большая работа, которая не всегда отслеживается и оценивается» [9, с. 179]. При этом, по мнению А. В. Тараненко и Х. Н. Кушваха, «не последнюю роль в приобщении студентов к НИР играет сама личность научного руководителя, стиль педагогического руководства и методы обучения» [8, с. 224]. Кроме того, очень важно и то, «что приносит научно-исследовательская деятельность самому студенту, как она повлияет на процесс его обучения и развития в дальнейшем» [9, с. 179].

В. Ф. Габдулхаков, А. Ф. Зиннурова, А. М. Гарифуллина, З. Р. Егорова, Л. Д. Павлова выделяют пять разных ролей преподавателей в их работе по развитию инновационной деятельности обучающихся: 1) «транслятора инновационных идей или инновационного содержания; роль куратора, закрепляющего за студентами направления инновационной деятельности»; 2) «консультанта, разъясняющего возникшие у студентов вопросы по организации и проведению исследовательской деятельности»; 3) «тьютора, обеспечивающего научно-методическое сопровождение проектирования и реализации исследовательской деятельности студентов»; 4) «наставника, корректирующего и направляющего исследовательскую деятельность на основе имеющегося у него передового опыта»; 5) «ментора, обеспечивающего достижение студентами значительных результатов в исследовательской



деятельности», и замечают, что «если в университетах стран Запада эти роли дифференцированы и строго регламентированы, то в российских вузах они могут совмещаться в одном лице и иметь в результате формальный характер», также отмечают, что «системная организация этих ролевых структур (от полученных знаний к деятельности) обеспечивает технологизацию подготовки студентов к творческой исследовательской деятельности», при этом «эффективность такого менторинга существенно возрастает, если эти роли выполняет не один преподаватель, а несколько (от шести до двенадцати человек), среди которых есть сильные ученые, практики, новаторы» [10].

Проектное обучение отличается тем, что деятельность обучаемых направлена на получение конкретного (практического) результата и его публичного предъявления. Самостоятельное производство знания, создание образовательной продукции по изучаемым дисциплинам, самостоятельное решение возникающих проблем проясняют основные цели проектного метода [6, 11]:

- стимулирование мотивации обучения;
- включение обучающихся студенческой группы в режим самостоятельной групповой (или индивидуальной) работы, связанной с проектированием какого-либо объекта или процесса;
- развитие разнообразных способностей (познавательных, коммуникативных, организаторских, профессиональных и других);
- реализация принципа сотрудничества обучаемых и преподавателя при решении учебных и (или) проблемных задач.

Работа в проекте учит следующим элементам проектной деятельности [12]:

- поиску и систематизации необходимой информации;
- умению ставить цель и задачи;
- методике анализа и применения информации с целью решения поставленной задачи путем практического применения полученных знаний, умений и навыков в разных ситуациях, а также при решении творческих проблемных нестандартных задач;
- умению обосновывать свой подход к решению задачи и представить полученные результаты.

В статье У. Сы [13] рассмотрены проекты в вузах Китая, ориентированные на презентацию достижений обучающихся и преподавателей, а также демонстрирующие особенности подготовки будущих педагогов-художников через призму формирования независимости их творческих решений.

Е. Н. Шуранова и Г. В. Фокина проектную деятельность по начертательной геометрии рассматривают как педагогическую технологию, позволяющую глубже освоить предметную область и способствующую повышению мотивации обучающихся к учебе. По мнению авторов, главной проблемой для участия в проектной деятельности студенты считают сложность совмещения с учебной (требует много времени, не совпадает с жестко зафиксированным учебным планом и рабочей программой дисциплины) [12].

Организация сопровождения самостоятельной работы над проектом, как отмечают И. Д. Столбова, Л. В. Кочурова и К. Г. Носов, требуют достаточного набора технологических, коммуникативных и информационных инструментов, которые обеспечивали бы управление процессом учебного проектирования, опираясь на принцип персонализации, учитывающий личностные особенности и предпочтения обучающихся. По мнению авторов, данный раздел цифровой образовательной среды должен включать «все требуемые информационные и справочные ресурсы, в том числе узкоспециализированные библиотеки; систему управления ходом выполнения задания; а также оперативный консультативный канал связи с преподавателем, необходимый, если в результате проектирования у студента возникают затруднения» [1, с. 58]. Кроме того, при организации проектной деятельности необходимо учитывать и тот момент, что изучение начертательной геометрии и инженерной графики происходит на первом курсе, когда идет адаптация обучающихся к вузовской системе обучения.

Хотелось отметить, что синтез научно-исследовательского и проектного методов также довольно широко применяется в образовательной сфере. Так, проектная деятельность включает в себя исследовательские методы в связи с тем, что для эффективного решения поставленной задачи необходим комплексный подход, предполагающий научно-исследовательский и учебно-исследовательский поиск. В свою очередь, научно-исследовательские методы в процессе подготовки

проекта способствуют более гибкому подбору исследовательского инструментария, например, помогают «анализировать возможные варианты развития проекта, вносить необходимые изменения для получения качественного результата» [6, с. 520].

В данной работе исследовательская и проектная деятельность рассмотрена как одни из важных подвидов воспитательной деятельности, способствующие как развитию личностных ценностных компетенций обучающихся, так и формированию научно-исследовательской и профессионально-проектной культуры будущего специалиста, начиная с первого года обучения в вузе.

Современное воспитание и развитие личности будущего специалиста предполагает реализацию идей педагогики сотрудничества и педагогики поддержки, нацеленных на адаптацию обучающихся к новой вузовской среде, раскрытию потенциальных возможностей и способностей личности студента, активизацию его творческой активности, продуктивного самообразования и социальной деятельности. Организация воспитательной деятельности в вузе отличается многообразием практикуемых методов по воздействию на интеллектуальную, мотивационную, волевую, экзистенциальную, предметно-практические сферы, сферу саморегуляции; вариативностью форм обучения: групповых и индивидуальных, формальных и неформальных, аудиторных и внеаудиторных, традиционных и инновационных [14]. Побудителем к деятельности является установка, готовность индивида на основе мотивации к соответствующему виду активности [15].

Кроме того, хотелось сделать акцент еще на том, что в психолого-педагогической литературе в последнее время очень часто поднимается вопрос о недостаточном уровне геометро-графической подготовки выпускников инженерных направлений подготовки технических вузов. Среди существенных причин этой проблемы Е. В. Любчинов, Ю. Е. Ионова и Т. М. Мясоедова выделяют, во-первых, сокращение часов на изучение графических дисциплин, отмечая, что в некоторых вузах дисциплине «Инженерная и компьютерная графика» («Начертательная геометрия и инженерная графика», «Инженерная графика», «Начертательная геометрия и компьютерная графика» и другое – название в вузах варьируется) отводится всего три или даже две зачетные единицы, и таким образом, со слов авторов, дисциплина переходит из разряда изучаемых «на ознакомление и многое из того, что ранее давалось в рамках аудиторных занятий сейчас уходит на исключительно самостоятельное изучение, либо не дается вовсе». Во-вторых, «не стоит забывать и том, что преподаватели высшей школы уже давно сталкиваются с проблемой отсутствия у поступающих базовых знаний по школьному курсу черчения, что еще больше усугубляет проблему». В-третьих, эти причины связывают с поглощением кафедр начертательной геометрии и инженерной графики выпускающими кафедрами, слиянием их с другими кафедрами, а также нехваткой высококвалифицированных научно-педагогических кадров по данному профилю. Авторы справедливо отмечают, что в совокупности эти факторы дают низкий уровень приобретенных знаний выпускников, и если с отсутствием черчения в школах и сокращением кафедр инженерной графики рядовой преподаватель ничего сделать не может, то нехватка аудиторного времени заставляет преподавателей искать новые подходы, методики и способы обучения, которые бы позволили сохранить должное качество образования в рамках изучения графической дисциплины в текущих ограниченных условиях [5].

### Результаты экспериментов

Первой дисциплиной графического цикла инженерных направлений подготовки в ФГБОУ ВО «Поволжский государственный технологический университет» является «Начертательная геометрия и инженерная графика». Она рассчитана на изучение в течение от одного семестра до трёх студентами различных инженерных направлений подготовки. Так, например, обучающиеся по направлению подготовки 09.03.04 «Разработка программных систем» факультета информатики и вычислительной техники изучают весь курс дисциплины за один семестр, обучающиеся по направлению подготовки 08.03.01 «Строительство» института строительства и архитектуры – за два семестра, а обучающиеся по направлению подготовки 13.03.01 «Промышленная теплоэнергетика» института механики и машиностроения – за три семестра. В электронном образовательном портале вуза на каждый семестр читаемой дисциплины создан отдельный курс. Поэтому важно еще на первом занятии акцентировать внимание: на сколько по времени рассчитан весь курс дисциплины; как и где можно задать текущие

вопросы; где искать теорию, если нет лекций; где брать домашние задания и методические указания по их выполнению? Если пропущено занятие по уважительной причине, как подготовиться к зачету или экзамену, можно ли получить оценку по системе РИТМ и каким образом, как и за что можно получить дополнительные баллы? Кроме того, каждый преподаватель как автор своего курса обогащает курс читаемой дисциплины по семестрам необходимыми методическими материалами по своему творческому усмотрению. Поэтому студентам рекомендовалось, чтобы каждый из них зашёл на электронный курс еще во время первого занятия. При этом преподаватель не только давал разъяснение о том, как работать с курсом, но и вместе со студентами просматривал все позиции в нем для конкретной группы определенного направления подготовки. Все это содействовало тому, что студенты с большей ответственностью начинали относиться к учебным занятиям по курсу, ознакомились, посетив соответствующие папки, с обязательными видами практических и лабораторных работ, а также с олимпиадными заданиями, примерными темами докладов для конференций, лучшими исследовательскими проектами, тезисами и презентациями докладов с конференций и мастер-классов прошлых лет.

Организация самостоятельной работы при изучении графических дисциплин в ФГБОУ ВО «Поволжский государственный технологический университет» была рассмотрена в работе [16].

Воспитательная работа кафедры начертательной геометрии и графики ФГБОУ ВО «Поволжский государственный технологический университет» проводилась в соответствии с утвержденным планом. Ежегодно преподавателями кафедры разрабатываются олимпиадные и конкурсные задания, а также организовываются и проводятся следующие предметные олимпиады и конкурсы по дисциплинам «Начертательная геометрия и инженерная графика» и «Начертательная геометрия и компьютерная графика»:

- ✓ Межрегиональный студенческий конкурс «Моделирование и 3D-прототипирование» Всероссийского студенческого Форума «Инженерные кадры – будущее инновационной экономики России» (ноябрь);

- ✓ I (внутривузовский) II (всероссийский) и III (международный) тур открытой интернет-олимпиады по дисциплине «Начертательная геометрия и инженерная графика» (декабрь–апрель);

- ✓ Внутривузовская предметная олимпиада по начертательной геометрии (февраль–март);

- ✓ Внутривузовская предметная олимпиада по проекционному черчению (март–апрель);

- ✓ Внутривузовская предметная олимпиада «Перспектива и тени» (апрель–май);

- ✓ Внутривузовская предметная олимпиада по компьютерной графике (май–июнь).

Кроме того, преподаватели кафедры начертательной геометрии и графики готовят обучающихся и к внешним конкурсам. Так, например за последние пять лет, с 2021 года обучающиеся 1–4 курсов принимали участие во Всероссийской олимпиаде с международным участием по геометрическому моделированию, проводимой ФГБОУ ВО «Оренбургский государственный университет» (май); с 2024 года обучающиеся 1–2 курсов – во Всероссийской Интернет-олимпиаде по комплексу графических дисциплин, проводимой ФГБОУ «Сибирский государственный университет путей сообщения» (май); с 2025 года – в конкурсе методических и научных публикаций по инженерной, компьютерной и презентационной графике, проводимом ФГБОУ «Сибирский государственный университет путей сообщения» (февраль).

Стоит отметить, что олимпиада по начертательной геометрии и инженерной графике играет важную роль в организации исследовательской и проектной деятельности студентов, хотя сама по себе не является научным исследованием. Ее значение можно раскрыть через несколько аспектов:

- *стимулирование исследовательского подхода.* Так, олимпиадные задания часто выходят за рамки стандартных учебных задач, требуя анализа и поиска нестандартных решений (например, построение сложных пересечений поверхностей), применения алгоритмов 3D-моделирования, оптимизации графических методов (например, выбор наиболее рационального способа построения). Это формирует у студентов навыки, необходимые для научной работы;

- *развитие проектного мышления.* Участие в олимпиаде помогает структурировать инженерные задачи (от анализа условия до визуализации решения), работать с пространственными моделями (что важно для научных исследований в области компьютерной графики, машинного обучения и робототехники);

- *подготовка к научной деятельности.* Так, олимпиада может стать отправной точкой для исследований: сравнения эффективности графических алгоритмов (например, анализ точности разных методов проецирования), разработка новых методик обучения (на основе анализа ошибок участников), интеграция с цифровыми технологиями (использование цифровых моделей для визуализации задач);
- *отбор талантливых студентов для НИР.*

Поэтому олимпиада служит мотивационным инструментом, тренировочной площадкой для развития исследовательских компетенций и стартовой точкой для реализации проектных и научных работ.

Кроме того, преподаватели кафедры ежегодно организуют и проводят конкурсы студенческих докладов в подсекции «Прикладная геометрия», в секции «Фундаментальные науки» внутривузовской студенческой научно-технической конференции «Молодежь и наука» и в секции «Прикладная геометрия и компьютерная графика» международной молодежной научной конференции по естественнонаучным и техническим дисциплинам «Научному прогрессу – творчество молодых» (апрель), а также готовят к участию и на внешних конференциях (например, за последние два года доклады студентов в соавторстве с руководителями – преподавателями кафедры начертательной геометрии и графики были опубликованы в сборнике трудов международной научно-практической конференции «Инновационные технологии в инженерной графике: проблемы и перспективы», Брест, Новосибирск).

Таким образом, при изучении дисциплины «Начертательная геометрия и инженерная графика» в ФГБОУ ВО «Поволжский государственный технологический университет» обучающимся предоставляется большой выбор для реализации своих творческих идей в научно-исследовательской и учебно-исследовательской деятельности.

Учебным планом и рабочей программой курсовые проекты по дисциплине «Начертательная геометрия и инженерная графика» не предусмотрены. Поэтому для организации исследовательской и проектной деятельности были использованы средства внеучебной деятельности (консультации, дополнительные занятия), а также возможности цифровой среды вуза (наполнение электронного курса дисциплины необходимыми информационными папками на образовательном портале вуза, использование в нем чатов для общения «обучающийся-педагог»).

Тему исследовательского проекта обучающимся предлагалось выбрать по желанию в начале 1-го семестра. Тема должна была быть оригинальной и близкой либо к изучаемой дисциплине, либо, немного забегаая вперед, приближенная к будущей профессиональной деятельности, по видению самих обучающихся.

Сразу оговоримся, что обучающимся регулярно напоминалось, что за участие в олимпиадах, конференциях и разработку творческого или исследовательского проекта начисляются дополнительные баллы к системе РИТМ. Для отметки «отлично» необходимо за семестр набрать баллы от 90 до 100, «хорошо» – 75...89 баллов, «удовлетворительно» – 50...74 балла, из которых 10 баллов можно набрать только за научно-исследовательскую и проектную деятельность.

Как показала практика, в течение первых двух месяцев единицы среди обучающихся заинтересовались этими видами деятельности. Чаще всего, первые темы были связаны с теоретическими выкладками. Например, «История начертательной геометрии», «История возникновения чертежа», «Сравнение российских и зарубежных правил оформления чертежа» и другие.

Но по мере углубления в дисциплину во время аудиторных практических работ, решения тематических задач, домашних самостоятельных расчетно-графических работ, лабораторных работ по компьютерной графике в Компас-3D обучающиеся столкнулись с некоторыми трудностями в изучении некоторых аспектов. Одни из студентов втянулись в проект, чтоб решить эти проблемы. Другие заинтересовались проектами для компенсации потерянных баллов за обязательные виды работ, предусмотренных рабочей программой и технологической картой по системе РИТМ. Третьим хотелось окунуться в мир будущей профессии.

В связи с этим темы творческих проектов стали варьироваться. Например, обучающимся стали интересны исследования способов создания моделей деталей в Компас-3D, выбора среди них оптимального; сравнения инструментов по моделированию, визуализации и анимации в различных современных программах (например, Компас-3D, T-FLEX CAD, Blender, 3D Max и другие); параметризации и использования библиотек современных автоматизированных систем проектирования и другие.

После выбора темы рекомендовалось определиться с целью и задачами проекта, затем начать активно его прорабатывать. Презентации проектов были запланированы к первому прослушиванию в конце декабря внутри своей группы, в результате чего студенты поделились со сверстниками своими творческими решениями, задавали интересующие вопросы, лучшим из них было предложено доработать проект и представить результаты на конференциях.

Некоторые лучшие проекты (при согласии автора) были продемонстрированы в мастер-классах для школьников, что послужило еще одной репетицией публичного выступления перед конференцией, в которой экспертами были опытные педагоги кафедры начертательной геометрии и графики.

### Обсуждение результатов

Организация исследовательской и проектной деятельности представляет собой комплекс управленческих и педагогических действий, направленных на планомерное руководство процессом вовлечения студентов в научно-практическую работу. Данная система включает четкую этапность реализации, научно обоснованные методы руководства, систему мотивационных механизмов, критерии оценки эффективности.

Организация исследовательской и проектной деятельности в Поволжском государственном технологическом университете реализовалась через последовательность следующих этапов:

- 1) *диагностико-проектировочный этап*: анализ исходного уровня подготовки, формулировка целей и задач, проектирование индивидуальных образовательных маршрутов обучающихся;
- 2) *содержательно-технологический этап*: разработка системы проектных заданий, подбор цифровых инструментов, определение критериев оценки;
- 3) *рефлексивно-оценочный этап*: анализ достигнутых результатов, корректировка методики, разработка рекомендаций.

Методический аппарат организации процесса этих видов деятельности был основан на принципах системности и последовательности, индивидуального подхода, практической ориентированности и интеграции с учебным процессом. При этом применялись такие методы руководства, как консультативно-ориентирующие, стимулирующие, контрольно-оценочные и рефлексивные.

Методом включенного наблюдения удалось зафиксировать основные признаки познавательной деятельности студентов при выполнении проектов, подготовке презентаций и докладов к мастер-классам и конференциям, а также тезисов докладов для публикации, а именно выраженный познавательный интерес к теме исследования, высокую степень познавательной активности, инициативности и самостоятельности, стремление к успеху.

Кроме того, в качестве обратной связи был проведен опрос 49 окончивших 1 курс студентов инженерных направлений подготовки ФГБОУ ВО «Поволжский государственный технологический университет», вовлеченных в исследовательскую и проектную деятельность при изучении дисциплины «Начертательная геометрия и инженерная графика» в форме эссе – отзыва в свободной форме. Респонденты отметили следующие плюсы (тезисно представлены некоторые отзывы студентов):

✓ «Начертательная геометрия и инженерная графика» – это важная дисциплина, которая закладывает основы для дальнейшего обучения и профессиональной деятельности. Преподаватели сыграли важную роль в нашем обучении, отвечая на вопросы и проводя консультации, что помогло нам разобраться в непонятных моментах»;

✓ «Курс «Начертательная геометрия и инженерная графика» представляет возможность развить необходимые навыки, которые помогут в дальнейшем профессиональном развитии: навыки пространственного мышления и визуализации; создание чертежей и моделей; знание международных стандартов и норм»;

✓ «Этот курс оказался для меня очень удобным и информативным. Вся необходимая информация была доступна, что значительно облегчало процесс обучения. Особенно хочу отметить презентации и видеоуроки, которые были представлены в курсе. Они помогли мне лучше усвоить материал и разобраться в сложных темах. Одним из полезных аспектов курса стали тесты для самоконтроля, которые позволяли проверять свои знания и закреплять пройденный материал. За успешное прохождение тестов можно было получить дополнительные баллы, что стало хорошей мотивацией для более глубокого изучения тем»;

✓ «После походов на дополнительные занятия я стала лучше разбираться в видении деталей в трехмерном измерении. А после этого я стала лучше понимать другие темы, например, сечения»;

✓ «Интересным опытом было посещение олимпиад и конференции. Я посетила 2 олимпиады, за что получила дополнительные баллы, которые помогли мне поднять оценку. Конференция, на которую я ходила, тоже была интересной. Люди представляли интересные проекты и примеры их исполнения. За участие в конференции получила дополнительные баллы, которые тоже подтянули оценку»;

✓ «По возможности студенты принимали участие в олимпиадах. Олимпиады проходили на региональном, муниципальном, всероссийском и международных уровнях. Полученных знаний хватало, чтобы достигать призовые места не только на внутривузовском, но и выше»;

✓ «Участие в конференции стало для нас настоящим испытанием. Создание моделей бревен и сборка сруба в Компас-3D были сложными и трудоемкими задачами, но в то же время очень интересными. Этот опыт оказался для нас ценным. Мы узнали много нового и приобрели полезные навыки проектной работы»;

✓ «Мастер-класс для школьников мне понравился. Было, конечно, страшно выступать перед школьниками 10-го класса. Если говорить про сам конкурс, было видно, что участники вложили душу в свои проекты и видна их творческая натура. Результаты участников получились интересными и творческими. Участникам тоже понравилось работать в Компас-3D. Задумка, чтобы ребята нашли ошибки в созданной 3D-модели деревянного дома, довольно интересная и заинтересовала, вроде бы, учеников. С моей стороны было интересно поучаствовать в мероприятии как рассказчик-помощник для ребят»;

✓ «В процессе обучения мы активно занимались решением практических проектных заданий, что помогло значительно улучшить наши навыки, необходимые как в учебе, так и в будущей карьере. Одним из наиболее запоминающихся заданий было командное: мы работали над созданием модели углового крана. Сначала мы разрабатывали чертежи и модели деталей, а затем осуществляли сборку. Работа в команде была особенно увлекательной, ведь необходимо было создать детали, которые идеально сочетались бы с элементами, разработанными другими участниками группы. Это не только углубило наши знания по курсу, но и развило навыки коллективной работы»;

✓ «Самым интересным в обучении была практика в программе Компас-3D. При выполнении заданий выдавались пошаговые инструкции в методичках, которые были хорошо понятны, но, даже если и возникали затруднения при выполнении задания, всегда была возможность связаться с преподавателем, которая всегда была готова помочь даже во внеучебное время, а также поставить дополнительные занятия, если мы что-то не понимали или не успевали»;

✓ «Спасением стали дополнительные занятия построения зданий и предметов интерьера и мебели из дерева в Компас-3D, которые давали время и место для разбора команд и программ в Компас-3D»;

✓ «Мне выпала возможность пройти дополнительное обучение с последующим участием в конференции. Мы моделировали деревянный сруб бани в Компас-3D. Из плюсов дополнительного курса могу выделить: 1) Изучение функций программы Компас-3D, что в дальнейшем помогло в учебном процессе; 2) Интересное задание; 3) Данные, используемые в работе, были собраны с настоящих строительных объектов; 4) Свобода фантазии: сборку мы делали по своему собственному эскизу. Можно было изменить его в любой момент; 5) Если возникали трудности в работе, преподаватель всегда объясняла и помогала разобраться в проблеме; 6) Приобретение новых знакомых и друзей в процессе прохождения курса... В целом курс «Начертательная геометрия и инженерная графика» довольно трудоёмкий и требует много усилий, но полученные теоретические и практические знания необходимы для моей будущей профессии. Обучение было сложным, но очень интересным. Создание моделей в программе Компас-3D после прохождения курса теперь не вызывает затруднений».

Из минусов респондентами было отмечено: 1) «неудобное время проведения дополнительных занятий»; 2) «главная загвоздка, на мой взгляд – не хватает системного разбора того, как именно подходить к олимпиадным заданиям. Очень не хватает объяснений самих подходов и методов решения, разбора типовых ловушек, в которые все попадают. Было бы просто здорово, если бы добавили: во-первых, подробные разборы решений задач с прошлых олимпиад. Не просто ответы, а именно пошаговое объяснение: почему решали так, а не иначе, какие есть варианты, где подвохи.

Во-вторых, специальные семинары или дополнительные материалы, которые сфокусированы не на новой теме, а именно на стратегии решения. Как думать, когда задача нестандартная? С чего начать? Как не утонуть в условии? Вот такие штуки были бы крайне полезны для реальной подготовки к олимпиадам».

Опрос позволил оценить эффективность образовательного процесса: выявить, какие элементы курса (видеоуроки, тесты, дополнительные занятия) наиболее полезны для студентов и где проблемные зоны; определить, как участие в исследовательской и проектной деятельности влияет на понимание важности изучения дисциплины; оценить качество методической поддержки (доступность материалов, помощь преподавателей). Ответы студентов подтверждают ценность проектной деятельности для профессионального развития, демонстрируют положительное влияние цифровых инструментов на обучение, а также обосновывают необходимость дальнейшего развития исследовательской деятельности.

Таким образом, опрос стал ценным инструментом для анализа и улучшения системы организации исследовательской и проектной деятельности обучающихся.

### **Заключение**

Вопросы, связанные с привлечением обучающихся к исследовательской и проектной деятельности, являются актуальными в научно-психолого-педагогическом сообществе. Эти виды деятельности необходимы и с точки зрения образовательной и воспитательной функции, так как способствуют повышению интеллектуально-творческой деятельности среди обучающихся, формируют у них коммуникативные компетенции, развивают навыки социальной адаптации.

На основе изучения психолого-педагогической литературы, собственного опыта воспитательной работы с обучающимися при чтении графических дисциплин, анализа результатов опроса и включенного наблюдения разработаны следующие рекомендации по повышению мотивации обучающихся к исследовательской и проектной деятельности в условиях цифровизации высшего образования:

- для включения обучающихся в исследовательскую (в том числе НИР) и проектную деятельность с первого курса необходимо активно использовать образовательный портал вуза и социальные сети для популяризации научно-исследовательских и творческих проектных работ студентов, приглашать выступать перед ними или проводить мастер-классы активистов-старшекурсников, выпускников и работников с производства в офлайн-формате (или представить лучшие творческие работы и видеопрезентации профессионалов в онлайн-формате также в образовательном портале в отдельных папках «Примеры творческих проектов» и «О чем говорят на производстве»);
- в конце 1 семестра необходимо запланировать и организовать проведение одного занятия в виде конференций, чтоб каждый желающий мог попробовать себя в роли исследователя и автора проекта, а также поделился своей исследовательской работой или проектом в своей академической группе, выявить среди них лучшие работы и, доработав их, представить авторам возможность выступления на внутривузовских, региональных и всероссийских конференциях среди обучающихся;
- при выборе темы для исследовательской и проектной работы акцентировать внимание, чтобы она была близкой к их будущей профессиональной деятельности, но при этом дать свободу выбора обучающимся;
- при выполнении исследовательских и проектных работ активно использовать современные цифровые инструменты, в том числе САПР;
- при возникновении вопросов по выбору темы, трудностей по решению исследовательской задачи и проектной работы предоставлять своевременную консультацию обучающимся как в очном формате, так и онлайн;
- при разработке технологической карты по системе РИТМ предусмотреть дополнительные баллы в качестве поощрения обучающихся за участие в олимпиадах, конференциях, а также подготовку оригинальных творческих исследовательских и проектных работ.

Таким образом, представленная методика организации исследовательской и проектной деятельности демонстрирует системный подход к управлению данным процессом, где особое



внимание уделяется четкой последовательности организационных этапов, научно обоснованным методам руководства, комплексной системе мотивации.

Разработка рекомендаций по повышению мотивации студентов выступает как важный, но не единственный результат совершенствования организации процесса в целом. Основной акцент делается на создании целостной системы управления исследовательской и проектной деятельностью, обеспечивающей достижение поставленных образовательных целей.

### Список литературы

1. Столбова И. Д., Кочурова Л. В., Носов К. Г. К вопросу о цифровой трансформации предметного обучения // Информатика и образование. 2020. № 9. С. 53–63.
2. Маметьева О. С., Супрун Н. Г., Халикова Д. А. Научно-исследовательская работа студентов вуза: результативность и проблемы организации // Современные проблемы науки и образования. 2018. № 1. URL: <https://science-education.ru/article/view?id=27362>
3. Профильные аспекты графического образования в политехническом вузе / И. Д. Столбова, Е. П. Александрова, Л. В. Кочурова, К. Г. Носов // Высшее образование в России. 2019. № 3. С. 155–166.
4. Использование 3D-технологий для развития инновационного мышления / Т. В. Машарова, Н. А. Бушмелева, М. С. Перевозчикова, И. Ю. Хлобыстова // Перспективы Науки и Образования. 2020. № 3 (45). С. 426–440.
5. Любчинов Е. В., ИONOVA Ю. Е., Мясоeдoвa Т. М. Об опыте проведения занятий в рамках проекта «Цифровая кафедра» для студентов инженерного профиля // GRAPHICON 2024: материалы 34-й Международной конференции по компьютерной графике и машинному зрению, Омск, 17–19 сентября 2024 года. Омск: Омский государственный технический университет, 2024. С. 988–995.
6. Решетникова Е. В. Организация самостоятельной работы студентов, направленной на развитие публикационной деятельности, с использованием проектного и научно-исследовательского методов // Педагогика. Вопросы теории и практики. 2023. Т. 8, № 5. С. 517–524.
7. Бирюкова Н. В. Педагогическая модель формирования мотивации студентов к изучению непрофильных дисциплин в рамках контекстного обучения в вузе // Педагогика. Вопросы теории и практики. 2022. Т. 7, № 10. С. 990–996.
8. Тараненко А. В., Кушвахa X. H. Вовлечение студентов, обучающихся по направлению «Реклама и связи с общественностью» в научно-исследовательскую деятельность // Педагогика. Вопросы теории и практики. 2023. Т. 8, № 2. С. 224–228.
9. Коган Е. А. Отношение студентов вузов к научно-исследовательской работе // Человеческий капитал. 2020. № 8 (140). С. 179–187.
10. Менторинг в работе по развитию инновационной деятельности студентов / В. Ф. Габдулхаков, А. Ф. Зиннурова, А. М. Гарифуллина, З. Р. Егорова, Л. Д. Павлова // Современные проблемы науки и образования. 2023. № 1. URL: <https://science-education.ru/article/view?id=32401>
11. Шарипов Ф. В. Технология проектного обучения // Педагогический журнал Башкортостана. 2012. № 2 (39). С. 87–94.
12. Шуранова Е. Н., Фокина Г. В. Использование метода проектов при изучении начертательной геометрии // Современное педагогическое образование. 2023. № 5. С. 320–323.
13. У Сы. Выставочные проекты в художественно-педагогических вузах Китая как средство активизации творческой самостоятельности обучающихся // Педагогика. Вопросы теории и практики. 2023. Т. 8, № 9. С. 967–971.
14. Организация воспитательной работы в вузе: опыт анализа формирования института кураторов (на примере Санкт-Петербургского государственного химико-фармацевтического университета) / А. О. Волгушева, С. А. Воробьева, Н. А. Завершинская, Т. Ф. Черных // Педагогика. Вопросы теории и практики. 2023. Т. 8, № 7. С. 771–779.
15. Ложников А. Е., Маренко В. А. Саморегуляция учебной деятельности: когнитивный подход // GRAPHICON 2024: материалы 34-й Международной конференции по компьютерной графике и машинному зрению, Омск, 17–19 сентября 2024 года. Омск: Омский государственный технический университет, 2024. С. 1003–1009.
16. Турецких С. О. Организация самостоятельной работы по дисциплине «Начертательная геометрия и инженерная графика» // Современные проблемы технического образования: материалы XXII Всероссийской научно-методической конференции, Йошкар-Ола, 18–19 марта 2022 года. Йошкар-Ола: Поволжский государственный технологический университет, 2022. С. 189–192. URL: <https://science.volgatech.net/nm/Conferences/sovremennyye-problemy-tekhnicheskogo-obrazovaniya/1SPTO20221.pdf>

УДК 004.942: 378.147

DOI: 10.25686/978-5-8158-2474-4-2025-1063-1069

## Проблемы внедрения методов цифровой поддержки в фундаментальные физико-математические дисциплины

О. И. Ведяйкина, П. А. Хазов, С. С. Шилов

Нижегородский государственный архитектурно-строительный университет,  
Нижний Новгород, Россия

**Аннотация.** В статье рассматриваются актуальные проблемы внедрения методов цифровой поддержки в фундаментальные физико-математические дисциплины высшего образования на примере строительных вузов. Особое внимание уделяется вопросам интеграции цифровых технологий и виртуальных лабораторных работ на ранних этапах обучения, что способствует повышению эффективности образовательного процесса и подготовке специалистов, ориентированных на современные технологические требования. Анализируются особенности использования цифровых моделей и инструментов в таких дисциплинах, как физика и теоретическая механика. Обосновывается необходимость более ранней и системной цифровизации учебных программ для формирования компетенций, соответствующих современным стандартам науки и инженерии.

**Ключевые слова:** цифровизация образования, цифровая лабораторная работа, цифровая поддержка установки, высшее образование, цифровая поддержка исследования, физика, теоретическая механика, сопротивление материалов

## Problems of implementing digital support methods in fundamental physical and mathematical disciplines

O. I. Vediaikina, P. A. Khazov, S. S. Shilov

Nizhny Novgorod State University of Architecture and Civil Engineering,  
Nizhny Novgorod, Russia

**Abstract.** The article discusses the current problems of implementing digital support methods in fundamental physical and mathematical disciplines of higher education, using the example of construction universities. Special attention is paid to the integration of digital technologies and virtual laboratory work at the early stages of education, which helps to improve the efficiency of the educational process and prepare specialists who meet modern technological requirements. The article analyzes the features of using digital models and tools in disciplines such as physics and theoretical mechanics. The article substantiates the need for earlier and more systematic digitalization of educational programs to develop competencies that meet modern standards of science and engineering.

**Keywords:** digitalization of education, digital laboratory work, digital installation support, higher education, digital research support, physics, theoretical mechanics, material resistance

### Введение

В современном мире цифровизация проникает во все сферы жизни, неуклонно меняя привычные действия и открывая новые горизонты. Стремительное развитие вычислительной техники и компьютерного моделирования позволило достичь уровня, когда создание надежных методик цифровой поддержки (ЦП) экспериментальных исследований становится реальностью. Эта задача распространяется на все области технических наук и промышленного производства – от микроэлектроники до космической индустрии, требуя разработки универсальных и адаптируемых решений [1–4].

Создание и внедрение эффективных методик ЦП для широкого спектра технических исследований открывает возможности для оптимизации производственных процессов, валидации экспериментальных установок, поддержки их жизненного цикла, а также значительного сокращения затрат на исследования и разработки. Это достигается путем использования виртуальных моделей объектов, полностью отражающих геометрию реальных, а также имитирующих все физические процессы, происходящие внутри него или с ним взаимодействующие – цифровых двойников (ЦД). ЦД позволяет учитывать при проведении виртуального эксперимента весь спектр параметров – от температурных режимов и напряжений в материалах до динамических нагрузок и сложных химических реакций, в зависимости от специфики моделируемого объекта. Кроме того, такое внедрение даст возможность осуществлять ЦП испытаний, моделируя различные сценарии работы объекта или системы и анализируя их последствия, не прибегая к дорогостоящим и длительным физическим экспериментам.

Таким образом, развитие технологий ЦД и ЦП исследований является ключевым фактором повышения эффективности разработок во всех областях науки и техники, способствуя созданию более надежных, безопасных и экономичных методик проектирования.

### **Постановка задачи**

Цифровые технологии играют всё более важную роль в научных исследованиях и разработках, поэтому их внедрение в процесс обучения в системе высшего образования становится всё более актуальным вопросом [5]. Однако, если говорить о подготовке инженеров, например, в строительных вузах, то обучение использованию инструментов ЦП зачастую начинается уже слишком поздно. Вместо того чтобы интегрировать эти технологии в учебный процесс на всех этапах обучения, студенты знакомятся с цифровым моделированием, как правило, только на старших курсах, когда начинаются проектирование зданий и сооружений. То есть значительная часть фундаментальных технических дисциплин проходит без использования современных цифровых инструментов, хотя их применение могло бы значительно повысить эффективность обучения и подготовить будущих специалистов к работе с передовыми технологиями [6].

В статье поднимается вопрос внедрения средств ЦП исследований в физико-математические дисциплины первого курса обучения, такие как физика и теоретическая механика [7, 8].

### **Теория**

Фундаментальные физико-математические науки традиционно строятся на глубоком теоретическом понимании базовых концепций и строгих доказательствах, которые часто требуют значительных усилий и временных затрат. Внедрение цифровых технологий в процесс изучения таких дисциплин, как физика, теоретическая механика, сопротивление материалов и др., способствует их более эффективному и доступному освоению. Например, создание и проведение виртуальных экспериментов делает их более наглядными и доступными для детального изучения с возможностью многократного повторения сложных моментов. Применение виртуальных лабораторий позволяет студентам лучше понимать принципы функционирования механических систем и физических процессов, происходящих в них, а также свойства веществ, сред и материалов.

При этом, несмотря на прогресс цифровых технологий, внедрение современных методов ЦП в эти области сталкивается с такими проблемами и ограничениями, как:

- недостаток инфраструктуры: аудитории и лаборатории недостаточно технически оснащены;
- недостаточная интеграция новых подходов: современные образовательные программы часто отстают от быстроразвивающихся технологий;
- нехватка квалифицированного персонала: требуется внедрение курсов повышения квалификации для преподавателей для освоения новых методик обучения;
- бюджетные ограничения вузов: финансовые возможности многих российских вузов ограничены, что снижает темпы внедрения инновационных решений и замедляет модернизацию образовательного процесса.

Для преодоления перечисленных препятствий необходима комплексная стратегия развития высшего образования, предусматривающая обеспечение финансовой поддержки университетских инициатив по внедрению цифровых технологий, повышение квалификации профессорско-преподавательского состава, обновление технических возможностей вузов, формирование новой культуры внедрения методов ЦП в процесс обучения с первых курсов.

### **Результаты экспериментов**

Для проведения анализа существующего формата обучения были отобраны рабочие программы дисциплин (РПД) «Физика» и «Теоретическая механика» российских строительных университетов по профилю «Промышленное и гражданское строительство» очной формы обучения уровня образования «бакалавриат» [9–18]. На основании рабочих программ, которые находятся в общем доступе на официальных сайтах вузов, проанализированы данные по количеству часов, выделенных

на изучение дисциплины, в том числе лабораторных работ, и получаемых в результате компетенциях (табл. 1, 2).

Таблица 1. Показатели из рабочих программ дисциплины «Физика» строительных вузов

Наименование вуза	Общее количество часов, ч	Количество лабораторных, ч	Получаемые компетенции
Астраханский государственный архитектурно-строительный университет (АГАСУ)	180	34	ОКП-1.1 ОКП-1.2 ОПК-1.4 ОПК-1.5 ОПК-1.11
Казанский государственный архитектурно-строительный университет (КГАСУ)	216	18	ОПК-1 ОПК-2
Национальный исследовательский Московский государственный строительный университет (НИУ МГСУ)	216	32	ОКП-1.1 ОКП-1.2 ОПК-1.4 ОПК-1.5 ОПК-1.11
Нижегородский государственный архитектурно-строительный университет (ННГАСУ)	180	16	ОКП-1.1 ОКП-1.2 ОПК-1.4 ОПК-1.5 ОПК-1.11
Томский государственный архитектурно-строительный университет (ТГАСУ)	180	26	ОКП-1.1 ОКП-1.2 ОПК-1.4 ОПК-1.5 ОПК-1.7

Данные по университетам из рабочих программ в таблицах близки. На изучение физики тратится большее количество часов, при этом на лабораторные работы приходится до 34 часов (см. табл. 1). Согласно РПД, лабораторные работы проводятся на стендах, без внедрения цифровых технологий. По дисциплине «Теоретическая механика» лабораторные работы отсутствуют (см. табл. 2).

Таблица 2. Показатели из рабочих программ дисциплины «Теоретическая механика» строительных вузов

Наименование вуза	Общее количество часов, ч.	Количество лабораторных, ч.	Получаемые компетенции
Астраханский государственный архитектурно-строительный университет (АГАСУ)	144	0	ОПК-1.1 ОПК-3.2 ОПК-6.11
Казанский государственный архитектурно-строительный университет (КГАСУ)	216	0	ОПК-6 ОПК-7
Национальный исследовательский Московский государственный строительный университет (НИУ МГСУ)	144	0	ОПК-1.1 ОПК-6.11
Нижегородский государственный архитектурно-строительный университет (ННГАСУ)	144	0	ОПК-1.1 ОПК-3.2 ОПК-6.11 УК-2.6
Томский государственный архитектурно-строительный университет (ТГАСУ)	144	0	ОПК-3.2 ОПК-1.9

Расшифровка кода компетенций, получаемых студентом в результате освоения программы дисциплин «Физика» и «Теоретическая механика» в строительных вузах, представлены в таблице 3.

Таблица 3. Код и наименование компетенции

Код	Наименование компетенции
<i>ОПК</i>	<i>Общепрофессиональные компетенции</i>
ОПК-1	Способность использовать основные законы естественнонаучных дисциплин в профессиональной деятельности, применять методы математического анализа и математического (компьютерного) моделирования, теоретического и экспериментального
ОПК-1.1	Выявление и классификация физических и химических процессов, протекающих на объекте профессиональной деятельности
ОПК-1.2	Определение характеристик физического процесса (явления), характерного для объектов профессиональной деятельности, на основе теоретического (экспериментального) исследования
ОПК-1.4	Представление базовых для профессиональной сферы физических процессов и явлений в виде математического(их) уравнения(й)
ОПК-1.5	Выбор базовых физических и химических законов для решения задач профессиональной деятельности
ОПК-1.7	Решение уравнений, описывающих основные физические процессы, с применением методов линейной алгебры и математического анализа
ОПК-1.9	Решение инженерно-геометрических задач графическими способами
ОПК-1.11	Определение характеристик процессов распределения, преобразования и использования электрической энергии в электрических цепях
ОПК-2	Способность выявить естественнонаучную сущность проблем, возникающих в ходе профессиональной деятельности, привлечь их для решения соответствующий физико-математический аппарат
ОПК-3.2	Выбор метода или методики решения задачи профессиональной деятельности
ОПК-6	Способность использовать основные законы естественнонаучных дисциплин в профессиональной деятельности, применять методы математического анализа и математического (компьютерного) моделирования, теоретического и экспериментального
ОПК-6.11	Составление расчётной схемы здания (сооружения), определение условий работы элемента строительных конструкций при восприятии внешних нагрузок
ОПК-7	Способность выявить естественнонаучную сущность проблем, возникающих в ходе профессиональной деятельности, привлечь их для решения соответствующий физико-математический аппарат
<i>УК</i>	<i>Универсальные компетенции</i>
УК-2.6	Составление последовательности (алгоритма) решения задачи

Совместный анализ РПД «Физика» и «Теоретическая механика» выявил существенный пробел в использовании цифровых технологий студентами первых-вторых курсов. Овладение решением задач методами математического (компьютерного) моделирования представлено только в ОПК КГАСУ, но при более внимательном изучении программы курса можно выделить только математическую составляющую этой компетенции в применимости к рассматриваемым фундаментальным дисциплинам.

В остальных изученных программах отсутствуют компетенции, посвященные цифровому моделированию или использованию специализированного программного обеспечения. Лабораторные работы, как правило, проводятся исключительно на реальных экспериментальных стендах либо же сводятся к упрощенным виртуальным аналогам, где от студента требуется лишь элементарное взаимодействие, например нажатие одной кнопки. Такой подход не способствует развитию глубокого понимания физических явлений и ограничивает возможности для формирования необходимых компетенций – студенты не получают достаточного опыта работы с современными программными пакетами для численного моделирования и обработки данных.

### Обсуждение результатов

Цифровые технологии могут значительно улучшить преподавание фундаментальных физико-математических дисциплин, таких как физика и теоретическая механика. Их внедрение на ранних этапах обучения, особенно на уровне лабораторных работ, позволит студентам освоить методы моделирования физических явлений. ЦП, выступая в роли виртуальных экспериментов, помогает изучать процессы, дополняя реальные лабораторные установки, при этом обучая основам компьютерного моделирования. Например, в молекулярной физике, помимо традиционного изучения давления в газах, можно исследовать аэродинамическое обтекание как реальный физический процесс, полезный в инженерной практике (рис. 1).

ЦП может быть адаптирована под нужды конкретных направлений и специальностей, в строительных вузах – технических инженеров. Создание ЦД реальных лабораторных стендов позволит не только проводить эксперименты как в физическом, так и в виртуальном пространстве, но получить навыки цифрового проектирования обучающимся (рис. 2).

В курсе теоретической механики (раздел «Кинематика») решаются элементарные задачи движения материальной точки, такие как траектории тел, брошенных под углом к горизонту. Такого плана задачи идеально подходят для освоения основ ЦП исследований, поскольку результаты их компьютерного моделирования легко проверяются «на бумаге».

Такой поэтапный подход включения в цифровую среду, начиная с простых опытов, готовит будущих специалистов к моделированию в дальнейшем сложных физических процессов.

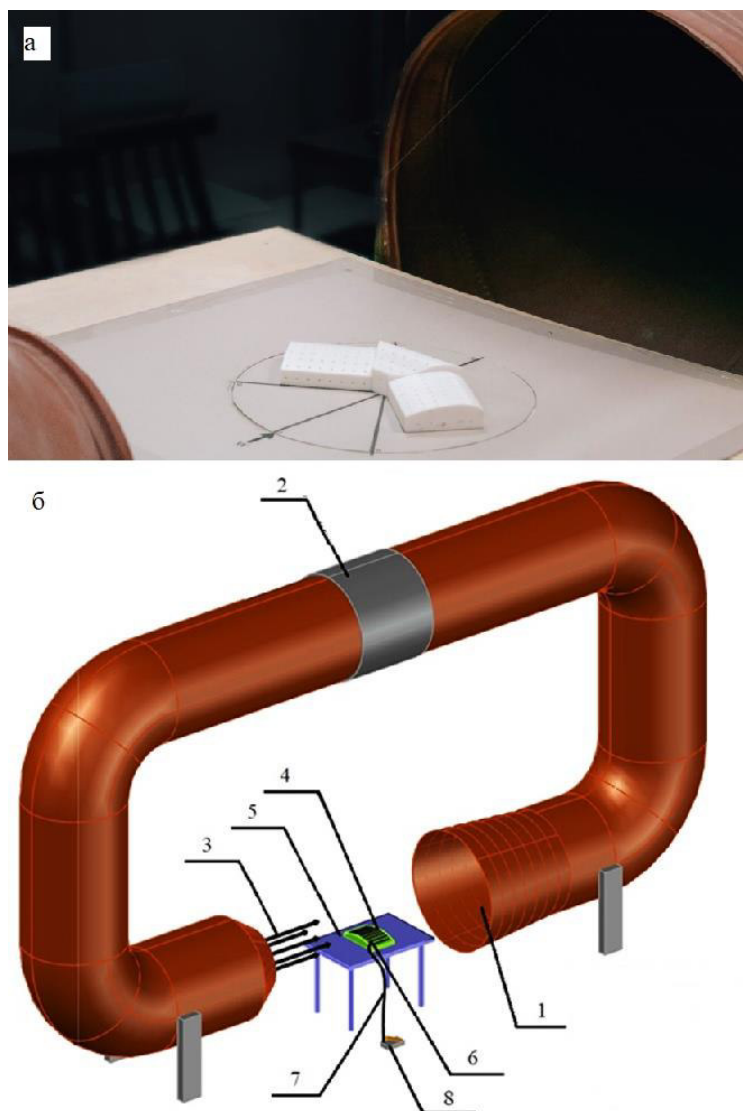


Рисунок 1. Пример проведения аэродинамических параллельных физических (а) и цифровых экспериментальных (б) лабораторных работ

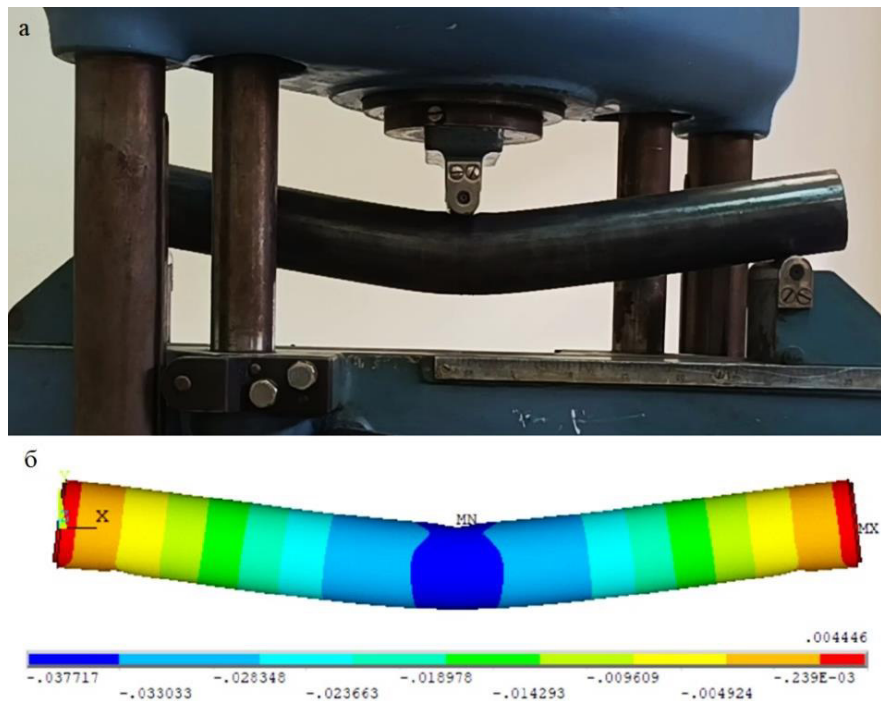


Рисунок 2. Пример проведения параллельных физических (а) и цифровых экспериментальных (б) лабораторных работ по сопротивлению материалов

### Выводы

На основе анализа компетенций и рабочих программ по физике и теоретической механике сделано заключение о необходимости цифровизации образовательного процесса по фундаментальным физико-математическим наукам. Отсутствие адекватной цифровой поддержки научных исследований на ранних курсах обучения является серьезным недостатком, препятствующим формированию конкурентоспособных специалистов. Внедрение цифровых инструментов может быть оптимально реализовано в первую очередь путем разработки и внедрения интерактивных виртуальных лабораторий, позволяющих студентам самостоятельно моделировать физические явления, изменять параметры экспериментов и анализировать результаты в режиме реального времени, что позволит глубже вникать в физический процесс происходящего эксперимента.

Интеграция ЦП на начальных этапах обучения способствует повышению качества образовательного процесса и готовит студентов к эффективному решению реальных инженерных задач с применением современных технологий. Несмотря на то что задания по компьютерному проектированию широко распространены, начинать их освоение целесообразно раньше, внедряя цифровые технологии в базовые дисциплины, традиционно решаемые на бумаге. Это обеспечит постепенный переход к активному использованию методов ЦП именно тогда, когда оно станет необходимым компонентом специализированных технических (строительных) курсов.

Успешное введение цифрового обучения продемонстрировано на примере графических предметов, где параллельно изучаются традиционное ручное черчение и работа с системами автоматизированного проектирования (такими как Компас, AutoCAD).

### Список литературы

1. Февральских А. В. Компьютерное моделирование аэродинамики пропеллера в российском пакете программ Flowvision // Русский инженер. 2024. № 4 (85). С. 32–35.
2. Ведяйкина О. И., Хазов П. А., Шилов С. С. Алгоритм цифровой поддержки работоспособности малогабаритных аэродинамических экспериментальных установок // Вестник Южно-Уральского государственного университета. Серия: Строительство и архитектура. 2025. Т. 25, № 2. С. 62–71.
3. Сапелкин Р. И., Матренинский С. И., Мищенко В. Я. Цифровой двойник процесса вулканизации материала защитного покрытия из каучукового бетона на поверхности трубного металла // Научный журнал строительства и архитектуры. 2025. № 1 (77). С. 102–110.



4. Рыбкин С. В. Анализ возможностей платформы Delta Design для проектирования интегральных микросхем // Наноиндустрия. 2021. Т. 14, № S7 (107). С. 273–274.
5. Grinshkun V. V., Itinson K. S. Learning digital medical environment as a tool of teaching computer science to medical students // RUDN Journal of Informatization in Education. 2024. Vol. 21, iss. 2. Pp. 169–180.
6. Разработка цифрового двойника туннельной печи / И. В. Салов, И. А. Щербатов, Ю. А. Салова и др. // International Journal of Open Information Technologies. 2022. Т. 10, № 12. С. 102–109.
7. Машиньян А. А., Кочергина Н. В., Бирюкова О. В. Образовательная среда по общей физике в техническом университете // Мир науки. Педагогика и психология. 2022. Т. 10, № 4. URL: <https://mir-nauki.com/PDF/35PDMN422.pdf>
8. Нагимуллина С. С. Интеграция систем искусственного интеллекта в лабораторный практикум по физике: опыт и перспективы // Вестник педагогических наук. 2025. № 4. С. 318–324.
9. Рабочая программа дисциплины «Теоретическая механика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Квалификация выпускника – бакалавр. Форма обучения – очная. Астрахань: АГАСУ, 2022.
10. Рабочая программа дисциплины «Теоретическая механика» по специальности 08.05.01 «Строительство уникальных зданий и сооружений». Квалификация выпускника – инженер-строитель. Форма обучения – очная. Казань: КазГАСУ, 2018.
11. Рабочая программа дисциплины «Теоретическая механика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Уровень образования – бакалавриат. Форма обучения – очная. Москва: МГСУ, 2023.
12. Рабочая программа дисциплины «Теоретическая механика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Уровень образования – бакалавриат. Форма обучения – очная. Нижний Новгород: ННГАСУ, 2024.
13. Рабочая программа дисциплины «Теоретическая механика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Квалификация выпускника – бакалавр. Форма обучения – очная. Томск: ТГАСУ, 2021.
14. Рабочая программа дисциплины «Физика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Квалификация выпускника – бакалавр. Форма обучения – очная. Астрахань: АГАСУ, 2022.
15. Рабочая программа дисциплины «Физика» по специальности 08.03.01 «Строительство». Профиль – все профили. Квалификация выпускника – бакалавр. Форма обучения – очная. Казань: КазГАСУ, 2018.
16. Рабочая программа дисциплины «Физика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Уровень образования – бакалавриат. Форма обучения – очная. Москва: МГСУ, 2023.
17. Рабочая программа дисциплины «Физика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Уровень образования – бакалавриат. Форма обучения – очная. Нижний Новгород: ННГАСУ, 2024.
18. Рабочая программа дисциплины «Физика» по направлению подготовки 08.03.01 «Строительство». Профиль «Промышленное и гражданское строительство». Квалификация выпускника – бакалавр. Форма обучения – очная. Томск: ТГАСУ, 2021.

УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-1070-1075

## Анализ кластеров как инструмент в управлении образовательной деятельностью

Д. Ф. Кучкарова<sup>1</sup>, Д. А. Ачилова<sup>2</sup>

<sup>1</sup> Национальный исследовательский университет «Ташкентский институт инженеров ирригации и механизации сельского хозяйства», Ташкент, Узбекистан

<sup>2</sup> Ташкентский международный университет Кимё, Ташкент, Узбекистан

**Аннотация.** Статья посвящена вопросам управления качеством образовательного процесса в высших учебных заведениях с акцентом на систематический мониторинг мотивационных установок и уровня предметной подготовки студентов. Особое внимание уделено проблеме адаптации первокурсников к академической среде и необходимости разработки индивидуализированных образовательных траекторий. В качестве инструмента диагностики предложена анкетно-тестовая методика, включающая задания по инженерной графике и вопросы, направленные на выявление доминирующих типов мотивации. Полученные данные служат основой для кластеризации студентов по параметрам мотивации и знаний, что позволяет оптимизировать управление учебным процессом и способствует повышению его эффективности и качества.

**Ключевые слова:** кластерный анализ, мотивация, анкета-опросник, корреляция

## Cluster analysis as a tool in educational management

D. F. Kuchkarova<sup>1</sup>, D. A. Achilova<sup>2</sup>

<sup>1</sup> "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers"  
National Research University, Tashkent, Uzbekistan

<sup>2</sup> Kimyo International University in Tashkent, Tashkent, Uzbekistan

**Abstract.** The article addresses the issues of quality management in the educational process within higher education institutions, with a particular focus on the systematic monitoring of students' motivational orientations and subject-specific knowledge levels. Special attention is given to the challenges of first-year students' adaptation to the academic environment and the need for developing individualized learning trajectories. As a diagnostic tool, a questionnaire-based methodology is proposed, incorporating test items in engineering graphics and questions aimed at identifying dominant types of motivation. The data obtained serve as a basis for clustering students according to their motivation and knowledge parameters, thereby enabling more effective educational process management and contributing to the enhancement of its efficiency and overall quality.

**Keywords:** cluster analysis, motivation, questionnaire, correlation.

### Введение

Качество образования в высших учебных заведениях представляет собой комплексную характеристику, отражающую степень соответствия образовательных результатов установленным нормативным стандартам, общественным ожиданиям и запросам обучающихся. В зависимости от уровня интерпретации, а также варьирующихся требований и индикаторов данное понятие может служить ориентиром при оценке эффективности системы высшего образования.

Общее качество образовательного процесса в университете формируется на основе деятельности отдельных сотрудников и структурных подразделений, а также уровня обеспеченности ресурсами и условий функционирования образовательной организации. При этом высокая результативность отдельных компонентов системы не всегда гарантирует достижение целостного качественного результата.

Итоговое качество образования находит отражение в уровне подготовки выпускника, включающем совокупность знаний, умений и компетенций, необходимых для выполнения профессиональных функций в соответствии с нормативными и ожидаемыми требованиями.

Система управления качеством образования в высшем учебном заведении охватывает широкий спектр функций и взаимодействий между структурными элементами, эффективность которых определяет вариативность возможных образовательных результатов.

Одним из ключевых компонентов управления образовательным процессом в высшем учебном заведении является систематическое наблюдение за учебной активностью студентов на протяжении всего периода обучения. Начальный этап освоения образовательной программы представляет собой критически важную фазу, сопряжённую с адаптацией студентов к новым академическим условиям и требованиям.

Эффективное управление учебным процессом предполагает наличие диагностического инструментария, обеспечивающего возможность комплексной оценки его параметров и оперативного реагирования на выявленные затруднения. Важным этапом разработки такого инструментария является определение исходного уровня знаний и мотивационных установок студентов на начальной стадии их обучения.

Современная модель высшего образования характеризуется акцентом на индивидуализацию учебных траекторий, расширением возможностей для самообразования, а также разнообразием содержания и методических подходов, обеспечивающих гибкость и демократичность образовательной среды.

Мотивационная структура обучающегося, включающая как внутренние, так и внешние компоненты, формирует его личностную направленность и определяет стратегию образовательной активности, влияя на глубину и устойчивость учебной вовлечённости.

Мотивационную сферу студента, на которую влияют убеждения, взгляды, интересы, приоритеты, уровень интеллекта, окружение, можно разделить по разным критериям оценки:

- профессиональные мотивы (стать квалифицированным специалистом для успешной будущей деятельности);
- мотивы избегания (нежелание отставать от товарищей, избежать осуждения и наказания за плохую учёбу);
- мотивы престижа (добиться одобрения окружающих, иметь высокий рейтинг среди студентов и преподавателей);
- коммуникативные мотивы (иметь широкий круг знакомых, обрести полезные связи);
- мотивы внешнего воздействия (требования родителей получить высшее образование, влияние окружения).

Мониторинг двух ключевых компонентов учебной деятельности студентов – мотивационной направленности и структурного уровня знаний – позволяет формировать индивидуализированную образовательную траекторию в процессе обучения в высшем учебном заведении, а также выбирать оптимальные стратегии её управления. Взаимодействие указанных компонентов носит динамический характер и определяется множеством факторов, включая социально-культурные условия и индивидуально-психологические особенности обучающихся [1, 2].

Мотивация представляет собой многокомпонентную систему, в которой основными формами выступают внутренняя мотивация (ориентированная на процесс и результат) и внешняя мотивация (связанная с поощрением или избеганием). Внешняя мотивация функционирует как инструмент достижения цели, тогда как внутренняя отражает механизмы саморегуляции и личностного контроля [3].

### **Постановка задачи**

В рамках исследования авторами разработан диагностический опросник, включающий задания, направленные на выявление уровня и структуры знаний по инженерной графике, а также вопросы, позволяющие определить преобладающие типы мотивации студентов на начальном этапе их обучения в вузе.

Предлагается в первую неделю учёбы студентам первых курсов.

Анкета-опросник представляет собой 40 заданий тестового характера, включающим задание на решение проекционных и конструктивных задач инженерной графики и определение типа мотивации. В результате каждая академическая группа студентов условно делится на подгруппы (кластеры), где каждый кластер объединяет студентов с близкими по уровню знаниями и типам мотивации.

Авторы поставили перед собой цель – выявить взаимосвязь между типом мотивации, первоначальным уровнем знаний и трансформацией этих параметров в процессе обучения в вузе.

### **Теория**

Были выбраны следующие три группы мотивации:

- 1) стать квалифицированным специалистом;
- 2) иметь высокую зарплату;
- 3) получить диплом;

- 4) требования родителей получить высшее образование;
- 5) не отставать от товарищей;
- 6) избежать осуждения близкого окружения;
- 7) познакомиться со студенческой жизнью;
- 8) иметь широкий круг знакомых для дальнейшего общения;
- 9) советы знакомых и близких;
- 10) добиться одобрения близкого окружения.

В анкете-опроснике также имеются вопросы по таким разделам инженерной графики, как определение нужной проекции предмета, изометрической проекции, графические задания на логику и конструктивные задачи по программе средней школы.

Каждый правильный ответ по тестам инженерной графики оценивается в 1/3 балла.

Для определения типа мотивации и определения уровня и структуры знаний строится два графика в декартовой системе координат.

На рисунке 1 представлены результаты определения типа мотивации в академической группе из 10 студентов (количество студентов может быть произвольным и типов мотивации может быть больше трёх групп).

Каждый студент имеет свой порядковый номер. На рисунке 1 отмечены точками соответствие порядкового номера студента и номера мотивации.

Академическая группа в результате анкеты-опроса разделилась на несколько кластеров по типу мотивации и знаниям предмета. На следующем этапе эти кластеры путём анализа состава перераспределяются на новые кластеры, которые объединяются по близкому уровню знаний и мотивации.

Мотивация: 1 кластер [V, II, IX], 2 кластер [I, VIII, IV], 3 кластер [III, VI, VII, X] (рис. 2).

Знания: 1 кластер [V, IX, III, VI], 2 кластер [I, III, VIII, IV], 3 кластер [VII, X] (рис. 3, 4).

Объединяем кластеры с одновременным учётом совпадающих значений мотивации и знаний каждого студента.

Мотивация+знания: 1 кластер [V, IX], 2 кластер [I, VIII, IV], 3 кластер [VII, X], 4 кластер [II, III, VI] (рис. 5).

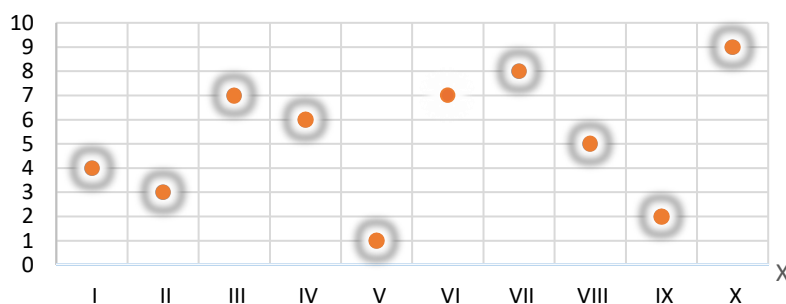


Рисунок 1. Каждому студенту соответствует его тип мотивации

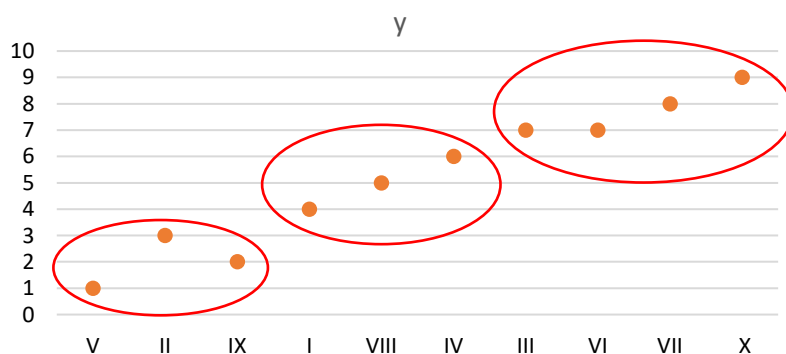


Рисунок 2. Разделение студентов на кластеры по типу мотивации

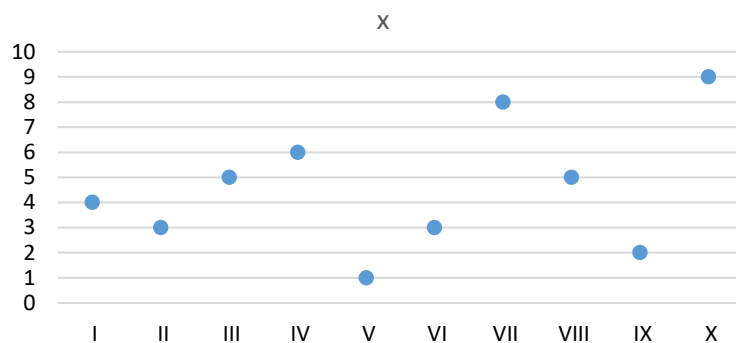


Рисунок 3. Каждому студенту соответствует его уровень знаний

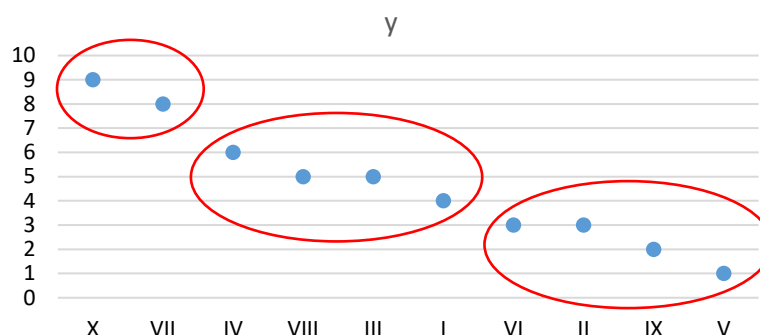


Рисунок 4. Разделение студентов на кластеры по знаниям предмета

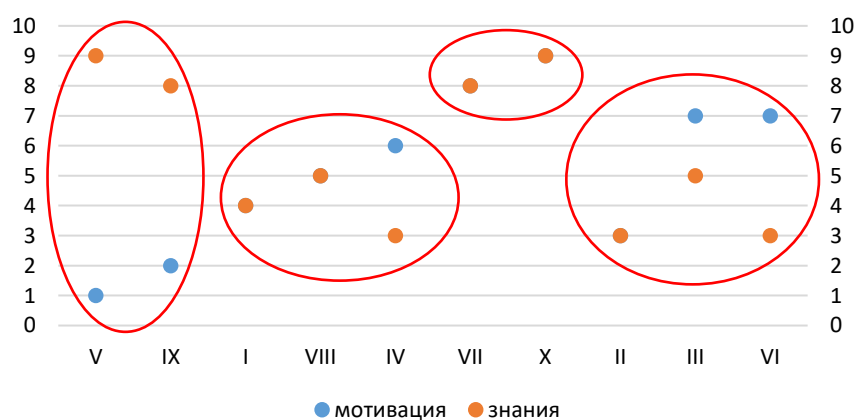


Рисунок 5. Разделение на кластеры по двум факторам (мотивация+знание)

В течение семестра в процессе адаптации студентов к академической среде, под влиянием различных внутренних и внешних факторов происходят изменения как в мотивационной структуре, так и в уровне и содержании усваиваемых знаний. Данные, полученные в ходе текущего, промежуточного и итогового контроля, позволяют сформировать целостное представление о динамике образовательных достижений студентов в условиях высшего учебного заведения [4].

Очевидно, что на результаты обучения оказывает влияние широкий спектр факторов, включая личностные особенности, учебную среду и организационные условия. При допущении, что исходная мотивация студентов остаётся стабильной на протяжении семестра, результаты итогового контроля могут служить основанием для выявления корреляции между начальной мотивационной установкой и достигнутыми учебными результатами.

В таблице представлены нормированные оценки промежуточного и итогового контроля для ранее рассмотренной группы из 10 студентов, выраженные по 10-балльной шкале.

## Нормированные оценки промежуточного и итогового контроля

Студенты	Результаты первого рубежного контроля	Результаты второго рубежного контроля	Результаты итогового контроля	Сумма баллов
I	6	6	5	17
II	7	6	5	18
III	5	5	6	16
IV	4	4	6	14
V	8	7	7	22
VI	2	5	4	11
VII	7	7	6	20
VIII	7	5	5	17
IX	8	7	8	23
X	6	6	5	17

Для обработки результатов контроля знаний применялись два метода – дисперсионный однофакторный анализ по критерию F (Фишера) для связанных выборок и метод  $\chi^2$  Фридмана.

Были сформулированы гипотезы:

**H<sub>0</sub>** – индивидуальные различия между студентами являются не более выраженными, чем различия, обусловленные случайными причинами;

**H<sub>1</sub>** – индивидуальные различия между студентами являются более выраженными, чем различия, обусловленные случайными причинами [5].

## Результаты экспериментов

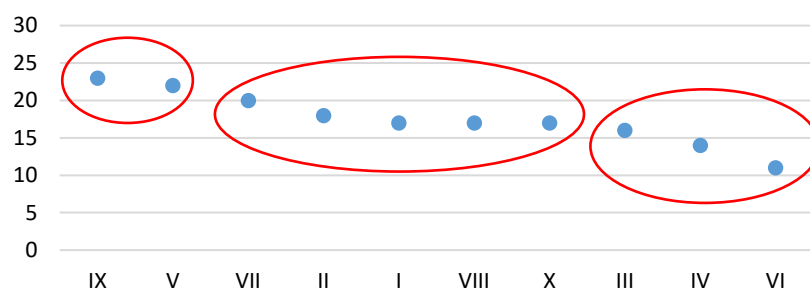


Рисунок 6. Кластеры по общему баллу в конце семестра

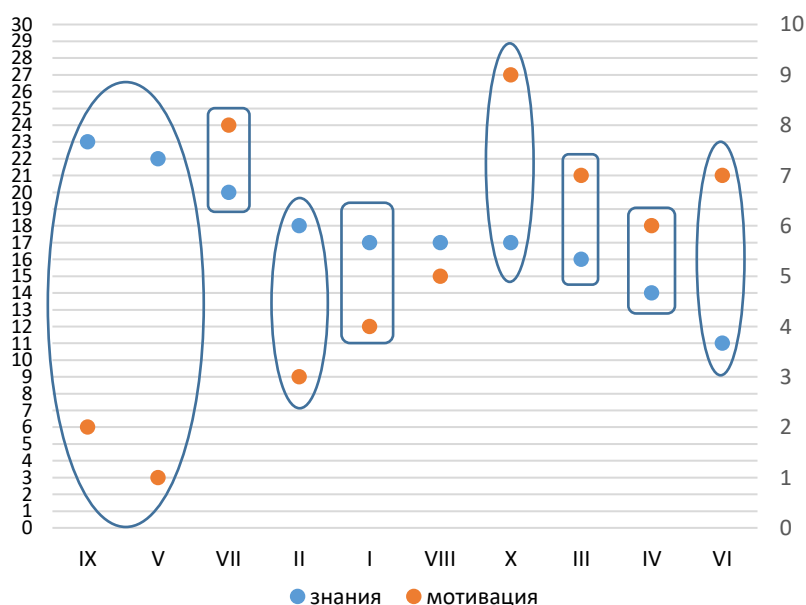


Рисунок 7. Новые кластеры мотивация+знания

По результатам применения двух методов можно сделать вывод, что студенты 1-го кластера мотивации проявляют большую интеллектуальную настойчивость по сравнению с остальными участниками. Также результаты контроля показывают, что студенты 2-го кластера проявляют среднюю интеллектуальную настойчивость, не оправдывая первоначального ожидания. Гипотеза  $H_0$  подтверждается.

Студенты 3-го кластера показали более высокие баллы по сравнению со студентами 2-го кластера. В 4-м кластере студенты набрали средние баллы и лишь результаты одного студента оказались ниже среднего баллов. Гипотеза  $H_1$  не подтверждается.

### Заключение

Результаты проведённых исследований показывают, что два основных фактора – мотивация и уровень знаний студентов первокурсников – являются динамично изменяющимися и взаимодополняемыми. Для общих выводов необходимо иметь большую репрезентативную выборку.

Для конкретной вышеприведённой группы студентов можно сделать вывод о том, что задачей преподавателя, имеющего данные об уровне знаний и типе мотивации, является разработка методики индивидуального личностно ориентированного подхода.

Учитывая специфику каждой дисциплины в вузе, следует предлагать студентам дифференцированные задания, а также вовлекать наиболее мотивированных студентов в различные проекты, ориентированные на будущие специальности.

### Список литературы

1. Дюран Б. Кластерный анализ. М.: Книга по Требованию, 2012. 128 с.
2. Бердникова Д. В., Виндекер О. С. Исследование взаимосвязи мотивации достижения с учебной мотивацией и другими личностными характеристиками // Психологический вестник Уральского федерального университета. 2013. Вып. 10. С. 6–14.
3. Виндекер О. С., Луканин М. Е. Исследование структуры мотивации достижения у студентов в контексте их представлений об интеллекте и целях обучения // Известия Уральского федерального университета. Сер. 1: Проблемы образования, науки и культуры. 2020. Т. 26, № 2 (197). С. 111–123.
4. Гордеева Т. О., Сычев О. А. Мотивационные профили как предикторы саморегуляции и академической успешности студентов // Вестник Московского университета. Серия 14: Психология. 2017. № 1. С. 67–87.
5. Kuchkarova D., Achilova D. Quality Management of Teaching Graphic Disciplines in a Technical University // ICGG 2024 – Proceedings of the 21st International Conference on Geometry and Graphics, Kitakyushu, Japan, August 5–9, 2024. Springer Cham, 2024. Vol. 1. Pp. 207–214.



УДК 378:004

DOI: 10.25686/978-5-8158-2474-4-2025-1076-1082

## Историческое наследие как компонент современной информационно-образовательной платформы

Е. Е. Швейгерт, Е. В. Князева

Санкт-Петербургский политехнический университет Петра Великого,  
Санкт-Петербург, Россия

**Аннотация.** В целях определения возможностей использования исторического подхода в создании дизайна цифровых продуктов для образовательных учреждений, авторами было проведено исследование, результаты которого представлены в статье.

**Ключевые слова:** образовательная платформа, цифровизация образования, культурно-исторический контент, образовательный контент.

## History as a component of a modern information and educational platform

E. E. Schweigert, E. V. Knyazeva

Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia

**Abstract.** In order to determine the possibilities of using a historical approach in creating the design of digital products for educational institutions, the authors conducted a study, the results of which are presented in this article.

**Keywords:** educational platform, digitalization of education, cultural and historical content, educational content.

### Введение

Глобальная цифровизация является основной тенденцией последних лет. Она коснулась многих сфер жизнедеятельности человека, в том числе образования. В 2019 году был запущен федеральный проект «Цифровая образовательная среда», цель которого оптимизировать образовательный процесс, обновить информационно-коммуникационную инфраструктуру [1].

Сайт образовательного учреждения является основной площадкой для представления учреждения в цифровом пространстве. Впечатление, которое он производит на абитуриентов и их родителей, оказывает прямое влияние на формирование имиджа училища. Помимо этого, сайт решает задачи, связанные с коммуникацией со студентами и преподавателями и задачи, связанные с организацией образовательного процесса.

В рамках исследования была изучена информация об истории училища, которая находится в свободном доступе. Исследование истории художественного училища имени Н. К. Рериха и составление исторического обзора было основано на нескольких ключевых источниках. Основную информацию об этапах становления и развития учреждения удалось получить с официального сайта училища, где представлены краткие исторические справки, перечень прежних названий, а также сведения о значимых личностях, связанных с учебным заведением [2]. Существенное расширение и углубление исторического контекста даёт книга «Художественной училище имени Николая Рериха. Страницы истории» (автор Э. М. Романовская), в которой детально описаны периоды преобразований, педагогическая деятельность, а также роль Н. К. Рериха и других директоров, сделавших существенный вклад в развитии училища [3].

### Постановка задачи

Цель исследования – создать сайт, посвященный истории Санкт-Петербургского художественного училища имени Н. К. Рериха. Наличие на сайте раздела, посвященного истории создания училища, связано с решением нескольких задач.

Во-первых, это необходимо, чтобы сохранить историю училища и рассказать ее новому поколению, представив в современной визуальной форме. Во-вторых, проанализировав сайты учреждений высшего и среднего специального образования, мы обнаружили, что большинство из них придает особое значение истории учреждения. Это может быть связано не только с принципиальной позицией руководства – рассказывать историю учреждения. Заведения с богатой историей могут использовать это как преимущество в конкурентной борьбе за абитуриентов.

## Теория

Выделено несколько основных этапов, в которые с учреждением произошли сильные изменения. За всю историю существования училище сменило много названий, несколько раз менялось главное здание. Каждый новый руководитель вносил что-то новое в организацию образовательного процесса. Все эти изменения неизбежно оставляли след в идеологии училища.

Училище было основано в 1839 году в соответствии с Указом Государя Императора Николая Первого. Открытие в Петербурге специальной Рисовальной школы для вольноприходящих учеников связано с именами «русских немцев» – Е. Ф. Канкрин, на тот момент министра финансов Российской империи, и К. Х. Рейссига, видного ученого, которого также следует назвать теоретиком и организатором художественного образования. Организации Рисовальной школы предшествовало появление нескольких учебных классов, возникновение которых также инициировали Канкрин и Рейссиг. «Рисовальная Школа для вольноприходящих» или «Рисовальная Школа на Бирже», как её называли в соответствии с местом расположения на Стрелке Васильевского Острова, предоставляя бесплатное художественное образование для всех желающих [3–5].



Рисунок 1. Основатели Рисовальной школы для вольноприходящих К. Х. Рейссиг и Е. Ф. Канкрин [6, 7]

Особого внимания заслуживают начинания К. Х. Рейссига, относящиеся к первым годам существования школы: открытие класса гальванопластики и женских рисовальных классов. Класс гальванопластики, как представляется сегодня, явился своего рода техническим воплощением просветительской идеи единства научного и художественного познания мира.

В 1878 году школа переехала в здание Общества поощрения художеств на Большой Морской улице, 38, где продолжила свою деятельность до 1918 года. Особую роль в становлении и развитии учреждения сыграл Николай Константинович Рерих – выдающийся художник, философ, общественный деятель. В период с 1906 по 1916 годы он возглавлял Школу Императорского общества поощрения художеств, придав ей новое направление и гуманистическую философию, основанную на идее народного просвещения [4].

Рерих налаживает работу мастерских по фарфору и фаянсу, керамике, резьбе по дереву и создает новые мастерские: иконописную, чеканки, рукодельную, ткацкую. Организует классы графики, композиции, которую преподает сам, рисования животных, медальерного искусства и многие другие [8].

После революции 1917 года школа продолжает работать как бесплатные «Курсы рисования и черчения», находящиеся на Литейном проспекте. В конце 1918 года произошла реорганизация учебного заведения, в результате которой курсы были трансформированы в «Художественную школу-мастерскую». После переезда на Таврическую улицу, учебное заведение получило неофициальное название «Таврическое художественное училище».

Ощутимый вклад в развитие училища сделал Ян Константинович Шабловский. В 1934 году он совершил поистине героический поступок, взявшись за возрождение бывшей Школы Императорского общества поощрения художеств, которая к тому времени практически прекратила свою деятельность.

Особого внимания заслуживает деятельность учебного заведения в годы Великой Отечественной войны. Даже в условиях блокады Ленинграда «Школа Шабловского» продолжала работать и выполнять свою образовательную миссию. В предвоенном учебном году в училище обучалось более двухсот тридцати студентов. В тот момент оно находилось на Таврической улице. Зимой 1942 года училище было «законсервировано», но занятия не прекращались полностью. В ограниченном режиме, но училище продолжало работать, об этом свидетельствуют записи Елены Оскаровны Мартилла, советской и российской художницы, которая поступила в училище на Таврической в ноябре 1941 года [9, 10]. Летом 1942 года занятия в училище прекратились, но уже осенью в 1943 году был объявлен новый набор учащихся.

Значительным событием в истории училища стало присвоение ему имени выдающегося деятеля искусства Владимира Александровича Серова. Это решение было принято Верховным Советом Советского Союза в 1968 году [11].

В 1992 году по приказу Комитета по культуре мэрии Санкт-Петербурга учебное заведение было переименовано в «Санкт-Петербургское художественное училище имени Н. К. Рериха». В настоящее время училище предлагает обучение на отделениях живописи, скульптуры, дизайна и реставрации. Логотип Санкт-Петербургского художественного училища имени Н. К. Рериха был утвержден, когда учебное заведение получило своё современное название.

Художественное училище имени Н. К. Рериха имеет богатую историю, восходящую к Школе Императорского общества поощрения художников. Училище оказало значительное влияние на развитие отечественного и мирового искусства – здесь учились или преподавали выдающиеся художники, среди которых И. Репин, М. Шагал, Ф. Васильев и многие другие известные личности.



Рисунок 2. Хронология изменения расположения училища

Помимо информации о различных преобразованиях в училище, расположенных в хронологическом порядке, на сайте с исторической справкой об училище должен быть раздел, посвященный известным личностям, – художникам, деятелям культуры, которые являлись студентами или преподавателями, то есть имели прямую связь с учреждением.

В разные годы в училище преподавали такие признанные деятели искусства, как И. Я. Билибин – художник и график, известный благодаря своим иллюстрациям к русскому фольклору, И. Н. Крамской – русский живописец и художественный критик, П. П. Чистяков – русский живописец и график, А. В. Шусев – русский и советский архитектор, П. К. Клодт – русский скульптор немецкого происхождения, П. А. Мансуров – живописец, график и стенограф, один из представителей русского авангарда [13, 14]. Под руководством этих выдающихся педагогов в стенах училища прошли становление многие художники, чьи имена впоследствии вошли в историю отечественного и мирового

искусства. Среди них пейзажист Фёдор Васильев, известный тонкостью цветовых решений и глубиной настроения в работах [14]. Василий Верещагин, посвятивший творчество военной теме. Илья Репин, оказавший значительное влияние на развитие русской реалистической школы. В числе учеников также был Марк Шагал, оставивший заметный след в искусстве 20-го века [15]. В архиве библиотеки училища до сих пор хранится мужской портрет, выполненный Шагалом.

### Результаты исследования

Собранный материал был проанализирован. Ориентируясь на историческую справку, был сделан вывод о том, что дизайн страницы сайта может содержать отсылки к знаковым периодам в истории училища. Для формирования совокупности визуальных характеристик, из которых будет складываться образ училища в восприятии абитуриентов, были проанализированы здания, в которых располагалось училище, а также элементы фирменного стиля: логотип, фирменные цвета, рекламно-информационные материалы. Особое внимание было уделено формам, декоративным элементам, цветовым решениям. Это позволило выделить характерные особенности, которые можно использовать в дизайне сайта.

Училище начало свой путь в здании Императорского общества поощрения художеств. Оно имеет черты, которые предвосхитили появление такого стиля, как модерн. Например, ассиметричный фасад, декоративные элементы, мозаичные украшения [16]. Доходный дом Дернова на Таврической улице, в котором училище располагалось в 1934–1961 годах, имеет черты, характерные для поздней эклектики и модерна [17]. Особняк на улице Пролетарской диктатуры стал домом для художественного училища в 1961 году. Это здание было построено в 1902–1903 годах по проекту архитектора Василия Константиновича Вейса, который работал в стилях модерна и эклектики.

Логотип училища представляет собой стилизованную заглавную букву «Р», совмещенную с декоративным элементом, визуально напоминающим орнамент в стиле архитектуры модерна, – характерными скругленными, природными линиями. Логотип – это графическое изображение или символ, который представляет собой уникальную идентификацию бренда, компании или продукта. Символически логотип отсылает к имени самого Н. К. Рериха, имя которого было присвоено училищу в 1992 году. Элемент, напоминающий архитектурную деталь, может быть интерпретирован как отсылка к классическому академическому образованию. Сопоставление логотипа и архитектурного элемента представлено на рисунке 3.



Рисунок 3. Логотип училища имени Н. К. Рериха [2]

Целевой аудиторией сайта являются абитуриенты, студенты, а также просто люди, интересующиеся историей искусства, культурным наследием и выставочной деятельностью училища.

Чтобы понять, как пользователь взаимодействует с таким ресурсом, можно рассмотреть типичный путь заинтересованного посетителя – от первого контакта с информацией до формирования устойчивой лояльности и вовлеченности. Для этого была разработана карта пользовательского пути (Customer Journey Map, CJM). Это важный маркетинговый инструмент, который визуализирует весь процесс взаимодействия пользователя с продуктом или услугой [18]. В контексте исследования



продуктом является сайт. В основе карты пользовательского пути лежит детальный анализ поведения пользователя, его эмоций, потребностей и препятствий на каждом этапе взаимодействия. Этот инструмент позволяет рассмотреть процесс с позиции пользователя и выявить ключевые моменты, влияющие на его решение. При создании цифровых продуктов инструмент играет важную роль, потому что помогает разработчикам и дизайнерам понять, как пользователи будут взаимодействовать с сайтом, приложением, платформой.

На первом этапе пользователь случайно сталкивается с упоминанием училища – через публикации в социальных сетях, афиши или разговоры. У него возникает любопытство: что это за учреждение, чем оно отличается от остальных? Этот интерес приводит к следующему шагу – пользователь начинает поиск информации, вводит в поисковую строку запросы вроде «история художественного училища Рериха» или «выставки СПб художественное училище». Полная карта пути пользователя представлена на рисунке 4.



Рисунок 4. Карта пользовательского пути

Таким образом, сбор информации об истории училища был необходим в рамках работы над сайтом, потому что в ходе составления краткого исторического обзора были выявлены ключевые события (изменение названия и смена руководства), а также известные личности, которые напрямую или косвенно связаны с училищем. Полученные данные будут использованы при наполнении сайта.

По результатам визуального анализа было принято решение использовать в оформлении страницы тонкие декоративные элементы, шрифты с засечками в заглавном тексте и добавить в систему шрифтов курсивный в качестве дополнительного. Такое решение позволило сохранить связь с сайтом училища.

Варианты цветовых решений и шрифтов представлены на рисунке 5.



Рисунок 5. Варианты цветовых решений и шрифтовых пар

Были разработаны два варианта цветовых решений. В первом преобладает черный, желто-коричневый для выделения важных элементов и кнопок. Такой выбор цветов создает сдержанную и сосредоточенную атмосферу, которая подходит для официальной подачи.

Во втором варианте в качестве доминирующего выступает кремовый. Дополняющие цвета – оксид красный и черновато-красный – используются точно: для выделения ключевых элементов, заголовков и навигации. Такое сочетание придаёт выразительность и динамику, не нарушая общей сдержанности.

### Выводы и заключение

Полученные в ходе анализа данные о целевой аудитории и пользовательском поведении будут использованы при создании сайта. Поскольку основная аудитория ориентирована на быстрое получение информации через цифровые каналы и склонна к визуальному восприятию, оптимальным решением становится создание одностраничного сайта. Такой формат обеспечивает простую и логичную навигацию, позволяет компактно разместить ключевые блоки: историю училища, галерею архивных фото, биографии выдающихся выпускников.

Исторически сложившиеся визуальные и смысловые символы, такие как герб, архитектурный стиль зданий, имя учредителя, традиционные мероприятия, могут быть переосмыслены и включены в цифровую оболочку сайта как элементы визуального наследия. Их использование способствует не только эстетической целостности интерфейса, но и повышает узнаваемость бренда учреждения.

### Список литературы

1. Федеральный проект «Цифровая образовательная среда». URL: <https://edu.gov.ru/national-project/projects/cos/> (дата обращения: 04.09.2024).
2. Санкт-Петербургское художественное училище имени Н. К Рериха: официальный сайт. URL: <https://www.rerihspbhu.ru> (дата обращения: 01.06.2025).
3. Романовская Э. М. Художественное училище имени Николая Рериха: страницы истории. СПб.: Лань, 2001. 192 с.
4. Боровская Е. А. К 200-летию Императорского общества поощрения художеств. Художественные мастерские Школы ИОПХ // Научные труды Санкт-Петербургской академии художеств. 2023. № 64.
5. Боровская, Е. А. Художественное образование в Петербурге: традиции и современность. Рисовальная школа ОПХ / Е. А. Боровская // Научные труды Санкт-Петербургской академии художеств. – 2006. – № 3 (дата обращения: 07.06.2025);
6. Von Reissig, Cornelius August: электронное изображение // Wiki2: сайт. URL: [https://wiki2.org/ru/Файл:Von\\_Reissig,\\_Cornelius\\_August.jpg](https://wiki2.org/ru/Файл:Von_Reissig,_Cornelius_August.jpg) (дата обращения: 20.07.2025).
7. Сементковский Р. И. Е. Ф. Канкрин. Его жизнь и государственная деятельность. М.: ЖЗЛ от Павленкова, 2008. 131 с.
8. Беликов П. Ф. Н. К. Рерих: биографический очерк. URL: <https://rerih.org/library/5045> (дата обращения: 07.06.2025).
9. Магаева С., Мартилла Е. Мученики ленинградской блокады. М.: Благовещение, 2007. 112 с.
10. Мартилла Э. О. Виртуальный Русский музей: электронный каталог // Виртуальный Русский музей: сайт. URL: [https://rusmuseumvrm.ru/reference/classifier/author/marttila\\_eo/index.php](https://rusmuseumvrm.ru/reference/classifier/author/marttila_eo/index.php) (дата обращения: 01.07.2025).

- 
11. PastVu: фотоархив исторических фотографий. URL: <https://pastvu.com/p/998692> (дата обращения: 01.07.2025).
  12. Владимирова Л. С., Леонидович М. В. Художественная сила Е. К. Эвенбах, ученицы «Школы Рериха» // Сервис. 2024. № 1. С. 38–50.
  13. Боровская Е.А. Два пути творческой реализации художника-педагога. П. Чистяков и И. Крамской // Научные труды Санкт-Петербургской академии художеств. 2019. № 48. С. 123–135.
  14. Одегов И. И. П. Чистяков – основоположник школы русского реалистического рисунка // Известия ЮФУ. Технические науки. 2004. №1 (36). С. 249.
  15. Лиховид Ю. Г., Керимова А. Т. Марк Шагал в современном искусстве // Молодой исследователь Дона. 2018. № 1 (10). С. 45–52.
  16. Citywalls: архитектурно-исторический портал Санкт-Петербурга: сайт. URL: <https://www.citywalls.ru/house2151.html> (дата обращения: 20.07.2025).
  17. Citywalls: архитектурно-исторический портал Санкт-Петербурга: сайт / материал о доходном доме Дернова. URL: <https://www.citywalls.ru/house620.html?highlight=доходн%20дом%20дернов> (дата обращения: 20.07.2025).
  18. Дулепинская Д. А. Основные понятия и инструменты онлайн маркетинга // Вестник науки. 2023. № 9 (66). С. 45–52. (дата обращения: 20.07.2025).
  19. Тарумова Н. Т. Презентация объектов отечественного культурного наследия в интернете: контекст музейных практик // Трансформации музеев-библиотек-архивов и информационное обеспечение исторической науки в информационном обществе. 2017. № 2017. С. 213–222.
  20. Матвеева В. А., Шаталова Н. А. К 175-летию Санкт-петербургского художественного училища имени Н. К. Рериха: страницы истории библиотеки // Петербургская библиотечная школа. 2014. № 1 (45). С. 62–69.



УДК 004.92

DOI: 10.25686/978-5-8158-2474-4-2025-1083-1090

## Организация научно-исследовательской работы обучающихся с использованием системы параметрического моделирования КОМПАС-3D

Ю. М. Булдакова

Поволжский государственный технологический университет, Йошкар-Ола, Россия

**Аннотация.** Одним из важнейших элементов системы высшего образования является подготовка высококвалифицированных кадров, обладающих не только профессиональными знаниями и умениями, но и способных к активной научно-исследовательской деятельности. Представлено содержание научно-исследовательской работы обучающихся первого курса по дисциплине «Инженерная графика». На примере моделирования процесса работы устройства регулирования давления в системе трехмерного проектирования КОМПАС-3D v21 обучающиеся знакомятся с параметрическими возможностями САД-системы как на уровне деталей, так и на уровне сборки, с приемами работы в приложении «Механика: Анимация».

**Ключевые слова:** научно-исследовательская работа студентов, компьютерная графика, твердотельное моделирование, параметризация, сборка, анимация, моделирование процессов, система КОМПАС-3D

## Organization of students' research work using the KOMPAS-3D parametric modeling system

Y. M. Buldakova

Volga State University of Technology, Yoshkar-Ola, Russia

**Abstract.** One of the most important elements of the higher education system is the training of highly qualified personnel with not only professional knowledge and skills, but also capable of active research activities. The content of the research work of the first-year students in the discipline «Engineering graphics» is presented. Using the example of modeling the operation of a pressure control device in the KOMPAS-3D v21 three-dimensional design system, students get acquainted with the parametric capabilities of the CAD system both at the detail and assembly levels, and with the working techniques in the «Mechanics: Animation application».

**Keywords:** students' research work, computer graphics, solid-state modeling, parameterization, assembly, animation, process modeling, KOMPAS-3D system.

### Введение

Подготовку квалифицированных специалистов, занимающихся проектированием изделий машиностроения и технологией их изготовления, в современных условиях невозможно представить без применения систем автоматизированного проектирования (САПР) [1]. Современные САПР предлагают широкие возможности для создания параметрических моделей различных объектов [2]. Использование механизма параметризации САПР в рамках научно-исследовательской работы (НИР) обучающихся по дисциплине «Инженерная графика» позволяет развивать инженерно-конструкторские навыки – модифицировать довольно сложные сборочные единицы, например изменять размеры деталей и управлять их формой, управлять положением детали в сборочной единице за счет изменения параметров привязок, автоматически перестраивать всю сборку при изменении одного из параметров. Также одной из наиболее актуальных областей применения компьютерной графики является система виртуальной реальности. Она позволяют имитировать объекты реального мира, визуализировать и обрабатывать большие объемы информации. По этой причине обучение основам реалистичной визуализации 3D-моделей является обязательным условием графической подготовки будущих инженеров. В современные САПР интегрированы приложения и библиотеки, которые упрощают задачу конструктора по разработке проектов любой сложности [3]. К примеру, в среде проектирования КОМПАС-3D имеется приложение «Механика: Анимация» [4], которое позволяет наглядно продемонстрировать работу различных механизмов и взаимодействие их составляющих друг с другом [5–7].

### Постановка задачи

Основной целью изучения дисциплины «Инженерная графика» является приобретение обучающимися знаний и навыков, необходимых для выполнения и чтения чертежей. Процесс обучения

целесообразно проводить на простых базовых задачах с поэтапным повышением уровня сложности заданий. Выполнение одних и тех же заданий вручную и с использованием САПР демонстрирует обучающимся широкие возможности и преимущества последних. Именно раздел «Чтение и детализация чертежей сборочных единиц» как нельзя лучше подходит для рассмотрения некоторых аспектов параметрического моделирования в САПР. Параметрическое моделирование существенно отличается от обычного трехмерного моделирования. В случае параметрического проектирования создается математическая модель объекта с параметрами, при изменении которых происходят изменения конфигурации детали, взаимные перемещения деталей в сборке и т. п. На примере сборочной единицы «Пневмоклапан давления» [8] обучающимся по сборочному чертежу (чертежу общего вида) в системе КОМПАС-3D требуется создать параметрическую модель сборки [9]. Для имитации работы устройства необходимо в параметрической модели сборочной единицы определить степень свободы элементов регулятора давления, а затем создать анимацию его открывания и закрывания под действием давления.

### Теория

Моделирование сборки «Пневмоклапан давления» в системе КОМПАС-3D начинается с создания уникальных деталей (рис. 1): корпуса, крышки, упора, пружин, тарелки, диафрагмы, плунжера, толкателя, заглушки, клапана, вкладыша, трубки, прокладки. Построение трехмерной твердотельной модели деталей заключается в многократном добавлении и вычитании объемов и наиболее часто требует использования операций выдавливания и вращения эскиза. В процессе моделирования деталей используются различные инструменты САПР (рис. 2): фаска, условное изображение резьбы, массив по концентрической сетке и т. д., средства библиотеки «Стандартные изделия», которые значительно сокращают процесс формирования моделей.

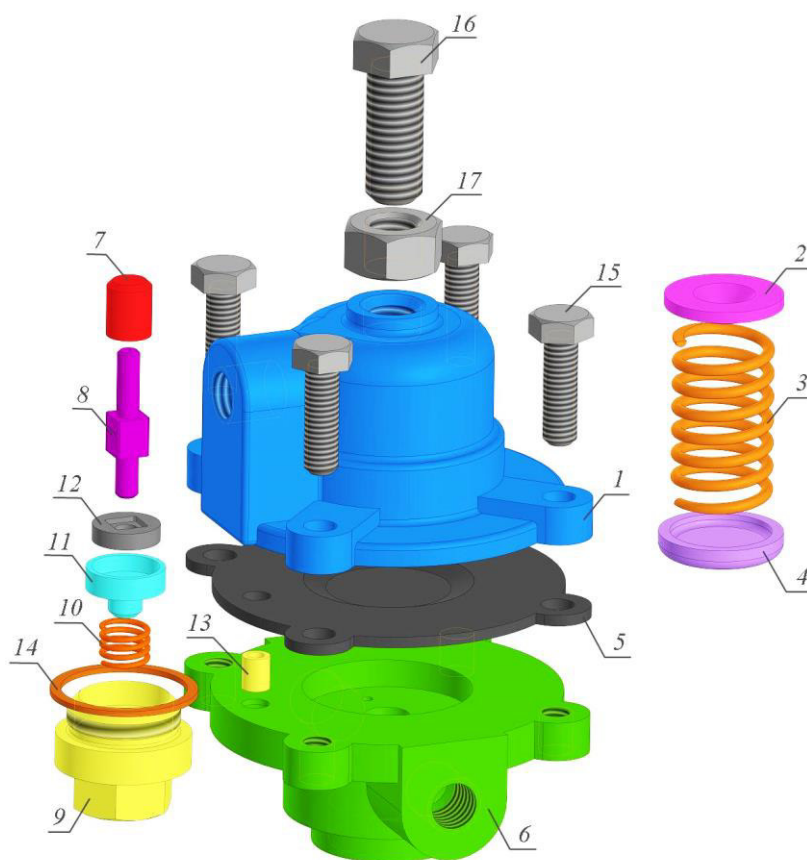


Рисунок 1. Детали, входящие в сборочную единицу «Пневмоклапан давления»:

1 – крышка, 2 – упор, 3 – пружина, 4 – тарелка, 5 – диафрагма, 6 – корпус, 7 – плунжер, 8 – толкатель,  
9 – заглушка, 10 – пружина, 11 – клапан, 12 – вкладыш, 13 – трубка, 14 – прокладка,  
15 – болт М6×20 ГОСТ 7805-70, 16 – болт М10×40 ГОСТ 7805-70, 17 – гайка М10 ГОСТ 5915-70

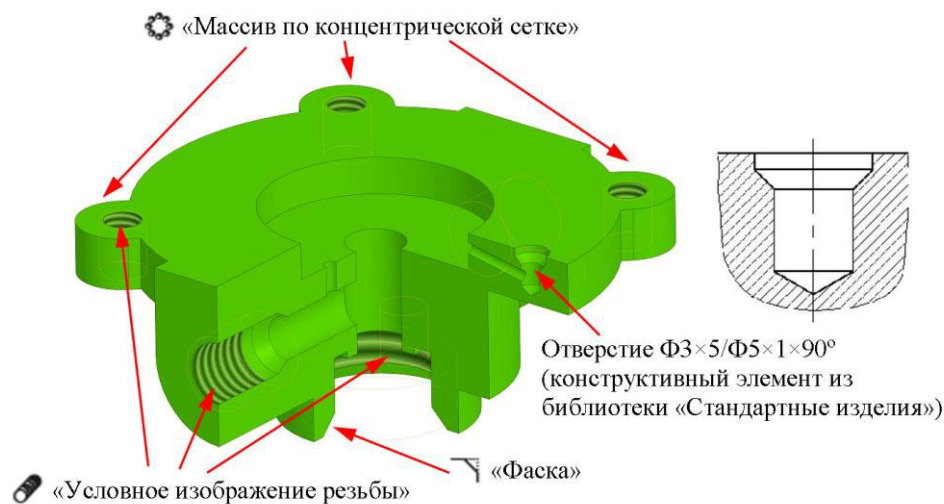


Рисунок 2. Модель уникальной детали «Корпус»

Согласно требованиям ГОСТ 2.109-73 [10], на сборочных чертежах (чертежах общего вида) для упрощения можно не изображать фаски, проточки для выхода резьбонарезного инструмента и т. д. Следовательно, чтение чертежа не должно сводиться к простому перечерчиванию изображений указанных деталей со сборочного чертежа (чертежа общего вида) сборочной единицы. Обучающиеся должны понимать, что для обеспечения стандартного профиля резьбы по всей длине стержня или отверстия необходима проточка для выхода резьбообразующего инструмента, форму и размеры которой выбирают согласно ГОСТ 10549-80 [11], при этом любая цилиндрическая поверхность, если будет сопрягаться с другой поверхностью, должна содержать на входе фаску для облегчения процесса сборки [12].

Функцией предохранительного клапана является поддержание постоянного давления на выходе, несмотря на колебания давления на входе. Для имитации процессов растяжения и сжатия пружин создаются их параметрические модели (рис. 3). С помощью команды «Спираль цилиндрическая» строится кривая, при этом следует указать способ построения «По числу витков и шагу». В поле «Количество витков» для пружины (поз. 3) требуется ввести переменную  $n$ , в поле «Шаг» – выражение  $H/n$ , где  $H$  – высота пружины (поз. 3). Аналогично для пружины (поз. 10) указать «Количество витков» –  $t$ , «Шаг» –  $h/t$ , где  $h$  – высота пружины (поз. 10). Для выдавливания сечения проволоки по созданной пространственной кривой обращаемся к команде «Элемент по траектории».

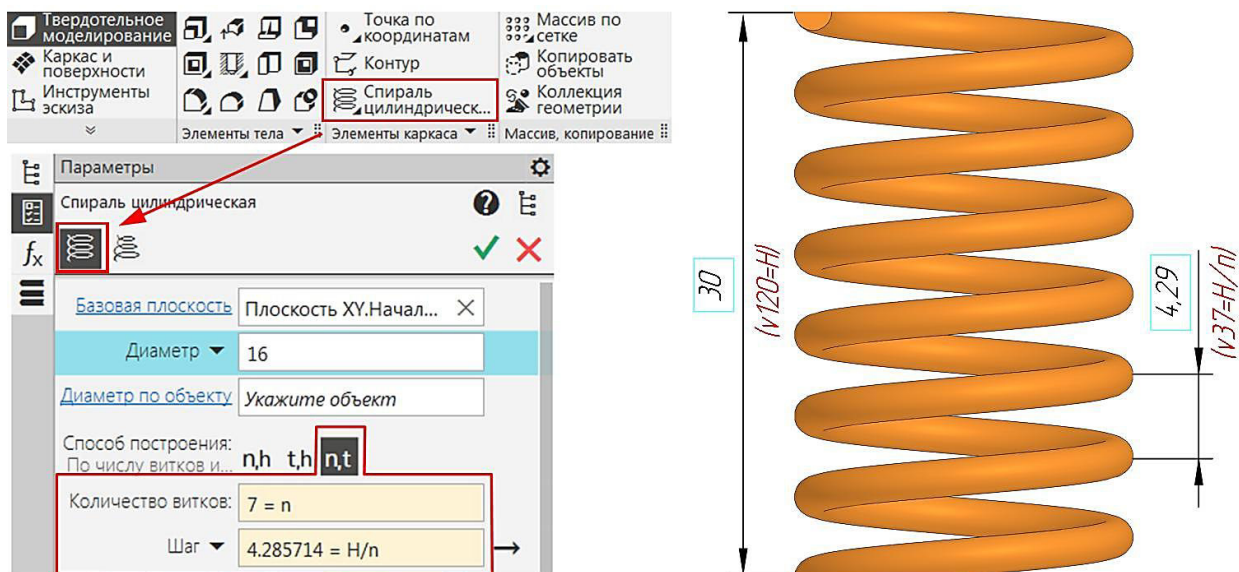


Рисунок 3. Параметрическая модель пружины (поз. 3)

При создании эскиза подвижной части диафрагмы (поз. 5), которая регулирует и перекрывает поток рабочей среды за счет движения вверх–вниз, необходима параметризация ее геометрии: взаимосвязь между объектами и наложенными на них ограничениями, простановка управляющих размеров углового  $a$  и линейного  $c$  (рис. 4). Признаком управляющего размера является прямоугольная рамка вокруг его значения, а соответствующие размерам имена переменных отображаются под размерными линиями.

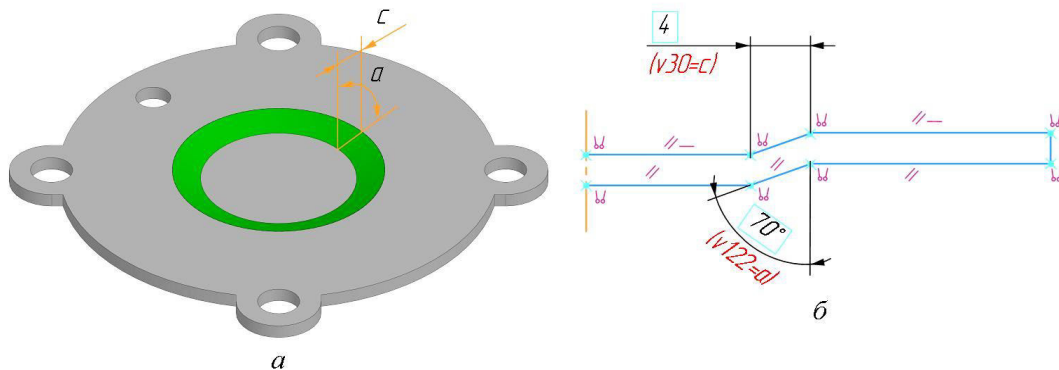


Рисунок 4. Параметрическая модель детали «Диафрагма»:  
а – 3D-модель детали; б – простановка управляющих размеров в эскизе

Источником значений переменных в моделях деталей – пружины (поз. 3), диафрагмы (поз. 5) и пружины (поз. 10) – будут служить ссылки на значения переменных в сборке, что позволит, изменяя параметры сборки, изменять параметры деталей, входящих в эту сборку. С целью управления параметрами сборки в таблице «Панели переменных» (рис. 5) вводим зависимости одних параметров от других, например, изменение угла  $a$ :

$$da = (90 - a) + a \tan \left[ \frac{dh - c \cdot \tan(90 - a)}{c} \right], \quad (1)$$

где  $dh$  – изменение высоты пружины (поз. 10),  $\tan$  – тангенс с аргументом в градусах,  $\text{atand}$  – арктангенс с результатом в градусах.

Теперь с изменением хотя бы одного из значений входного параметра  $a$ ,  $c$ ,  $dh$  будет меняться угол на величину  $da$  в соответствии с формулой (1). Для перестроения параметризованных моделей деталей необходимо обратиться к команде «Перестроить» или нажать клавишу «F5».

Аналогично вводятся выражения:

для высоты пружины (поз. 3) при деформации –

$$H = H_0 - dH - dh,$$

для высоты пружины (поз. 10) –

$$h = h_0 + dh,$$

где  $H_0$  – высота пружины (поз. 3) в свободном состоянии;  $dH$  – изменение высоты пружины (поз. 3);  $h_0$  – высота пружины (поз. 10) в свободном состоянии.

Переменным  $H$ ,  $h$ ,  $da$  автоматически будет присвоен статус «информационная», признаком которого является значок **i** (рис. 5).

Стандартные изделия: болт М6×20 ГОСТ 7805-70 (поз. 15), болт М10×40 ГОСТ 7805-70 (поз. 16), гайка М10 ГОСТ 5915-70 (поз. 17), не требуется моделировать как уникальные детали. Их модели вставляются в сборку из библиотеки «Стандартные изделия»: «Приложения» → «Стандартные изделия» → «Вставить элемент» → «Стандартные изделия».

Взаимное расположение компонентов сборки определяется создаваемыми позиционирующими сопряжениями (рис. 6): соосность, совпадение, на расстоянии и т. д. Сопряжения играют большую роль в создании анимации. Главное не перенасыщать сборку ненужными связями. Например, если на два компонента одновременно наложены сопряжения «Параллельность» и «Совпадение», то это может привести к ошибке пересопряжения сборки. Когда связи между компонентами настроены правильно, то процесс анимации проходит без ошибок.



	Имя	Выражение	Значение	Параметр	Комментарий
▼ Сборка (Тел-0, Сборочных единиц-0, Деталей-20)					
	h0	30	30		высота пружины 3 в свободн...
	dH	0	0		изменение высоты пружины 3
i	H	h0-dH-dh	30		высота пружины 3
	h0	6.5	6.5		высота пружины 10 в свобод...
	dh	0	0		изменение высоты пружины 10
i	h	h0+dh	6.5		высота пружины 10
	a	70	70		угол (в градусах)
	c	4	4		размер
i	da	$90-a+ATAND((dh-c*TAND(90-a))/c)$	0		изменение угла a (в градусах)

Рисунок 5. Фрагмент панели переменных в сборке

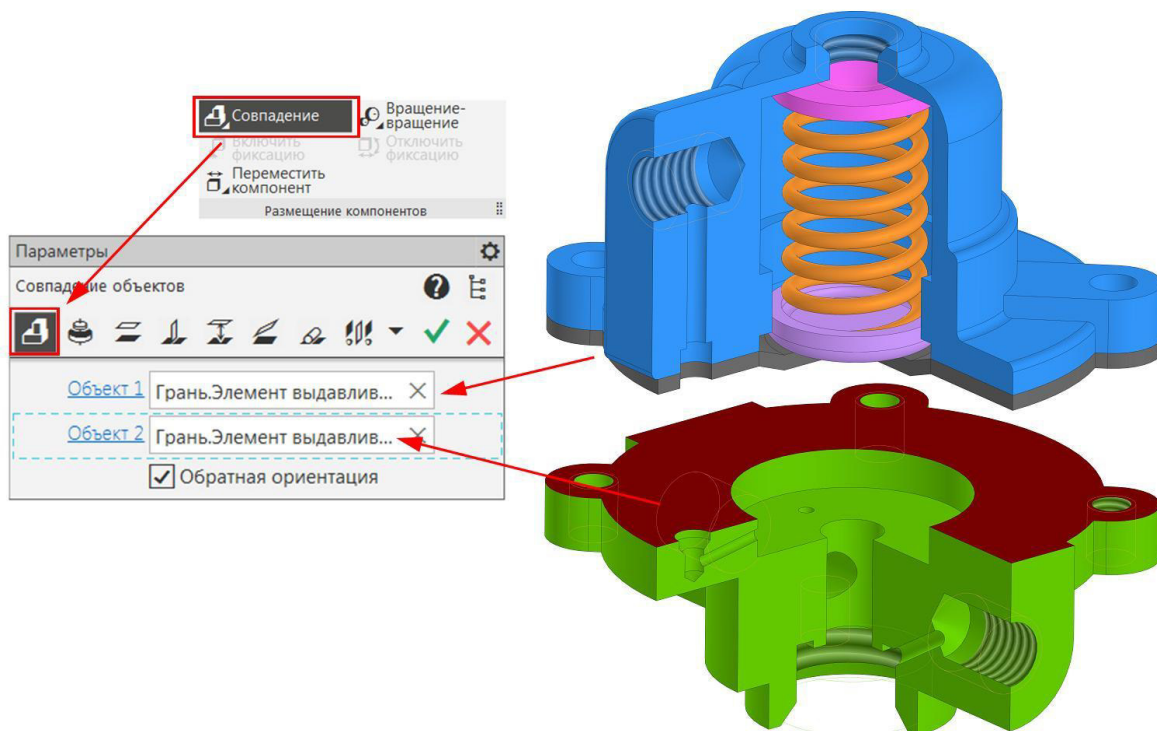


Рисунок 6. Позиционирующее сопряжение «Совпадение»

На основании полученной трехмерной параметрической модели сборки разрабатывается анимация процесса работы устройства регулирования давления. В системе КОМПАС-3D можно применить модуль анимации для создания поэтапного показа работы клапана, что поможет увидеть, как изменяется его положение при различных значениях давления. Сценарий анимации будет включать в себя два шага (рис. 7): шаг №1 – изменение степени поджатия пружины (поз. 3) с помощью болта (поз. 16); шаг №2 – имитация повышения давления под диафрагмой, при этом пружина (поз. 3) будет сжиматься и клапан (поз. 11) под действием пружины (поз. 10) уменьшает или прекращает проход газа из зоны А в зону Б.

На первом шаге анимации необходимо выбрать переменную dH: «Компоненты и переменные» → «Выбрать переменную». Появится окно выбора переменных. Интервал изменения переменной и время деформации пружины необходимо откорректировать, используя команды меню: «Параметры» → «Переменные» → «Редактировать параметры». На шаге №2, аналогично, выбираем переменную dh.

### Результаты экспериментов

После создания сценария анимации можно воспроизвести работу устройства. Для этого надо выполнить команду меню «Воспроизведение». На экране появляется управляющая панель (рис. 8)

с кнопками «Пуск» («▶»), «Стоп» («■»), «Пауза» («||») и «Создавать видеоролик» («●»). Функция «Создавать видеоролик» позволяет записать анимацию в видеофайл, доступный для просмотра в видеопроигрывателе без использования среды КОМПАС-3D.

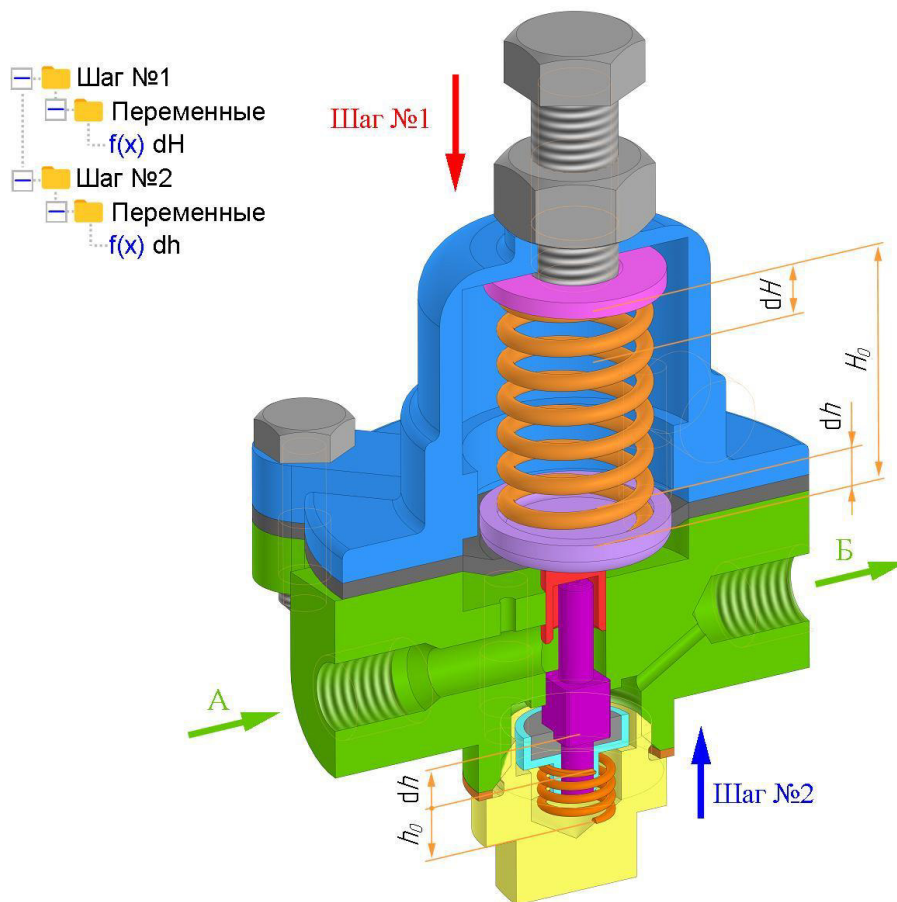


Рисунок 7. Сценарий анимации работы устройства «Пневмоклапан давления»

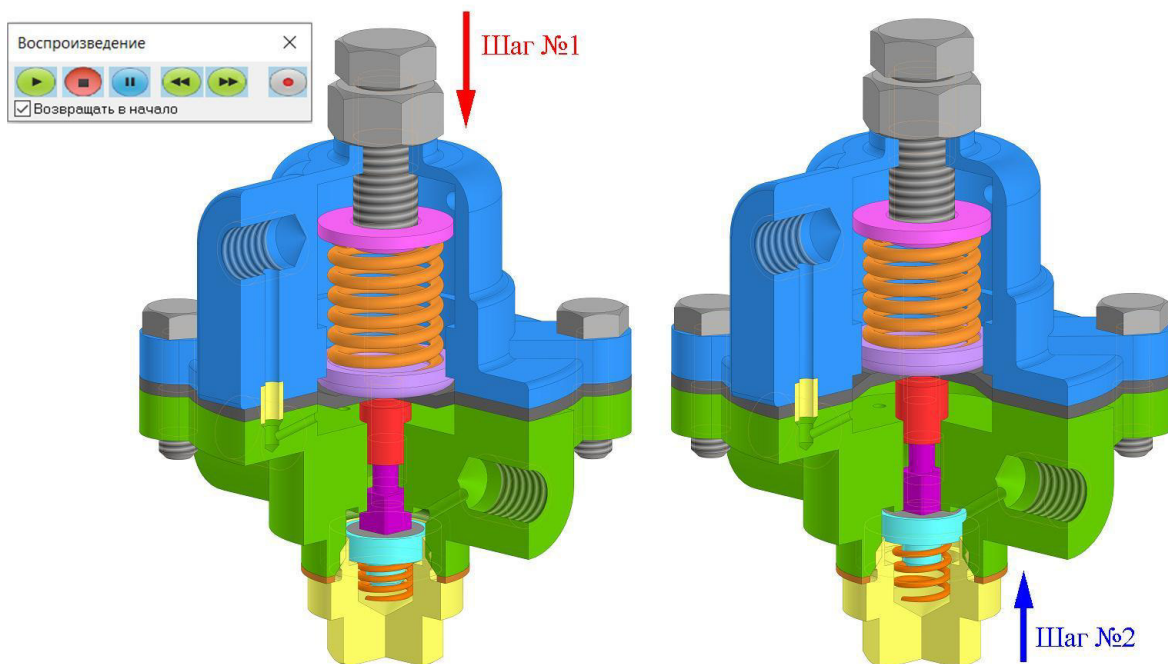


Рисунок 8. Воспроизведение сценария анимации работы устройства «Пневмоклапан давления»

### Обсуждение результатов

На основании вышеприведенной работы обучающиеся под руководством преподавателя выполняют индивидуальный проект. Результатом выполнения индивидуального задания является не только подготовка теоретического доклада, но и активное обсуждение на студенческих научных конференциях возможных решений практических проблем. При этом выступление перед значительной аудиторией слушателей способствует совершенствованию универсальных, общепрофессиональных и профессиональных компетенций.

На базе ПГТУ ежегодно проводится международная молодежная научная конференция по естественнонаучным и техническим дисциплинам «Научному прогрессу – творчество молодых», на которой рассматриваются и обсуждаются теоретические, экспериментальные, прикладные вопросы по актуальным проблемам естественнонаучных и технических дисциплин на современном этапе их развития. В секции «Прикладная геометрия и компьютерная графика» на XX конференции представлены результаты (рис. 9) в рамках НИР по изучению и реализации возможностей параметризации и анимации в системе КОМПАС-3D [13, 14] обучающимися первого курса.

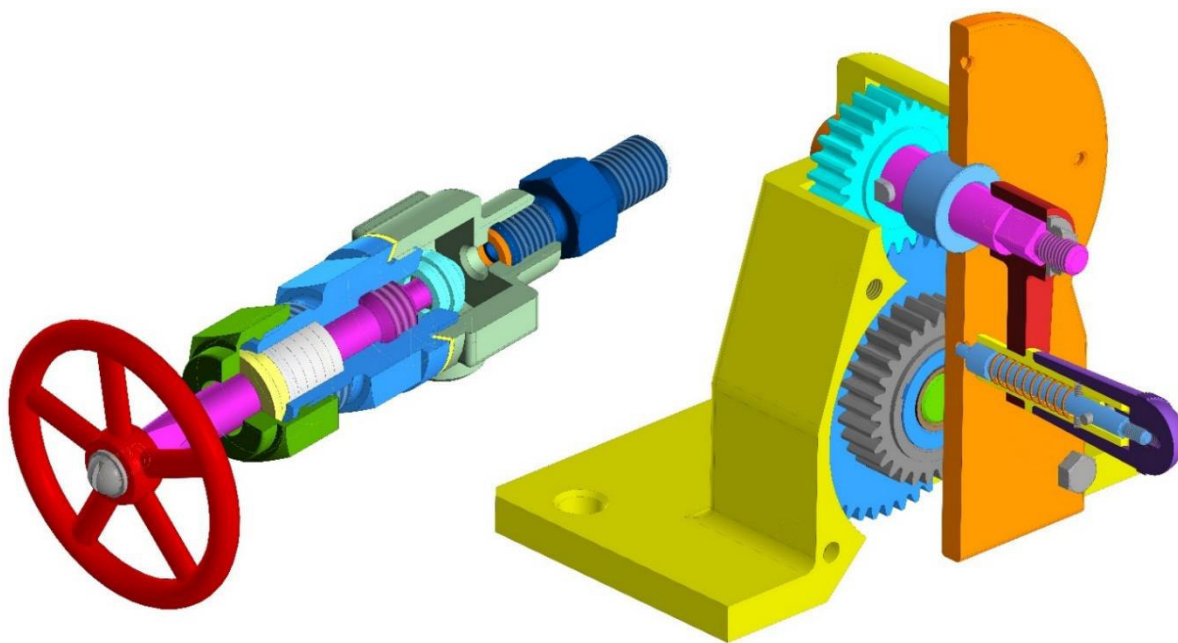


Рисунок 9. Параметрические модели сборок, выполненные обучающимися в рамках НИР

По результатам данной работы обучающиеся знакомятся с основами использования технологии параметризации, реализованной в КОМПАС-3D применительно к различным деталям и сборочным единицам, а при выстроенной единой междисциплинарной схеме обучения создаются условия для более эффективного освоения сквозной параметризации на основе использования CAD/CAM/CAE систем – от разработки до изготовления изделия, что позволяет адаптировать выпускников высших учебных заведений к условиям современного производства.

### Выводы

Параметрическое моделирование обеспечивает возможность вариантного проектирования в сжатые сроки, а также позволяет обнаружить ошибки на ранних стадиях проектирования и оперативно внести необходимые изменения. Также, безусловно, параметрическое 3D-моделирование помогает вовлечь обучающихся в научно-исследовательскую работу, делает процесс обучения интересным и понятным, многократно повышает качество образования.

Внедрение практико-ориентированных технологий в образовательный процесс позволит на этапе обучения в вузе смоделировать процесс применения полученных знаний в условиях работы реального производства, тем самым поднимется уровень конкурентоспособности будущих специалистов на рынке труда.



**Список литературы**

1. Постановление Правительства РФ от 26.12.2017 № 1642 (ред. от 21.02.2025) «Об утверждении государственной программы Российской Федерации «Развитие образования» // Docs.cntd.ru – электронный фонд нормативно-технической и нормативно-правовой информации Консорциума «Кодекс». URL: <https://docs.cntd.ru/document/556183093> (дата обращения: 30.07.2025).
2. Управление параметризацией сложных машиностроительных изделий в среде средних САПР / В. В. Шишкин, Т. П. Пилюгина, А. П. Грошев, Н. А. Зотов // Вестник УлГТУ. 2005. № 1 (29). С. 50–52. URL: <https://cyberleninka.ru/article/n/upravlenie-parametrizatsiey-slozhnyh-mashinostrotelnyh-izdeliy-v-srede-srednih-sapr> (дата обращения: 30.07.2025).
3. Залогова Л. А. Принципы трехмерной графики в программах трехмерного моделирования и системах виртуальной реальности // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2010. № 3. С. 84–90. URL: <https://cyberleninka.ru/article/n/printsiipy-trehmernoy-grafiki-v-programmah-trehmernogo-modelirovaniya-i-sistemah-virtualnoy-realnosti> (дата обращения: 30.07.2025).
4. Механика: Анимация. Инструкция по работе // КОМПАС-3D – официальный сайт САПР КОМПАС. URL: [https://kompas.ru/source/info\\_materials/user-manuals/rukovodstvo-polzovatelya-mekhanika-animaciya.pdf](https://kompas.ru/source/info_materials/user-manuals/rukovodstvo-polzovatelya-mekhanika-animaciya.pdf) (дата обращения: 30.07.2025).
5. Паньков М. С. Оживляя механизмы: Работа с анимацией в системе автоматизированного проектирования КОМПАС-3D (часть 1) // КОМПАС-3D – официальный сайт САПР КОМПАС. URL: [https://kompas.ru/source/articles/SiG\\_04-2014\\_animation.pdf](https://kompas.ru/source/articles/SiG_04-2014_animation.pdf) (дата обращения: 30.07.2025).
6. Паньков М. С. Оживляя механизмы: Работа с анимацией в системе автоматизированного проектирования КОМПАС-3D (часть 2) // КОМПАС-3D – официальный сайт САПР КОМПАС. URL: [https://kompas.ru/source/articles/2015-01\\_-\\_SiG-animation-part-2.pdf](https://kompas.ru/source/articles/2015-01_-_SiG-animation-part-2.pdf) (дата обращения: 30.07.2025).
7. Паньков М. С. Оживляя механизмы: Работа с анимацией в системе автоматизированного проектирования КОМПАС-3D (часть 3) // КОМПАС-3D – официальный сайт САПР КОМПАС. URL: [https://kompas.ru/source/articles/2015-02\\_Animaciya.pdf](https://kompas.ru/source/articles/2015-02_Animaciya.pdf) (дата обращения: 30.07.2025).
8. Аксарин П. Е. Чертежи для детализования: учеб. пособие для машиностроит. спец. вузов. Москва: Машиностроение, 1978. 132 с.
9. Булдакова Ю. М., Кузьмина О. В., Полушина Т. А. Чтение и детализование чертежа общего вида (сборочного чертежа): учебно-методическое пособие. Йошкар-Ола: Поволжский государственный технологический университет, 2025. 52 с.
10. ГОСТ 2.109-73 Единая система конструкторской документации (ЕСКД). Основные требования к чертежам (с Изменениями № 1-11) // Docs.cntd.ru – электронный фонд нормативно-технической и нормативно-правовой информации Консорциума «Кодекс». URL: <https://docs.cntd.ru/document/1200001992> (дата обращения: 30.07.2025).
11. ГОСТ 10549-80 Выход резьбы. Сбеги, недорезы, проточки и фаски (с Изменением № 1) // Docs.cntd.ru – электронный фонд нормативно-технической и нормативно-правовой информации Консорциума «Кодекс». URL: <https://docs.cntd.ru/document/1200012239> (дата обращения: 30.07.2025).
12. ГОСТ 10948-64 (СТ СЭВ 2814-80) Радиусы закруглений и фаски. Размеры (с Изменением № 1) // Docs.cntd.ru – электронный фонд нормативно-технической и нормативно-правовой информации Консорциума «Кодекс». URL: <https://docs.cntd.ru/document/1200011919> (дата обращения: 30.07.2025).
13. Булдакова Ю. М., Парфенов С. А. Создание анимации работы делительного приспособления фрезерного станка в КОМПАС-3D // Научному прогрессу – творчество молодых: материалы XX международной молодежной научной конференции по естественнонаучным и техническим дисциплинам, Йошкар-Ола, 25–26 апреля 2025 г. Йошкар-Ола: ПГТУ, 2025. С. 135–137. URL: <https://science.volgattech.net/upload/documents/sbornic20.pdf> (дата обращения: 30.07.2025).
14. Булдакова Ю. М., Черезов Н. С. Создание анимации работы пневмоаппарата клапанного в КОМПАС-3D // Научному прогрессу – творчество молодых: материалы XX международной молодежной научной конференции по естественнонаучным и техническим дисциплинам, Йошкар-Ола, 25–26 апреля 2025 г. Йошкар-Ола: ПГТУ, 2025. С. 138–140. URL: <https://science.volgattech.net/upload/documents/sbornic20.pdf> (дата обращения: 30.07.2025).

## СОДЕРЖАНИЕ

Предисловие .....	8
-------------------	---

## ПЛЕНАРНЫЕ ДОКЛАДЫ

<i>Пряничников В. Е., Ястребов В. В., Кувишинов С. В., Харин К. В.</i> Графические технологии расширенного взаимодействия с сервисными роботами .....	10
<i>Иванов В. А., Конкин Н. А., Овчинников В. В., Рябова М. И.</i> Метод обнаружения принимаемых мод сигнала на изображении ионограммы радиозондирования ионосферы с использованием машинного обучения .....	13

## Секция 1. РЕАЛИСТИЧНАЯ КОМПЬЮТЕРНАЯ ГРАФИКА И ВЫЧИСЛИТЕЛЬНАЯ ОПТИКА

<i>Милин Н. А., Родионов Р. О., Галактионов В. А., Фролов В. А.</i> Сравнительный анализ выполнения нейронных моделей для графики .....	23
<i>Родионов Р. О., Приходько Е. В., Фролов В. А., Волобой А. Г.</i> Метод спектрального рендеринга с использованием рядов Фурье .....	30
<i>Николаев А. А., Родионов Р. О., Фролов В. А.</i> Трассировка лучей нейронных геометрических моделей .....	41
<i>Кунц М. С., Гарифуллин А. Р., Коротаев И. О., Фролов В. А.</i> Приближенный рендеринг В-сплайновых поверхностей с триммингом в реальном времени .....	53
<i>Ершов С. В., Санжаров В. В., Фролов В. А., Галактионов В. А.</i> Метод вычисления полиномиального разложения для высокоточной трассировки лучей в оптических системах .....	61
<i>Санжаров В. В., Ершов С. В., Фролов В. А., Волобой А. Г.</i> Выбор начального приближения в трассировке лучей оптических систем с асферическими поверхностями .....	71
<i>Будак А. С., Гарифуллин А. Р., Галактионов В. А., Фролов В. А.</i> Кроссплатформенный дифференцируемый рендеринг функций расстояний .....	83
<i>Майоров Н. Д., Гарифуллин А. Р., Волобой А. Г., Фролов В. А.</i> SDFMultiOctree: неявное представление поверхностей на основе функций расстояния со знаком .....	94
<i>Зубатов К. С., Щербаков А. С., Фролов В. А., Санжаров В. В.</i> Кодирование глобального освещения гауссианами в приложениях реального времени .....	105
<i>Raenchuk Maksim</i> Enhancing 3D Gaussian Splatting with diffusion models: a survey .....	113
<i>Жирков А. О., Родионов Р. О., Копанев П. Д., Фролов В. А.</i> T3D: Модульный программно-аппаратный комплекс реконструкции 3D-объектов .....	133
<i>Дав Я. Н.</i> Темпоральный и пространственный ресэмплинг для отрисовки с использованием физического переноса света .....	146
<i>Султанов Р. Р., Лях Г. С., Богданов М. К., Суворов А. М.</i> Гибридный метод отображения теней в реальном времени для систем с несколькими видеокартами .....	154
<i>Богданов М. К., Суворов А. М., Булаев А. П.</i> Гибридный алгоритм отображения облаков для систем с несколькими графическими адаптерами .....	164
<i>Богданов М. К., Ивашечкина М. Е., Суворов А. М., Булаев А. П.</i> Гибридный метод расчета окружающего затенения для систем с несколькими видеоадаптерами .....	172
<i>Галеев Т. Р., Богданов М. К.</i> Разработка и анализ метода отражений в экранном пространстве для систем с дискретным и интегрированным GPU .....	183
<i>Жданов Д. Д., Куприянов С. И.</i> Метод восстановления текстуры объектов сцены .....	192
<i>Ван Янь, Жданов Д. Д., Потемин И. С., Жданов А. Д.</i> Алгоритм расчета бликов в оптических системах .....	206
<i>Ван Ч., Жданов Д. Д., Жданов А. Д.</i> Исследование методов отслеживания зрачков для систем автономного вождения .....	215
<i>Вяткин С. И., Долговесов Б. С.</i> Объемно-ориентированная визуализация каустик и динамических преломляющих объектов со сложными свойствами материала в реальном времени .....	227
<i>Макаров Д. Н., Слюсарев Д. К., Будак В. П.</i> Расчет показателя дискомфорта с помощью технологий компьютерной графики .....	239
<i>Будак В. П., Макаров Д. Н., Смирнов П. А.</i> Модель отражения света от реальных поверхностей: матричный подход с учетом анизотропии рассеяния .....	250

Шуткин В. Н., Морозкин Н. К., Семенов В. А., Тарлапан О. А. Эффективная генерация иерархических уровней детализации: параллельная обработка, инкрементальные обновления, полигональные упрощения на основе видимости .....	261
Будак В. П., Гримайло А. В., Желтов В. С. Расчетное ядро программы «Свет в ночи»: методы и сравнение с коммерческими аналогами .....	276
Тимохин П. Ю., Михайлюк М. В. Трассируемые мультиобъекты на основе 3D-гауссианов .....	286
Барладян Б. Х., Дерябин Н. Б., Волобой А. Г., Шапиро Л. З., Денисов Е. Ю., Галактионов В. А. Многослойная визуализация дисплея пилота гражданского воздушного судна .....	298
Бирюков Е. Д., Копылов М. С., Соколов В. Г., Потемин И. С., Волобой А. Г. Верификация системы оптического моделирования Lumiscapt по тестовым сценариям CIE 171:2006 .....	305

## Секция 2. НАУЧНАЯ ВИЗУАЛИЗАЦИЯ И ВИЗУАЛЬНАЯ АНАЛИТИКА

Поповская А. П., Бодрунова С. С., Васев П. А., Стародубцев И. С. Сравнение пользовательской эффективности традиционных и нейросетевых методов UX/UI-проектирования корпоративных порталов .....	320
Якунин А. В., Бодрунова С. С. Влияние графического интерфейса на когнитивный контроль пользователей при решении стрессогенных задач .....	335
Алексеев А. К., Бондарев А. Е. О неевклидовой визуализации многомерных функций .....	344
Джейранян А. Д., Ларионова Я. А., Лядова Л. Н. Проектирование инструментов визуализации данных на основе интеграции возможностей предметно-ориентированного моделирования и генеративного искусственного интеллекта .....	353
Бондарев А. Е., Кувшинников А. Е. Сравнительная оценка солверов пакета OpenFOAM при моделировании цепи косых скачков сжатия .....	367
Петров М. А., Крутина Е. В., Баходирий М. А. Р. Оценка корректного восстановления технологии штамповки при помощи инструментов цифрового проектирования и моделирования .....	374
Forghani Majid, Forghani Kazem, Bolkov Mikhail. Low-dimensional embedding for exploring the phylogenetic characteristics of tick-borne encephalitis virus .....	385
Forghani Kazem, Forghani Majid, Bolkov Mikhail. Visualization of phylogenetic self-similarity of genomic sequence .....	397
Максименко Л. А. Современные технологии визуализации для создания и управления строительной документацией .....	407
Решетников М. А., Замятина Е. Б. Моделирование социальных сетей: визуализация и расчет метрик .....	415
Терещенко Т. М., Щур С. Ю. Исследование и разработка алгоритма создания инклюзивного сайта в сфере государственных услуг .....	424
Бондарева Н. А. Автоматизированный поиск в видеопотоке: практический инструмент для исследований и контент-анализа видеофрагментов .....	434
Третьяков С. А., Стародубцев И. С. Синергетическая интеграция методов роевого интеллекта в задачах визуализации и аналитики .....	442
Петров М. А., Корнюшин М. В., Смирнов А. В., Рыбальченко В. В. Исследование процесса литья керамического шликера под низким давлением .....	451
Петров М. А., Исаков В. В. Применение метода корреляции цифровых изображений для оценки деформации листовой заготовки .....	459
Попов Е. В., Юрченко П. В. Применение искусственного интеллекта для визуализации контуров природных объектов на спутниковых снимках .....	466
Соколов И. Д., Ионкина К. В., Улизко М. С., Антонов Е. В., Бажанова Е. Н., Артамонов А. А. Метод идентификации неявных связей в решении аналитических задач .....	472
Константиновская Т. В., Борисов В. Е., Луцкий А. Е. Визуализация вихревой системы треугольного крыла .....	487
Молотков А. А., Сапрыкин Д. Л., Третьякова О. Н., Тужилин Д. Н., Шамордин А. А. Разработка методики позиционирования изделия при многокоординатной лазерной обработке .....	494
Лабутин И. А. A unified approach for EEG visualization on a mobile devices .....	498
Конкин Н. А., Вершинин М. В. Машинное обучение для анализа доступности коротковолнового канала связи .....	505

<i>Aleksandrov Victor, Kniaz Vladimir V., Gomozev Pavel, Novikov Daniil. Visualization of unrealistic details in synthetic images using feature maps of a neural network .....</i>	514
<i>Манаков Д. В., Васёв П. А. Стабилизация взаимодействия или автоволны в параллельных вычислениях?.....</i>	517
<i>Коган Л. П., Вольвач А. Е. О возможности визуализации процессов подготовки катастрофических событий... 524</i>	
<i>Бондарева Н. А. Интеграция генеративных нейросетей в математическое и трехмерное моделирование: современное состояние и перспективы развития .....</i>	535
<i>Бондарева Н. А. Информационная плотность объектов в цифровой среде: теоретические основы и методология оценки .....</i>	544
<i>Бондарев А. Е. Цифровой двойник как область научной визуализации: математическое моделирование узла ветроэнергетической установки .....</i>	555

### Секция 3. КОМПЬЮТЕРНОЕ ЗРЕНИЕ

<i>Поляков Р. В. Способ оптико-информационного обеспечения обнаружения артефактов на сложном фоне роботизированным комплексом с элементами нейронных сетей .....</i>	568
<i>Лаптев В.В., Гергет О. М. Методика полуавтоматической разметки данных: снижение трудоемкости формирования масок семантической сегментации.....</i>	575
<i>Вишневский М. Р., Фёдоров И. Ю. Разработка метода повышения плотности и качества лидарного сканирования на основе зеркального усиления механических LiDAR-систем .....</i>	586
<i>Загайнов С. Д., Коршунов Д. М., Хвостиков А. В. Сегментация изображений геологических аншлифов с использованием дообученных визуально-языковых моделей .....</i>	595
<i>Зубова Е. О., Шакин М. Н., Рыбаков А. В. Модернизация системы управления для обнаружения края металлической полосы, свернутой в рулон, на агрегате продольной и поперечной резки листа с применением компьютерного зрения.....</i>	607
<i>Морохин Д. В., Иванов А. В. Разработка модульной системы технического зрения для соревновательного колесного робота .....</i>	619
<i>Ходаков Н. А., Жгутова Е. С., Бабаян П. В. Алгоритм оценки точности совмещения изображений после компенсации проективного преобразования .....</i>	625
<i>Буянкина Я. А., Бабаян П. В. Пороговая сегментация изображений на основе генетического алгоритма.....</i>	631
<i>Белов Е. А., Замятина Е. Б. Опыт автоматического распознавания объектов на видеоконтенте, распространяемых в социальных сетях, с применением YOLOv12.....</i>	641
<i>Потапов Д. С. Метод быстрого обнаружения неверного поворота фотографий детей для сайта «усыновите.рф».....</i>	649
<i>Скопин Н. Ю., Мосалев М. С., Королев Д. А., Пискунова А. М. Распознавание рукописных графических элементов и мультязычного текста: обзор современных подходов, базовый метод и перспективы развития....</i>	659
<i>Левитин А. В., Мудров Е. П. Обнаружение клубней картофеля на фоне грунта .....</i>	670
<i>Местецкий Л. М., Смирнова В. Сегментация строк в изображениях рукописных документов.....</i>	675
<i>Дорошенко И. А., Знаменская И. А. Моделирование, распознавание и анализ на основе машинного обучения движущихся в газе твердых частиц различных масштабов .....</i>	685

### Секция 4. СИСТЕМЫ АВТОМАТИЗАЦИИ ПРОЕКТИРОВАНИЯ – РАЗРАБОТКА, ВНЕДРЕНИЕ И ПРИОРИТЕТЫ РАЗВИТИЯ

<i>Зеар Аунг, Щидловский С. В., Ну Ну Вар. Автоматизация движения мобильного робота за стеной TurtleBro с помощью ПИД-регулятора ROS .....</i>	693
<i>Аберясева И. Ю., Волков Д. П., Денисова Н. А. Аналитическое сравнение выполнения сборок разной сложности в СПЖЦ «САРУС» .....</i>	704
<i>Маркина Ю. Д. Автоматизация расчёта несущей способности антенно-мачтовых сооружений.....</i>	715
<i>Костина А. С., Макаров И. И., Денисова Н. А. Оптимизация выполнения криволинейных поверхностей в САД «САРУС» путем параметризации .....</i>	724
<i>Аладкин Д. А., Михайленко И. В., Денисова Н. А. Оптимизация построения 3D-модели в СПЖЦ «САРУС» с использованием таблиц переменных .....</i>	734
<i>Быкова А. С., Денисова Н. А. Исследование срабатывания функций группы «Ограничения» при построении фасонных поверхностей в программном обеспечении «САРУС» .....</i>	744

Макаров И. И., Денисова Н. А. Жизненный цикл детали от проектирования сложных криволинейных поверхностей в ПМ САД до создания управляющей программы в ПМ САМ «САРУС» .....	755
Клюева О. А., Денисова Н. А. Разработка пользовательского сценария создания 3D-модели детали «Вал» в программном модуле САД СПЦЖ «САРУС» .....	768

## Секция 5. ОБРАБОТКА И АНАЛИЗ БИМЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Аксенов К. Д., Аксенова Л. Е. Комплексная система анализа фундус-изображений с применением методов глубокого обучения .....	781
Ларионов А. И., Турлапов В. Е. Сети KAN и MLP в био- и медицинской диагностике на примере стресса растений .....	790
Локишин Н. Д., Крылов А. С. Адверсативная аугментация физически информированных нейронных операторов для задачи одномерного уравнения диффузии .....	811
Куракин М. Э., Соколов Н. А., Серебрякова А. А. Обзор наборов данных электронной микроскопии мозга для задач сегментации .....	816
Миронов А. И., Соколов Н. А., Серебрякова А. А. Влияние объема синтетически сгенерированных изображений данных в наборе данных для задачи сегментации аксонов в электронной микроскопии мозга .....	828
Трыкин А. М., Васильев Е. П., Турлапов В. Е. Exploring several ResNet-like models for atypical mitotic figure classification .....	835
Трыкин А. М., Кривоносов М., Калинина Е. Cell Localization with YOLO-based Models in Live-Dead Cell Viability Assay .....	844
Рехамния В., Трыкин А. М., Васильев Е. П., Турлапов В. Е. Доверенная самообучающаяся полимасштабная сегментация полнослайдовых гистологических изображений для диагностики опухолей .....	852

## Секция 6. ЦИФРОВАЯ ЗЕМЛЯ И БОЛЬШИЕ ДАННЫЕ

Белова А. В., Черницын И. А., Кычкин А. В. Динамическая визуализация в решении задач инжиниринга ГИС-данных экологического мониторинга атмосферного воздуха промышленных предприятий .....	867
--	-----

## Секция 7. ГЕОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ. КОМПЬЮТЕРНАЯ ГРАФИКА

Федотов И. А. Извлечение геометрических объектов из пространства чертежа САД-модели посредством изоморфизма графов .....	878
Фёдоров Д. Б. Определение рабочих зон механизмов рук антропоморфных роботов с учётом синтеза движений по вектору скоростей .....	899
Симонов А. В., Симонов М. В. Генерация параметрических поверхностей для моделирования рельефа с псевдослучайными геометрическими характеристиками .....	906
Ганчук С. Н. Гладкие преобразования В-сплайнов .....	912
Рустамян В. В. Развитие теории $\alpha\beta$ -триангуляции в трехмерном евклидовом пространстве .....	920
Ченарани С. Динамическая персонализация интерфейсов с использованием эргономического анализа, закона Фиттса и обучения с подкреплением .....	930
Щур С. Ю. Моделирование структуры «звезда Паскаля» с применением генеративного метода .....	934
Яров А. Б., Толок А. В. Алгоритм построения рельефа с учётом затенения функционально-воксельным методом для моделирования зон обнаружения радиолокационных станций .....	942
Головкина В. Б., Бобров М. А. Обоснование эффективности разработки игровых локаций с помощью процедурной генерации .....	945
Дворецкий А. Т. Геометрическое моделирование солнцезащитных устройств .....	954

## Секция 8. ГРАФИЧЕСКИЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ИММЕРСИВНЫЕ ТЕХНОЛОГИИ НА СТАДИЯХ ЖИЗНЕННОГО ЦИКЛА ПРОДУКТА

Серова М. А. Формализация коэффициента совокупных потерь при каскадной трансляции геометрических моделей .....	960
Шанин Е. Д., Спиринов А. А., Шутков А. А. Влияние степени интерактивности приложения на эффективность экспозиционной терапии в виртуальной реальности .....	968

Глумова Е. С. Алгоритм адаптивного разбиения облака точек с использованием иерархии ограничивающих объемов.....	976
Курушин Е. А. Анализ функционально ориентированной оценки передачи и восстановления геометрической модели.....	985
Николаев А. Е., Сахно Е. В., Филинских А. Д., Томчинская Т. Н. Информационная система инвентаризации вуза с применением BIM-технологий.....	995

#### Секция 9. ПОДГОТОВКА КАДРОВ ВЫСШЕЙ КВАЛИФИКАЦИИ В ОБЛАСТИ ИНЖЕНЕРНОЙ ГЕОМЕТРИИ, КОМПЬЮТЕРНОЙ ГРАФИКИ И ЦИФРОВОЙ ПОДДЕРЖКИ ЖИЗНЕННОГО ЦИКЛА ИЗДЕЛИЙ

Замятина Е. А. Геометрические модели и алгоритмы проектирования отражающих экранов в акустике.....	1006
Заяц Е. Е., Февральских А. В. Разработка программного обеспечения для построения октаэдрических сеток.....	1014
Ведяйкина О. И., Хазов П. А. Цифровые двойники в задачах валидации экспериментальных исследований и установок.....	1020
Шилов С. С. Алгоритмы автоматизации определения аэродинамических показателей поверхностей отрицательной гауссовой кривизны.....	1030

#### Секция 10. ЦИФРОВАЯ ТРАНСФОРМАЦИЯ ОБРАЗОВАНИЯ

Моисеева О. А. Возможности систем искусственного интеллекта в решении тестовых заданий по начертательной геометрии и инженерной графике.....	1041
Турецких С. О. Организация исследовательской и проектной деятельности при изучении графических дисциплин в условиях цифровизации высшего образования.....	1052
Ведяйкина О. И., Хазов П. А., Шилов С. С. Проблемы внедрения методов цифровой поддержки в фундаментальные физико-математические дисциплины.....	1063
Кучкарова Д. Ф., Ачилова Д. А. Анализ кластеров как инструмент в управлении образовательной деятельностью.....	1070
Швейгерт Е. Е., Князева Е. В. Историческое наследие как компонент современной информационно-образовательной платформы.....	1076
Булдакова Ю.М. Организация научно-исследовательской работы обучающихся с использованием системы параметрического моделирования КОМПАС-3D.....	1083