



GraphiCon 2017

27-я Международная конференция
по компьютерной графике
и машинному зрению

Труды конференции

GraphiCon 2017

27th International Conference
on Computer Graphics and Vision

Conference Proceedings

Пермский государственный национальный исследовательский университет,

Пермь, 24–28 сентября 2017 года

<http://2017.graphicon.ru/>

УДК 004.92
ББК 32.973.26-018.2
G73

G73 GraphiCon 2017: труды 27-й Междунар. конф. по компьютерной графике и машинному зрению (Пермь, 24–28 сент., 2017 г.) / Перм. гос. нац. исслед. ун-т. – Пермь, 2017. – 380 с.

ISBN 978-5-7944-2963-3

Настоящий сборник содержит материалы 27-й Международной конференции по компьютерной графике и машинному зрению. В её подготовке и проведении участвовали специалисты ведущих организаций, отраслевых и академических научно-исследовательских институтов, а также высшей школы.

Цикл «GraphiCon» представляет собой регулярные ежегодные конференции. Он служит международным форумом для исследователей и разработчиков со всего мира, целью которого является обмен новыми идеями, знаниями и опытом в области компьютерной графики, машинного зрения, обработки изображений, систем визуализации и виртуального окружения, систем поддержки принятия решений, облачных вычислений и обработки больших массивов данных.

В сборник включены приглашенные, пленарные и секционные доклады конференции.

Адресуется сотрудникам научно-исследовательских и образовательных организаций, предприятиям ИТ-индустрии, аспирантам, студентам, а также широкому кругу читателей.

УДК 004.92
ББК 32.973.26-018.2

*Печатается по решению
кафедры математического обеспечения вычислительных систем
Пермского государственного национального исследовательского университета*

ISBN 978-5-7944-2963-3

© ПГНИУ, 2017

Организационный комитет

Макарихин Игорь Юрьевич
ректор ПГНИУ
*Почётный Председатель
Оргкомитета*

Чуприна Светлана Игоревна
зав. каф. математического
обеспечения вычислительных
систем механико-математического
факультета ПГНИУ
Председатель Оргкомитета

Кузнецов Андрей Геннадьевич
декан механико-математического
факультета ПГНИУ
Сопредседатель

Рябинин Константин Валентинович
ПГНИУ
*Заместитель Председателя
Оргкомитета*

Шарапов Юрий Альбертович
ПГНИУ
*Помощник Председателя
Оргкомитета*

Постаногов Игорь Сергеевич
ПГНИУ
*Помощник Председателя
Оргкомитета*

Ветров Андрей Леонидович
проректор по научной работе и
инновациям ПГНИУ

Шкарапута Александр Петрович
заместитель декана
механико-математического
факультета ПГНИУ по научной
работе

Александров Вассил
Суперкомпьютерный центр
Барселоны

Деменева Ольга Ивановна
специалист по УМР
механико-математического
факультета ПГНИУ

Решетникова Елизавета Алексеевна
председатель студенческого
научного общества
механико-математического
факультета ПГНИУ

Галактионов Владимир Александрович
ИПМ им. М.В. Келдыша РАН

Клименко Станислав Владимирович
ИФТИ

Столбов Валерий Юрьевич
ПНИПУ

Карпов Владимир Ефимович
ИСП РАН

Масленникова Алла Александровна
МГУ

Программный комитет

Рябинин Константин Валентинович
ПГНИУ
*Председатель Программного
комитета*

Волобой Алексей Геннадьевич
ИПМ им. М.В. Келдыша РАН

Бондарев Александр Евгеньевич
ИПМ им. М.В. Келдыша РАН

Конушин Антон Сергеевич
МГУ

Барина Ольга Вячеславовна
МГУ

Вишняков Борис Ваисович
ГосНИИАС

Выголов Олег Вячеславович
ГосНИИАС

Горбачев Вадим
ГосНИИАС

Дышкант Наталья Фёдоровна
МГУ

Ерофеев Михаил Викторович
МГУ

Копылов Андрей Валерьевич
ТулГУ

Кустикова Валентина Дмитриевна
ННГУ

Николаев Дмитрий Петрович
ИППИ РАН

Середин Олег Сергеевич
ТулГУ

Сидякин Сергей Владимирович
ГосНИИАС

Чернявский Алексей
Samsung R&D Institute Rus

Фаворская Маргарита Николаевна
СибГУ им. М.Ф. Решетнева

Крылов Андрей Серджевич
МГУ

Турлапов Вадим Евгеньевич
ННГУ

Ерёмченко Евгений Николаевич
ООО «Неогеография»

Тикунов Владимир Сергеевич
МГУ

Игнатенко Алексей Викторович
МГУ

Барладян Борис Хаимович
ИПМ им. М.В. Келдыша РАН

Михайлюк Михаил Васильевич
НИИСИ РАН

Фролов Владимир Александрович
МГУ

Дебелов Виктор Алексеевич
ИВМиМГ СО РАН

Жданов Дмитрий Дмитриевич
ГОИ

Потемин Игорь Станиславович
НИУ ИТМО

Шапиро Лев Залманович
ИПМ им. М.В. Келдыша РАН

Галактионов Владимир Александрович
ИПМ им. М.В. Келдыша РАН

Клименко Станислав Владимирович
ИФТИ

Бобков Валерий Александрович
ИАПУ ДВО РАН

Ватолин Дмитрий Сергеевич
МГУ

Долговесов Борис Степанович
ИАиЭ СО РАН

Иевлева Ольга Тихоновна
ЮФУ

Местецкий Леонид Моисеевич
МГУ

Никитин Игорь Николаевич
ИФТИ

Никитина, Ляля Дамировна
ИФТИ

Пилюгин Виктор Васильевич
МИФИ

Ротков Сергей Игоревич
ННГАСУ

Толок Алексей Вячеславович
ИПУ РАН

Алешин Владимир Петрович
МФТИ

Астахов Юрий Сергеевич
ИФТИ

Афанасьев Валерий Олегович
МФТИ

Брусенцев Петр
AVILab

Васин Юрий Григорьевич
ННГУ

Вельбицкий Игорь Вячеславович
Фонд Глушкова, Украина

Венгринович Валерий Львович
ИПФ НАН, Белоруссия

Гуревич Игорь Борисович
РАН

Дижевский Андрей Юрьевич
ИФТИ

Жиляков Евгений Георгиевич
БелГУ

Захарушкин Владимир Федорович
МФТИ

Копылов Андрей Борисович
ТулГУ

Массель Людмила Васильевна
ИСЭ СО РАН

Муратова Галина Викторовна
ЮФУ

Сметанин Юрий Геннадиевич
РФФИ

Судейкин Михаил Иванович
МФТИ

Насонов Андрей Владимирович
МГУ

Сорокин Дмитрий Васильевич
МГУ

Ветров Дмитрий Петрович
МГУ

Захарова Алёна Александровна
ТПУ

Шкляр Алексей Викторович
ТПУ

Знаменская Ирина Александровна
МГУ

Ринкевичюс Бронюс Симович
МЭИ

Столбова Ирина Дмитриевна
ПНИПУ

Международный программный комитет

Timur Paltashev
AMD, USA

Marina Gavrilova
University of Calgary, Canada

Karol Myszkowski
MPI, Germany

Martin Reiser
NTU, Singapore

Vladimir Savchenko
Hosei University, Japan

Alexei Sourin
NTU, Singapore

Olga Sourina
FhG IGD, Germany

Tom Sederberg
Brigham Young University, USA

Andre Gagalowicz
INRIA, France

Sabine Coquillart
INRIA, France

Andres Iglesias
University of Cantabria, Spain

Wolfgang Heiden
BRS, Germany

Andre Hinkenjann
IVC BRS, Germany

Gianluca Mura
Politecnico Di Milano, Italy

Ekaterina Prasolova-Forland
NTNU, Norway

Samet Refik
Ankara Uni, Turkey

Bill Seaman
Duke Uni, USA

Skala Vaclav
University of West Bohemia, Czech
Republic

Iryna Wetzel
Intergraph, Switzerland

Alexander Wolodchenko
Institute for Cartography, Germany

Адрес: **ГрафиКон-2017, Кафедра математического обеспечения вычислительных систем,
Пермский государственный национальный исследовательский университет (ПГНИУ),
614990, Россия, г. Пермь, ул. Букирева, д. 15.**
Тел.: +7 (342) 239-64-35 E-mail: 2017@graphicon.ru Web Site: <http://2017.graphicon.ru/>

Организаторы и спонсоры



АНО Научное Общество «ГрафиКон»
Организатор



Пермский государственный
национальный исследовательский
университет
Локальный организатор



Пермский национальный
исследовательский
политехнический университет
Соорганизатор



Лаборатория компьютерной графики
и мультимедиа ф-та ВМК МГУ
им. М.В. Ломоносова
Соорганизатор



Институт прикладной математики
им. М.В. Келдыша РАН
Соорганизатор



Исследовательская группа
«Неогеография»
Соорганизатор



ФАНО России
Финансовая поддержка



Лаборатория мультимедийных
решений
Официальный спонсор



ООО «Ньюлана»
Спонсор



Пермский естественнонаучный
форум 2017 «Наука и глобальные
вызовы XXI века»
*Информационно-коммуникационная
поддержка*



Компания Macroscop
Media-партнёр

Информация о спонсорах



Леонид Юрьевич Морозов

Леонид Юрьевич Морозов родился в 1983 году в Перми. В 2005 году окончил Пермский государственный университет по специальности «политология». В 2006 году получил диплом Государственного университета Высшая школа экономики по специальности «финансы и кредит». Окончил Школу управления «Сколково» (г. Москва) и получил степень Executive MBA. Проходил зарубежные стажировки в США (Silicon Valley) и Китае (Шанхай). Официально признан одним из лучших выпускников ПГНИУ и ВШЭ.

С 2003 года по 2014 год работал в ОАО АКБ «Урал ФД», где прошел путь от специалиста до заместителя председателя правления банка, члена правления. В 2009 году Леонид Морозов стал членом Совета директоров Управляющей компании «Парма – менеджмент».

С 2014–2017 гг. работал в Правительстве Пермского края в должности заместителя председателя правительства — министра экономического развития Пермского края. Л.Ю. Морозов входит в состав кадрового резерва управленческих кадров Пермского края. Является Членом Наблюдательного Совета Финансово-банковской ассоциации ЕвроАзиатского сотрудничества (г. Москва) и Членом рабочей группы по разработке пространственного развития РФ при Минэкономразвитии РФ. Леонид Юрьевич Морозов также является членом Попечительского совета Уральского Фонда «Белая гора». Имеет награды и благодарности федеральных и региональных органов государственной власти. В 2015 году получил награду «За личный вклад в развитие банковского дела в России».

С 5 июня 2017 года Леонид Морозов является первым заместителем генерального директора ЗАО «Инвестиционная компания «Финансовый дом».



Лаборатория мультимедийных решений

ООО «Лаборатория мультимедийных решений» уже более 6 лет специализируется на разработке мобильных приложений, технологий виртуальной и дополненной реальности, экспозиционных мультимедийных решениях.

Компания поддержана Фондом Развития Интернет-Инициатив, Фондом Бортника, неоднократно воплощала в жизнь лучшие музейные проекты, поддержанные Фондом Потанина.

Мы реализовали более 400 проектов по всей России и за рубежом. Среди них: Правительство Москвы, Кунсткамера, Музеи Московского Кремля, Лаборатория Касперского, Правительство Чукотского АО, Метафракс и др.



*Директор компании «Лаборатория мультимедийных решений»
А.В. Дураков демонстрирует президенту России В.В. Путину
приложения дополненной реальности 8 сентября 2017 года*



ООО «Ньюлана» – российский разработчик кроссплатформенного программного обеспечения. С момента основания в 2008 году и по сей день компания создаёт удобные пользовательские приложения, которые повышают производительность работы профессионалов по всему миру. Продуктами Ньюланы пользуются Microsoft, IBM, Tesla Motors, Broadcom, Hewlett-Packard, Citi Bank и др.

Усилия компании сосредоточились на двух актуальных областях разработки. Первая область – удаленное управление (remote desktop), где многоплатформенные продукты компании, включающие в себя облачный сервис, составляют конкуренцию лидерам рынка. Стремление упаковать набор мощных функций в удобный пользовательский интерфейс привело к возможности использовать получившиеся приложения как для задач небольшого офиса, так и для удалённого администрирования целого парка компьютеров и серверов, распределенного на нескольких континентах.

Вторая область приложения глубоких знаний и универсальных навыков сотрудников Ньюланы – это визуализация данных. Компания создала мультиплатформенную библиотеку для построения интерактивных двумерных и трёхмерных диаграмм и графиков различных типов, а также библиотеку для одномерных и многомерных данных в табличном виде для iOS. Разработчики использовали аппаратное ускорение графики средствами OpenGL(ES) и Direct3D для достижения высокой производительности и максимального качества генерируемого изображения. Программное обеспечение Ньюланы может быть использовано для визуализации произвольных числовых данных в задачах бизнес-аналитики, естественных и гуманитарных наук.



Macroscop – российская инновационная компания, разрабатывающая программное обеспечение для IP-камер.

Под управлением продуктов Macroscop работают 300 000+ IP-камер в 23 000+ видеосистем. Партнерская сеть компании объединяет 3 000+ дистрибьюторов и системных интеграторов в 27 странах мира: России, Франции, Великобритании, США, Австралии, ОАЭ, Нигерии и др.

История Macroscop началась в 2008 году, когда у основателей компании появились две научные идеи, которые сегодня позволяют решать проблему анализа огромного объема данных, получаемых от камер видеонаблюдения. Сегодня Macroscop – это команда специалистов в области видеоанализа, разработки, маркетинга и продаж, которые ежедневно работают над реализацией миссии компании: **«Воплощать то, что казалось невозможным, заставляя видеосистемы приносить максимальную пользу людям».**

Дорогие коллеги!

Организационный комитет 27-ой Международной конференции по компьютерной графике и машинному зрению ГрафиКон-2017 рад приветствовать Вас в Перми!

Пермь, город-миллионер с 295-летней историей, является столицей Пермского края и всего Прикамья. Город расположен на берегах реки Камы, крупнейшего притока р. Волги. В честь нашего города назван отрезок истории Земли длиной в 47 миллионов лет.

27 лет назад, в феврале 1991 в Москве прошла 1-я международная конференция по компьютерной графике и машинному зрению ГрафиКон'91. ГрафиКон объединяет исследователей и разработчиков из многих стран мира с целью сближения науки, индустрии и бизнеса, обмена новыми идеями и результатами, знаниями и опытом в области компьютерной графики, машинного зрения, обработки изображений, систем визуальной аналитики и виртуального окружения, поддержки принятия решений и обработки Больших Данных. В 2005 году на базе этой ежегодной м/н конференции создана автономная некоммерческая организация Научное общество «Графикон», которая активно помогает организаторам конференции из разных городов.

За свою уже многолетнюю историю география конференции значительно расширилась: помимо Москвы и Московской области, ГрафиКон принимали такие города, как Санкт-Петербург, Нижний Новгород, Ростов-на-Дону, Новосибирск, Владивосток. Организатором конференции ГрафиКон в 2017 году впервые выступает Пермский государственный национальный исследовательский университет (ПГНИУ). В конференции принимают участие представители 7 стран. Конференция будет проходить на территории ПГНИУ и Пермского национального исследовательского политехнического университета (ПНИПУ). ПГНИУ входит в пятерку лучших классических университетов России.

В прошлом году ПГНИУ исполнилось 100 лет!

Сотрудники нашего ВУЗа и ПНИПУ выполняют широкий спектр научных исследований в области компьютерной графики, машинного зрения, визуальной аналитики и дополненной реальности, имеющих весомый практический результат, о чем свидетельствуют результаты инновационной и коммерческой деятельности, нашедшие широкое признание, в том числе и на уровне первых лиц государства. 8 сентября 2017 года во время своего рабочего визита в Пермь президент России Владимир Путин высоко оценил уровень программно-аппаратного обеспечения в указанных областях, разработанного сотрудниками и выпускниками ПГНИУ и ПНИПУ.

Организационный и программный комитеты 27-й международной конференции ГрафиКон-2017 желают участникам плодотворной работы, творческого обсуждения научных докладов и идей, успехов и новых встреч!

*От имени организационного и программного комитетов ГрафиКон-2017
Светлана Игоревна Чуприна,
Константин Валентинович Рябинин*

Дорогие участники конференции ГрафиКон-2017!

Разрешите приветствовать Вас на пермской земле в стенах старейшего вуза Урала!

Развитие Российской экономики и российской высшей школы неразрывно связано с цифровыми технологиями. В будущем это станет залогом успешного развития практически любой сферы бизнеса. Крайне важно, закладывая основу российской цифровой экономики, экономики высоких технологий, сформировать общее видение и подходы к тому, как готовить и какие навыки вырабатывать у студентов вузов, какова должна быть экосистема для создания новых и совершенствования уже существующих ИТ-технологий.

Развитие новых ИТ-направлений, таких как Интернет вещей (IoT), Ubi-Computing, интеллектуальный анализ Больших Данных и др. немыслимо без современных средств компьютерной графики. Такие конференции, как ГрафиКон, где презентуются и обсуждаются результаты исследований известных учёных и молодых специалистов в этой области, несомненно способствуют прогрессу цифровых технологий.

Проведение конференции такого высокого уровня в Пермском государственном национальном исследовательском университете подчеркивает признание научным сообществом как высокого профессионального уровня его профессорско-преподавательского состава, так и тот вклад, который ученые и сотрудники университета вносят в фундаментальные и прикладные исследования в области компьютерной графики и машинного зрения. Это в полной мере относится также и к ученым Пермского национального исследовательского политехнического университета, на территории которого будет проходить секция «Геометрическое моделирование. Компьютерная графика в образовании».

Желаю конференции ГрафиКон-2017 успеха, а всем ее участникам плодотворной работы и новых научных достижений на благо своих регионов и нашей страны в целом!

*Первый заместитель генерального директора
ЗАО «Инвестиционная компания «Финансовый дом»,
экс-министр экономического развития Пермского края,
Executive MBA Skolkovo
Леонид Юрьевич Морозов*



Содержание

Пленарные доклады

<i>V. Alexandrov, J. Espinosa, I. Rudomin</i> Scientific Visualization and Visual Analytics Challenges	17
<i>S.V. Klimenko</i> Contemporary Information Technology: Information Visualization, Virtual Environment, Visual Perception and Analysis	18
<i>Е.Н. Ерёмченко</i> Digital Silk Road Alliance: общенаучные вызовы и перспективы	19
<i>С.И. Ротков, Е.В. Попов, Т.В. Мошкова, В.А. Тюрина, Д.Ю. Васин, С.А. Роменский, Н.Л. Макаров, В.Л. Чепкасов</i> Проблема преобразования бумажных архивов чертежно-конструкторской документации в электронную модель изделия и связанные с ней геометро-графические задачи	20
<i>К.А. Sidorov</i> Paleographic Dating of Birch Bark Manuscripts	21
<i>Д.Д. Жданов, А.Г. Волобой, В.А. Галактионов</i> Методы стохастической трассировки лучей как средство создания систем дополненной реальности	22
<i>А.А. Захарова, Е.В. Вехтер, А.В. Шкляр</i> Решение задач анализа данных с помощью аналитических визуальных моделей	23
<i>Д.В. Сорокин</i> Методы совмещения микроскопических изображений в задачах обработки и анализа изображений живых клеток	24
<i>А.В. Дураков</i> Виртуальная и дополненная реальность в экспозиционной деятельности	25
<i>А.В. Разумков</i> МАКРОСКОП: практика применения алгоритмов видеоанализа для решения задач безопасности и бизнес аналитики	26

Реалистичная графика

<i>V.F. Frolov, V.S. Sangarov, V.A. Galaktionov</i> A concept for database oriented 3D graphics engine infrastructure	28
<i>V.I. Gonakhchyan</i> Comparison of hierarchies for occlusion culling based on occlusion queries	32
<i>V.G. Sokolov, D.D. Zhdanov, I.S. Potemin, N.N. Bogdanov, A.D. Zhdanov, E.U. Denisov</i> Reconstruction of BSDF based on optimization of microrelief normal distribution	37
<i>D.D. Zhdanov, I.S. Potemin, A.A. Kishalov, A.D. Zhdanov, N.N. Bogdanov</i> Stochastic ray tracing methods in problems of photorealistic image synthesis for augmented reality systems	42
<i>Б.Х. Барладян, А.Г. Волобой, В.А. Галактионов, В.В. Князь, И.В. Ковернинский, Ю.А. Солоделов, В.А. Фролов, Л.З. Шапиро</i> Программная реализация OpenGL SC для авиационных встраиваемых систем	47
<i>Е.В. Боревич, С.В. Мещеряков, В.Э. Янчус</i> Эффективные методы и модели цифровой обработки киноматериала	51
<i>В. Будаков, В. Желтов, В. Чембаев</i> Компьютерная графика и качество освещения реальных осветительных установок в светотехнике	55
<i>В.А. Дебелов, К.Г. Кушнер</i> Программная модель линзы для обработки лучей линейно поляризованного света	58
<i>А.С. Щербаков, В.А. Фролов</i> Ускорение алгоритма излучательности на графических процессорах	63

Научная визуализация и визуальная аналитика

<i>V.L. Averbukh, M.O. Bakhterev, Z.S. Chayka, P.A. Vasev</i> The Brush Metaphor: an Example of Metaphor Adaptation	69
<i>M. Forghani, P. Vasev, V. Averbukh</i> A Visualization System for Binary Rooted Trees	71
<i>Yu.A. Konstantinov, K.P. Latkin, F.L. Barkov, A.S. Smirnov, M.M. Poskrebyshev, V.P. Pervadchuk, D.B. Vladimirova, Yu.A. Konin, A.I. Garanin, V.V. Burdin</i> Special Optical Fiber Preform Layers 3D-Reconstruction	74
<i>E. Malikova, A. Pasko, V. Adzhiev, V. Pilyugin</i> Sound Matrices approach in multisensory analytics: for multidimensional scalar fields data analysis	79
<i>В.Л. Авербух, Н.В. Авербух</i> Исследование психологических состояний пользователей сред виртуальной реальности	87
<i>С.В. Андреев, А.Е. Бондарев, А.В. Бондаренко, Ю.В. Визильтер, В.А. Галактионов, А.В. Гудков, С.Ю. Желтов, В.Т. Жуков, Е.Б. Иловайская, В.А. Князь, К.В. Мануковский, Н.Д. Новикова, М.В. Ососков, Н.Ж. Силаев, О.Б. Феодоритова</i> Приближенная оптимизация формы узла лопастей энергоустановки	92
<i>Р.А. Багутдинов, С.Г. Небаба, А.А. Захарова</i> Алгоритм обработки разнородных данных для мультисенсорной СТЗ на примере анализа температуры и концентрации газа	97
<i>О.В. Бехтольд, В.Ю. Гудков</i> Модификация алгоритма анализа области частот для отпечатков пальцев	101
<i>А.Е. Бондарев, В.А. Галактионов, Л.З. Шапиро</i> Визуальный анализ и обработка многомерных данных	103
<i>В.Н. Горбачев, И.К. Метелёв, Е.М. Кайнарова, М.А. Полякова</i> Стеганографическая защита изображений из PDF документов на основе конвертора PDF-SVG	108
<i>Д.А. Завьялов, А.А. Захарова, А.В. Шкляр</i> Визуальные методы оценки и модели данных при проектировании разработки месторождений углеводородов ..	112
<i>А.А. Захарова, Е.В. Вехтер, А.В. Шкляр</i> Решение задач анализа данных с помощью аналитических визуальных моделей	116
<i>И.А. Знаменская, Е.Ю. Коротева, В.В. Шишаков, А.В. Хахалин, Е.А. Кузьмичева, С.А. Исайчев, А.М. Черноризов</i> Анализ видеопоследовательностей и термограмм лица для дистантного съема физиологических индикаторов психо-эмоциональных состояний человека	121
<i>В.В. Князь, О.В. Выголов, В.В. Федоренко, В.С. Горбацевич, В.А. Мизгинов, Д.Г. Клевцов</i> Применение глубоких свёрточных нейронных сетей для визуализации препятствий на взлётно-посадочной полосе	125
<i>Д.В. Манаков</i> Модели восприятия визуальной информации	129
<i>Б.С. Ринкевичюс, И.Н. Павлов, И.Л. Расковская, А.В. Толкачев</i> Лазерная визуализация тепловых процессов в прозрачных твердых телах	133
<i>К.В. Рябинин, Д.А. Баранов, К.И. Белоусов</i> Интеграция инструментария научной визуализации SciVi с информационной системой Семограф	138
<i>К.В. Рябинин, С.И. Чуприна</i> Адаптация систем научной визуализации к задачам мониторинга данных от легковесных аппаратных робототехнических устройств	142
<i>И.С. Стародубцев, Г.А. Зуева</i> Особенности реализации естественных интерфейсов для прикладных задач	147

Компьютерное зрение

<i>S. Arseev, A. Konushin</i> Gait Recognition Based on Relative Distance and Motion Features	151
<i>A. Kukleva, V. Konushin, A. Konushin</i> Abandoned Objects Detection in Video Sequences	154
<i>F. Morozov, A. Konushin</i> Background subtraction with convolutional neural network and oversegmentation	158
<i>K.A. Sidorov</i> Paleographic Dating of Birch Bark Manuscripts	162
<i>S. Slyadnev, V. Turlapov</i> Towards cylindrical primitive extraction from industrial laser-scan data by local approximation with refinement	169
<i>I.A. Valuiskaia, E.V. Shalnov, A.S. Konushin</i> Automatic calibration of surveillance video camera	175
<i>Г.В. Аристов, А.В. Ключев</i> Распознавание и классификация микроструктуры металлов и сплавов с использованием глубоких нейронных сетей	180
<i>Д.С. Белобородов, Л.М. Местецкий</i> Выделение связного переднего плана на картах глубины с помощью медиального представления сцены	184
<i>В. Бобков, А. Кудряшов</i> Создание единой текстурной карты для объемной модели, полученной с помощью трехмерного сканера	189
<i>И. Гайниров, Н. Хлебников, И. Обабков</i> Метод захвата движений как средство естественного интерфейса	193
<i>А.В. Горевой, А.С. Мачихин</i> Оценка погрешности измерений геометрических параметров, выполняемых с использованием призмочно-линзовых оптических систем	197
<i>Е.А. Долотов, В.Д. Кустикова</i> Сравнение некоторых методов решения задачи детектирования лиц на изображениях	202
<i>Р.А. Клестов, В.Ю. Столбов</i> Гибридный метод распознавания контуров на изображении на основе технологий компьютерного зрения	208
<i>Л.И. Лебедев, Ю.Г. Васин, М.П. Осипов</i> Автономная навигация в закрытых помещениях с использованием адаптивного описания эталонных изображений	212
<i>В.С. Лютов, А.С. Конушин</i> Глубинные двоичные дескрипторы изображения человека для его повторной идентификации и сопровождения в видео	217
<i>А.А. Михайличенко, А.Б. Клеценков</i> Способ бесконтактного измерения размеров крупногабаритных объектов	223
<i>Л.В. Новоторцев, А.Г. Волобой</i> Выделение отрезков на аэроснимках в рамках задачи обнаружения строений	228
<i>И.Г. Ханьков, М.В. Харинов</i> Кластеризация пикселей цветового изображения методом Уорда	232

Обработка и анализ биомедицинских изображений

<i>A. Khvostikov, J. Benois-Pineau, A. Krylov, G. Catheline</i> Classification methods on different brain imaging modalities for Alzheimer disease studies	237
<i>Н.А. Аношина, Т.Б. Сагиндыков, Д.В. Сорокин</i> Метод синтеза реалистичных данных крио-электронной микроскопии одной частицы и крио-электронной томографии	243

<i>Т.В. Голубятников, С.В. Поршнев, Н.И. Клиорин</i> Компьютерный синтез эхокардиографических изображений сердечной мышцы человека: проблемы и их решения	248
<i>Ф.А. Гурьянов, А.С. Крылов</i> Оптимизация корреляционных методов регистрации биомедицинских изображений	253
<i>В.В. Данилов, И.П. Скирневский, О.М. Гергет</i> Метод поиска эффективной последовательности алгоритмов обработки медицинских данных при сегментации анатомических структур сердца	258
<i>О.О. Евсютин, Р.В. Мещеряков, А.В. Ращупкина</i> Алгоритм встраивания информации в пары блоков ДКП-коэффициентов сжатого цифрового изображения	264
<i>Д.А. Коваленко, В.С. Гнатюк</i> Ассоциация сцен в эндоскопических видео	269
<i>Д.А. Лачинов, А.А. Белокаменская, В.Е. Турлапов</i> Точный автоматический алгоритм детектирования цефалометрических точек на КТ-изображениях	275
<i>А.А. Мельникова, А.С. Крылов</i> Метод оценки ядра размытия движением изображений на основе анализа хребтовых структур	279
<i>Д.В. Сорокин</i> Совмещение изображений флуоресцентной микроскопии для компенсации движения живых клеток: обзор	284

Цифровая Земля и Большие Данные

<i>А.Я. Аноприенко, Е.Н. Ерёмченко, С.В. Клименко</i> Digital Earth как метод визуализации	290
<i>Ю.В. Брылкин</i> Создание цифровых моделей рельефа в трёхмерном пространстве методами фрактальной геометрии	295
<i>Т.П. Варшанина</i> Пространственно-временная визуализация и перспективы прогнозирования геофизических процессов	298
<i>Д.Ю. Васин, М.А. Редькин</i> Распознавание символов на большеформатных растровых изображениях документов со слабо формализованным описанием объектов	303
<i>Н.Ю. Дудаков, И.Н. Мерзляков, А.Р. Ханнанов, Т.Н. Томчинская</i> Разработка системы динамических тестов для начинающих и опытных водителей с применением технологий виртуальной и дополненной реальности	308
<i>А. Кудряшов, И. Соловьёв</i> Построение рельефа при реконструкции сцены городского пространства с помощью сервисов Google Maps	312
<i>Н.С. Пономарев, А.С. Тарасенко, Т.Ш. Утешева</i> Автоматическая генерализация графических объектов с учетом иерархии классов	315

Геометрическое моделирование. Компьютерная графика в образовании

<i>А. Malyshev, S. Slyadnev, V. Turlapov</i> Graph-based feature recognition and suppression on the solid models	319
<i>S. Slyadnev, A. Malyshev, V. Turlapov</i> CAD model inspection utility and prototyping framework based on OpenCascade	323
<i>А.С. Бакенов</i> Т-сплайны: геометрическая гибкость и возможность локального изменения представления поверхности	328
<i>И.Г. Балюба, Е.В. Конопацкий</i> Конструирование дуг обвода из кривых одного отношения	332
<i>А.А. Бойков, А.М. Федотов</i> Использование полифигур для валидации и верификации вычислительных моделей в системе компьютерной проверки решений задач инженерной геометрии	335

<i>Ю.В. Брылкин</i>	
Визуализация различных модификаций алгоритма броуновского движения для построения фрактальных поверхностей	339
<i>Н.Н. Василюк</i>	
Особенности обучения компьютерной графике в рамках курса «Информатика» для непрофильных специальностей и направлений	342
<i>Д.В. Волошинов, Е.С. Казначеева, Е.С. Хайбарахманова</i>	
Обобщенный конструктивный алгоритм преобразования инверсии относительно окружности	346
<i>Н.Д. Жилина, М.В. Лагунова, Т.В. Мошкова, С.А. Роменский, С.И. Ротков, М.М. Смычѣк, В.А. Тюрина</i>	
Синтез параметрической конструктивной модели 3D-объекта	350
<i>В.А. Короткий</i>	
Визуальная компьютерная реконструкция квадратичных соответствий	355
<i>А.Л. Липкина, Н.А. Ломов, Л.М. Местецкий</i>	
Распознавание шрифтов по изображениям на основе анализа структуры букв	359
<i>И.О. Михайлов, И.С. Стародубцев, В.С. Провков</i>	
Виртуальный испытательный полигон для роботизированных охранных комплексов	367
<i>В.А. Семенов, С.Н. Абросимов, Д.Е. Тихонов-Бугров</i>	
Геометро-графическая подготовка в рамках специалитета в русле всемирной инициативы CDIO	370
<i>И.Д. Столбова, А.Б. Шахова</i>	
Интегративная модель геометро-графической подготовки студентов	373



Пленарные доклады

Plenary Talks

GraphiCon 2017

24–28 сентября 2017
Пермь, Россия

Scientific Visualization and Visual Analytics Challenges

V. Alexandrov, J. Espinosa, I. Rudomin
vassil.alexandrov@bsc.es
Barcelona Supercomputing Center, Barcelona, Spain

Abstract

With the latest advanced of Data and Computational Science, the need to be able to visualize and analyze scientific results, as well as represent and deal with vast amount of data (Big Data), becomes of unquestionable importance. This talk focuses on current trends and needs for analyzing and visualizing Big Data. First it analyses the interplay between High Performance Computing (HPC), Information Visualization and Virtual Reality (VR). In particular, the needs of visual analytics for data and compute intensive problems. Some examples combining HPC and high-end visualization and VR approaches will be drawn from areas of Marketing, Customer Relations and Security areas.

Then, it will address different techniques for visualizing Big Data, considering the visualization process as a complex and greedy task. This part will show in particular that, when it comes to Big Data, current tools do not scale well and specific techniques should be applied. Some examples will be presented for processing and visualizing urban data produced in real-time.

Finally we will show how to use this data for predicting particular behavior, using crowd simulation and Machine Learning techniques in combination with visual analytics methods. Sometimes this predicted behavior will be used for prevention, for example, in the case of undesirable behavior in security applications. In other cases, it will be used to strengthen the said behavior in some cases of marketing and customer relations.



Vassil Alexandrov
*(ICREA Research Professor
in Computational Science
at Barcelona Supercomputing Center)*

Contemporary Information Technology: Information Visualization, Virtual Environment, Visual Perception and Analysis

S.V. Klimenko

stanislav.klimenko@gmail.com

Moscow Institute of Physics and Technology, Moscow, Russia

Abstract

The ever increasing complexity of the physical phenomena studied in scientific and engineering disciplines, requires the development of new approaches and powerful technique for processing and analysis of complex and Big data. Scientific Visualization developing methods and tools for understanding the problems to be solved by bringing in a person's ability to see and perceive the image.

The report deals with the preconditions of occurrence of this scientific field, the stages of its formation as a scientific discipline and major achievements. Shows the transition from technology to render virtual environment. Currently, scientific visualization is quite complete scientific discipline. Purpose of virtual environments is to provide users with a virtual workspace, in which they can observe and investigate create the virtual real-time data, models and the scenes. Presents selected results of the development of scientific visualization and virtual environment from the Moscow Institute of Physics and Technology and the Institute of Computing for Physics and Technology in cooperation with research centers in Germany and Singapore.

In conclusion ascertained advantages of visualization technologies and virtual environments to improve the efficiency of fundamental research, to meet the challenges of science, education and industry.

Promising areas of research in this subject area are information visualization, visual analytics, a new man-machine interfaces based on haptic devices and tangible images.

The contents of the report: An Historical Excursus, What is Visualization, From Visualization to Virtual Environment, Contemporary VE Installations, Current State-of-the-Art, Available Base and Research Experience, Where We Go, Support for Decision-Making, Virtual Simulation of Tangible Interaction, Conclusions.



Stanislav V. Klimenko

(Scientific Director

*of The Institute of Computing for Physics and Technology,
full professor and scientific leader*

of The Department CPT at Moscow Institute of Physics and Technology)

Цифровая Земля и визуализация в управлении: вызовы и перспективы

Е.Н. Ерёмченко
eugene.eremchenko@gmail.com
Группа «Неогеография», Протвино, Россия

Аннотация

В докладе даётся общая информация об альянсе «Цифровой Шёлковый Путь» (Digital Silk Road Alliance, DSRA), учреждённом под эгидой Международного Общества Цифровой Земли (International Society of Digital Earth, ISDE) в апреле 2017 года в качестве системы поддержки принятия решений континентального масштаба в рамках проектов «Шёлкового Пути и Шёлкового Пояса». Идеология DSRA базируется на трёх принципах: Цифровая Земля, широкое использование космических данных дистанционного зондирования, и Big Data.

Обосновывается трактовка ЦЗ как научной революции (в терминах Т. Куна) в области геовизуализации. Рассматривается хронология развития ЦЗ в 2017 году.

Обозначаются перспективы вовлечения методов научной визуализации и визуальной аналитики в решение практических управленческих задач.



Евгений Николаевич Ерёмченко

*(Руководитель группы
«Неогеография»,
член совета ISDE,
соучредитель DSRA,
научный сотрудник
Московского государственного
университета им. Ломоносова)*

Проблема преобразования бумажных архивов чертежно-конструкторской документации в электронную модель изделия и связанные с ней геометро-графические задачи

С.И. Ротков, Е.В. Попов, Т.В. Мошкова, В.А. Тюрина, Д.Ю. Васин,
С.А. Роменский, Н.Л. Макаров, В.Л. Чепкасов
gotkov@nngasu.ru

Нижегородский государственный архитектурно-строительный университет,
Нижний Новгород, Россия

Аннотация

Переход на цифровое производство и экономику (программа «Индустрия 4.0») предполагает повсеместное внедрение соответствующих информационных систем и технологий. Системы класса CAD/CAM/CAE/PDM/PLM etc, в своей основе требуют наличия 3D модели объекта исследования и производства, с одной стороны, а с другой, генерируют чертежно- конструкторскую и технологическую документацию, необходимую для физического изготовления объекта. Поскольку проектирование, как правило, идет на основе аналогов и прототипов, то для работы упомянутых систем требуется наличие достаточно большого электронного архива моделей и чертежей. В то же время на предприятиях и организациях за большой промежуток времени накопился огромный архив чертежной документации на бумажной носителе, в объеме доходящий до десятков миллионов листов, которые необходимо ввести в ЭВМ и согласовать с различными информационными системами. В докладе рассматривается организация и создание информационной технологии потокового преобразования данных с бумажного носителя в электронную модель изделия, включающую в себя, в том числе, и преобразование геометрических данных 2D→ 3D. Технология включает в себя решение задачи преобразования «растр-вектор», ориентированной на распознавание и анализ элементов чертежа, формирование промежуточной базы данных, направленной на согласование форматов данных с информационными системами и создание новых функциональных операторов решения прикладных проблем пользователя, не реализуемых имеющимися в системах средствами, генерацию 3D модели объекта. Рассматриваемая технология является инвариантной по отношению к отраслям науки и промышленности.

Работа выполнена по грантам РФФИ № 15-07-05110, 15-07-01962, 17-07-00543.



Сергей Игоревич Ротков

*(Доктор технических наук, профессор,
Заведующий кафедрой инженерной геометрии,
компьютерной графики и автоматизированного проектирования
Нижегородского государственного архитектурно-строительного университета)*

Paleographic Dating of Birch Bark Manuscripts

K.A. Sidorov
sidorovk@cardiff.ac.uk
Cardiff University, Cardiff, UK

Abstract

We address the problem of estimating the age of an important class of historical documents: mediaeval manuscripts on birch bark, found in Novgorod and other cities in Northwest Russia. These birch bark manuscripts are one of the most valuable corpora of Old Russian texts, and a crucially important source of information about mediaeval history and evolution of the language. Accurate dating of BBMs is necessary in order to place them in a correct historical context, before historians and linguists can take advantage of their valuable contents.

In this work, we investigate how the age of birch bark manuscripts can be automatically estimated from their appearance (paleographically), using computer vision and machine learning techniques. Our method achieves mean absolute accuracy of 18.9 years which is comparable to or surpasses the performance of human experts and of other computational paleography studies. Therefore, our results may corroborate and refine existing paleographic analysis methods.



Kirill Sidorov
(*PhD, Assistant Professor,*
School of Computer Science and Informatics,
Cardiff University)

Методы стохастической трассировки лучей как средство создания систем дополненной реальности

Д.Д. Жданов¹, А.Г. Волобой², В.А. Галактионов²
ddzhdanov@mail.ru|voloboy@gin.keldysh.ru|vlgal@gin.keldysh.ru

¹НИУ ИТМО, Санкт-Петербург

²ИПМ им. М.В. Келдыша РАН, Москва

Аннотация

С развитием компьютерной техники и средств коммуникации человека с цифровым миром все большую популярность приобретают технологии дополненной реальности (augmented reality). Если в 90-х годах прошлого века ставилась задача генерации реалистичной картинки с изображением виртуальной модели, встроенной в фото- или видео-окружение, то современные задачи касаются широкого круга приложений – от создания систем отображения информации на лобовом стекле автомобиля до генерации тестовых изображений для обучения систем беспилотного вождения. Создание современных комплексов виртуальной и дополненной реальности требует разработки высококачественных оптических систем и согласования работы данных систем с процессами конвергенции, аккомодации и световой адаптации глаза. Это порождает новые задачи их виртуального прототипирования и анализа распространения света.

В докладе рассматривается круг современных задач, связанных с системами дополненной реальности, и основные подходы к их решению. Авторы много лет работают в этой области, создавая соответствующее программное обеспечение. Методы стохастической трассировки лучей предлагаются как одно из наиболее универсальных средств анализа, моделирования и виртуального прототипирования таких систем. В частности, рассматривается решение, описанное в статье «Stochastic ray tracing methods in problems of photorealistic image synthesis for augmented reality systems», представленной на конференции.



**Дмитрий Дмитриевич
Жданов**

*(Кандидат
физико-математических наук,
заведующий кафедрой
Технологий визуализации НИУ ИТМО)*



**Алексей Геннадьевич
Волобой**

*(Доктор физико-математических наук,
ведущий научный сотрудник
Отдела Компьютерной графики
и вычислительной оптики
ИПМ им. М.В. Келдыша РАН)*



**Владимир Александрович
Галактионов**

*(Доктор физико-математических наук,
профессор, заведующий
Отделом Компьютерной графики
и вычислительной оптики
ИПМ им. М.В. Келдыша РАН)*

Решение задач анализа данных с помощью аналитических визуальных моделей

А.А. Захарова, Е.В. Вехтер, А.В. Шкляр
zaa@tpu.ru|vehter@tpu.ru|shklyarav@tpu.ru

Институт кибернетики Томского политехнического университета, Томск, Россия

Аннотация

Дается обзор результатов исследований, направленных на изучение потенциальных преимуществ визуального анализа. Обозначаются слабые места в современном состоянии средств визуального анализа данных, связанные с недостаточным уровнем систематизации и обобщения знаний о возможностях визуализации. Предлагаются некоторые базовые определения, связанные с использованием средств визуализации для решения прикладных задач, в том числе задач анализа больших объемов данных, обучения пользователей и управления их восприятием. Обсуждаются возможности перехода от визуальных образов данных к методам представления информации, аналогичным использованию текстовых форм различного уровня абстракции.

Демонстрируются возможности и преимущества использования авторского метода комплексного визуального анализа, возникающие при решении практических задач. Сформулирован подход к построению визуальных моделей как инструментов визуальной аналитики. Предложена базовая методика проведения аналитического исследования произвольных данных, позволяющая ввести понятие формальной визуальной модели.



Алена Александровна Захарова
(Доктор технических наук,
Институт кибернетики,
Томский политехнический университет)

Методы совмещения микроскопических изображений в задачах обработки и анализа изображений живых клеток

Д.В. Сорокин^{1,2}
dsorokin@cs.msu.ru

¹Лаборатория математических методов обработки изображений,
факультет вычислительной математики и кибернетики,
Московский Государственный Университет имени М.В. Ломоносова, Москва, Россия;

²Centre for Biomedical Image Analysis,
Faculty of Informatics, Masaryk University, Brno, Czech Republic

Аннотация

В докладе представлен обзор методов совмещения микроскопических изображений, представляющих собой важнейший шаг в задачах обработки и анализа изображений живых клеток. Качественное совмещение изображений необходимо для компенсации движения клеток и последующего изучения локального движения клеточных субструктур. Несмотря на большое разнообразие методов совмещения микроскопических изображений, все они имеют свои особенности и границы применимости. Выбор метода для реальных приложений обусловлен спецификой биологического эксперимента, и, как следствие, полученных данных. В докладе основное внимание уделено обсуждению достоинств и недостатков различных методов совмещения микроскопических изображений и границ их применимости в реальных приложениях.



Дмитрий Васильевич Сорокин

*(Кандидат физико-математических наук,
старший научный сотрудник
лаборатории математических методов обработки изображений
факультета Вычислительной математики и кибернетики
МГУ имени М.В. Ломоносова,
researcher at Centre for Biomedical Image Analysis,
Masaryk University, Brno, Czech Republic)*

Виртуальная и дополненная реальность в экспозиционной деятельности

А.В. Дураков
adurakov@gmail.com
Лаборатория мультимедийных решений,
Пермь, Россия

Аннотация

Технологии дополненной и виртуальной реальности появились много лет назад, предлагая миру массу полезных применений во многих сферах. Однако на текущий момент по-настоящему массовое применение эти технологии находят в сфере развлечений. Именно в этой сфере Лаборатория Маугри реализовала более десятка крупных проектов с использованием технологий дополненной и виртуальной реальности. Наиболее интересные проекты будут представлены на докладе, а также будут рассмотрены перспективы массового применения этих технологий в других сферах.



Андрей Викторович Дураков
(Директор компании «Лаборатория мультимедийных решений»,
старший преподаватель ПГНИУ)

МАКРОСКОП: практика применения алгоритмов видеонализа для решения задач безопасности и бизнес аналитики

А.В. Разумков
razumkov@macroscop.com
Macroscop, Пермь, Россия

Аннотация

Macroscop разрабатывает коробочное ПО для видеонаблюдения, которое поставляется неограниченному кругу пользователей со стандартным для всех них функционалом. Это существенно образом отражается на требованиях к разрабатываемым для такого ПО алгоритмам видео анализа. В отличие от алгоритма, разрабатываемого для решения конкретной задачи одного или нескольких пользователей, где есть возможность адаптировать его под конкретные условия использования, алгоритм для ПО с неограниченным кругом пользователей должен обеспечивать целевую точность при различных параметрах анализируемой сцены, освещенности и пр. Вследствие этого наборы видео данных, используемые для обучения алгоритмов, оценки точности и тестирования на предмет ее деградации при доработках алгоритмов должны быть значительно более широкими. При этом размер указанных наборов – лишь один из ключевых параметров. Несмотря на то, что коробочное ПО для видеонаблюдения позволяет определить требования к параметрам установки камер, анализируемой сцены и освещенности, которые должны соблюдаться пользователями для обеспечения определенной точности, сложность формализации таких требований также обуславливает необходимость расширения спектра условий, в которых должна обеспечиваться целевая точность.



Артем Владимирович Разумков
(Генеральный директор компании Macroscop)



Реалистичная графика

Realistic Graphics

GraphiCon 2017

24–28 сентября 2017
Пермь, Россия

A concept for database oriented 3D graphics engine infrastructure

V.A. Frolov^{1,2}, V.S. Sangarov³, V.A. Galaktionov¹

¹Keldysh Institute of Applied Mathematics (Russian Academy of Sciences), Moscow, Russia;

²Moscow State University, Moscow, Russia;

³Gubkin russian state university of oil and gas, Moscow, Russia

Integration layer between digital content creation software (DCCS) and rendering software in a form of specialized database is proposed in this paper. In our approach, we focus on providing fast 3D-scene updates, ability to work with large digital assets (not fitting into memory), importing and exporting arbitrary parameters, serialization, convenient debugging tools and distributed rendering. Such database can be used as means to integrate different rendering engines with DCCS and also to transfer data between different DCCS.

Keywords: software architecture, rendering, process interoperation.

1. Introduction

Fast growth of computing power in the past 20 years gave rise to new research and industrial fields. Data volumes have grown from megabytes to gigabytes and performance has raised from megaflops to teraflops. But it's not only about the numbers, more important is how it influences interaction with the user. As technology matures, it generally becomes more accessible and friendly for the users. This also holds true for digital content creation. Since rendering is inevitably tied with such applications, rendering engines' developers need to answer the challenges created by the need to improve user experience. In our work we aim to answer these challenges with a database oriented approach to rendering engine infrastructure design.

1.1. Modern digital content creation process requirements

1. *Interactivity* of digital content creation process (WYSIWYG paradigm). Working in the interactive and non-interactive modes can be thought of as editing a text document in MS Word and in \LaTeX . The latter has certain advantages, however, creating a document in MS Word is faster.
2. *No restrictions on memory*. The whole 3D scene may not fit into RAM of a single computer, but content creation process and rendering (at the very least in preview mode) should not have delays [1].
3. *Parameters variability* and extensibility of both DCCS and rendering systems. Different projects may require different computer graphics algorithms to be used (like for hair and fur). It's quite common for post-production studios to create their own rendering plugins both for DCCS and renderers [2].
4. *Serialization*, import and export. Artists reuse variety of digital assets from their previous works or content repositories and therefore it's necessary to import and export everything. Moreover, in the visual effects and animation industry pipeline content needs to be passed between different software back and forth [3, 4].
5. *Debugging* and testing. While working on complex projects it is almost inevitable for some errors and bugs to appear. Some of these errors appear only for a certain order of actions and thus can be hard to reproduce.

To isolate and fix these errors, it's important to track changes in the scene.

6. *Distributed rendering* [4] and explicit transfer of changes. Changes made in the DCCS should be visible as the same for every other computer participating in rendering. It's unacceptable to send the whole scene over the network, we need to *track and transfer changes only*.

2. Previous work

2.1. Wavefront OBJ, Filmbox FBX and others

The most basic way to transfer digital content between applications is to use binary or text files with strictly defined format. For example, simple but limited OBJ file format [5] or more complex and flexible FBX [6]. In the case of strictly defined file formats there's always a trade-off between flexibility and complexity. The problem is that it's impossible to predict what features DCCS or rendering engine developers will need in the future - what parameters would materials have, will there be new forms of geometry and light sources, etc.

2.2. OpenCollada and Alembic

OpenCollada is a step forward compared to rigid file formats. The main difference is that OpenCollada defines only the standard for storing objects (called «COLLADA») in XML [7] and allows developers to *add new parameters*. It's possible since OpenCollada is an open source project and uses XML descriptions both for DCCS and rendering engine. More recent format with similar ideology is **Alembic**, presented on SIGGRAPH in 2011 [8]. The main focus of Alembic is to efficiently store complex animation - key framed or a product of any kind of simulation (fluid, cloth, etc.). However, support for materials in Alembic is still not available [9], so one needs to use additional file formats like MaterialX [10] to transfer such data.

2.3. Limitations of «just files»

In the best case, only 2 or 3 of the requirements for modern digital content creation can be met with common file formats (most likely parameters variability, serialization and debugging/testing). To move forward there needs to be a way for working with large scenes, tracking and logging changes, and, most importantly,

prompt transfer of changes between different software (without rewriting the whole file).

2.4. DRAM and DLL plugins

The opposite of using files for the integration purpose is to transfer data structures through memory or shared memory via serialization as in [11]. Dynamic Loading Library (DLL) plugins closely integrated with DCCS are worth mentioning specifically. These plugins usually directly call virtual functions of DCCS and thus don't import anything. For example, rendering plugin can call a function called *Shade* for material evaluation. While possessing all of the advantages of being closely integrated, such approach limits the performance, scalability and does not guarantee correctness. You can never be sure that function *Shade* is implemented sufficiently effective and that it does *what render developers want* it to do. Moreover, this type of interaction between rendering system and DCCS strongly depends on DCCS in question [12] and that in turn complicates debugging process and integration into any other DCCS.

Both of these approaches, that we just described, are simple and usually provide sufficient performance. However, they can meet only two requirements - interactivity and parameters' variability.

2.5. SDB

The comprehensive explanation of some of the requirements we listed in the beginning of our paper can be found in [4]. Relying on their experience, authors of [4] conclude that rendering engine should exist in the form of an API to the Scene Database (SDB) and describe functionality of such database. The implementation was not specified thus leaving an open space for future research.

2.6. USD

Universal Scene Description (USD) developed by Pixar was revealed to the public in 2016 [3]. This technology was primarily designed to address issues arising when different artists work on digital content for a big project like an animated feature film. USD assumes that digital content is created by *different people in different applications*.

Basically, USD is represented by a set of files in a JSON-like format which are organized together by the means of references and compositions. For example, a composition of a location scene and a scene with animated character. This approach allows group of artists to work efficiently and simultaneously on a film shot components which are called layers and can be combined into a final result or reused in other shots. The resulting hierarchy of digital content created in various software allows tracking the history of scene creation and making several versions of the whole scene or it's components. Also, by using references and delayed loading USD makes it possible to work with large scenes [3].

While USD satisfies a lot of requirements mentioned in the beginning of this article, it's difficult to use this technology directly as an integration layer with rendering

system. The main problem is the absence of mechanism for fast updates. USD does not use global identifiers for objects and *with every change* one needs to recursively check all of the file hierarchy. Moreover, the fact that any file can be overwritten complicates the implementation of distributed rendering, since some of the scene files on different rendering nodes can have different versions.

2.7. Multiverse

Product called Multiverse being developed by J-Cube [13] was initially designed to tackle the problem of loading large scenes (alembic files in particular) and working with them in Autodesk Maya. To achieve this Multiverse takes over the job of loading and otherwise accessing geometry from the editor and implements delayed loading. The scene data is streamed directly into the rendering system or into OpenGL-based interactive viewport of the editor (i.e. Maya).

The downside of this approach is that Multiverse becomes too closely integrated with DCCS and *reimplements* many of it's functions. If system like Multiverse is *already integrated* with a certain DCCS then it could be used as integration layer between the DCCS in question and a rendering system. Otherwise, amount of work required to integrate Multiverse with a DCCS can be tenfold more than any other integration approach. Currently, Multiverse is available only for Maya and Katana, and does not support 3Ds Max, Blender or any of the CAD/CAM systems like Rhino or CATIA.

2.8. Bunsen

Project Bunsen being developed by The Foundry [14] is a cloud-based software for assembling the final scene from different digital assets which supports import from various DCC and CAD/CAM software.

Bunsen provides users with the ability to process imported data in the most suitable for the task at hand way through a *data processing node graph*. The nodes in this graph can perform variety of operations such as converting splines into polygons, polygon mesh optimization, materials assignment, applying level-of-detail (LOD) techniques. If any of the assets used in a particular scene is changed, user needs to re-export it. Bunsen will recognize that one of the asset files has changed and will reload it and execute all nodes dependent on that asset again. The scene data is then prepared and streamed into selected imaging software – interactive or «offline» renderer.

Since Bunsen is in the stage of active development, many details about it are not yet known to the public [14]. Nevertheless, the announcement of such system shows that there's a demand for novel integration solutions oriented on user-friendliness and emerging technologies as we pointed out in the begging of this paper.

3. Suggested approach

Our concept can be thought of as a hybrid of object oriented database and a version control system, (like Git or

Subversion) with no-overwrite strategy of making changes. We believe that between the Editor and the rendering engine should exist special API. This API needs to fulfil the requirements listed in the beginning of the article in a simple and transparent manner (Fig. 1).

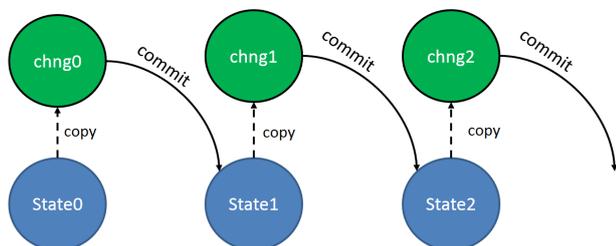


Fig. 1. Our intermediate layer (API) treats 3D content creation like working with source code – make changes and explicitly commit them.

A single state of the scene is a single XML file called «state_001.xml» in some directory «myscene». This file references several text or binary files in subdirectory «data» (usually geometry and textures). The internal XML structure is subdivided into «library» and «scenes». The library contains references to all external files in subdirectory «data» and describes materials, lights and camera. A single scene is just a list of geometry instances. Instance is a reference to geometry object with custom transformation matrix and material remap list (if one should have instances with different materials).

Neither XML files, nor geometry or textures (thinking of them as external files in subdirectory «data») should be actually saved to the hard drive. When the Editor passes them to the render, they are transferred through Operating System (OS) shared memory (Fig. 2). We will discuss this issue more specifically when we talk about «Virtual Buffer».

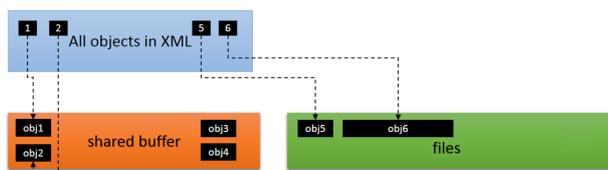


Fig. 2. New or changed objects are always placed in the OS shared memory cache.

3.1. Making changes

A distinctive feature of the proposed technology is explicit tracking and recording the changes. Our API has 3 methods for each object (geometry, material, light and others) – *Create*, *Open* and *Close* (like files in OS). *Create* method makes an empty object. The pair of *Open* / *Close* allows to change the object. When user code (from the Editor side) calls *Open* for some existing object, the copy of the object's XML parameters will be automatically created. Next, user can work with this copy, changing its

XML parameters in RAM via pugixml [15]. When the work is finished *Close* method should be called. From this point on, the new state of the object is considered as «ready to commit», but it is still stored as a separate copy (Fig. 3, right).

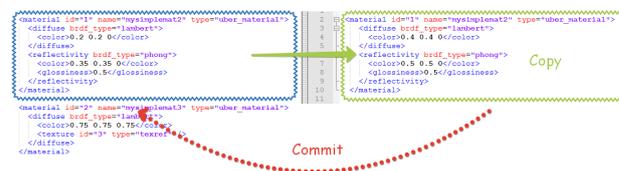


Fig. 3. User of our API can only change an XML copy of an object (right). After «commit» this copy will replace original XML description (left) and form new state file.

We can say that if the current state of the object is stored in the file «state_001.xml», then the new state of the object is stored in a separate file «change_001.xml» (Fig. 1). All these files do not need to be saved to the hard drive – they can be stored in RAM in some dynamic structures (used by the pugixml library in our case). The user may change any number of any objects (including changing one object several times). The file «change_001.xml» will contain only last changes for these objects. Thus it is important to note that file «change_001.xml» will not contain any information about objects that were not actually changed.

Finally, the Editor calls *Commit* to pass the new scene state to the renderer engine. The *Commit* operation creates file «state_002.xml» in which old objects from «state_001.xml» are replaced by their copies from «change_001.xml». It should be clarified that during the execution of the *Commit* operation, only new, modified objects and their XML nodes will be passed to the renderer to update their states inside render engine. Thus, the new render state will be consistent with file «state_002.xml», however, the state file itself is not analyzed. It should be mentioned that it does not matter in which order user changes objects. After *Commit* operation has been called, the API will pass all changes to the render in a fixed, well-known order. Thus the rendering engine developers may rely on fixed and *well-known* sequence of calls from our API.

3.2. Virtual Buffer

To handle big data (geometry and textures) we use a concept of virtual append-buffer with infinite size. The buffer can append linear data blocks to its end. But only last *N* Megabytes are put into RAM. The rest of the buffer is flushed to a hard drive as a set of chunks – binary files in *data* subdirectory (Fig. 2). It should be outlined that any new or changed object is *always* appended to the end of the buffer. This is due to the no-overwrite strategy. If you need to change a texture placed in «chunk_036.bin», you'll have to create a copy of this texture and place it into «chunk_037.bin». The XML description will change the reference from 36 chunk to 37, but both chunks will exist in the buffer «until the end of time». This way we can

be sure that most of the objects that user works with are placed in RAM. For common execution scenario any object that was flushed to a hard drive (in the green rectangle in Fig. 2) may only be in 2 states inside renderer. First state: this object has already been passed to the renderer and has valid state inside of it. Second: the object will not be passed to the renderer at all since it is not needed anymore (details further).

3.3. Why no-overwrite

We use no-overwrite strategy due to support for network rendering. Allowing to overwrite any file would create a possibility that on one machine this file will have an old state, and on the other a new one. No-overwrite strategy guarantees this can never happen. If the texture file is not on the local machine, it means that the file has not been transferred to this machine yet (and then the renderer waits for the transfer of this file), or it would not be transferred, because the system has a new state for the same texture. In the latter case, the renderer must go to the next state of the scene and wait for the new file of the same texture to be transferred and ignore the old one.

4. Results and discussion

The API is integrated with a freeware rendering system Hydra Renderer and two DCC applications – 3ds Max and Fabric Engine. The suggested concept of DCC application and renderer integration meets all of the requirements that we listed in the beginning of the paper. Among the drawbacks of the proposed approach, one can note an obvious overrun of disk memory caused by the need to store complete copies of different versions of the same object. Nevertheless, effective network transfer of such copies is possible (for example, by using paged memory for virtual buffer and COW [16] for each page) but wasn't considered by us.

It should be noted that in the proposed approach the rollback to some previous system state in most cases is equivalent to loading this state «from scratch», – a complete analysis of some file «state_N.xml» and loading most of the *needed* (i.e. we don't have to scan the whole virtual buffer and can load only necessary chunks) data for geometry and textures from the hard drive.

5. Acknowledgments

This work is sponsored by RFBR 16-31-60048 «mol.a.dk» and 16-01-00552.

6. References

- [1] Ingo Wald Andreas Dietrich and Philipp Slusallek. An Interactive Out-of-Core Rendering Framework for Visualizing Massively Complex Models. // Eurographics Symposium on Rendering (2004).
- [2] Marsel Khadiyev. Ornatix mentalRay shaders. // third party plugin for mental ray for hairs.
- [3] Pixar USD. URL: <https://graphics.pixar.com/usd/docs/index.html>.

- [4] Khodulev A.B., Kopylov E.A., Zdanov D.D. Requirements to the Scene Data Base // Proc. 8th International Conference on Computer Graphics and Visualization, Moscow, 1998, p. 189-195.
- [5] Wavefront obj file format. URL: https://en.wikipedia.org/wiki/Wavefront_obj_file.
- [6] Autodesk FBX format. URL: <https://www.autodesk.com/products/fbx/overview>.
- [7] OpenCollada. URL: <http://www.opencollada.org>.
- [8] Siggraph 2011, Alembic talk.
- [9] Alembic. URL: <http://www.alembic.io/>.
- [10] MaterialX An XML standard for export and import a shader graph. URL: <http://www.materialx.org/>.
- [11] Дерябин Н.Б., Денисов Е.Ю. Объектно-ориентированная инфраструктура систем компьютерной графики // Graphi'Con 2007, Россия, МГУ июнь 23-27, 2007, с. 289-292.
- [12] Барладян Б.Х., Волобой А.Г., Шапиро Л.З. Построение реалистичных изображений в системах автоматизированного проектирования // Графи-кон 2013, 16-20 сентября 2013 года, с.148-151.
- [13] Multiverse. URL: <http://multi-verse.io/>.
- [14] Adam Glick, George Matos. Scalable Enterprise Visualization, GPU Technology Conference 2017.
- [15] Arseny Kapoulkine. A light-weight C++ XML processing library. URL: <http://pugixml.org>.
- [16] Bovet, Daniel Pierre; Cesati, Marco. Understanding the Linux Kernel. // O'Reilly Media, Inc. p. 295. 2002.

Comparison of hierarchies for occlusion culling based on occlusion queries

V.I. Gonakhchyan
pusheax@ispras.ru

Ivannikov Institute for System Programming of the RAS, Moscow, Russia

Efficient interactive rendering of large datasets still poses a problem. Widely used algorithm frustum culling is too conservative and leaves a lot of hidden objects in view. Occlusion culling with hardware occlusion queries is an effective technique of culling of hidden objects inside view. In the paper, we perform the comparative analysis of popular indexing techniques in conformity to occlusion culling.

Keywords: Occlusion culling, occlusion query, opengl draw call.

1. Introduction

Occlusion culling algorithms remove occluded objects to enhance frame buffer composition speed. In the paper, object is defined as a mesh that can be rendered using one draw call. Two categories of invisible objects are illustrated in fig. 1: occluded objects and objects beyond camera frustum. Frustum culling quickly discards objects beyond camera frustum but leaves occluded objects inside the camera. We consider the algorithm to find occluded objects. The goal of this paper is to analyze how different hierarchies for scene organization affect occlusion culling and rendering performance.

Occlusion culling algorithms are described in-depth in seminal papers [2][5]. We are interested in online algorithms based on hardware occlusion queries that are widespread today.

Some of the previous algorithms were implemented using OpenGL 2 which is using the immediate rendering mode [10]. In immediate mode all of the geometry is sent each frame resulting in CPU-GPU bandwidth bottleneck and driver overhead caused by excessive number of commands in driver queue. Display lists were used to compile multiple rendering commands once and reduce driver overhead. OpenGL 3 introduced retained rendering mode and immediate mode techniques such as display lists were deprecated and later removed from specification. It raised the question of effective use of object indexing techniques in modern OpenGL. Previously developers had to worry about the amount of geometry passed to GPU. Now developers have to think about driver overhead and state changes as well [18]. In this paper, we analyze performance considerations when using different hierarchies in retained rendering mode.

Although Vulkan API is out of the scope of this paper, it was shown that it gives a better multithreaded performance by utilizing all of the cores of CPU for rendering [3]. It is still based on the same hardware and using retained rendering mode so all of the results in the paper stand.

Hardware occlusion query is GPU technique to find visible faces of a polyhedron [6]. Visibility check stops when the first visible face is found. Checking query result in the frame queries were sent is a blocking function call and it causes CPU starvation. Visibility query result is available without noticeable delay in the next frame because by that time all draw calls required for query execution were sent. So visibility information in given frame is based on the previous frame.

Space coherency is a relationship between nodes in which one node's visibility determines the visibility of other nodes. For example, if building is invisible then objects inside building are also invisible. Time coherency determines visibility in the future by visibility at given time. For example, if object is visible in the given frame then we can consider it visible some number of frames and avoid sending expensive queries.

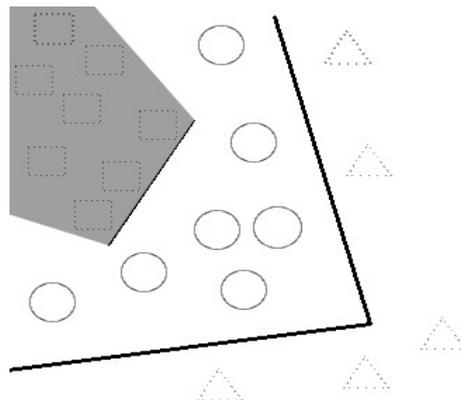


Fig. 1. Different types of culled objects. Frustum culled objects are depicted as triangles. Occluded objects are depicted as rectangles.

2. Previous work

Greene et al. introduced occlusion culling algorithm based on hierarchical z-buffer [7]. Z-buffer is stored as pyramid structure. Z-buffer is divided into 4 components, each one having maximum z value in its region. The process is repeated until pixels are reached. Z-pyramid allows for quick triangle culling by comparing minimum z value of triangle and z value in the corresponding region. This method can be implemented in hardware and software. Software implementation requires expensive rasterisation and pyramid structure updates on CPU. In hardware implementation (ATI Hyper-Z) triangles still have to be transformed and rasterized on GPU which does not replace quick method for geometry culling with a hierarchical structure.

Bittner et al. considered ways for optimal usage of occlusion queries for hierarchical scenes based on extension NV_occlusion_query [1]. Previous frame visibility results are used in the next frames. Main performance problems come from CPU starvation and GPU starvation. Visibility results are checked in the next frame to remove CPU starvation. Previously visible objects are rendered at the beginning of the current frame to remove GPU starvation. Authors used kD-tree constructed according to the surface-area heuristic [11]. Visibility queries are sent for leaf nodes and results are propagated to upper hierarchy levels.

Guthe et al. observed that in many cases visibility queries make performance worse than frustum culling, proposed probability criterion to minimize the number of queries, performance model which helps to avoid queries when rasterisation is cheaper [9]. Mattausch et al. suggested further ways of minimizing the number of queries like sending one query for group [12]. Authors used p-hbvo (polygon-based hierarchical bounding volume decomposition) [13], which is

well-suited for static scenes and expensive to maintain for dynamic scenes.

Software rasterisation and visibility checks can be used instead of hardware occlusion queries [4]. First, triangles of significant occluders are rendered on CPU, hierarchical z-buffer of specified resolution is created. Then bounding boxes of objects or hierarchy nodes are rasterized to determine their visibility against created z-buffer. This helps to avoid expensive occlusion query read-back but requires occluder selection that is best done manually. Also, occluder rasterisation can be expensive and low level-of-detail models give only approximate results.

Scene preprocessing can be effectively applied to static scenes. Teller et al. proposed to use BSP tree with decomposition along axes for architectural scenes [17]. Achieved hierarchy corresponds to room structure in a building. Visibility information is stored as the graph with rooms and portals. Room visibility can be determined by rasterizing portals. However, that graph is expensive to compute and works effectively only for static scenes. Commercial solution Umbra computes voxel representation of a scene [14]. Empty voxels serve as portals between different parts of a scene. Software rasterisation of portals determines the visibility of parts of a scene. That algorithm effectively finds occluded objects for static 3d scenes and is widely used in video games.

Greene introduced image space algorithm based on precomputed occlusion masks [8]. As input, it takes a list of polygons in front-to-back order. It recursively subdivides image space into quadtree until visibility of polygon can be determined for each quadrant. Main advantages of this approach are small memory requirements and no pixel overwrites. However, it requires special hardware to implement efficiently. Zhang et al. proposed visibility culling based on hierarchical occlusion maps, which is better suited for modern hardware [19]. It constructs occlusion map hierarchy by rendering chosen occluders and then traverses bounding volume hierarchy of the model database to perform visibility culling. The algorithm allows for approximate visibility when opacity threshold is set to value lower than one. The main disadvantage of the algorithm is the sophisticated process of occluder selection which favors large objects with small polygon count for faster construction of occlusion map hierarchy.

3. Occlusion culling algorithm

Hierarchical occlusion culling algorithm in this paper is based on Coherent hierarchical culling [1]. Although our implementation does not include query batching, tight bounding volumes, probabilistic estimation of visibility, it allows comparing benefits and limitations of different subdivision hierarchies.

Pseudocode of the algorithm:

```
function RenderFrame(rootNode, frustum,
queries, sentNodes)
  PerformFrustumCulling(frustum, rootNode)

  for i in 0..queries.size-1
    vis <- GetQueryResult(queries[i])
    SetNodeCulled(sentNodes[i], vis == 0)
  end for
  PropagateVisibilityUpHierarchy()

  nodes <- GetVisibleNodesInFrustum()
  for n in nodes
    for inst in n
      if !InstRendered(inst)
        Render(inst)
        SetInstRendered(inst)
      end for
    end for
  end for
```

```
end for

sentNodes <- GetLeafNodesInFrustum()
for i in 0..sentNodes.size-1
  SendQuery(queries[i], sentNodes[i]
    .boundingBox)
end for
```

end function

Function "RenderFrame" renders one frame. Function "PerformFrustumCulling" recursively sets frustum culled bit for every node outside the frustum. First for loop checks visibility results of occlusion queries sent in the previous frame. Second for loop renders objects of visible nodes in the frustum. Last for loop sends queries for all leaf nodes in the frustum.

During the first frame, all objects inside frustum are rendered, and all hierarchy leafs inside frustum are queried. During the second frame, query results are checked, and only visible objects are rendered. Hierarchy leafs inside frustum are queried each frame. We propagate visibility up the hierarchy to optimize performance of hierarchy traversal. Space decomposition hierarchies allow multiple nodes per object. Rendered state of each instance is stored in bit array to make sure that all objects are rendered only once.

Frame buffer composition time for a large number of objects can be approximated by the formula:

$$T_{frame} = T_{check} + T_{render} + T_{queries},$$

where T_{check} — time it takes to check query results,

T_{render} — time to render visible objects,

$T_{queries}$ — time it takes to send queries for leafs nodes inside frustum.

When the number of queries is small T_{render} is the bottleneck. When the number of queries is large $T_{queries} + T_{check}$ is the bottleneck. Let's rewrite the formula by expanding the terms:

$$T_{frame} = c_1 N_q + c_3 N_{obj} + c_2 N_q,$$

where N_q — number of queries (leaf nodes inside frustum),

N_{obj} — number of visible objects inside the frustum. N_{obj} is the function of camera position and hierarchy height for the given object distribution.

Let's consider a common case where a scene is indexed by octree and camera is positioned outside the scene. In the worst case, three sides of the bounding box of the scene are visible. Assuming that objects are distributed uniformly across the bounding box of the scene, the number of objects per octree leaf equals $N_{total}/2^{3h}$, where N_{total} is the total number of objects in the scene, h is octree height. Then the number of visible objects approximately equals $3N_{total}/2^h$. We get the formula for frame duration that depends only on octree height and constants:

$$T_{frame} = (c_1 + c_2)2^{3h} + c_3 3N_{total}2^{-h}.$$

Optimal octree height for given scenario is $\frac{1}{4} \log_2 \frac{N_{total}c_3}{c_1 + c_2}$.

For example, for dataset 1 calculated octree height $h = 4$ gives the best performance in practice because dataset 1 can be described by that theoretical model.

LBVH and BVH SAH perform better on scenes where the density of objects is uneven. Better clustering helps to lower the number of queries to get visible objects inside the frustum.

4. Hierarchies

4.1 Octree

Octree is uniform space decomposition structure that uses three axis-perpendicular planes to simultaneously split the scene's bounding box into eight regions at each step [15]. When object's bounding box intersects the splitting plane, it is either

assigned to the internal node (single reference octree) or propagated below and assigned to multiple leaf nodes (multiple reference octree). Storing geometry in leafs increases clustering quality and as a result reduces number of visible objects. The downside is the increased number of occlusion queries, which are sent for every leaf in the frustum. We performed rendering performance comparison to find out which technique is more effective. Dataset 1 is small enough that GPU can handle rendering and queries quite efficiently (fig. 2). As a result, multiple reference octree gives the best time because of efficient clustering. Many objects intersect upper levels of octree resulting in redundant draw calls in case of single reference octree. Dataset 2 has non-uniform object distribution where several planes occupy half of the scene. It produced a lot of redundant leaf nodes and occlusion queries that degraded performance when storing objects in leafs. Dataset 3 has many buildings with tightly packed objects inside buildings. Even though the number of objects is large, it can be very efficiently subdivided requiring only small number of nodes. Overall, multiple reference octree provides the most efficient occlusion culling of large architectural scenes.

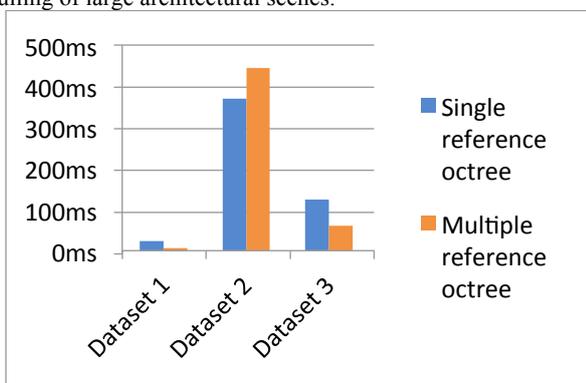


Fig. 2. Average frame rendering time for single and multiple reference octrees.

When rendering a visible node, contained object is skipped if it was already rendered in the current frame. Out of all hierarchical structures considered in the paper, octree gives the least effective clustering because of wasted space without any objects. Octree allows dynamic scenes because visibility results of octree nodes can be used for subsequent frames as they have fixed position in space. Also, it does not need to be rebalanced as other space decomposition hierarchies like kd-tree. Maximum octree level can be restricted depending on GPU performance.

4.2 LBVH

LBVH is primitive space decomposition hierarchy that is based on sorting along space filling curve [15]. All centers of object bounding boxes are sorted along Z space filling curve and grouped hierarchically from bottom to top [16]. LBVH achieves tighter clustering than octree and as a result less number of occlusion queries. Estimating the number of queries is simple because the number of leafs is determined at the start of the construction. LBVH cannot handle dynamic scenes because moving objects make occlusion queries for previously constructed LBVH useless in the current frame.

4.3 BVH

BVH with surface area heuristic for choosing the splitting plane to minimize the number of ray and bounding box intersection tests was developed for ray tracing, but it also gives efficient clustering of primitives for occlusion culling [13][15]. BVH construction is a top-down recursive process; on each step, we create two axis aligned bounding boxes. Triangles are sorted by the longest scene dimension and splitting plane with minimum cost is taken according to surface area heuristic.

Because of top-down construction, BVH SAH sometimes creates clusters that cannot be subdivided into two nodes. We

try to subdivide such cluster for each axis in order, and in case of failure leave it as a leaf node.

We compare rendering performance when using BVH SAH for storing three types of primitives: objects, subdivided objects, triangles. Object is set of triangles that can be rendered with one draw call. Subdivided object is an object that was subdivided into multiple objects to achieve better triangle clustering. Storing each triangle as an object generates hierarchy with the best clustering. For rendering efficiency, large number of triangles is stored in a leaf node, render state changes are avoided when encountering triangles of the same object. Storing triangles in BVH gives many additional draw calls creating a CPU bottleneck (fig. 3). Storing subdivided objects gives much faster performance. However, clustering efficiency increase is not enough to cover for additional draw calls for considered datasets using simple shader. Let's consider the difference in rendering time of dataset 1 for objects (8.7ms) and subdivided objects (45.6ms). Even though object subdivision helped to lower average number of query calls N_q from 192 to 171, it raised the average number of draw calls N_{obj} from 4546 to 19452 (tests were conducted for $bvh\ height = 10$).

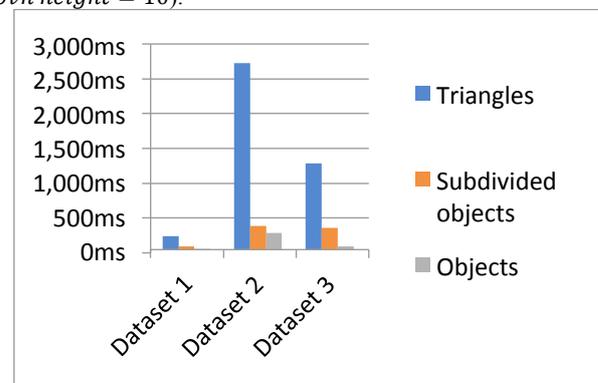


Fig. 3. Average frame rendering time for different types of primitive clusters when using occlusion culling on BVH SAH.

5. Performance comparison

5.1 Datasets

For rendering performance comparison we took three large datasets (figs. 4–6):

1. Dataset 1: 5,012,582 triangles, 50,521 objects. Building is tightly packed with objects having small variation in size.
2. Dataset 2: 10,827,713 triangles, 71,961 objects. Scene has many relatively large objects, half of the scene's volume is occupied by several planes.
3. Dataset 3: 10,154,304 triangles, 221,796 objects. Artificial test scene with 36 buildings. Each building has cluster of objects that can be culled after rendering exterior consisting of small number of objects.

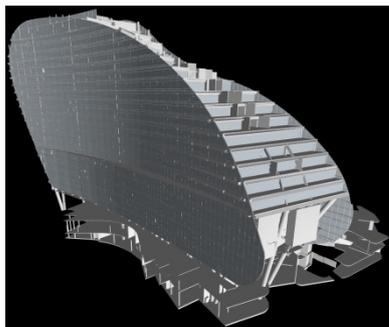


Fig. 4. Dataset 1 – architectural scene with 5 million triangles.

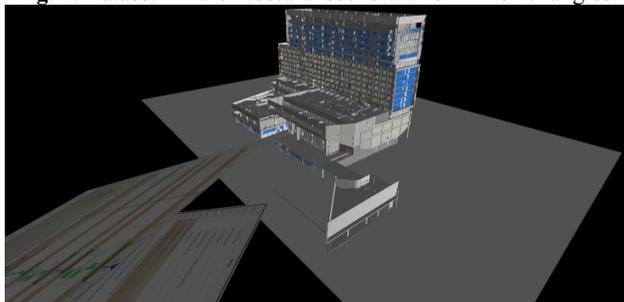


Fig. 5. Dataset 2 – architectural scene with 10.8 million triangles.

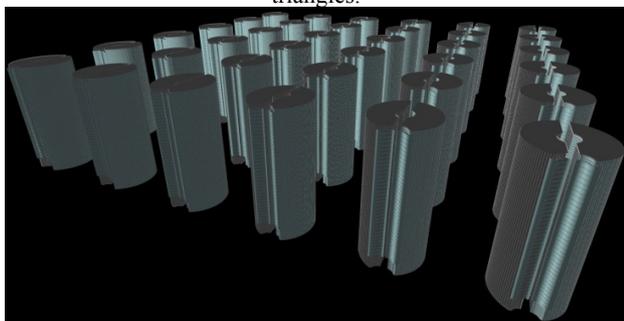


Fig. 6. Artificial test scene with buildings with 10.2 million triangles.

5.2 Clustering

Let's calculate the average number of rendered objects to compare object clustering efficiency of considered hierarchies for occlusion culling. Better object clustering should result in fewer visible nodes and fewer draw calls. During tests, we fixed the number of leafs for all hierarchies. BVH has the most efficient clustering of objects (fig. 7). It could be better but top-down subdivision process leads to the scenario where relatively long objects are gathered in a node and cannot be subdivided efficiently. In dataset 2 we encountered clusters with 40–70 objects where subdivision by any axis produced singleton. Algorithm based on octree issues more draw calls when space decomposition gives bounding volumes with a lot of empty space. Octree is more efficient for dataset 1 because it has very little empty space. It gives worst clustering in spacious dataset 3 because it cannot decompose it as efficiently using fixed number of leafs.

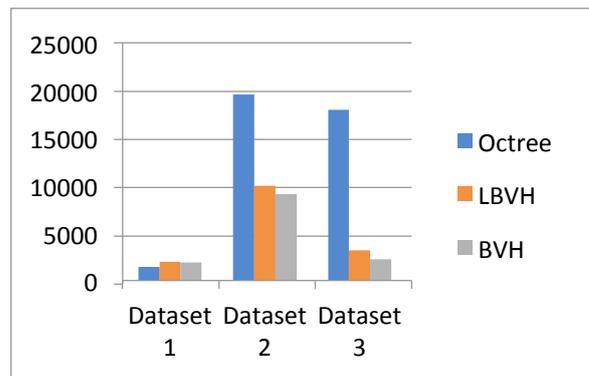


Fig. 7. Average number of rendered objects during scene walkthrough for considered hierarchies.

5.3 Frame rendering

All scene geometry is uploaded once at the beginning, shader with one directional light is used. Test results were produced on the system: AMD FX 8320 Processor, 24GB DDR3 RAM, AMD Radeon HD 6770 1GB.

Camera walkthrough is performed diagonally from the lower left to the upper right corner of a scene. Average and maximum frame rendering times are measured along the camera path (figs. 8, 9). Fig. 8 shows average rendering performance of all considered hierarchies on three datasets. Octree showed the fastest time for dataset 1 because it produced the best clustering (fig. 7). It performed better than expected for dataset 3 because most of the objects can be culled with relatively small number of queries. Dataset 2 was problematic for all hierarchies because it has most of the objects in one building. For efficient rendering careful balance of draw calls and occlusion queries is required. BVH SAH showed the fastest time because of efficient clustering. LBVH is close in performance to BVH SAH for all datasets.

Occlusion culling may give worse performance than frustum culling when GPU can efficiently render all of the objects inside the frustum. However, frustum culling shows worst performance on datasets 2 and 3 because of the large number of visible objects inside the frustum. Note that occlusion culling algorithm in the paper is not state-of-the-art and can be improved further to reduce the number of queries using visibility prediction and multiqueries [9][12].

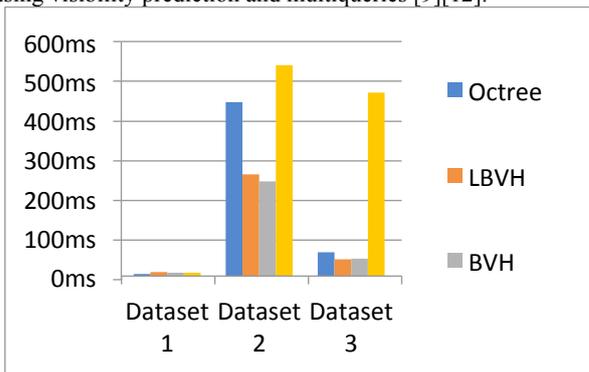


Fig. 8. Comparison of average frame rendering times for all datasets and hierarchies.

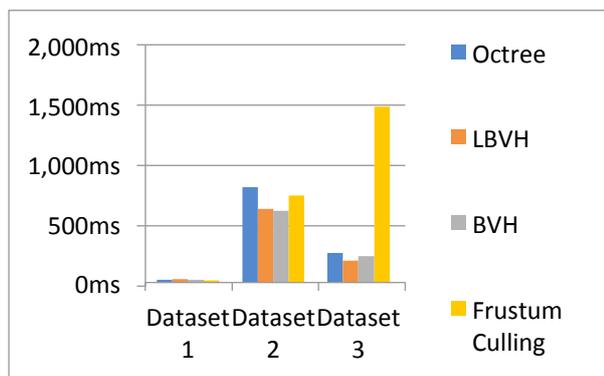


Fig. 9. Comparison of maximum frame rendering times for all datasets and hierarchies.

6. Conclusion

We performed the comparison of frame rendering performance when using different types of primitives and found that using objects instead of subdivided objects is more effective (fig. 3).

Octree efficiently handles datasets where most of the scene's volume is occupied by objects (fig. 7, dataset 1). Although storing objects in interior nodes of octree helps to select large objects and get better performance (fig. 2, dataset 2), storing objects in leafs is overall more effective and can be used to determine the number of leafs by scene's volume.

BVH SAH gives the most effective clustering of objects (fig. 7), and it positively affects frame rendering time (fig. 8). LBVH is close in performance to BVH SAH. Also, it is faster to construct, and bottom-up construction is better suited to get the optimal number of leafs.

7. References

- [1] Bittner, J. et al, 2004. Coherent hierarchical culling: Hardware occlusion queries made useful. In Computer Graphics Forum, Vol. 23, No.3, pp. 615–624.
- [2] Bittner, J. and Wonka, P., 2003. Visibility in computer graphics. Environment and Planning B: Planning and Design, Vol. 30, No.5, pp.729–755.
- [3] Blackert, A., 2016. Evaluation of multi-threading in Vulkan.
- [4] Chandrasekaran C. et al, 2013–2016. Software Occlusion Culling. <https://software.intel.com/en-us/articles/>.
- [5] Cohen-Or D. et al, 2003. A survey of visibility for walkthrough applications. IEEE Transactions on Visualization and Computer Graphic, Vol. 9, No. 3, pp. 412–431.
- [6] GLAPI/glBeginQuery. <https://www.opengl.org/wiki/GLAPI/glBeginQuery>.
- [7] Greene, N. et al, 1993. Hierarchical Z-buffer visibility. Proceedings of the 20th annual conference on Computer graphics and interactive techniques. ACM.
- [8] Greene, N., 1996. Hierarchical polygon tiling with coverage masks. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. pp. 65–74.
- [9] Guthe, M. et al, 2006. Near Optimal Hierarchical Culling: Performance Driven Use of Hardware Occlusion Queries. In Eurographics Symposium on Rendering. pp. 207–214.
- [10] Legacy OpenGL - OpenGL Wiki. https://www.khronos.org/opengl/wiki/Legacy_OpenGL.
- [11] Macdonald, J. D., Booth, K. S., 1990. Heuristics for ray tracing using space subdivision. Visual Computer, Vol. 6, No. 6, pp. 153–165.
- [12] Mattausch, O. et al, 2008. CHC++: Coherent Hierarchical Culling Revisited. EUROGRAPHICS, Vol. 27, No. 3.
- [13] Meissner, M. et al, 2001. Generation of Decomposition Hierarchies for Efficient Occlusion Culling of Large Polygonal Models. In Vision, Modeling, and Visualization, Vol. 1, pp. 225–232.
- [14] Next Generation Occlusion Culling. http://www.gamasutra.com/view/feature/164660/sponsored_feature_next_generation_php?print=1.
- [15] Pharr, M. et al, 2016. Physically based rendering: From theory to implementation. Morgan Kaufmann.
- [16] Samet, H., 2006. Foundations of multidimensional and metric data structures. Morgan Kaufmann.
- [17] Teller, S., Sequin, C., 1991. Visibility preprocessing for interactive walkthroughs. Computer Graphics (Proceedings of SIGGRAPH 91), Vol. 25, No. 4, pp. 61–69.
- [18] Wloka, M., 2003. Batch, batch, batch: What does it really mean. Presentation at game developers conference.
- [19] Zhang, H. et al, 1997. Visibility culling using hierarchical occlusion maps. Proceedings of the 24th annual conference on Computer graphics and interactive techniques. pp. 77–88.

About the authors

Gonakhchyan Vyacheslav Igorevich, junior researcher at the department of System integration and multi-disciplinary collaborative environments of Ivannikov Institute for System Programming of the RAS. His email is pusheax@ispras.ru.

Reconstruction of BSDF based on optimization of microrelief normal distribution

V.G. Sokolov¹, D.D. Zhdanov², I.S. Potemin², N.N. Bogdanov², A.D. Zhdanov², E.U. Denisov¹
 sokolovv1969@gmail.com|ddzhdanov@mail.ru|ipotemin@yandex.ru
 nnbogdanov@corp.ifmo.ru|adzhdanov@corp.ifmo.ru

¹ The Keldysh Institute of Applied Mathematics RAS, Moscow, Russia;

² ITMO University, Saint Petersburg, Russia

The article is devoted to the elaboration of the method of reconstruction of rough surface scattering properties. The rough surface, in this case, is considered as a boundary between dielectric and air. Usually, these properties are described with bi-directional scattering distribution function (BSDF). Direct measurement of such function is either impossible or very expensive. The BSDF reconstruction method, based on the distribution of the surface microrelief heights, requires a complex fitting procedure and often yields not very good results. In the suggested solution the rough surface is simulated with a parametric function emulating distribution density of normals to facets of the surface microrelief. The optimization of distribution density of normals to facets of the surface microrelief show good agreement with desired output.

Keywords: Microrelief, BSDF, Rough surface, Diffusivity, Rendering, LGP, TIR, Wave optics, Ray optics.

1. Introduction

The light guiding optical elements with rough surfaces are widely used in devices with complex light propagation. As a rule, there are two main applications of rough surfaces: either to obtain the specific goniometric diagram of the light scattering or to obtain desired spatial luminance distribution for various light guiding devices like illuminating system of displays, car dashboards, LED luminaries, etc. When simulating the light propagates inside of material we are in need of optical properties of the rough surface boundary between two media while optical properties of the whole optical element are senseless. Moreover, the properties are individual from each side of the light incidence to the rough surface. So the correct light simulations have to apply the properties taking into account the side the light does incident on.

An example of rough surface usage is presented on Figure 1. Dots with microrelief are distributed on the bottom face of the light guiding plate (LGP) [1]. They are rough scattering surfaces. A light ray is propagated in LGP due to total internal reflection. After the scattering on dots, the ray deviates from mirror reflection direction and can leave LGP [2]. Variable density of dots allows obtaining a uniform light emitting along output surface.

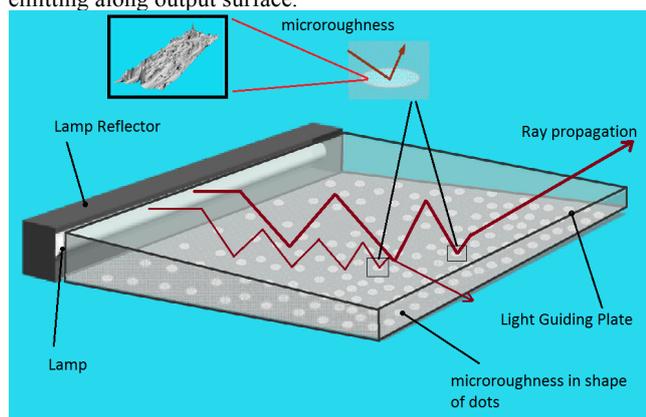


Fig. 1. An example of the application of the LGP with rough dots.

The scattering properties of the rough surface are described by a bidirectional scattering distribution function (BSDF). This function has complex multi-dimensional representation and depends on many parameters: incident light direction, output observation direction, and spectrum (color). For flat thin samples, like shown in Fig. 2a (“surface” model), the BSDF

can be measured by any goniophotometer. If the thickness of object with microroughness can be ignored and the BSDF is assigned to single surface the measured BSDF model will be physically correct. The model can be applied to various diffuse sheets and filters. Unfortunately, the model is not applicable if the thickness of the element with the roughness is important for the light propagation inside of the transparent element. The “solid” model presented in Fig. 2b should be used in this case. It means that we need two BSDFs of the rough surfaces, one from the air to the glass and another one from the glass to the air.

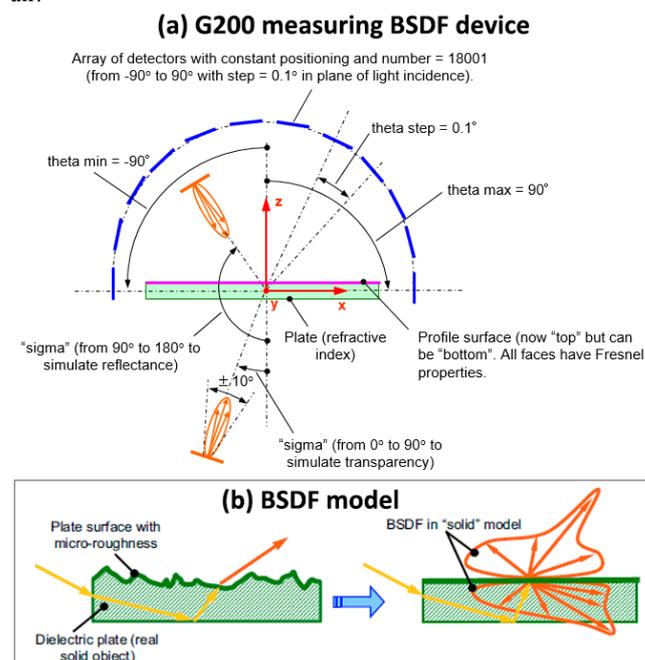


Fig. 2. “Surface” vs. “Solid” BSDF application models.

The main problem is that BSDF of the rough surface cannot be measured directly. There is a number of reasons. The first one is multiple reflections between the rough surface and other surfaces of the measured sample. Another reason — it is impossible to illuminate the sample or detect light under grazing angles of the light incidence (observation) to the rough surface. The solution of the problem is very expensive and requires special equipment to exclude multiple reflections between sample faces and refraction on the side opposite to the measured rough one.

An alternative way of BSDF reconstruction is computer simulation of light scattering on the microrelief boundary of the sample media [5]. This indirect way has a number of disadvantages as well. In particular, variations of surface profile can be comparable with a wavelength of illuminating light. It means that computations have to use the wave optics approach which, at first, is very complex and, at second, can be inaccurate due to not sufficient accuracy of the surface profile measurement. In the current article, a combined approach is proposed. It utilizes optimization of BSDF shape based on an approximation of the shape by Cauchy and Gauss functions with a limited number of parameters. This approach provides a more accurate reconstruction of BSDF than the method proposed in [5].

Many of scientists are engaged in solving of the problem of difficult reconstruction BRDF [6-13]. Many of articles [6-10, 12] are devoted to accurate and physical correct reconstruction in comparing with MERL BRDF database [14]. This database contains reflectance functions of 100 different materials. Authors of MERL BRDF database describe in article [15], their method of gating BRDF, but question about measurement correctness is appeared. In our research the certificated measurement equipment GCMS-4 was used [16], which allows us, making physical accurate measurement of BRDF. It's difficult to say that measuring BRDF in MERL database was accurate, because there is not information about certificated equipment. That's why, question about validity of measurements is appeared. That is also great, that some authors could make BRDF reconstruction and have a not bad results [6-10, 12], but they compared results only with measurements from MERL BRDF database. In our research we reconstructed BRDF and compare results with BRDF measurements from GCMS 4, and we can be sure that our result are physical accurate.

Many authors consider issue only about reconstruction BRDF, but not about BSDF in general. As a rule, BRDF is applied only for surfaces, but it does not enough for accurate modelling, for example, frosted/opal glass.

We suggest method for BSDF reconstruction, which allow making accurate modeling for difficult scenes with frosted/opal glass. Our research in BRDF reconstruction has good results and the modelling accuracy is approved by comparing with measurements from GCMS-4 [16].

2. Numerical methods of BSDF reconstruction

There are several numerical approaches of BSDF calculation for rough surface based on an approximation of the ray optics as well as the wave optics. In the previous article, we describe the solution, where surface microrelief is represented as the height distribution of the representative sample area [5].

Reconstruction of BSDF of the plate with the rough surface was based on two sets of measured data: the microprofile height distribution and BSDF of whole sample (transparency and/or reflectance). Often results of the reconstruction were not quite good and require complex optimization of the microprofile (reducing to scaling and filtration of the profile). However, the filtration cannot guarantee the success.

The new approach is based on the only kind of data: measured transparency or/and reflectance of a plane sample measured as a one-sheet element. In spite of the difference with previous algorithm base model of the new approach is the same. The source of the reconstructed BSDF is an intensity distribution calculated after the ray transformations on the microfacets boundary of two media. The only difference is that microfacets are defined as a distribution density of normals. Application of the OPTOS MicroRelief tool [18] of Lumicept [17] provides correct calculations of the intensity distribution, scattered on the microrelief.

An initial distribution of the normal facets necessary for light simulations can be restored from the measured BRDF of the sample. Without shadowing of one facet by another one the normal distribution is about 2 times narrower than BRDF. Of course, it is a rough approximation but can be used as an initial step of the whole BSDF reconstruction.

For BSDF reconstruction we use a real flat sample (plate) in which one of the surfaces is smooth and the other is rough. The plate is illuminated by a collimated beam of light. For each incident light direction, the reflected and transmitted light intensities are measured by the Integra's spectral scatterometer [3, 4]. For the simplification, the measurements are performed in a single plane—in the plane of light incidence. The simulation scheme is similar to measurements. Parallel light illuminates a plate with the same angular deviation and aperture. The rough surface of a sample is simulated with help of BSDF calculated by Lumicept BSDF generator using the distribution of normals. Light scattered with the plate was collected on circular detectors placed on some angular grid. The distance from detectors to the measured sample and radius of detectors corresponds to the characteristics of measuring device: mutual position of the sample and sensor of goniophotometer, angular and spatial resolution of a goniophotometer.

Figure 3 presents measured and simulated angular distribution of light transmitted through the plate sample with one rough surface. The combined graph contains light transmittance for all measured incident light directions. The measurements and simulations were fulfilled for the following directions of the light incidence: 0°, 15°, 30°, 45°, 60°, 75° (angle between normal and incident light direction). They are marked with different colors. Note that all measurements and simulations are made in the plane of light incidence. The solid plots present result of the real sample measurements. Dash-line plots correspond to simulated sample with reconstructed BSDF. It is seen that there is the essential difference between simulated and measured results. The same tendency can be observed on graphs with reflectance data (omitted in the article).

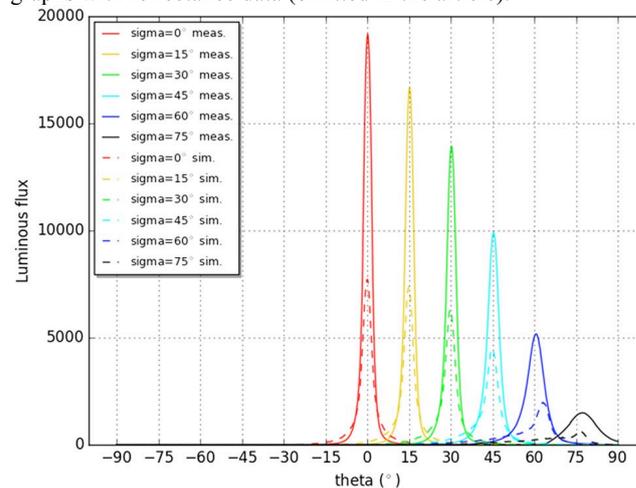


Fig. 3. Comparison of measured and calculated transparency.

3. Optimization of BSDF reconstruction procedure based on the distribution density of normals

Figure 3 shows the high deviation of measured BSDF from the reconstructed one. The main reason for the difference, the initially reconstructed model of deviation of normals does not fit to the real model of the light scattering on the sample. On the other hand, the distribution density of normals is the way of the indirect BSDF definition. Thus, an optimization of the angular

distribution density of normals to facets allows reach target BSDF of the sample.

The main idea of proposed optimization method is to use only one set of data: sample transparency characteristics. Figure 4 illustrates the optimization procedure step by step. The rough surface is defined by the distribution density of normals to the surface facets. Optimization procedure consists of the following step:

1. At first input data about sample sizes, refractive index, sample transparency, initial parameters for the function description of the distribution density of normals.

2. The second step consists of tuning scene, generation of tabular function for microfacets based on initial parameters. After that, the microfacets distribution is added the OPTOS MicroRelief plugin of Lumicept BSDF simulator [18].

3. The third step calculates an intensity distribution for the prepared sample.

4. Further, the optimizer compares measured and simulated results and calculates deviations (as RMS).

5. The next step is analyzing of deviations between optimized and measured results to take the decision to stop or to continue optimization process.

5.1. If the optimizer does not reach the desired deviation, then the optimizer changes parameters of the distribution density of normals and goes to the step 2 to continue the process.

5.2 Afterwards, if deviations are suitable, the final BSDF is generated with the help of the “BSDFCalculator” tool.

6. Finally, the optimizer plots graphs of BSDF of the measured sample vs. BSDF of the reconstructed sample.

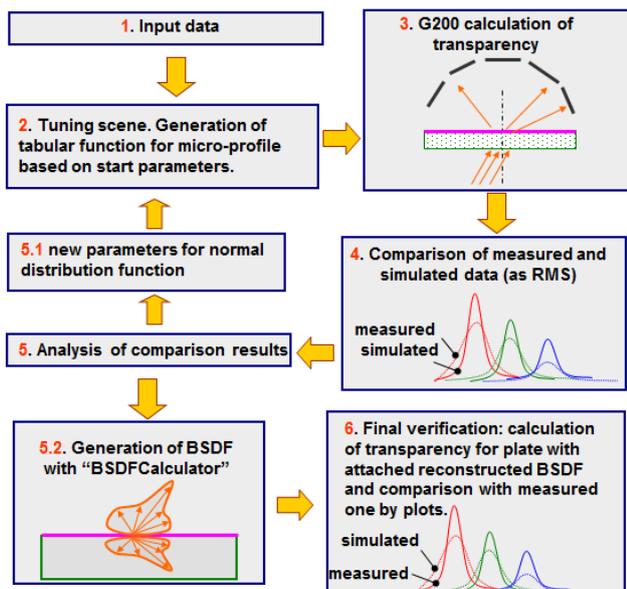


Fig. 4. Optimization scheme of BSDF reconstruction.

It is important that BSDF is reconstructed on the basis of the distribution density of normals to the surface facets. However, a tabular definition of the distribution is not suitable for most of the optimization tools (the multiparametric procedure is very time consuming one) and the most suitable representation of the law of distribution is an analytical function with a minimal number of parameters. Experiments allowed selecting two base kinds of functions: “Gauss-like” and “Cauchy-like”. In the most of the cases, the “Cauchy-like” distribution gives better output while for some microreliefs the “Gauss-like” approximation seems better. It seems that “Gaussian” approximation gives good agreement in BTDF zones of high transparency (at least from viewpoint of RMS between simulated and measured results). So it is reasonable to use both types of function in the optimization process. The

general view of Gauss and Cauchy functions are shown in figure 5. One can see that Cauchy distribution is wider in zones of far theta angles (x). The parameter x_0 specifying a shift of distribution peak along theta angles is rather formal parameters because of the most of distributions density of normals have a maximum for $x_0 = 0$. But this parameter was reserved for “advanced” optimization.

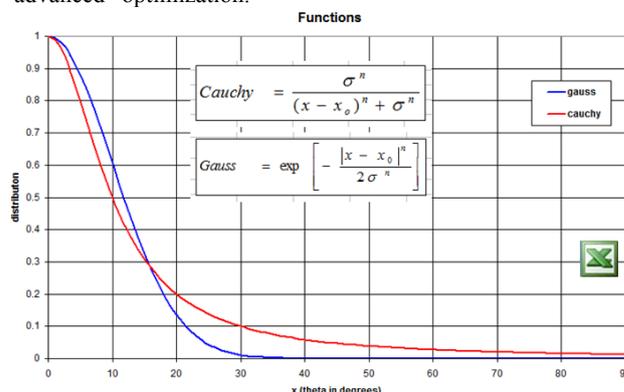


Fig. 5. Cauchy distribution and Gauss distribution.

Taking into account that the general tabular function representation of the distribution density of normals is not a good solution for optimization procedure an alternative “mixing” solution was selected. Base function of the distribution density of normals can be specified with “Gauss-like” or “Cauchy-like” approximation while some areas of the function can be exchanged to locally tabular one. It can be explained with graphs below (figure 8 and 9). Short description of the algorithm.

1. Let’s suppose that optimization procedure with the analytical function of distribution density of normals cannot fit the BSDF in the area close to zero theta. It means that the distribution density of normals in the area of zero angular deviation have to be changed with the tabular function.

2. Then the optimizer adds a number of points to the tabular representation of the distribution density of normals in the area and continues optimization of the mixed function. If a number of added points is not high the optimization procedure can find a solution.

4. Comparison of the BSDF reconstruction methods based on the distribution of heights and based on the distribution density of normals

To test new tool several problematic examples from the past [5] have been selected. These samples required complex fitting procedure based on filtration and scaling of measured profiles, for some of them artificial profile data were used (measured for different samples, for example). See results achieved in the past in figure 6.

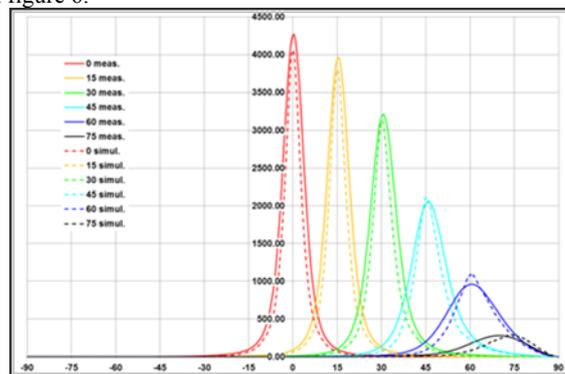


Fig. 6. Results of BSDF reconstruction based on microprofile fitting.

Results of BSDF reconstruction based on Cauchy-like function are shown in figure 7.

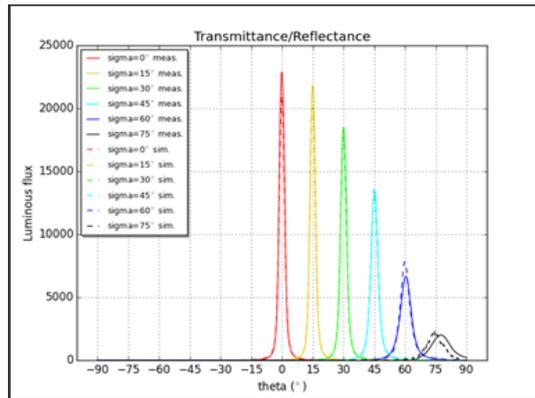


Fig. 7. Results of BSDF reconstruction based on Cauchy-like function.

Results of BSDF reconstruction based on Gauss-like function are shown in figure 8.

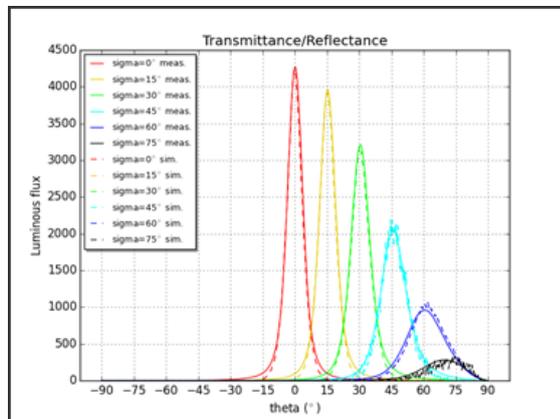


Fig. 8. Results of BSDF reconstruction based on Gauss-like function.

We can finally say that the optimization results of distribution density of normals show good agreement with desired output (at least in the scope investigation samples). In the most of the cases, the Cauchy-like function gives acceptable results at least not worse than a design with measured microrelief [5]. The Gauss-like function in some cases is useful too. All this allows going to a conclusion that precise measurements of the microrelief are not necessary at all.

The usage of OPTOS MicroRelief plugin [18] allows excluding BSDFGenerator of Lumicept [17] from the optimization procedure. It accelerates optimization process as it is not required to generate BSDF on each optimization step that requires significant calculation time.

Attempt to apply the tabular function of the distribution density of normals to facets as a parameter optimization was failed. Optimization of the multiparametric function is the very time-consuming procedure and all benefit caused by the freeform shape of distribution density of normals function is killed by slowdown and general divergence of the optimization procedure.

We can see good agreement between measured and simulated results for incident angles close to normal direction and acceptable agreement for other incident angles. In the article, we demonstrate the results for light transmittance only. However, the optimization procedure can be applied to reflectance as well. Usually, optimization of transparency results improves the reflectance too.

Also, we made a photorealistic rendering of the plate with a rough surface. The visual appearance of the plate with the rough surface BSDF before optimization (i.e. when initially measured profile was used) is presented in Fig. 9a. The visual appearance of the plate with the optimized rough surface BSDF is presented in Fig. 9b.

The images on Fig. 9 were synthesized with physically accurate rendering tool based on path tracing integrated to the Lumicept software package [17]. The scene consists of a plate with BSDF assigned to the outer plate surface. The plate is placed over chessboard-like substratum and illuminated with a set of light sources creating complex diffuse illumination.

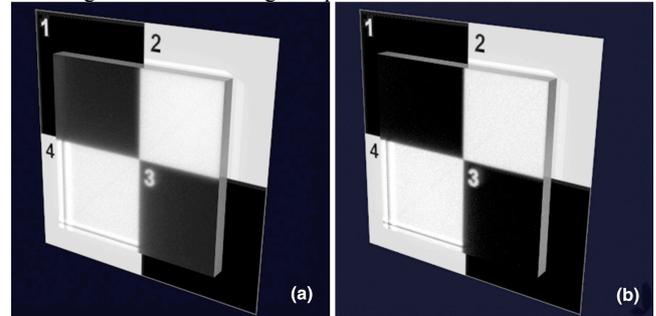


Fig. 9. Visual appearance of a plate with rough surface.

5. Conclusion

The results of optimization method of the distribution density of normals for BSDF reconstruction show good agreement with desired output (at least in the scope investigated samples).

In the most of the cases, the Cauchy-like function gives acceptable results at least not worse than a design with measured microrelief [5]. Moreover, the Gauss-like function in some cases is also useful. That allows making a conclusion about the possibility to exclude measurements of the microprofile at all to reconstruct BSDF accurately.

6. Acknowledgements

This research has been supported by the RFBR grants No. 17-01-00363 and No. 15-01-01147, Financial support of the leading universities of Russian Federation (subsidy 074-U01), as well as by Integra Inc.

7. References

- [1] Zhdanov, D.D., Zhdanov, A.D., Potemin, I.S., Venediktov, V.Yu., Galaktionov, V.A. A Method of Fast Building of Locally Equidistant Distributions of Geometrical Objects of Illumination Systems// Optics and Spectroscopy, 2015, Vol. 119, No. 2, pp. 311–317. DOI: 10.1134/S0030400X15080263.
- [2] Zhdanov, D.D., Garbul, A.A., Maiorov, V.A., Potemin, I.S., Sokolov, V.G. Indeterminate ray tracing in problems of the analysis of light scattering and the design of illuminating systems // J. Opt. Technol. – 2014, , Vol. 81, Issue 6, pp. 322-326. DOI:10.1364/JOT.81.000322.
- [3] Zhdanov, D., Sokolov, V., Potemin, I., Voloboy, A., Galaktionov, V., Kirilov, N.// Opt. Rev.– 2014. – 21(5), – 642 (2014).
- [4] Letunov, A.A., Barladian, B., Galaktionov, V.A., Ershov, S.V., Voloboy, A., Zueva, E.// Proc. 22nd Int. DAAAM Symp., pp. 1459 (2011).
- [5] Sokolov, V.G., Zhdanov D.D., Potemin, I.S., Garbul, A.A., Voloboy, A.G., Galaktionov, V.A., Kirilov N. Reconstruction of scattering properties of rough air-dielectric boundary // Optical Review. – 2016. – 23(5). – 834-841, DOI: 10.1007/s10043-016-0250-6.

- [6] Nurcan Seylan, Serkan Ergun, Aydın Öztürk “BRDF Reconstruction Using Compressive Sensing”// 21st International Conference on Computer Graphics, Visualization and Computer Vision 2013. – pp. 88-94. ISBN 978-80-86943-74-9.
- [7] Jannik Boll Nielsen, Henrik Wann Jensen, Ravi Ramamoorthi “On Optimal, Minimal BRDF Sampling for Reflectance Acquisition”// ACM TOG 34(6). – 2015. – pp. 1-11.
- [8] Doris Antensteiner and Svorad ˇ Stolz “Full BRDF Reconstruction Using CNNs from Partial Photometric Stereo-Light Field Data” // Workshop on Light Fields for Computer Vision at ECCV-2017 – 2017. – pp. 13-21.
- [9] Feng Lu, Xiaowu Chen, Imari Sato, and Yoichi Sato “SymPS: BRDF Symmetry Guided Photometric Stereo for Shape and Light Source Estimation” // IEEE Transactions on Pattern Analysis and Machine Intelligence. – Issue: 99. – pp. 1-14.
- [10] Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi “On Differential Photometric Reconstruction for Unknown, Isotropic BRDFs”// IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2013. – Vol. 35. – No. 12. – pp. 2941-2954.
- [11] Sing Yee Chuua, Xin Wanga*, Ningqun Guoa, Ching Seong Tan “Performance of Range Gated Reconstruction: A Theoretical Analysis”// Proc. of SPIE. – Vol. 10250. – pp. 1-5.
- [12] Jiri Filip, Michal Havlí, Radomír Vávra “Adaptive highlights stencils for modeling of multi-axial BRDF anisotropy”// The Visual Computer. – 2017. – Volume 33. – Issue 1. – pp. 5–15.
- [13] Guojun Chen, Yue Dong, Pieter Peers, Jiawan Zhang, Xin Tong “Reflectance Scanning: Estimating Shading Frame and BRDF with Generalized Linear Light Sources”// OOPSLA '94. – 1994. – Volume 5. – Issue 4, Oct. 1994 – pp. 67-71.
- [14] MERL BRDF database
<http://people.csail.mit.edu/wojciech/BRDFDatabase/>
- [15] Wojciech Matusik, Hanspeter Pfister, Matt Brand and Leonard McMillan “A Data-Driven Reflectance model”// ACM Transactions on Graphics 22, 3(2003), 759-769.
- [16] Muracami Color Research Laboratory,
<http://www.mrcr.co.jp/english/index.html>.
- [17] Lumicept – Hybrid Light Simulation Software,
<http://www.integra.jp/en>.
- [18] Ershov S.V., Zhdanov D.D. Efficient application of Optical Objects in light simulation software // Proc. 15th International Conference on Computer Graphics and Applications - GraphiCon-2005, Russia, Novosibirsk, June 20-24, 2005, p. 284-288.

E.U. Denisov, research officer at the Keldysh Institute of Applied Mathematics RAS.
His e-mail: eed@gin.keldysh.ru.

About the authors

V.G. Sokolov, research officer at the Keldysh Institute of Applied Mathematics RAS.

His e-mail: sokolovv1969@gmail.com.

D.D. Zhdanov, PhD. Associate professor at ITMO University. Head of the Visualization Technology chair.

His e-mail: ddzhdanov@mail.ru.

I.S. Potemin, PhD. Associate professor at ITMO University. Assistant of the Visualization Technology chair.

His e-mail: ipotemin@yandex.ru.

N.N. Bogdanov, PhD student at ITMO University, Visualization Technology chair.

His e-mail: nnbogdanov@corp.ifmo.ru.

A.D. Zhdanov, PhD student at ITMO University, Visualization Technology chair.

His e-mail: adzhdanov@corp.ifmo.ru.

Stochastic ray tracing methods in problems of photorealistic image synthesis for augmented reality systems

D.D. Zhdanov¹, I.S. Potemin¹, A.A. Kishalov², A.D. Zhdanov¹, N.N. Bogdanov¹
 ddzhdanov@mail.ru|ipotemin@yandex.ru|grinfo@mail.ru|adzhdanov@corp.ifmo.ru|nnbogdanov@corp.ifmo.ru
¹ITMO University, Saint Petersburg, Russia;

²JSC 'Scientific Industrial Enterprise of Fiber Optics and Laser Equipment' (NPP VOLO), Saint Petersburg, Russia

The article is devoted to the investigation of stochastic ray tracing methods in problems of photorealistic image synthesis for augmented reality systems and in particular for Head-Up Display (HUD). As a result of the study, we were convinced that the method of forward stochastic ray tracing can be successfully used for virtual prototyping of augmented reality systems. We extended the model of bidirectional stochastic ray tracing by forward ray tracing to simulate image formation in HUD systems without diffuse components.

Keywords: stochastic ray tracing, photorealistic rendering, image synthesis, augmented reality, head-up display.

1. Introduction

In computer graphics, photorealistic visualization methods can be used not only to construct images of 3D scenes but also to synthesize images in complex optical systems, such as augmented reality systems, in particular, to synthesize images for the head-up displays (HUD). A large number of articles is devoted to the design of the optical part of the display and evaluation of resulting image quality, for example, [5, 8]. However quality evaluation is commonly reduced to aberration analysis and contrast function calculation, while problem of visual quality evaluation including host analysis or rear projection effect are left behind. This work is focused mainly on visual quality analysis of designed HUD system. A principal scheme of the augmented reality image formation in a typical HUD is shown in Fig. 1.

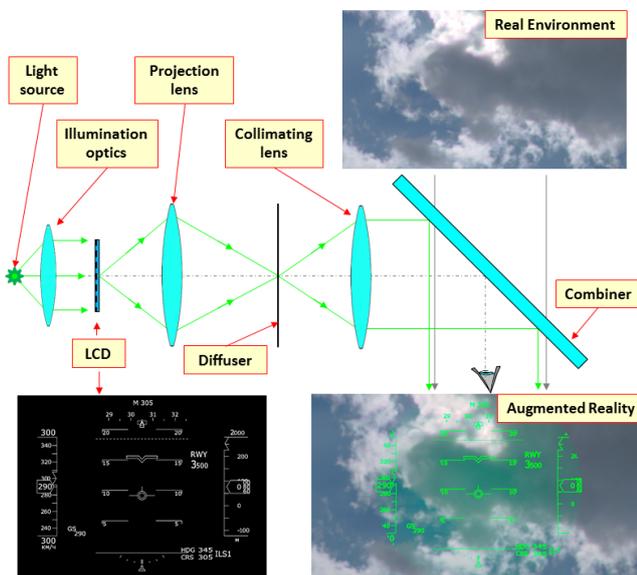


Fig. 1. Principal scheme of HUD with LCD source.

The light source together with the lighting optics illuminates the LCD, which shows some useful information in the form of a transparent slide. The projection lens forms an enlarged image of the slide in the plane of the diffuser. Then collimating optics creates the infinitely distant image of the diffuser. Finally, the beam splitter (combiner) redirects the light from the diffuser in such a way that the image formed on the diffuser and the surrounding reality become visible to the eye simultaneously. Thus, the eye observes the real environment augmented with text or graphic information formed on the LCD.

Using the methods of photorealistic visualization when designing such devices is very important because it makes it possible to build a virtual prototype and to reconcile the luminance of augmented reality objects with the luminance of real environment. Also, it allows choosing the optimal parameters of light sources, illumination optics and a projection lens.

The task of an image formation created by an optical system differs from the classical problem of constructing an image existing in computer graphics. The main difference is that the pinhole camera used in computer graphics forms an image in terms of luminance corresponding to the luminance of the observed scene objects, while in the optical system the image is the illuminance collected from the entire area of the output surface of the optical system. For the visual optical system, the eye is a part of the optical system (its position and pupil size are important components of the visualization system) and the eye retina perceives the illuminance as a luminance. Moreover, the optical system is part of the overall visualized scene and can contain surfaces with both diffuse and specular properties, and all components of luminance (direct vision, primary, caustic and secondary luminance) can be formed both in the observed scene and in the optical system itself. Therefore, the simulation was based on the method of bidirectional stochastic ray tracing, taking into account the caustic and secondary luminance using photon maps. There a large number of methods that allows to synthesize a physically correct image of 3D scene. Starting with Kajiya [4] who presented a solution of rendering equation with backward ray tracing method a huge number of approaches were invented to numerically solve that equation. In [10] Veach described the base principles of the photorealistic image synthesis methods. They include stochastic bidirectional ray tracing methods [1, 7], photon maps based methods [2, 6, 9], Vertex Connection and Merging method [3]. However, we should take into account that these methods are used to synthesize an image of scene that have no complex optical system consisting of dozens of specular faced between scene and virtual camera. The use of this method has made it possible to improve the efficiency of image synthesis significantly and take optical and geometric parameters of the elements of the optical scheme of the lens (lenses and mirrors) into account in the synthesis of the image. In addition, the influence of the optical and geometric parameters of the mechanical design of this lens is taken into account.

However, not in all cases, the method based on the bidirectional stochastic ray tracing approach are the most effective. In particular, a decrease in efficiency can occur when using this method for virtual prototyping of the augmented reality systems. Therefore, the aim of this paper is to investigate the methods of stochastic ray tracing, which allow obtaining

maximum efficiency when forming physically correct images in the HUD.

2. HUD optical system models

In the scope of this investigation, we used two approaches to the formation of augmented reality image in HUD systems:

- with an intermediate diffuse screen (Fig.2);
- without an intermediate diffuse screen (Fig.4).

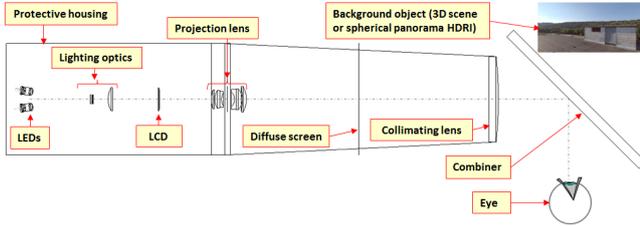


Fig. 2. Scheme with the intermediate diffuse screen.

LEDs with a very small emitting area were used as light sources. The size of the radiating surface is 0.2mm X 0.2mm. Lighting optics provide uniform illumination of the entire LCD surface. A slide with text or graphic information is formed on the LCD (Fig. 3).

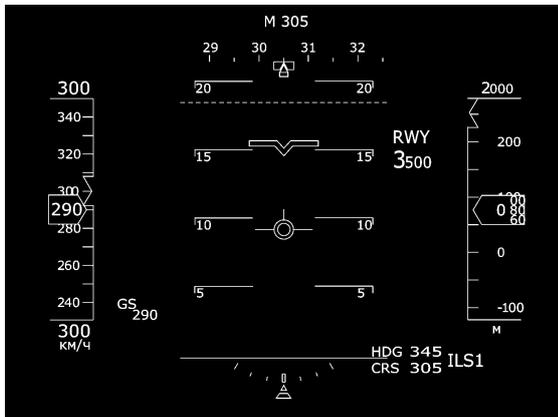


Fig. 3. An example of the image formed on the LCD.

On this slide, all useful information is presented in form of the transparent zones on a black background. However, the black background is not absolutely absorbing. In modern LCD the contrast between absolutely transparent areas and the opaque background is about 300: 1.

The image of the slide enlarged with the projection lens is formed on the diffuser, which is the source of the secondary radiation for the subsequent part of the imaging system in the HUD. The main disadvantage of the first approach is reduced visible luminance of the augmented reality object due to the light scattering on the diffuser.

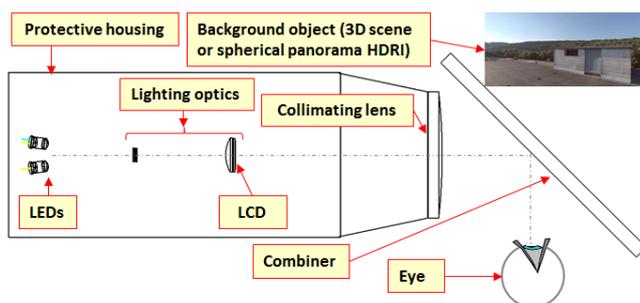


Fig. 4. Direct scheme of HUD image formation.

In the second approach, the lighting optics is the same as in the first one, but in this case, there is no intermediate projection on the diffuse screen. This gives us two important advantages:

- Reduction of the HUD optical system size;
- Reduction of energy loss due to light scattering.

However, together with the above advantages, a problem of the modeling of such systems arises. The problem relates to the choice of the most effective method of stochastic ray tracing.

3. Description of the method

In the case of a diffuse screen, the only source of luminance is the caustic luminance of the diffuse screen, which can be calculated using photon maps. Obviously, for such a scheme, the most effective calculation method is bidirectional stochastic ray tracing using photon maps.

Photorealistic rendering of optically complex scenes reduces to solving the rendering equation [4] for each point of the scene image. The rendering equation allows us to calculate the luminance of the color component c on the point \vec{p} and in a given direction \vec{v} . For static scenes the rendering equation can be written as follows:

$$L(\vec{p}, \vec{v}, c) = \tau(\vec{p}, \vec{v}, c) \left(L_0(\vec{p}, \vec{v}, c) + \frac{1}{\pi} \int_{4\pi} BPDF(\vec{p}, \vec{v}, \vec{v}', c) L(\vec{p}, \vec{v}', c) (\vec{n} \cdot \vec{v}') d\omega \right) \quad (1)$$

where:

$L_0(\vec{p}, \vec{v}, c)$ – the luminance of the object emission at the point of observation,

$\tau(\vec{p}, \vec{v}, c)$ – the transmittance (transparency) of the medium between the observer and the observation point,

$\frac{1}{\pi} \int_{4\pi} BPDF(\vec{p}, \vec{v}, \vec{v}', c) L(\vec{p}, \vec{v}', c) (\vec{n} \cdot \vec{v}') d\omega$ – the luminance

formed by primary and secondary illumination of the observed object.

where:

$BPDF(\vec{p}, \vec{v}, \vec{v}', c)$ – the luminance factor of the surface (or Bidirectional Scattering Distribution Function (BSDF)) from the source \vec{v}' in direction \vec{v} to observer,

$L(\vec{p}, \vec{v}', c)$ – the luminance of the ambient light in a solid angle $d\omega$ in the direction \vec{v}' to the observation point \vec{p} .

\vec{n} – the normal to the surface at the point of observation.

The rendering equation (1) is an equation with infinite recursion. If the contribution of the secondary radiation to the total value of the apparent luminance is significant, then the bidirectional stochastic ray tracing based on the "Russian roulette" method using photon maps for calculating the caustic and secondary luminance allows performing physically correct infinite integration in the most efficient manner [10].

Stochastic ray tracing method can be based on the stochastic forward and backward ray tracings. Each component of the ray tracing can be used for the image synthesizing. The main problem of the separate methods is a significant reduction of the simulation efficiency when the scene contains scattering elements. In the latter case the most effective solution is one of the variations of stochastic bidirectional ray tracing with photon maps method. This method allows us to collect illuminance from scatterings elements of the optical system. On the other hand, if the scene do not contain scattering elements the efficiency of the methods (vs. bidirectional stochastic ray tracing) significantly increases. As a result, to increase the efficiency of synthesizing an image formed by the optical systems without scattering elements the authors had extended the general approach of bidirectional stochastic ray tracing with photon maps. The extension allowed merging of an image

components formed with stochastic forward or backward ray tracing only.

Note that for realistic visualization and correct evaluation of image quality in HUD, we implemented the model of the eye as a lens camera.

To calculate the local illuminance value we can use well known dependence between the luminance of the small pupil area and illuminance from this light area:

$$dE(\vec{p}, c) = L(\vec{p}_i, \vec{v}_i, c) (\vec{n} \cdot \vec{v}_i) d\omega_i \quad (2)$$

where $L(\vec{p}_i, \vec{v}_i, c)$ – the luminance of the pupil area in the point \vec{p}_i and direction to the pupil area \vec{v}_i ,

$d\omega_i$ – the solid angle from the image point \vec{p} to the pupil area in the point \vec{p}_i ,

\vec{n} – the direction of the exit pupil normal.

The eye luminance can be recalculated from the illuminance (2) using the following equation.

$$L(\sigma, c) = \frac{E(\vec{p}, c)}{\cos(\sigma) \cdot \Omega} \quad (3)$$

where σ – the field of view angle.

Figure 5 illustrates the method of the local luminance calculation. The total illuminance is obtained by integrating all local illuminance values computed for a given image point.

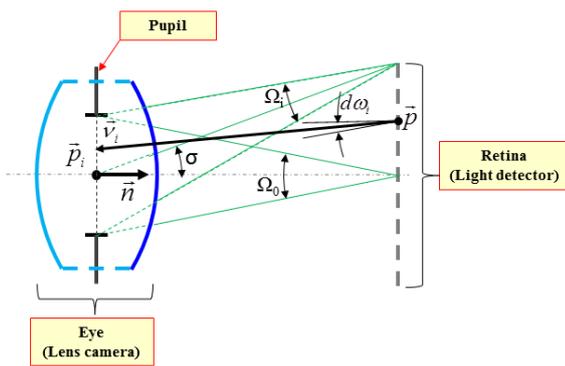


Fig. 5. Local illuminance calculation from the lens exit pupil area.

4. Investigation of the results for a scheme with a diffuse screen

Figure 6 shows the simulation results of an augmented reality photorealistic image (with closed channel of the real environment luminance). The main effect under simulation is the rear projection effect that forms a bright spot on top of an image with graphical and digital information. In this case, the most appropriate method is stochastic bidirectional ray tracing with photon maps. Since there was only one diffuse surface with dozens of specular surfaces on the path between light source and observation plane, all the illumination was reduced to caustic illumination and common methods based on important sampling solutions gave no advantage. Also methods of sole forward or backward ray tracing have not succeeded, that means that no distinct image was synthesized at the equivalent time. The latter is due to low effectiveness of collimation and projection lenses of this device. The most efficient solution was simple bidirectional ray tracing with photon maps.

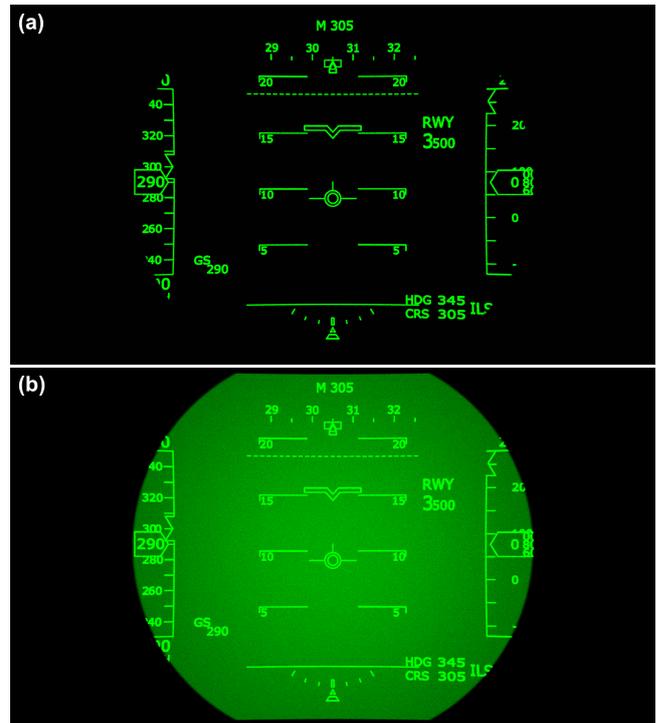


Fig. 6. The HUD image simulation result for case (a) LCD with absolute contrast and (b) LCD with real contrast.

Figure 7 shows the result of an image synthesis in HUD with an open channel to monitor the real environment.



Fig. 7. An example of AR image synthesis for the environment with a high level of luminance.

The HDRI of a spherical panorama was used as a model of the real environment. In this case, the HUD luminance is 1600 cd/m². The maximum and minimum luminance of HDRI is 21013 cd/m² and 292 cd/m², respectively. The eye is adapted to the average luminance, which level, in this case, is 2140 cd/m².

It is clearly seen that in the case of a very bright object in the field of view, the level of LED luminance is not enough for a good visualization of useful information in the HUD.

The next figure (Fig.8) shows a case with a much less level of the real environment luminance.

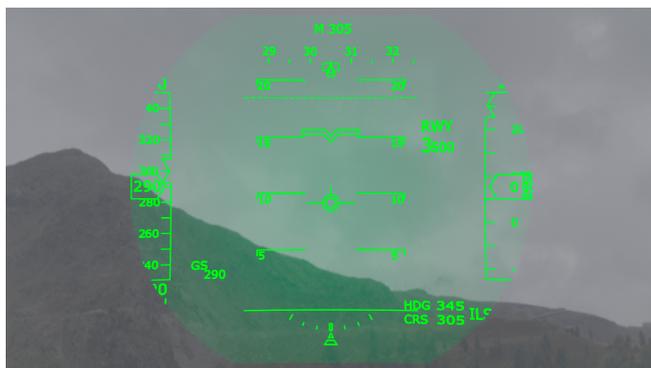


Fig. 8. An example of AR image synthesis for the environment with a low level of luminance.

In this case, the HUD luminance is equal to 10000 cd/m^2 and it is the maximum level of luminance in the entire image. The minimum luminance of HDRI is 34 cd/m^2 . The eye is adapted to the average luminance, which level, in this case, is 234 cd/m^2 .

From the example, it is seen that with a low level of the real environment luminance, all the additional useful information in the HUD is completely distinguishable. However, in this case, a disadvantage in the form of an undesirable bright spot in the field of view is clearly visible, which worsens the conditions for observing the real environment.

By reducing the LED luminance, one can achieve the optimal ratio between luminances of useful graphic information and an unwanted spot caused by the direct visibility of the light source through the non-absolutely black background of the LCD. Figure 9 shows the simulation result of augmented reality image with the optimal level of luminance in the AR channel of HUD.

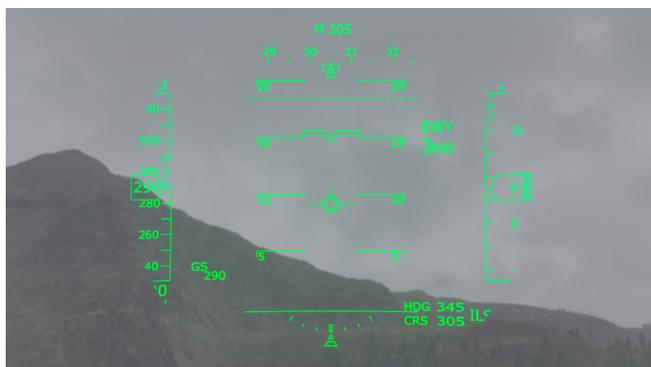


Fig. 9. Example of AR image synthesis for optimal selection of the LED brightness level.

In this example, the HUD luminance value is 1600 cd/m^2 and it is the maximum level of luminance in the entire image. The minimum brightness of HDRI has not changed and is 34 cd/m^2 . The eye is adapted to the average luminance, which value, in this case, has decreased to the level of 142 cd/m^2 .

5. The investigation results for a scheme without diffuse screen

The HUD scheme without a diffuse screen is the most promising, both in terms of dimensions and in terms of energy saving. However, simulation of such a scheme using the method of bidirectional stochastic ray tracing is practically impossible because, in this case, there are no sources of secondary and caustic radiation. In this case, the integral in the rendering equation (1) goes to zero and only the luminance of the direct vision L_0 remains. This circumstance, in turn, leads to a fatal

decrease in the effectiveness of the method of bidirectional ray tracing using photon maps, since we see the source directly, and this source has a very small area of the order of 0.04 mm^2 . In other words, we switch to the ordinary backward ray tracing, which reads the brightness of the primary light source.

Figure 10 shows the HUD images obtained in result of simulation of the scheme without a diffuse screen for the LCD matrix with ideal and real contrast respectively.

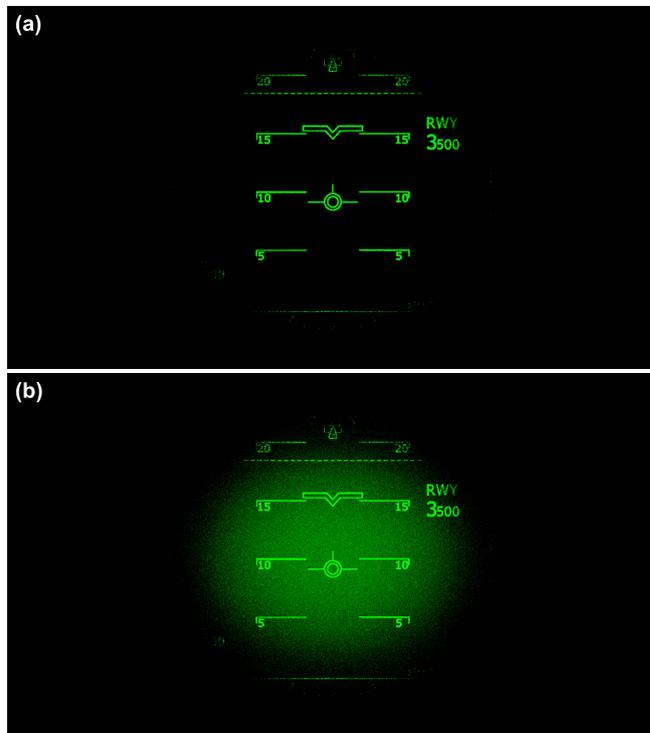


Fig. 10. The HUD image simulation result for case (a) LCD with absolute contrast and (b) LCD with real contrast.

The fact that only part of the LCD is in the field of view is explained by the fact that in this case we used a simplified version of the collimating lens (singlet) with a limited field of view. This circumstance does not hinder our research because the essence of the research is to find the most effective method of stochastic ray tracing for particular types of HUD, and not to design the optical system of the lens with a large field of view.

In first case modeling was performed by the pure Path Tracing method. It collected luminance from LEDs that illuminate the screen. It should be noted that the calculation of the HUD images shown in Fig. 10, takes two days. At the same time, the noise level is quite high, that noticeably reduces the image quality. This is because of the fact that the size of the LED is small and, accordingly, the effectiveness of the backward ray tracing method is low.

Since in this case the method of backward ray tracing turned out to be ineffective, we used the method of forward stochastic ray tracing. The idea of the method is that the light source emits rays with a density proportional to the angular intensity distribution, and the rays, propagated through the optical system, are accumulated on the radiation receiver, which is specified as the model of the eye. We implemented a method of forward stochastic ray tracing and carried out a similar simulation of an image formation in the HUD. The result of this simulation is shown in Fig. 11.

High image quality, obtained by the method of direct stochastic ray tracing, allowed visualization of image defects caused by the peculiarities of illumination optics design. The calculation time required to achieve the same quality is 2 to 3

orders of magnitude lower than in the case of backward ray tracing.

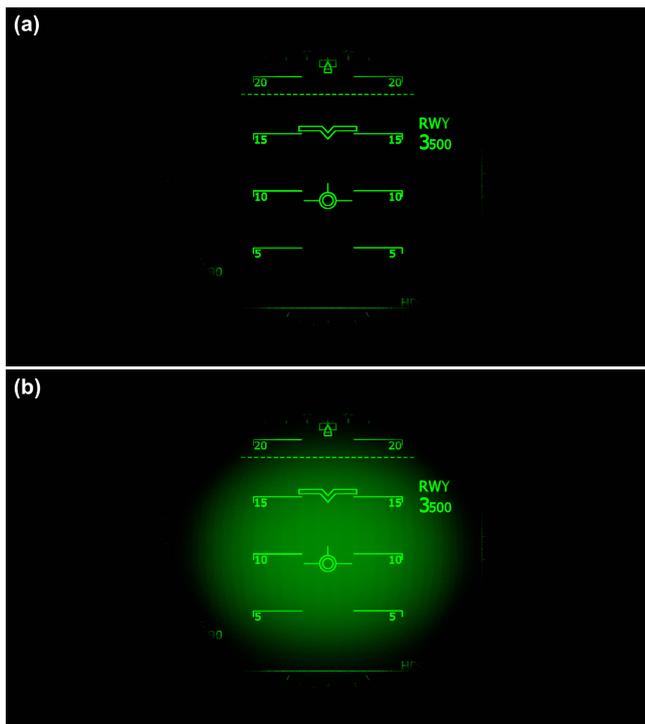


Fig. 11. Example of AR image synthesis with use of forward ray tracing.

To form the full image in real environments, we implemented a hybrid ray tracing method. In this method, the full image is synthesized from images obtained both by the method of forward ray tracing and by the method of bidirectional ray tracing using photon maps. The results of such a hybrid simulation are shown in Fig. 12. At the moment, the automatic generation of image obtained by the hybrid method is not possible and merging is performed manually by summing two synthesized images in real luminance.



Fig. 12. Example of AR image synthesis with use of hybrid ray tracing.

In this case, the luminance of HUD in the central part of the field of view is 10000 cd/m². The maximum and minimum luminance of HDRI is 21013 cd/m² and 292 cd/m², respectively. The eye is adapted to the average luminance, which level, in this case, is 2940 cd/m².

6. Conclusion

As a result of the current research, we found out that the method of forward stochastic ray tracing can be successfully used for virtual prototyping of augmented reality systems.

We extended the model of bidirectional stochastic ray tracing by separate forward ray tracing procedure which efficiently simulates image formed by HUD systems without diffuse components.

We are planning to improve the method of hybrid ray tracing by possibility of automatic selection of the type of ray tracing used to form various image components.

7. Acknowledgements

The work was carried out with partial state financial support of the leading universities of the Russian Federation (subsidy 074-U01), RFBR grant No. 16-01-00552, as well as by Integra Inc.

8. References

- [1] Eric P. Lafortune, Yves D. Willems. Bi-directional path tracing // Proceedings of Third International Conference on Computational Graphics and Visualization Techniques (Compugraphics '93), p.145-153, Alvor, Portugal, December 1993.
- [2] Hachisuka Toshiya, Jensen Henrik Wann. Stochastic progressive photon mapping // ACM Trans. Graph. 2009. dec. Vol. 28, no. 5. P. 141:1 141:8. URL: <http://doi.acm.org/10.1145/1618452.1618487>.
- [3] Iliyan Georgiev, Jaroslav Křivánek, Tomáš Davidovič, Philipp Slusallek, Light Transport Simulation with Vertex Connection and Merging // ACM TOG 31(6), SIGGRAPH Asia 2012.
- [4] J. T. Kajiya. The rendering equation // Computer Graphics (SIGGRAPH '86 Proceedings), 1986, vol. 20, p.143-150.
- [5] James Melzer "Head Mounted Displays", Mac Graw Hill (1997) (ISBN 978-1456563493).
- [6] Jensen Henrik Wann, Christensen Per. High quality rendering using ray tracing and photon mapping // ACM SIGGRAPH 2007 courses. SIGGRAPH '07. New York, NY, USA: ACM, 2007. URL: <http://doi.acm.org/10.1145/1281500.1281593>.
- [7] M.Pharr, G.Humphreys Physically Based Rendering. From theory to implementation // Morgan Kaufmann, 2004.
- [8] Ozan Cakmakci and Jannick Rolland, "Head-Worn Displays: A Review", Journal of Display Technology, Vol. 2, No. 3, September 2006
- [9] Toshiya Hachisuka, Jacopo Pantaleoni, and Henrik Wann Jensen. 2012. A path space extension for robust light transport simulation. ACM Trans. Graph. 31, 6, Article 191 (November 2012), 10 pages
- [10] Veach Eric. Robust monte carlo methods for light transport simulation: Ph. D. thesis. Stanford, CA, USA: Stanford University, 1998. AAI9837162.

About the authors

Zhdanov Dmitry Dmitrievich, PhD, Head of the Visualization Technology Chair of ITMO University.
His e-mail: ddzhdanov@mail.ru.

Potemin Igor Stanislavovich, PhD, Assistant of the Visualization Technology Chair of ITMO University.
His e-mail: ipotemin@yandex.ru.

Kishalov Anton Aleksandrovich, Head of the optoelectronics department of NPP VOLO.
His e-mail: grinfo@mail.ru.

Zhdanov Andrey Dmitrievich, Ph.D. student at ITMO University, Visualization Technology chair.
His e-mail: adzhdanov@corp.ifmo.ru.

Bogdanov Nikolay Nikolaevich, Ph.D. student at ITMO University, Visualization Technology chair.
His e-mail: nnbogdanov@corp.ifmo.ru.

Программная реализация OpenGL SC для авиационных встраиваемых систем

Б.Х. Барладян¹, А.Г. Волобой¹, В.А. Галактионов¹, В.В. Князь², И.В. Ковернинский², Ю. А. Солоделов², В.А. Фролов¹, Л.З. Шапиро¹

bbarladian@gmail.com|voloboy@gin.keldysh.ru|vlgal@gin.keldysh.ru|vl.kniaz@gosniias.ru|
ivkoverninsk@2100.gosniias.ru|yasolodelov@2100.gosniias.ru|vfrolov@graphics.cs.msu.ru|pls@gin.keldysh.ru

¹Институт прикладной математики им. М.В.Келдыша РАН, Москва;

²ФГУП Государственный Научно-Исследовательский Институт Авиационных Систем (ГосНИИАС)

В работе рассматриваются вопросы повышения эффективности программной реализации библиотеки OpenGL SC для ее использования в авиационных встраиваемых системах. Алгоритмы растеризации были оптимизированы с учетом специфики авиационных приложений. Для ускорения визуализации применялись многопоточковые вычисления.

Ключевые слова: OpenGL SC, ускорение визуализации, многопоточковые вычисления.

Software OpenGL SC for aviation embedded systems

B. Kh. Barladian¹, A.G.Voloboy¹, V.A. Galaktionov¹, V.V. Kniaz², I. V. Koverninskiy², Y. A. Solodelov², L.Z.Shapiro¹

¹The Keldysh Institute of Applied Mathematics Russian Academy of Science;

²State Research Institute of Aviation Systems (GosNIIAS)

The rendering efficiency problems of multipurpose OpenGL SC for specific applications of embedded aviation systems are considered. Rasterization algorithms were optimized taking into account aviation specific. The multithreading calculations were applied for rendering acceleration.

Keywords: OpenGL SC, rendering acceleration, multithreading calculations.

1. Введение

В [1] сформулированы требования к операционной системе реального времени, предназначенной для работы с интегрированной модульной авионикой. В частности, должна быть обеспечена поддержка стандартов OpenGL SC/ES [2] и ARINC 661 [3] и возможность сертификации по DO-178C[4]. Необходимость сертификации требует соблюдения корректных процессов разработки ПО, а также полного доступа к исходным кодам библиотеки OpenGL. Важным требованием является независимость от платформы – ядро библиотеки не должно содержать кода, специфичного для конкретной архитектуры или аппаратной платформы. Компания CoreAVI выпускает OpenGL драйверы, использующие графические процессоры [5], но сертификация таких драйверов невозможна без участия разработчиков графических процессоров. В силу этих требований разработка программной OpenGL является практически единственным решением.

Конечно, с помощью программных решений трудно превзойти по скорости визуализации аппаратную реализацию библиотеки. Однако программное решение существенно проще оптимизировать для специальных приложений [6]. Предполагается, что разрабатываемая библиотека OpenGL будет работать в операционной системе реального времени JetOS [7]. В настоящее время JetOS не реализована в полном объеме, в частности, не реализована поддержка потоковых вычислений, поэтому основные исследования проводились под операционной системой Linux.

В настоящее время в литературе описано достаточно много алгоритмов программной реализации OpenGL [8], [9], [10], реализация которых на современных мощных компьютерах, например Intel i7-4770 3.4 GHz, дает вполне приемлемые по скорости результаты для типичных авиационных приложений. Скорость близка к

показателям, показываемым OpenGL драйвером для видеокарты NVIDIA Quadro 410 - ~50 кадров в секунду. К сожалению, та же библиотека OpenGL не обеспечивает требуемой производительности на типичном авиационном компьютере на базе процессора PowerPC e500mc, 4 ядра, 1 ГГц [11].

Типичное авиационное приложение «Пилотажно-навигационный дисплей» (Primary flight display – PFD в англоязычной литературе) на рис. 1 работает со скоростью ~1.7 кадров в секунду, а визуализация рельефа на рис. 2 со скоростью ~1.4 кадров в секунду. Такие скорости визуализации неприемлемы для авиационных приложений.

Мы проанализировали специфику использования авиационными приложениями библиотеки OpenGL. Выяснилось, что большая часть этих приложений использует визуализацию только двумерных (2D) объектов. Эти приложения применяются, в основном, для визуализации значений различных приборов в цифровом и аналоговом виде, положения воздушного судна в пространстве, различных индикаторов состояния систем, метеорологической и картографической информации. Такая визуализация использует ограниченное число сочетаний вызовов OpenGL, учитывая которое мы сумели существенно увеличить скорость визуализации для специализированных приложений. При этом полное покрытие спецификации стандарта OpenGL SC было сохранено.



Рис. 1. Пилотажно-навигационный дисплей.

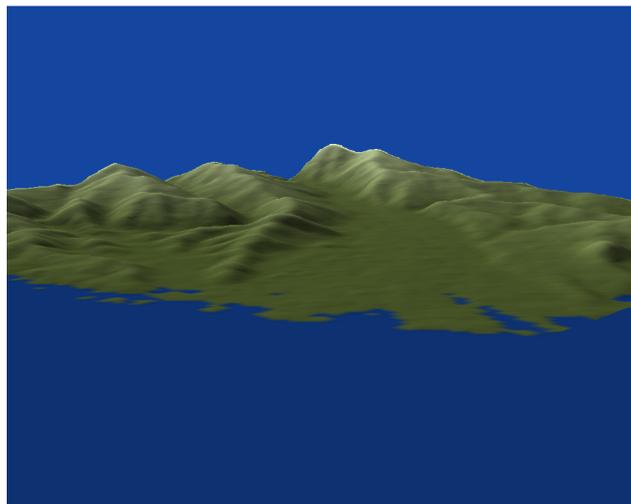


Рис. 2. Визуализация рельефа.

2. OpenGL алгоритмы – специфика использования

Алгоритмы растеризации занимают основное время при работе авиационных приложений. Большинство используемых в настоящее время приложений работают с двухмерными объектами, типичным представителем которых является PFD, приведенное на рис. 1.

Ускорение процедуры растеризации было проведено, в основном, путем применения следующих подходов:

1. Для вычисления значений величин (цвета или текстурных координат) внутри треугольника мы использовали линейную интерполяцию вместо применения барицентрических координат. Интерполяция производится сначала вдоль сторон треугольника, а затем вдоль строки пикселей. Использование линейной интерполяции вдоль

отрезка вместо применения барицентрических координат позволило ускорить растеризацию примерно в 1.9 раза.

2. Везде, где это возможно, мы использовали вычисления с фиксированной точкой. Это дало заметное ускорение в силу специфики архитектуры процессора PowerPC – несколько таких операций могут выполняться за один такт. Использование вычислений с фиксированной точкой позволило ускорить растеризацию примерно в 2.2 раза.
3. Для вычислений в строке пикселей внутри треугольника был использован набор специальных встраиваемых функций.

На последнем пункте следует остановиться отдельно. При анализе 2D авиационных приложений было выявлено, что при растеризации треугольников большая часть времени уходит на проверки условий обработки данного пикселя, в то время как эти условия одинаковы для всех пикселей треугольника. Мы выявили пять таких часто используемых случаев:

1. Заполняется только буфер трафарета (стенсил);
2. Тестируется буфер трафарета, а вершины треугольника имеют один и тот же цвет;
3. Тестируется буфер трафарета, текстура заменяет альфа канал и вершины треугольника имеют один и тот же цвет;
4. Буфер трафарета не используется, текстура заменяет альфа канал и вершины треугольника имеют один и тот же цвет;
5. Вершины треугольника имеют один и тот же цвет и никаких других условий проверять не надо.

При визуализации рельефа (3D приложение) для эффективности, как правило, не используются буфер трафарета, альфа канал, текстура и освещение. Освещение вычисляется один раз заранее и результат в виде цветов в вершинах используется при визуализации. Таким образом, для 3D приложения, где используется Z буфер, рассматриваем два дополнительных случая: вершины треугольника имеют один и тот же цвет или они имеют разные цвета. Если вершины одного цвета, то не нужна интерполяция и цвет, которым они представлены в экранном буфере цвета, можно заранее перед растеризацией упаковать в одну целую величину (тип int). Отсутствие необходимости интерполяции и упаковки цвета на каждом пикселе позволяет существенно повысить скорость растеризации, поэтому мы всегда выделяем этот случай (в сочетании с другими условиями, когда это необходимо). Таким образом, добавляются еще два специальных случая:

1. Используется Z-буфер, вершины треугольника имеют один и тот же цвет и никаких других условий проверять не надо.
2. Используется Z-буфер, вершины треугольника имеют разный цвет и никаких других условий проверять не надо.

Использование такого набора специальных встраиваемых функций позволило дополнительно ускорить растеризацию примерно в 1.7 раза.

Реализация предложенных подходов к алгоритмам растеризации, учитывающих специфику авиационных приложений и используемого процессора, позволила ускорить визуализацию PFD приложения на рис. 1 до ~12 кадров в секунду, а визуализацию рельефа на рис. 2 до скорости ~8 кадров в секунду. При этом все вычисления производились в одном потоке.

3. Использование многопоточковых вычислений

Использование многопоточковых вычислений для увеличения производительности библиотеки OpenGL является естественным развитием, поскольку большинство используемых процессоров являются многоядерными. Естественным подходом является разбиение экрана на прямоугольные области (тайлы) и обработка их различными потоками. Такой подход является классическим для использования многопоточковых вычислений при синтезе изображений с помощью трассировки лучей. Однако наши эксперименты показали неэффективность этого подхода. Причин здесь две. Первая – это дополнительные расходы на синхронизацию потоков. Вторая причина, не менее существенная для 2D авиационных приложений, это расходы на отработку отсечения треугольников границами прямоугольников. В случае отсутствия разбиения экрана на прямоугольники такой проблемы обычно не возникает, поскольку все треугольники, как правило, целиком попадают в область экрана.

В силу этих причин для распараллеливания растеризации мы использовали отдельные вычислительные потоки для построения изображения для разных кадров. В этом случае решаются обе вышеупомянутые проблемы падения производительности. Близкий подход использовался в работе [12]. Количество вычислительных потоков, используемых для растеризации, как правило, выбирается равным количеству ядер процессора. Дополнительно мы используем отдельный поток для копирования изображений, построенных вычислительными потоками, в буфер экрана (память видеокарты), поскольку затраты на эту процедуру являются критическими с точки зрения скорости визуализации. Поэтому ее следует выполнять параллельно с другими вычислениями, необходимыми для растеризации. Поскольку копирование изображения выполняется для кадра, для которого растеризация уже закончена, то никаких проблем с синхронизацией потоков не возникает. Синхронизация главного потока (main thread), вычислительных потоков непосредственно реализующих растеризацию и потока копирования изображений осуществляется с помощью событий. Используются следующие события:

1. N событий $s_ev[i]$ для синхронизации начала работы i -го вычислительного потока;
2. N событий $e_ev[i]$ сигнализирующих о завершении растеризации кадра i –ым вычислительным потоком;
3. N событий $c_ev[i]$ сигнализирующих о готовности изображения, построенного i –ым вычислительным потоком, для копирования на экран.

Как уже было упомянуто выше, N обычно выбирается равным количеству ядер процессора. На этапе инициализации создаются N событий s_ev , e_ev и c_ev . Затем события s_ev и c_ev сбрасываются, а e_ev взводятся. Создаются N вычислительных потоков для растеризации N контекстов $cont[i]$ для хранения информации о вызовах OpenGL и изображений, получаемых в результате растеризации вычислительными потоками. Создается также отдельный поток для копирования изображений, построенных вычислительными потоками, в буфер экрана. Индекс обрабатываемого кадра в главном потоке m_idx и индекс обрабатываемого кадра в потоке копирования изображений c_idx первоначально устанавливаются в 0. Нарастивание индексов в процессе визуализации будет происходить по модулю N .

Алгоритм работы главного потока приведен на рис. 3:

```

1. Wait( $e\_ev[m\_idx]$ )
2. Обработка OpenGL вызовов =>  $cont[m\_idx]$ 
3. Set( $s\_ev[m\_idx]$ )
4.  $m\_idx = (m\_idx++) \% N$  и возврат к п.1

```

Рис. 3. Алгоритм работы главного потока (main thread).

Поскольку все события e_ev взведены при инициализации $m_idx = 0$, то главный поток начинает сразу обрабатывать вызовы OpenGL для первого кадра и записывать результаты в контекст $cont[m_idx]$. По окончании вызовов для данного кадра он взводит событие $s_ev[m_idx]$ для запуска соответствующего вычислительного потока, переходит к следующему кадру по модулю N и снова ждет взведения события $e_ev[m_idx]$.

Алгоритм работы вычислительного потока idx ($idx = 0, 1, \dots, N$), производящего растеризацию, приведен на рис. 4:

```

1. Wait( $s\_ev[idx]$ )
2. Обработка  $cont[idx]$  => изображение
3. Set( $c\_ev[idx]$ )
4. Reset( $s\_ev[idx]$ ) и возврат к п.1

```

Рис. 4. Алгоритм работы вычислительного потока idx .

Работа вычислительного потока idx , производящего растеризацию, начинается после взведения события $s_ev[idx]$ главным потоком после заполнения контекста $cont[idx]$. Поток обрабатывает данные в контексте и строит из них путем растеризации треугольников, линий и точек изображение в памяти контекста. После этого он взводит событие $c_ev[idx]$, сбрасывает событие $s_ev[idx]$ и возвращается к п.1 ждать взведения события $s_ev[idx]$.

Алгоритм работы потока, производящего копирование изображений из памяти контекста в память видеокарты, приведен на рис. 5:

```

1. Wait( $c\_ev[c\_idx]$ )
2. Копирование изображения  $cont[c\_idx]$  => экран
3. Reset( $c\_ev[c\_idx]$ )
4. Set( $e\_ev[c\_idx]$ )
5.  $c\_idx = (c\_idx++) \% N$  и возврат к п.1

```

Рис. 5. Алгоритм работы потока копирования изображений

Работа потока копирования изображения начинается после взведения события $c_ev[c_idx]$ соответствующим вычислительным потоком. При инициализации индекс c_idx установлен в 0. После копирования изображения из контекста $cont[c_idx]$ на экран (в память видеокарты) поток сбрасывает событие $c_ev[c_idx]$, взводит событие $e_ev[c_idx]$, наращивает индекс кадра c_idx по модулю N и снова ждет взведения события $c_ev[c_idx]$.

Зависимость скорости растеризации (число кадров в секунду) от числа используемых потоков для описанного алгоритма для 4х ядерного процессора PowerPC приведена на рис. 6. Красный график показывает зависимость для 3D приложения (визуализация рельефа), синий график – для PFD приложения.

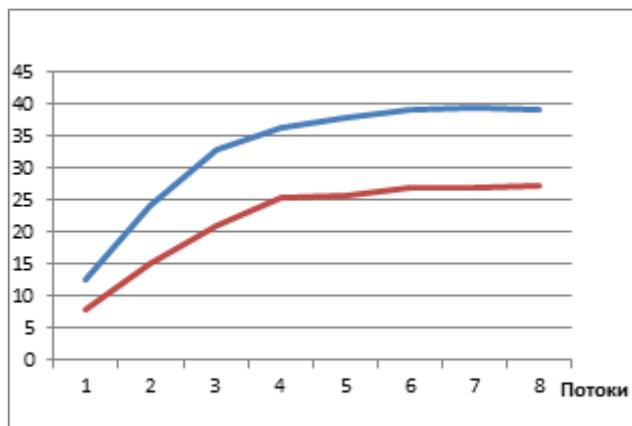


Рис. 6. Зависимость скорости растеризации (кадры в секунду) от числа используемых потоков.

Использование 4х потоков обеспечивает ускорение растеризации примерно в 2.9 - 3.2 раза.

Отдельно проведенное тестирование показало, что только копирование изображения из памяти на экран (в память видеокарты) занимает ~ 0.0165 сек для одного кадра, а работа приложения вместе с записью вызовов OpenGL в контекст и копированием посылаемых данных занимает ~ 0.006849 сек. Суммарное время ~ 0.023349 сек. Таким образом, получаем оценку сверху для производительности данного приложения ~ 42.83 кадра в секунду.

После оптимизации функций растеризации в п.2 и использования многопоточковых вычислений для них критическим, с точки зрения производительности, оказалась процедура копирования построенного изображения из памяти процессора в память видеокарты. По этой причине мы вынесли эту процедуру в отдельный поток с более высоким приоритетом. И если при реализации этой процедуры в главном потоке была достигнута скорость ~ 28 кадров в секунду, то при выделении для нее отдельного потока удалось достичь скорости ~ 36.5 кадров в секунду.

4. Результаты

Проведенные исследования показали возможность реализации программной OpenGL библиотеки на типичном авиационном компьютере на базе процессора PowerPC e500mc, 4 ядра, 1 ГГц, с производительностью, удовлетворяющей требованиям авиационных приложений. При этом библиотека полностью написана на языке C, не использует никаких специфических команд процессора PowerPC и переносима на любой компьютер, на котором поддерживается компилятор языка C. Процесс разработки библиотеки строго контролируется в соответствии с требованиями DO-178C, что в дальнейшем предоставляет возможность сертификации данной реализации OpenGL.

5. Благодарности

Работа поддержана грантами РФФИ 15-01-01147 и 16-01-00552.

6. Литература

- [1] Федосов Е.А., Ковернинский И.В., Кан А.В., Солоделов Ю.А., Применение операционных систем реального времени в интегрированной модульной авионике. OSDAY 2015, <http://osday.ru/solodelov.html>.
- [2] Safety Critical Working Group, <https://www.khronos.org/openglsc/>.
- [3] ARINC Specification 661-6: http://store.aviation-ia.com/cf/store/catalog_detail.cfm?item_id=2698.
- [4] DO-178C Software Considerations in Airborne Systems and Equipment Certification, http://www.rtca.org/store_product.asp?prodid=803.
- [5] ArgusCore SC™ OpenGL SC 1.0.1 / SC 2.0 graphics drivers for Safety Critical Systems, http://www.coreavi.com/sites/default/files/coreavi_product_brief_-_arguscore_sc1_sc2_rev_a_0.pdf.
- [6] P. M. a. J. Dudra, ADVANCED 2D RASTERIZATION ON MODERN CPUS // Applied Information Science, Engineering and Technology: Selected Topics from the Field of Production Information, т. 7, № 5, 2014.
- [7] Маллачиев К.М., Пакулин Н.В., Хорошилов А.В., Устройство и архитектура операционной системы реального времени // Труды ИСП РАН, том 28, вып. 2, 2016 г., стр. 181-192 (на английском). DOI: 10.15514/ISPRAS-2016-28(2)-12.
- [8] G. Wihlidal, Optimizing the Graphics Pipeline with Compute // GDC, 2016.
- [9] T. K. Samuli Laine, High-Performance Software Rasterization on GPUs // High-Performance Graphics, 2011.
- [10] The Mesa 3D Graphics Library, <http://www.mesa3d.org>.
- [11] Модуль универсального процессора данных МУПД/P3041-VPX 3U, <http://www.nkbvs.ru/products/elektronnie-modyli/vpx-3u/moduli-universalnogo-protssessora-dannix-mypd-p3041/>.
- [12] А.Н. Милов, Особенности построения архитектуры масштабируемой графической системы стандарта OpenGL на основе ЦПОС // Труды 17-ой международной конференции по компьютерной графике и зрению, Россия, Московский Государственный Университет, июнь 23-27, 2007, с. 281-284.

Эффективные методы и модели цифровой обработки киноматериала

Е.В. Боревиц, С.В. Мещеряков, В.Э. Янчус
 plasma5210@mail.ru|serg-phd@mail.ru|victorimop@mail.ru
 Санкт-Петербургский политехнический университет, Санкт-Петербург, Россия

В статье представлены новые эффективные методы и 3D модели для цифровой постобработки киноматериала, включая задачи захвата движения с использованием инфракрасных маркеров, воссоздания реалистичных теней динамических 3D объектов, а также цифровой цветокоррекции кинокадров с целью улучшить их восприятие зрителем в кинематографе. Выработаны практические рекомендации по цифровой обработке киноматериала, которые внедрены и успешно применяются в малобюджетной виртуальной видеостудии при создании короткометражных фильмов и видеоклипов.

Ключевые слова: кинокадр, цифровая цветокоррекция, 3D моделирование, захват движения, виртуальная реальность.

Effective methods and models for digital processing of the film contents

E.V. Borevich, S.V. Mescheryakov, V.E. Yanchus
 plasma5210@mail.ru|serg-phd@mail.ru|victorimop@mail.ru
 St. Petersburg Polytechnic University, St. Petersburg, Russia

The paper presents new effective methods and 3D models for digital film post-processing, including the problems of motion capture using infrared markers, recreation of realistic shadows of dynamic 3D objects, and digital colour correction of film frames for the purpose of better perception of the film contents by a viewer in the cinematography. Practical recommendations for digital film processing are worked out and are successfully implemented in a virtual visualization studio to produce short films and video clips having low budget.

Keywords: film frame, digital colour correction, 3D modelling, motion capture, virtual reality.

1. Введение

С развитием современных информационных технологий и компьютерной графики целиком изменился технологический процесс производства фильмов в кинематографе и на телевидении (рис. 1). Кроме замены классического монтажа на нелинейный, автоматизированные программно-аппаратные комплексы предоставляют большие возможности постобработки и улучшения отснятого киноматериала [8]:

- 1) добавлять визуальные эффекты совмещением видео и компьютерной графики для привлечения внимания;
- 2) создавать реалистичные тени как статических, так и динамических 3D объектов, что является главным отличием художественного кино от анимационного;
- 3) применять различные схемы цифровой цветокоррекции кинокадров с целью их художественной гармонизации и улучшения восприятия зрителем.

Данная работа посвящена исследованиям 2-го и 3-го этапов постобработки как менее изученных и наиболее затратных по времени и вычислительным ресурсам. Конечной целью постобработки является повышение эффективности всего процесса кинопроизводства, достижение максимальной реалистичности и визуальной привлекательности киноматериала на основе разработки новых моделей и методов графических преобразований.

2. Построение реалистичных теней в виртуальном пространстве

После захвата движения актера и перехода к виртуальной сцене все тени 3D объектов теряются, и необходимо их воссоздать с нуля (рис. 2). Построение теней является эффективным методом достижения максимальной реалистичности сцены в виртуальном пространстве. Тени, отбрасываемые от движущихся и

статичных объектов, создают объем, обеспечивают глубину сцены, показывают взаимное расположение объектов. Построение теней является достаточно сложной задачей, особенно для динамических 3D объектов, так как человек замечает «неправильные» тени и необходим их точный расчет.



Рис. 1. Структура современного кинопроизводства.

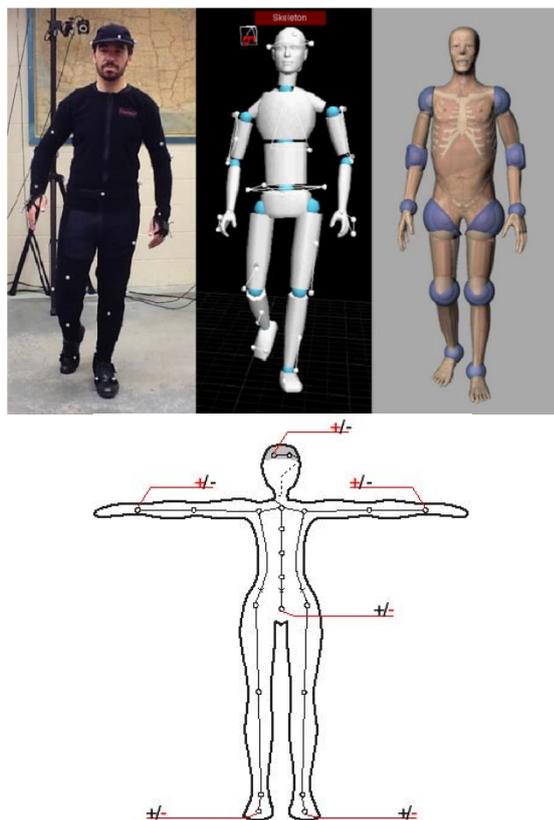


Рис. 2. Топология костюма с инфракрасными маркерами для захвата движения.

Существующие методы и модели захвата движения требуют больших вычислительных ресурсов [10]. С другой стороны, использовать высокоточную модель 3D объекта для расчета его теней не имеет смысла. Поэтому для задачи построения теней предложен новый метод на основе низкополигональной 3D модели, тени которой от источников освещения предполагается создавать такими же правильными и реалистичными, но объем их вычислений намного меньше. В случае создания теней от динамического 3D объекта модель анимируется, для чего создается скелет и система управления (рис. 3).

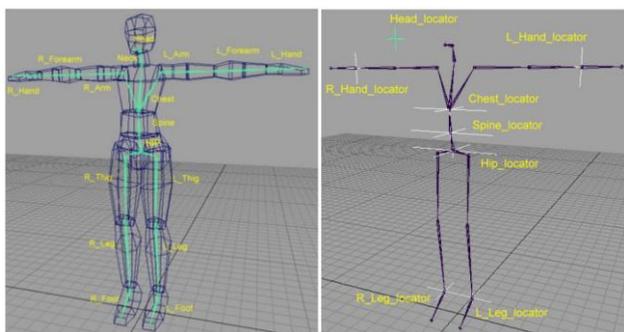


Рис. 3. Низкополигональная 3D модель и система ее управления.

Мощности современных компьютеров позволяют в реальном времени выполнять вычисления и рендеринг около 1,5 млн пикселей с максимальной глубиной цвета 24М и стандартной частотой 50 кадров в сек. Однако анимационный фильм может иметь тысячи 3D объектов, каждый из которых содержит от сотен до тысяч полигонов. Объем вычислений также зависит от разрешающей способности экрана (до 4К пикселей в кинотеатрах IMAX)

и его доли, которую занимает 3D объект. Учитывая, что низкополигональная модель насчитывает не более 10К полигонов в зависимости от размера 3D объекта, ее преимущество в объеме вычислений по сравнению с высокополигональными моделями оценивается минимум в 100 раз [11].

Другой проблемой постобработки отснятого материала является трудоемкость процесса удаления следов маркеров захвата движения, оставшихся после съемки инфракрасной (ИК) камерой. В традиционных студиях с декорациями и актером используются специальные костюмы с системой маркеров (рис. 2). Но эта технология не подходит для случаев, когда актер присутствует в виртуальной сцене, так как маркеры видны на отснятом материале [11].

Для решения этой проблемы разработана специальная система динамических ИК маркеров, которая представляет собой параллельное соединение трех цепочек диодов, соединенных последовательно как показано на рис. 2. В качестве маркеров использованы ИК светодиоды, которые фиксируются только специальными ИК камерами. Одежда, надетая поверх маркерного костюма, не поглощает свет в ИК диапазоне. В результате на главной камере маркеры не видны, а ИК камеры их фиксируют.

Наряду с динамическими маркерами, в виртуальной 3D сцене используются и статические калибровочные маркеры, которые необходимы для реконструкции камеры (рис. 4). Вначале определяются оси пространства, относительно которых будет понятно расположение виртуального окружения и положение актера. Калибровочные маркеры располагаются на зеленом фоне.

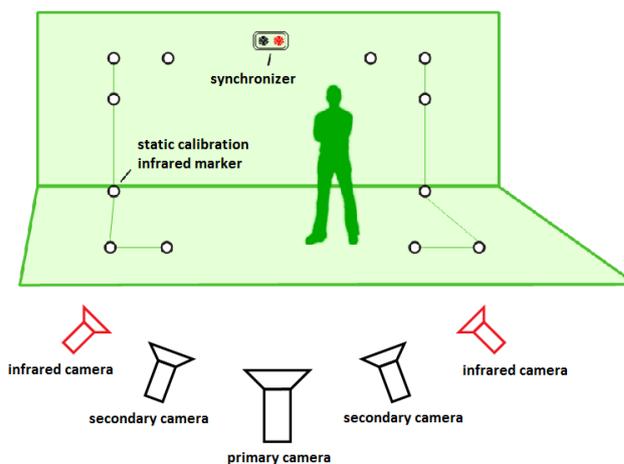


Рис. 4. Расположение маркеров и камер в виртуальном съемочном пространстве.

Для съемок используется пять синхронно записывающих с разных ракурсов камер. Все камеры статичны. Три из них снимают в видимом диапазоне. Главная камера стоит по центру, материал с этой камеры используется в создаваемом кадре. Две боковые камеры нужны для реконструкции камеры. Рядом с ними находятся еще две вспомогательные камеры, снимающие в ИК диапазоне.

Важно, чтобы подвижные маркеры были видны для всех ИК камер, так как они обеспечивают правильность траектории движения 3D образа персонажа. В случае, если при движении какой-либо маркер закрывается, его ставят на другое место для обеспечения видимости на всех камерах. Такая технология использования ИК маркеров значительно упрощает и ускоряет обработку отснятого киноматериала без ущерба его качеству.

3. Исследование влияния цветового решения на визуальную привлекательность кадра

Спектр задач, решаемых с помощью коррекции цвета, очень широкий. Однако особое место занимает художественная цветокоррекция как творческий инструмент режиссера. Она позволяет сделать кадр более выразительным и визуально привлекательным.

Одна из актуальных проблем принятия цветовых решений состоит в том, что в настоящее время не существует единицы количественного измерения визуальной привлекательности кинокадра, что не позволяет объективно сравнивать кадры между собой. Качество киноматериала оценивается субъективными методами экспертного анализа [1, 2].

На рис. 5 изображены элементы кинокадра, влияющие на его визуальную привлекательность. Одним из них является информативность киноматериала [5]. Это пассивная составляющая визуальной привлекательности, поскольку она зависит от замысла режиссера и не имеет прямых средств реализации. Активной составляющей визуальной привлекательности кинокадра является композиционное построение, так как это средство реализации кадра и наполнения его информативностью.

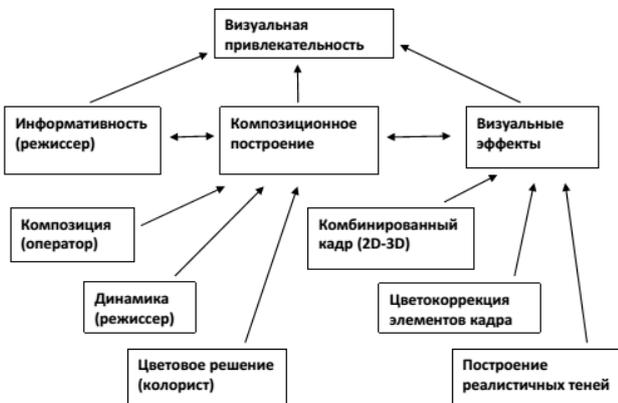


Рис. 5. Влияние на визуальную привлекательность кадра.

Композиционное построение кадра состоит из трех основных элементов:

- 1) композиция, которую контролирует оператор;
- 2) динамика, движение, которым управляет режиссер и которое является главным отличием кино от фотографии;
- 3) цветовое решение, которым занимается колорист.

Таким образом, на визуальную привлекательность кинокадра влияют информативность и композиционное построение. В свою очередь информативность кадра определяется замыслом режиссера, а реализация замысла происходит с помощью композиционного построения, состоящего из трех составляющих – композиции, динамики и цветового решения. Композиционное построение кадра, как его неотъемлемая составляющая, влияет на процесс рассматривания [6]. Тогда логично предположить, что информативность кинокадра тоже влияет на процесс рассматривания. Параметры рассматривания кинокадров, такие как количество и длительность зрительных фиксаций, являются количественными, и их можно объективно измерить с помощью ай-трекера – программно-аппаратного комплекса фиксации глазодвигательной активности человека, основанного на принципе отражения ИК лучей от зрачка [7].

Для проверки сделанных выводов поставлен эксперимент. В качестве стимульного материала выбраны

кадры из фильмов, имеющие два центра интереса и нейтральный фон. Каждый кадр изменен на основе пяти разных цветовых решений согласно теории цветовых контрастов Иттена [3]. В итоге была сформирована база из 120 кадров.

В эксперименте участвовало 30 испытуемых, которым демонстрировались 600 стимулов в пяти цветовых решениях. Всего с помощью ай-трекера собрано более 8000 фиксаций. Пример одного из графиков распределения параметров рассматривания кинокадров представлен на рис. 6.

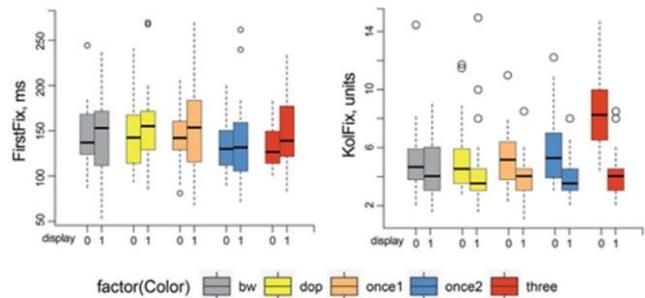


Рис. 6. График распределения выборок измеряемых параметров рассматривания кинокадров.

Для математической обработки статистических данных разработана программа на языке программирования R [4]. Для проверки и анализа выборок с нормальным распределением использовался критерий Шапиро-Уилка, а для ненормального – тест Уилкоксона. Для принятия гипотезы о статистической значимости сравниваемых выборок использован известный критерий p -value < 0,001 [4]. Такие значения выделены зеленым цветом для триадной схемы на рис. 7, где приведен пример расчета значений p -value для следующих параметров рассматривания кадров:

- Time – среднее время рассматривания одного стимула;
- KolFix – среднее количество зрительных фиксаций при рассматривании одного стимула;
- FirstFix – средняя длительность первой фиксации;
- SumFix – средняя длительность всех фиксаций.

p-value	achromatic scheme (bw)	complementary scheme (dop)	warm monochrome (once1)	cold monochrome (once2)	triadic scheme (three)
Time	0,227	0,041	0,015	0,004	5,738e-10
KolFix	0,218	0,029	0,009	0,002	1,26e-7
FirstFix	0,768	0,539	0,355	0,852	0,2311
SumFix	0,499	0,058	0,064	0,017	1,386e-6

Рис. 7. Сравнение параметров рассматривания кинокадров в различных цветовых решениях.

4. Практические рекомендации по цифровой цветокоррекции киноматериала

Полученные результаты объективно подтверждают, что фактор цветового решения имеет приоритетное значение при гармонизации кинокадра. Это позволило сформулировать практические рекомендации по цветокоррекции на этапе постобработки с целью правильной интерпретации кинокадра и улучшения его восприятия зрителем:

- 1) при работе с ахроматическим или монохромным кадром важным моментом является облегчение фона сцены, то есть кинокадр необходимо «вычищать», убирать мелкие незначимые элементы, которые перегружают кинокадр и, как следствие, увеличивают время его анализа;

2) использование комплементарного цветового решения (двух дополнительных противоположных цветов) является способом ускорения восприятия кадра, поэтому этот прием работы с цветом рекомендуется в жанрах кино с высокой динамикой сюжета и частой сменой кадров.

В результате исследования установлено, что фактор цветового решения имеет значимое влияние на процесс распознавания изображения человеком. Данный вывод полезен при композиционном построении кадра на начальном этапе проектирования фильма – раскадровке. В зависимости от режиссерской идеи можно использовать разное цветовое решение. На основании этого сформулированы практические рекомендации по цветовому решению кинокадра:

- 1) использование разных цветовых решений влияет на скорость чтения кадра зрителем, что влечет за собой возможность управления длительностью кадра, то есть в зависимости от режиссерского замысла можно цветовым решением управлять динамикой повествования фильма;
- 2) контрастные цветовые решения существенно облегчают задачу композиционного построения кадра, поскольку цвет передает дополнительную информацию о форме и идентификации объектов кадра, глубине сцены;
- 3) при использовании триадного цветового решения кадр получается наиболее читаемым в условиях его перегруженности, однако триада является наиболее сложным цветовым решением и требует грамотного сочетания и гармонизации трех основных цветов в кадре.

5. Практическая реализация и апробация результатов исследования

Новые методы цифровой постобработки киноматериала являются логическим продолжением исследований, представленных на международных научно-технических конференциях [1, 2] и опубликованных в журналах [8, 11]. Предложенные методы внедрены в лаборатории визуализации и компьютерной графики Санкт-Петербургского политехнического университета (СПбПУ) и успешно апробированы на некоторых низкобюджетных видеопроектах, например, воссоздание теней в фильме «Тень» Е. Шварца.

Сформулированные практические рекомендации по цветокоррекции применяются студентами и аспирантами при создании короткометражных фильмов и рекламных видеороликов для ежегодного международного фестиваля короткометражного кино Movie Art Fest [9].

6. Заключение

Таким образом, в работе получены следующие основные результаты, имеющие научную новизну:

- 1) новая схема размещения ИК маркеров, позволяющая целиком перейти на цифровую обработку в виртуальной реальности без макетов и декораций и исключить трудоемкий этап удаления следов маркеров из кинокадров;
- 2) новый эффективный метод поверхностного моделирования динамических 3D объектов при решении задачи воссоздания теней, который позволяет сократить объем вычислений в сотни раз без потери качества и реалистичности изображения;
- 3) методика проведения и результаты статистического и вычислительного экспериментов по исследованию влияния различных схем цветокоррекции на визуальную привлекательность кадров и их восприятие зрителем;

4) новые численные критерии объективной оценки информативности и привлекательности кадра, такие как количество и длительность зрительных фиксаций и пр., которые использованы совместно с технологией ай-трекинга для обработки результатов вычислительного эксперимента;

5) практические рекомендации по цветокоррекции с целью улучшения восприятия и визуальной привлекательности киноматериала, которые внедрены в лаборатории визуализации и компьютерной графики СПбПУ и используются студентами и аспирантами при создании короткометражных фильмов и рекламных видеороликов для международных конкурсов и кинофестивалей.

7. Литература

- [1] Борович Е.В., Мещеряков С.В., Янчус В.Э. Экспериментальное исследование компьютерной цветокоррекции на основе бинарной модели визуального восприятия / Компьютерное моделирование (КОМОД-2016): труды науч.-техн. конф., СПб, СПбПУ, 2016.
- [2] Борович Е.В., Янчус В.Э. Исследование влияния фактора цветовых гармоний в композиционном решении кинокадра на восприятие его зрителем / Неделя науки СПбПУ: Материалы научного форума с междунар. участием, СПб, 2015.
- [3] Иттен И. Искусство цвета. – М.: Аронов Д., 2014.
- [4] Кабаков Р.И. R в действии. Анализ и визуализация данных в программе R. – М.: ДМК Пресс, 2014.
- [5] Лотман Ю.М. Семиотика кино и проблемы киноэстетики. Об искусстве. – СПб., 1998.
- [6] Матюшин М.В. Справочник по цвету. Закономерность изменчивости цветовых сочетаний. – М.: Аронов Д., 2007.
- [7] Орлов П.А., Лаптев В.В., Иванов В.М. К вопросу о применении систем ай-трекинга // Научно-технические ведомости СПбГПУ. – 2014. – № 5 (205).
- [8] Янчус В.Э. Компьютерная обработка видеоматериала в кинематографической промышленности // Научно-технические ведомости СПбГПУ. – 2016. – № 2 (241).
- [9] Annual International Festival of Students Short Films (MovieArtFest-2017). <http://www.movieart.spbstu.ru>
- [10] Bodenheimer B., Rose C., Rosenthal S., and Pella J. The Process of Motion Capture: Dealing with the Data. Computer Animation and Simulation, Springer-Verlag, 1997.
- [11] Mesheryakov S.V., Shchemelinin D.A., and Yanchus V.E. Effective Technique to Reduce Big Data Computations in 3D Modeling of Dynamic Objects. Humanities and Science University Journal, Vol. 17, 2016.

Об авторах

Борович Екатерина Владиславовна, студентка кафедры инженерной графики и дизайна Санкт-Петербургского политехнического университета Петра Великого. Ее e-mail plasma5210@mail.ru.

Мещеряков Сергей Владимирович, д.т.н., профессор кафедры инженерной графики и дизайна Санкт-Петербургского политехнического университета Петра Великого. Его e-mail serg-phd@mail.ru.

Янчус Виктор Эдмундасович, старший преподаватель кафедры инженерной графики и дизайна Санкт-Петербургского политехнического университета Петра Великого. Его e-mail victorimop@mail.ru.

Компьютерная графика и качество освещения реальных осветительных установок в светотехнике

Владимир Будак, Виктор Желтов, Виктор Чембаев
budakvp@mpei.ru | zheltov@list.ru | chembervint@gmail.com

кафедра светотехники

Национальный исследовательский университет "МЭИ", Москва, Россия

Современные методы моделирования уравнения глобального освещения, в частности, основанные на локальных оценках метода Монте-Карло, позволяют определить пространственно-угловое распределение яркости в произвольной точке сцены. Это открывает новые возможности по определению качественных характеристик освещения сцены при светотехническом проектировании. В работе рассматривается новый подход к определению критерия качества освещения, и возможности перехода на новый уклад проектирования реальных осветительных установок на основе достижений в моделировании уравнения глобального освещения в компьютерной графике.

Ключевые слова: светотехника, компьютерная графика, критерий качества освещения, уравнение глобального освещения.

Computer graphics and illumination quality of real lighting systems in the lighting technology science

Vladimir Budak, Victor Zheltov, Victor Chembayev
budakvp@mpei.ru | zheltov@list.ru | chembervint@gmail.com

department of Lighting technology

Moscow Power Engineering Institute, Moscow, Russia

Modern methods of global illumination equation modelling including based on Local Estimations of a Monte Carlo method enable us to determine spatial angular radiance distribution in the specified point of the scene. It gives us new possibilities for determining the qualitative characteristics of a scene lighting in lighting design. This article considers the new approach to determine the illumination quality criteria and the possibility of switching to a new way of designing real lighting systems based on the achievements in modeling the global illumination equation in computer graphics.

Key words: lighting technology science, computer graphics, illumination quality criteria, global illumination equation.

1. ВВЕДЕНИЕ

Светотехника – это область науки и техники, предметом которой являются исследование принципов и разработка способов генерирования, пространственного перераспределения и измерения характеристик оптического излучения, а также преобразование его энергии в другие виды энергии и использование в целях освещения [4].

Одним из важнейших практических направлений светотехники является проектирование осветительных установок (ОУ). Основной задачей проектирования ОУ является определение типа, мощности, количества, месторасположения и ориентации световых приборов, обеспечивающих нормативные светотехнические требования к данной ОУ [4]. Нормативные требования зависят от освещаемого объекта, и могут, как отсутствовать вовсе в бытовом освещении, так и быть крайне строгими, как, например, в дорожном или спортивном освещении.

Задача проектирования ОУ появилась еще задолго до появления электричества, и решалась еще во времена освещения свечами. Однако не только задачи утилитарного освещения успешно решались до появления компьютерной техники, но и задачи дорожного и архитектурного освещения. Так, на рисунке 1 представлена фотография архитектурно-художественного освещения Московского Кремля в 1896 году.

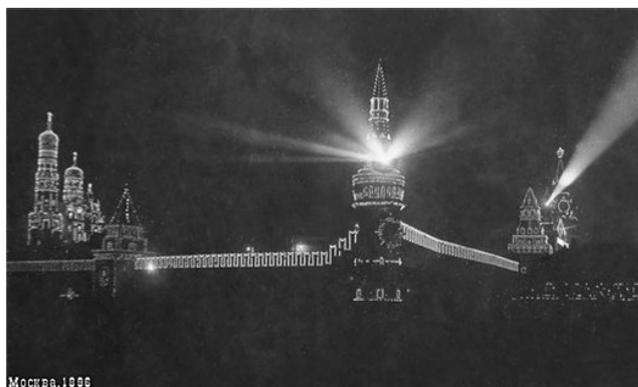


Рис. 1. Архитектурно художественная подсветка Московского Кремля в 1896 году

Однако проектирование ОУ до 90-х годов прошлого века представляло собой набор сложных изнурительных инженерных расчетов и очень часто опиралось на многолетний опыт и интуицию проектировщика.

2. КОМПЬЮТЕРНАЯ ГРАФИКА И СВЕТОТЕХНИКА

Появление компьютеров и развитие математических методов совершило подлинную революцию в светотехнике. Однако, эта революция была связана с развитием в другой области – трехмерной компьютерной графике. основополагающей работой тут является статья J. Kajia [2], в которой

формулируется уравнение глобального освещения (УГО), которое может быть записано в привычном нам виде

$$L(\mathbf{r}, \hat{\mathbf{I}}) = L_0(\mathbf{r}, \hat{\mathbf{I}}) + \frac{1}{\pi} \oint L(\mathbf{r}, \hat{\mathbf{I}}') \sigma(\mathbf{r}; \hat{\mathbf{I}}, \hat{\mathbf{I}}') |(\hat{\mathbf{N}}, \hat{\mathbf{I}}')| d\hat{\mathbf{I}}', \quad (1)$$

где $L(\mathbf{r}, \hat{\mathbf{I}})$ – яркость светового поля в точке \mathbf{r} по направлению $\hat{\mathbf{I}}$, $\sigma(\mathbf{r}; \hat{\mathbf{I}}, \hat{\mathbf{I}}')$ – коэффициент яркости элемента поверхности в точке \mathbf{r} в направлении $\hat{\mathbf{I}}$ при его направленном освещении параллельным пучком света в направлении $\hat{\mathbf{I}}'$, часто называемая двунаправленной функцией рассеяния (отражения или пропускания), L_0 – прямая компонента яркости, непосредственно от источников, $\hat{\mathbf{N}}$ – нормаль в точке \mathbf{r} к элементу поверхности сцены. Здесь и далее крышечкой над вектором будем обозначать единичные вектора. Контурный интеграл в уравнении означает интегрирование по полному телесному углу, элемент которого обозначен через $d\hat{\mathbf{I}}'$.

Именно на моделировании этого уравнения и построена современная трехмерная компьютерная графика, где степень реализма воспроизведения трехмерных сцен уже выходит на уровень постановки вопроса – является ли визуализация более реалистичной, нежели фотография [6]. Современное моделирование осветительных установок также неразрывно связано с решением уравнения глобального освещения, т.е. по сути, задача визуализации в трехмерной компьютерной графике эквивалентна проектированию ОУ. Однако, в сегодняшней инженерной практике моделирование базируется не на решении УГО, а на решении уравнения излучательности [1] методом конечных элементов [7]

$$M(\mathbf{r}) = M_0(\mathbf{r}) + \frac{\sigma(\mathbf{r})}{\pi} \int_{\Sigma} M(\mathbf{r}') F(\mathbf{r}, \mathbf{r}') \Theta(\mathbf{r}, \mathbf{r}') d^2\mathbf{r}', \quad (2)$$

где $M(\mathbf{r})$ – светимость в точке \mathbf{r} , $M_0(\mathbf{r})$ – начальная светимость в точке \mathbf{r} непосредственно от источников света в приближении одного отражения, $\Theta(\mathbf{r}, \mathbf{r}')$ – функция видности точки \mathbf{r} из точки \mathbf{r}' (1 если видна и 0 если не видна), $F(\mathbf{r}, \mathbf{r}')$ – элементарный форм-фактор [7], $\sigma(\mathbf{r})$ – коэффициент отражения.

Решение уравнения (2) методом конечных элементов получило собственное название – метод излучательности. Метод стал лишь небольшой вехой в компьютерной графике, оставив наиболее яркий след в виде программы моделирования – Lightscape. Однако, для светотехнического проектирования осветительных установок он стал основой на многие годы, выразившись в двух ключевых системах автоматизированного проектирования – DIALux [10] и Relux [11].

3. МОДЕЛИРОВАНИЕ СВОТТЕХНИЧЕСКИХ ХАРАКТЕРИСТИК

При светотехническом моделировании ОУ среди определяемых нормируемых характеристик выделяют два основных вида – количественные и качественные. К количественным величинам можно отнести, в первую очередь, освещенность и яркость, а также производные от них величины – средняя освещенность, неравномерность, отношения максимальной к минимальной и т.п. Среди нормируемых качественных характеристик в инженерной практике используется лишь объединенный показатель дискомфорта UGR [12]

$$UGR = 8 \lg \left[\frac{0.25}{L_a} \sum_{i=1}^N \frac{L_i^2 \omega_i}{p_i^2} \right], \quad (3)$$

где L_i – яркость блеского источника, кд/м², ω_i – угловой размер блеского источника, стер, p_i – индекс позиции блеского источника относительно линии зрения, L_a – яркость адаптации, кд/м².

Отметим, что глаз человека реагирует на яркость падающего излучения, но не на освещенность. При этом для большей части объектов нормируются характеристики на основе освещенности [13]. Исключение составляет дорожное и наружное освещение [14], где нормируется видимая глазу характеристика – яркость. При этом при расчетах в наружном освещении считается, что влияние многократных отражений света минимальное, и оно не учитывается.

Ситуация выглядит странно, глаз человека видит яркость, а нормируется в основном невидимая интегральная характеристика – освещенность. Но, тем не менее, она легко объяснима, если учесть, что нормативные документы разрабатывались во времена инженерных методов моделирования, когда был возможен лишь расчет освещенности в прямом свете и учет многократных отражений через использование приближенных расчетов, например, через коэффициенты использования [8].

Не удивительной тут становится и ситуация, когда в компьютерной графике созданы методы несмещенного моделирования уравнения глобального освещения, позволяя моделировать яркость. А в светотехнике, в это время, моделирование проводится на основе уравнения излучательности. Ведь действительно – зачем моделировать яркость, если нормируется освещенность?

Отметим, что в настоящее время начал активно развиваться новый продукт для моделирования ОУ – DIALux Evo, основанный на моделировании УГО с использованием фотонных карт. Что позволяет моделировать уже непосредственно яркость, однако широкого распространения в инженерной практике на сегодняшний день он не нашел. К тому же, его появление лишь первая веха в возможно начавшемся процессе пересмотра методологии проектирования ОУ и переходу к нормированию истинной видимой характеристики – яркости.

4. ЯРКОСТЬ – НА ПУТИ К КАЧЕСТВУ ОСВЕЩЕНИЯ

Феноменальное развитие математических методов и алгоритмов решения уравнения глобального освещения (1) в компьютерной графике позволяет уже сегодня перейти в светотехнике к моделированию непосредственно яркости и, как следствие, к возможности перехода от нормирования освещенности к яркости. Однако подобный переход сам по себе не имеет ценности. Это можно наблюдать на примере развития DIALux Evo. Программа позволяет считать яркость, но что дальше с ней делать? На наш взгляд, следующим шагом в развитии светотехнического моделирования должен стать переход к проектированию не на заданные количественные характеристики (среднюю освещенность, например), а на заданное качество. Учитывая, что как было сказано выше, в современном проектировании используется только объединенный показатель дискомфорта, в нашей предыдущей работе мы сформулировали новый подход к определению качества освещения [5]. Подход основан на представлении качества освещения в виде одного числа.

$$Q = \frac{1}{AK_{\text{пор}}} \int \frac{|\text{grad}L(x,y)|}{\bar{L}} p(x,y)h(x,y)dx dy, \quad (4)$$

где x, y – координаты точки на проекции сцены (пространство изображения), L – яркость данной точки в направлении наблюдения, \bar{L} – средняя по полю зрения яркость, $p(x,y)$ – некоторая весовая функция, учитывающая различный вклад в реакцию глаз точек, расположенных в центре поля зрения и на периферии, так как плотность колбочек наибольшая у зрительной оси [3], $h(x,y)$ – функция учитывающая значимость той или иной точки для качества освещения, $A = \int dx dy$, $K_{\text{пор}}$ – пороговый контраст, определяемый светотехнической задачей.

Предложенный критерий, в отличие от UGR , учитывает полное пространственно-угловое распределение яркости для заданной точки наблюдения. Так, например, в UGR войдет яркость L_i – непосредственно от точечного источника, попадающего в поле зрения, но вот блик на диффузно-направленно отражающей поверхности от этого источника уже не войдет в расчет коэффициента. Хотя очевидно, что влияние этого блика может быть существенным.

Рассмотрим вариант возможной методики проектирования на заданное качество в соответствии с предложенным критерием. В сцене выбирается ряд видовых точек, для которых будет определяться качество освещения. Если это рассматривать на примере учебной аудитории, то таковыми точками могут быть рабочее место преподавателя у доски, парты с видом на доску. После чего, для данных видовых точек рассчитывается угловое распределение яркости. Для всех поверхностей сцены задается весовой коэффициент $h(x,y)$, учитывающий степень значимости качества освещения для той или иной поверхности. Пример задания коэффициента значимости для аудитории приведен на рисунке 2.



Рис. 2. Задание коэффициента значимости h на примере учебной аудитории

Отметим, что коэффициент может быть задан как непосредственно в пространстве изображений (x,y) , так и непосредственно в пространстве трехмерной сцены, и потом переведен в пространство изображений для каждой видовой точки. После чего мы можем вычислить значение критерия качества Q для каждой из видовых точек и усреднить, возможно, с некоторым весом. Полученное значение будет способно ответить на вопрос – хорошо или плохо освещает установка, что в комплексе с нормировкой на количественные характеристики (например, среднюю освещенность) должно дать исчерпывающий ответ о качестве созданной установки.

5. БУДУЩАЯ РАБОТА

Сформулированный подход к критерию качества освещения, основанному на анализе пространственно-углового распределения яркости, неразрывно связан с наблюдателем, а, следовательно, и его психофизическим восприятием. Это требует экспериментальных исследований для определения пороговых величин и возможных нормировок. В рамках нашей работы мы начали исследование порогов в зоне комфорт-дискомфорт при наблюдении блестящих источников [9]. Однако, натурные эксперименты представляют очевидную сложность, поэтому мы видим дальнейшее развитие исследований критерия качества не только в их продолжении, но и в активном использовании компьютерного моделирования трехмерных сцен.

В этом случае мы предполагаем возможность искусственного создания заведомо «плохих» и «хороших» сцен с точки зрения качества освещения, основываясь на некоторых субъективных критериях. После чего в этих сценах может быть рассчитан предложенный критерий. А далее, предъявляя эти сцены наблюдателям для оценки качества освещения, может быть сопоставлена шкала, полученная при расчете критерия и шкала наблюдателя.

6. СПИСОК ЛИТЕРАТУРЫ

- [1] Goral C., Torrance K., Greenberg D., Battaile B. Modeling the interaction of light between diffuse surfaces // Computer Graphics, 1984. Vol. 18, No. 3. P.213-222.
- [2] Kajiy J.T. The rendering equation // Computer Graphics (Proc. SIGGRAPH'86), 1986. V.20, N4. P.143-150.
- [3] Lombardo M., Lombardo G., Lomoriello D.S., Ducoli P., Stirpe M., Serrao S. Interocular symmetry of parafoveal photoreceptor cone density distribution. // Retina, 2013. Vol. 33, No.8. P.1640-1649.
- [4] Айзенберг Ю. Б. Справочная книга по светотехнике // М. Энергоатомиздат, 1995, 528 с.
- [5] Будак В., Желтов В., Мешкова Т., Нотфуллин Р. Оценка качества освещения на основе пространственно-углового распределения яркости // Светотехника, 2017, №3
- [6] Дебелов В. А. Инструментальная верификация алгоритмов реалистического рендеринга // – Графикон-2016, Нижний Новгород, 2016, стр. 22-28.
- [7] Галлагер Р. Метод конечных элементов. Основы. — М.: Мир, 1984
- [8] Епанешников М. М. Расчет средней светности потолка и стен при проектировании осветительных установок // Свето-техника, 1959, №12.
- [9] Чембаев В., Будак В., Желтов В., Нотфуллин Р. Математическое моделирование осветительных установок и качественные характеристики освещения // 26-я Международная конференция GraphiCon2016, 2016, стр. 50-54
- [10] www.dialux.de
- [11] www.relux.biz
- [12] Свод правил СП 52.13330.2011 «Естественное и искусственное освещение. Актуализированная редакция», 2011
- [13] СНиП 23-05-95 Естественное и искусственное освещение
- [14] МГСН 2.06-99 Естественное, искусственное и смешанное освещение

Программная модель линзы для обработки лучей линейно поляризованного света

В.А. Дебелов^{1,2}, К.Г. Кушнер²

debelov@oapmg.sccc.ru | kushner.k.g@gmail.com

¹ Институт вычислительной математики и математической геофизики СО РАН, Новосибирск, Россия;

² Новосибирский государственный университет, Новосибирск, Россия

Линза – один из основных объектов в оптике, она в достаточной степени проработана в фотореалистической компьютерной графике, поскольку включается во многие 3D сцены. В статье рассматривается разработка модели линзы для конкретного применения, а именно: линзы в составе поляризационного микроскопа (или полярископа), когда надо учитывать поляризацию света и когерентность лучей при расчете интерференционных картин. Тем не менее, данная модель предусматривает более широкое, даже универсальное, применение. Хотя предлагаемый подход относится к реалистической компьютерной графике, при этом линза не подсцена (подмножество примитивов), а элемент сцены наряду с другими примитивами, ограниченный двумя «прозрачными» сферическими поверхностями и непрозрачной боковой поверхностью (цилиндр, конус). Разработанная программная модель линзы преобразует входной луч линейно поляризованного света в выходной луч с скорректированными интенсивностью, состоянием поляризации света и фазой. В данной работе моделируется линза в режиме, когда можно не учитывать эффект дифракции света.

Ключевые слова: линза, линейная поляризация света, оптическая дисперсия, фотореалистический рендеринг, трассировка лучей, поляризационный микроскоп.

Lens model to process linear polarized light

V.A. Debelov^{1,2}, K.G. Kushner²

debelov@oapmg.sccc.ru | kushner.k.g@gmail.com

¹ Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, 630090, Russia

² Novosibirsk State University, Russia

A lens is one of main objects in optics; it was quite sufficiently elaborated as a subscene of 3D scenes in the realistic rendering. In this paper, we consider a development of a computer model of a lens to be a part of a computer model of a polarizing microscope (or polariscope), i. e., a model that takes into account polarization and coherency of light rays to provide computations of interference pictures. Lens geometry is a volume bounded by two transparent spherical surfaces and an opaque side surface (cylinder or cone); it is implemented as a single primitive of 3D scene along with other primitives like surfaces. An implemented lens model transforms an incident linear polarized light ray to outgoing rays with corrected intensity, polarization state and phase. Note that an effect of diffraction is not taken into account.

Keywords: lens, linear polarization, optical dispersion, photorealistic rendering, ray tracing, polarizing microscope.

1. Введение

Разрабатываемая модель линзы ориентирована на ее применение в компьютерных моделях таких оптических устройств, как полярископ или поляризационный микроскоп [17]. В этих устройствах существенная субстанция – это поляризованный свет. Только учет поляризации света и физически корректное ее преобразование моделью линзы может обеспечить расчет интерференционных изображений (ортоскопические и коноскопические картины) прозрачных сред [2], для наблюдения которых эти устройства и предназначены.

На рис. 1 приведена оптическая схема типичного поляризационного микроскопа, из которой видно, что, начиная от поляризатора, и далее вверх по схеме распространяется линейно поляризованный свет. Набор линз достаточно разнообразен с геометрической точки зрения. Линза не одна, и они разные, поэтому для компьютерного моделирования микроскопа полезно иметь единую компьютерную модель для прозрачных, в основном, сферических линз. На практике встречаются и другие, например, асферические линзы, например, в качестве нижней линзы конденсора микроскопа. Но исследуемый в работе класс достаточно широк и удовлетворяет целям изучения и разработки компьютерной модели петрографического микроскопа.

Разрабатываемая модель линзы ориентирована на идеальное в некотором смысле стекло без каких-либо

включений и рассеивающих элементов, оптически однородное, бессвильное, без пузырей и включений, без двойного лучепреломления (т.е. без напряжений), но вполне реальное для поляризационного микроскопа, подробности см. в [14]. Работа линзы рассматривается в оптическом диапазоне электромагнитных волн 380–780 нм, стекло линзы считается однородным и в рамках модели учитываются только следующие определяемые маркой стекла параметры:

1. Спектр индекса рефракции (показателя преломления).
2. Спектр коэффициента пропускания стекла или поглощения.

Таким образом, задача данной разработки – создание компьютерной модели линзы, помещенной в оптически изотропную прозрачную среду, физически корректно обрабатывающей падающие на нее лучи линейно поляризованного света и обеспечивающей в дальнейшем необходимый и достаточный базис для расчета ортоскопических и коноскопических интерференционных картин.

2. Модель линзы

Разработаны геометрические модели линз: классические тонкие линзы [13], толстые линзы [5] и другие конкретные модели. В этих моделях линза рассматривается как единый объект, преобразующий проходящий через нее свет в основном по законам геометрической оптики. Тем не менее, при решении

конкретной прикладной задачи, как правило, обнаруживается, что нужная информация разбросана по источникам или, вообще, не полна, или требуемым параметрам линзы не уделялось внимания. Нас в связи с решаемой задачей интересуют линзы, имеющие в общем случае геометрию, показанную на рис. 2, т.е. две сферические поверхности и боковая цилиндрическая или коническая поверхность. Сферические поверхности могут вырождаться в плоскости.

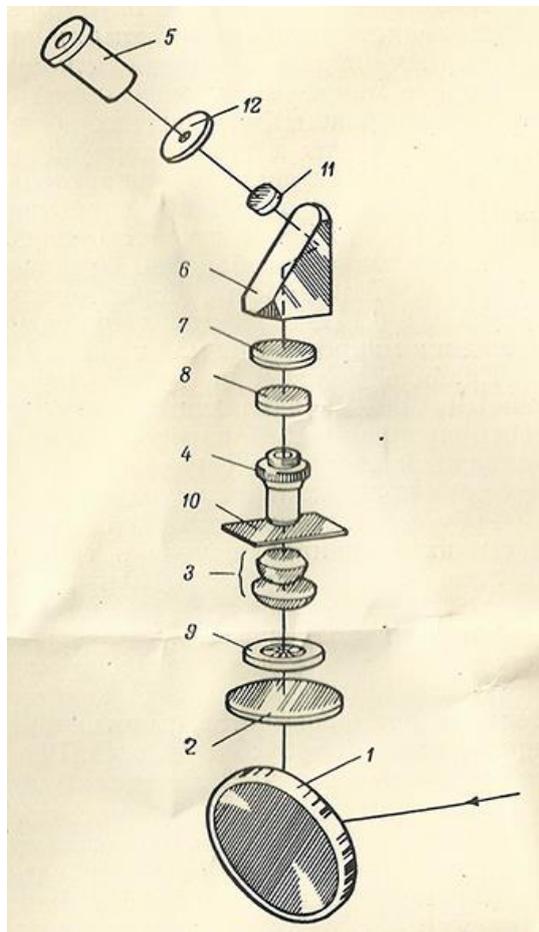


Рис. 1. Оптическая схема типичного микроскопа – поляризационного дорожного МПД-1: 1 – зеркало; 2 – поляризатор; 3 – конденсор; 4 – объектив; 5 – окуляр; 6 – призма; 7 – анализатор; 8 – компенсационная пластинка; 9 – ирисовая апертурная диафрагма; 10 – исследуемый образец (препарат); 11 – линза Бертрана; 12 – диафрагма.

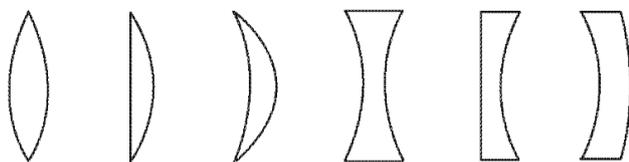


Рис. 2. Рассматриваемые типы линз.

В связи с тем, что наша основная задача в значительной мере относится к области компьютерной графики, мы остановились на геометрической модели линзы, описываемой набором поверхностей, ограничивающих среду с определенными оптическими характеристиками и представляющую подцену более общей 3D сцены. На аналогичном представлении остановились разработчики одной из самых мощных систем оптического дизайна ASAP [8]. Несомненно, в рамках ASAP было сделано многое из того, что потребовалось решать нам, а именно: трассировка в оптически изотропной прозрачной среде луча

поляризованного света [6]. Адаптировать ASAP не представляется возможным по следующим причинам. Система очень громоздка и дорогая, ориентирована на построение (геометрическое моделирование) и расчет небольшого числа путей и лучей света в сложной оптической системе, а не на получение изображений. Алгоритмы, применяемые в ней не описаны, поэтому нельзя воспользоваться их наработками по сбору необходимых материалов и принятых решениях.

В редакторах трехмерной обстановки можно сформировать линзу в качестве одного из объектов сцены, затем передать в систему реалистичной визуализации – рендер. Отметим, что в описании 3D сцены, передаваемом существующему рендеру, описание линзы будет развернуто в набор описывающих ее поверхностей, т.е. поверхностные примитивы, описывающие линзу, будут входить наряду с примитивами, описывающими другие объекты. В компьютерной графике линзы, как правило, не выделяются особо в сцене, например, см. работу [12], в которой исследуются оптические устройства, как обычные сцены. Самый главный фактор, не позволяющий использовать какие-либо рендеры открытые и коммерческие, заключается в том, что они вообще не учитывают поляризацию света при трассировке лучей.

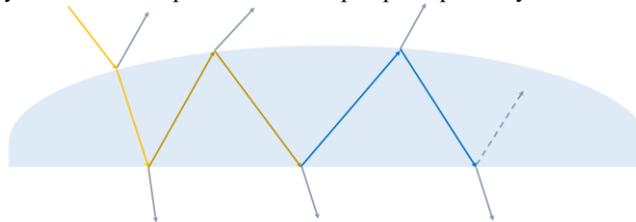


Рис. 3. Поддеревье трассировки лучей внутри линзы.

Здесь мы должны вернуться к основному назначению разрабатываемой модели линзы – линза в составе петрографического микроскопа, который, по сути, состоит в основном из линз – является конвейером линз, обрабатывающим лучи света. Блосс [2] назвал микроскоп оптической трубой. В связи с этим наблюдением и была поставлена задача разработать модель линзы, описание которой является традиционным для компьютерной графики, но линза при этом рассматривается как отдельный неделимый объект – примитив 3D сцены.

3. Общее описание функционирования модуля линзы

Рассмотрим процесс взаимодействия линзы с окружением при трассировке лучей. На рис. 3 показано, как падающий на линзу луч порождает целое дерево лучей. Одним из требований на разработку программного модуля линзы (ПМЛ) была универсальность его применения, т.е. единый модуль для использования в программах, осуществляющих как *прямую трассировку* лучей, так и *обратную*. Рассматривается два возможных варианта реализации: 1) ПМЛ, оформленный в виде самостоятельного программного модуля с четко специфицированным интерфейсом и управляемый некоторой прикладной программой (ПП), которая сама решает какие выходные лучи необходимо ей вернуть, а какие игнорировать; 2) ПМЛ является частью кода ПП, в нашем случае – программы, осуществляющей управление работой конвейера обработки лучей света. В дальнейшем описании не будем различать варианты реализации.

Прямая трассировка (лучи от источника в сцену). Под прямой трассировкой лучей при расчете изображений сцен понимается метод, когда исходные лучи генерируются на источниках света, далее отслеживаются их пути, возникающие за счет отражения и преломления на

поверхностях сцены до тех пор, пока они не достигнут плоскости изображения (экрана) либо другой регистрирующей поверхности. Прямая трассировка в связи с ее исключительной трудоемкостью в компьютерной графике не применяется для получения изображения, она используется часто в качестве вспомогательного этапа для первичного распределения освещенности по поверхностям сцены, например, в виде карт освещенности [1]. Микроскоп представляет специфическую сцену – геометрически это почти труба (рис. 1) с разными прозрачными перегородками (линзами). Все пути от источника, выходящие за пределы объема оптической трубы микроскопа часто можно игнорировать, что позволяет собрать лучи, дошедшие до окуляра, за приемлемое время. Конечно, необходимо дополнительно решать задачу влияния рассеянного света, если таковой случается, падающего на рассматриваемый препарат/образец. В целом прямая трассировка может служить для вычисления оценки различных эффектов и параметров оптических элементов. Например, при помощи прямой трассировки можно просчитывать такие эффекты как степень и зоны виньетирования объектива.

Обратная трассировка (от наблюдателя в сцену). Классический пример – это алгоритм обратной рекурсивной трассировки лучей Виттеда [9] – самый используемый для расчета реалистических изображений алгоритм. Работает в два этапа. На первом этапе из камеры через пиксель экрана выпускается луч, и строится дерево рендеринга аналогичное дереву прямой трассировки, листья которого – это источники. Часть ветвей обрывается в связи с уходом луча за пределы сцены или, когда исчерпана заданная глубина дерева. На втором этапе осуществляется сбор энергии, которая по исходному лучу придет в камеру. Для этого дерево проходит в обратном направлении – от листьев к корню, при этом в каждом узле объединяется энергия потомков и передается родителю.

Основной параметр алгоритмов трассировки – это глубина дерева рендеринга dT : чем она меньше, тем быстрее работает расчет. Интуитивно ясно, что чем дальше лист находится от корня дерева, тем меньше вклад этого листа в общую энергию, т.к. она уменьшается при каждом отражении или преломлении. С другой стороны, уменьшив глубину в угоду скорости расчета, т.е. оборвав ветви, начиная с некоторого уровня, можно потерять лист, которому соответствует очень мощный источник.

В самом общем случае ПМЛ на основе падающего луча и заданной глубины dT обрабатывает все лучи, выходящие вовне. На рис. 3 $dT = 5$. ПМЛ по каждому порожденному лучу вызовет специфицированные в ПП функции и запросит от нее информацию по всем. Влияние параметра глубины трассировки dT : на основе его значения уже на фазе построения дерева можно ограничить ветвление внутри линзы. Особенности различных применений: в контексте полярископа в случае прямой трассировки можно сразу игнорировать отраженный луч, направленный вовне, а также лучи, выходящие из линзы навстречу падающему лучу, поскольку они не будут участвовать в формировании целевого изображения. Конкретные управляющие режимы задаются модулю линзы из ПП, функцией которой является решение глобальной задачи.

Пока описание работы трейсера (рендер на основе трассировки лучей) выглядит аналогичным описанию работы трейсера Виттеда [9] при расчете фотореалистических изображений 3D сцен. В алгоритме Виттеда на родительский уровень передается обобщенная результирующая энергия, собранная со всех потомков данного узла – по сути одно число или три, если цвет. В

нашем случае необходимо отследить состояние поляризации каждой порции энергии, полученной от соответствующего отдельного листа-источника. При обратной трассировке на выходе из фазы сбора энергии модуль линзы предоставит целый «пучок» – список результирующих «возвращенных» лучей – минимум по одному для каждой из глубин $1, 2, \dots, dT$. Почему бы не объединить это в один луч? Мы придерживались принципа разделения функций: ПП решает основную задачу, ПМЛ отвечает за преобразование поляризованных лучей, проходящих через линзу. Например, решая задачу расчета интерференционной картины, ПП определит когерентность лучей и рассчитает интенсивность в пикселе, опираясь на пришедшие интенсивности, фазы и когерентности источников всего семейства лучей при обратном проходе дерева трассировки.

Стоит сделать замечание, что разработанный модуль линзы ПМЛ работает одинаково вне зависимости от того, какая трассировка реализуется в ПП: детерминированная [4, 9] или Монте-Карло [12], поскольку на границе даже в случае прозрачных оптически анизотропных сред падающий луч порождает не более двух отраженных и не более двух преломленных лучей, и все их можно и нужно отслеживать.

4. Свет, луч, поляризация, преобразования

В рамках решаемой задачи рассматриваются квазимонохроматические волны, представляемые при помощи лучей монохроматического света. Представление света в виде набора монохроматических лучей – это задача ПП. ПМЛ работает с одним конкретным лучом. Согласно [3] волна естественного (неполяризованного) света интенсивности I эквивалентна двум независимым линейно поляризованным волнам с интенсивностью каждой, равной $I/2$, и электрическими векторами, колеблющимися в двух взаимно перпендикулярных плоскостях, нормальных к направлению распространения – направлению луча, учитывая оптическую изотропность среды (здесь: стекла и воздуха). А частично поляризованный свет можно представлять в виде суммы полностью неполяризованной и полностью поляризованной частей, не зависящих друг от друга. Опираясь на эти допущения, модуль линзы работает только с лучами поляризованного света, не ограничивая общности применения. Предполагается, что эти факты учитываются в ПП, хотя можно создать оболочку ПМЛ, которая будет разбираться с вопросами неполной поляризации поступающих и выходящих лучей света. Таким образом, будем считать, что на вход рассматриваемого модуля всегда приходит луч линейно поляризованного света. В работе [10] рассматривается определенный взгляд на эволюцию представления луча в компьютерной графике. Так, *луч света* с любой степенью поляризации представляется как структура $Ray = \langle R, Payload \rangle$,

где $R = \langle P, \mathbf{dir} \rangle$ – это математический луч с начальной точкой P и направлением \mathbf{dir} , $\|\mathbf{dir}\| = 1$, $Payload$ (нагрузка) – это дополнительная информация, переносимая лучом. В данной работе для поляризации не используются какие-либо представления типа матриц когерентности, векторов Джонса или Стокса, матриц Мюллера [6, 16], оставляя свободу разработчикам ПП. В нашем случае используется следующее представление нагрузки:

$$Payload = \langle \lambda, I(\lambda), \mathbf{v}_p, k\varphi, O_p, ch \rangle,$$

где λ – длина волны света в вакууме, $I(\lambda)$ – интенсивность, переносимая лучом, \mathbf{v}_p – вектор по которому осуществляются колебания электрической компоненты волны в плоскости, перпендикулярной направлению луча \mathbf{dir} , $k\varphi$ – число скачков фазы (см. ниже), O_p – оптический путь, ch – идентификатор источника, используя который ПП может определять когерентность лучей, приходящих в некоторую точку. ch задается в ПП, а ПМЛ только передает это значение порожденным лучам. Электрическую компоненту световой волны линейно поляризованного света можно представить в виде $\mathbf{E} = \mathbf{E}_0 \cdot e^{-i\omega t - \varphi}$ [15], где E_0 – амплитуда, ω – частота, а φ – фаза. Запишем $\mathbf{E} = E_0 \cdot e^{-i\omega t - \varphi} \mathbf{v}_p$, тогда не надо хранить систему координат для задания линейной поляризации. Известно, что интенсивность пропорциональна квадрату амплитуды [11, 15], мы посчитали более практичным хранить в нагрузке луча именно интенсивность.

При *прямой трассировке* на вход ПМЛ подается луч в представлении *Payload*. Во всех узлах дерева трассировки будет выполняться однотипное преобразование G (см. ниже), которое генерирует два луча в таком же представлении за исключением случая полного внутреннего отражения, когда отсутствует преломленный луч. До точки входа в линзу за фазой и пройденным оптическим путем следит ПП, а начиная с этой точки входа до любого из выходов эти значения контролирует ПМЛ.

При *обратной трассировке* на этапе построения дерева также на вход ПМЛ подается луч, у которого в представлении задается только длина волны λ . После того, как из ПМЛ в ПП выданы все выходные лучи, модуль линзы ожидает, пока ПП достроит дерево трассировки, начнет второй этап сбора энергии и вернет в ПМЛ в ответ на *каждый* порожденный линзой луч информацию. Ответ может быть не один, а несколько полностью линейно поляризованных лучей в зависимости от величины dT – «пучок». При обратном проходе ПМЛ обрабатывает каждый из этих лучей во внутренних узлах линзы, применяя к каждому преобразование G .

Преобразование G – это не что иное, как применение формул Френеля в точках ветвления – пересечения луча с границей между прозрачными средами, см. [15], и подсчет изменения фазы при нахождении пути внутри линзы. Итак, пусть на границу двух прозрачных сред с индексами рефракции n_1 и n_2 падает полностью линейно поляризованный луч с электрическим вектором \mathbf{E} . Раскладываем его на две компоненты – $\mathbf{E} = \mathbf{E}_{\parallel} + \mathbf{E}_{\perp}$ – одна лежит в плоскости падения, другая перпендикулярна ей. Аналогичное разложение делаем для векторов $\mathbf{E}_r = \mathbf{E}_{r\parallel} + \mathbf{E}_{r\perp}$ отраженной и $\mathbf{E}_t = \mathbf{E}_{t\parallel} + \mathbf{E}_{t\perp}$ преломленной волны. Формулы Френеля позволяют найти соотношения между *амплитудами* падающей, отраженной и преломленной волн.

Подсчет оптического пути и скачков фазы. Оптический путь для расстояния длиной S в среде с индексом рефракции n вычисляется по формуле $O_p = nS$,

например, см. [10]. Непрерывное изменение фазы волны с длиной λ при проходе этого пути равно $2\pi \cdot O_p / \lambda$. Но в точках ветвления (на границах) возможны скачкообразные изменения фазы [11, 15], запишем правила таких изменений:

1. При преломлении падающий и преломленный лучи софазны.
2. E_{\perp} : при отражении фаза всегда меняется на $-\pi$.
3. E_{\parallel} : при отражении фаза меняется на $-\pi$, если $n_1 < n_2 \wedge \alpha < \beta \vee n_1 > n_2 \wedge \alpha > \beta$,

где β – угол Брюстера, т.е. $tg\beta = n_2/n_1$, когда n_1

– индекс рефракции среды, откуда падает луч, n_2 – индекс рефракции среды, на границу с которой он падает.

В представлении нагрузки луча параметр $k\varphi$ содержит число скачкообразных изменений фазы на пути луча внутри линзы, ПП должна подсчитывать полное изменение фазы внутри линзы по формуле $d\varphi = 2\pi \cdot O_p / \lambda - \pi \cdot k\varphi$.

5. Численные эксперименты

Доклад сопровождается иллюстративным материалом, первая часть которого получена рендерингом очень простых 3D сцен с линзой показывает известные факты:

1. Визуальное отличие рассчитанных изображений, полученных при рендеринге в цветовой модели RGB и при спектральном рендеринге. Неполаризованный свет.
2. Визуальное отличие рассчитанных изображений, полученных при спектральном рендеринге. Неполаризованный свет. В первом случае падающий непolarизованный свет раскладывается на параллельную и перпендикулярную поляризованные составляющие, к которым применяются формулы Френеля для границы двух прозрачных изотропных сред, т.е. как в преобразовании G . Во втором случае используется аппроксимация Шлика [7] для определения интенсивности порожденных линзой лучей. На основании этого делается вывод о неприменимости различных аппроксимаций и упрощений при преобразовании поляризованных лучей.

Вторая часть экспериментов посвящена примерам использования разработанного модуля линзы:

1. Расчет конуса K направлений начальных лучей дерева трассировки, узлы которого не попадают на бортик линзы с заданной геометрией. Бортик линзы, как правило, используется для размещения крепежных элементов, и попадание луча на него не позволяет оценить состояние поляризации после таких узлов.
2. Для обратной трассировки и начальных лучей, лежащих в конусе K . Расчет доли общей интенсивности, выходящей из линзы, для каждого луча из «пучка» в зависимости от сложности (уровня дерева трассировки) его пути внутри линзы.

6. Заключение

Для придания большей универсальности применения разработанного модуля спектр его применения расширен – можно задать дополнительные режимы работы: а) в рамках

RGB-модели представления света; б) в спектральной модели без учета поляризации. Поскольку разработана наиболее сложная физически корректная модель для расчета отражения и преломления световых лучей с учетом поляризации света, реализацию оказалось нетрудно настроить и на более простые предыдущие модели представления света и цвета – наиболее используемую трехстимульную RGB-модель и спектральную с аппроксимацией Шлика.

С одной стороны, модуль разработан так, что приложение, реализующее поляризационный микроскоп, будет последовательностью линз – конвейером, в начале которого источник поляризованного света, а в конце плоскость изображения. Образец минерала помещается внутри конвейера среди линз.

С другой стороны, конструкция модуля такова, что он достаточно просто может быть встроен в рендер, который позволяет расширять набор примитивов – типов объектов сцены. Например, в алгоритме Виттеда при построении дерева трассировки луч, попадая на прозрачную поверхность, образует узел с двумя потомками: ветвь отраженного луча и ветвь преломленного луча. В случае попадания луча на линзу образуется узел дерева с большим числом потомков, чем два.

В настоящее время реализован последовательный вариант с детерминированным поведением. Дальнейшее развитие разработки: а) параллельная реализация; б) более полный учет требований рендеров, рассчитывающих изображения по методу Монте-Карло.

7. Благодарности

Работа выполнена при частичной финансовой поддержке РФФИ по гранту 16-07-00762-а.

8. Литература

- [1] Arvo J. Backward Ray Tracing // ACM SIGGRAPH'86 Course Notes – Developments in Ray Tracing. 1986. P. 259-263.
- [2] Bloss F. D. Introduction to the Methods of Optical Crystallography. NY: Holt, Rinehart and Winston, 1961.
- [3] Born M., Wolf E. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. 6th ed. Cambridge Univ. Press, 1980.
- [4] Debelov V. A., Kozlov D. S. A Local Model of Light Interaction with Transparent Crystalline Media. // IEEE Transactions on Visualization and Computer Graphics. 2013. Vol. 19, No. 8. P. 1274–1287.
- [5] Heidrich W, Slusallek P., Siedel H.-P. An Image-Based Model for Realistic Lens Systems in Interactive Computer Graphics. Адрес доступа: <http://vccimaging.org/Publications/Heidrich1997IMR/Heidrich1997IMR.pdf> (дата обращения 04.05.2017).
- [6] Polarization in ASAP. Breault Research Organization, Inc. Адрес доступа: http://www.breault.com/sites/default/files/knowledge_base/brotg0923_polar_0.pdf (дата обращения 04.05.2017).
- [7] Schlick C. An Inexpensive BRDF Model for Physically-Based Rendering // Computer Graphics Forum. 1994. Vol. 13, No. 3. P. 233–246.
- [8] The ASAP Primer. Breault Research Organization, Inc. Адрес доступа: http://www.breault.com/sites/default/files/knowledge_base/broman0955_primer.pdf (дата обращения 04.05.2017).
- [9] Whitted, T. An Improved Illumination Model for Shaded Display // Commun. ACM. 1980. Vol. 23, No. 6. P. 343-349.
- [10] Васильева Л. Ф., Дебелов В. А. Эволюция модели луча света для рендеринга // Ситуационные центры и информационно-аналитические системы класса 4I для задач мониторинга и безопасности (SCVRT1516): Тр. Междунар. конф. Протвино: АНО ИФТИ, 2016. С. 178–184.
- [11] Дебелов В. А. Интерференция света, изотропные прозрачные объекты, трассировка лучей // ГРАФИКОН'2015: Тр. Юбилейной 25-й Междунар. конф. Протвино: АНО ИФТИ, 2015. С. 168–173.
- [12] Жданов Д. Д., Потемин И. С., Гарбуль А. А., Соколов В., Валиев И. Методы стохастической трассировки лучей в задачах построения изображений, формируемых реальными оптическими системами // Тр. 26-ой Междунар. конф. по компьютерной графике и зрению ГРАФИКОН'2016. Протвино: АНО ИФТИ, 2016. С. 29-33.
- [13] Ландсберг Г. С. Оптика. Учеб. пособие: для ВУЗов, 6-е изд. М.: Физматлит, 2003.
- [14] Латыев С. М., Егоров Г. В., Митрофанов С. С., Каракулев Ю. А., Тимошук И. Н. Конструирование типовых оптических деталей и сборочных единиц оптических приборов. Адрес доступа: http://de.ifmo.ru/bk_netra/page.php?tutindex=46&index=5&layer=1 (дата обращения 04.05.2017).
- [15] Матвеев А. Н. Оптика. М.: Высшая школа, 1985.
- [16] О'Нейл Э. Введение в статистическую оптику. М.: Мир, 1966.
- [17] Скворцов Г. Е., Панов В. А., Поляков Н. И., Федин Л. А. Микроскопы / Под ред. Н. И. Полякова. Л.: Машиностроение, 1969.

Об авторах

Дебелов Виктор Алексеевич, д.т.н., ведущий научный сотрудник Института вычислительной математики и математической геофизики СО РАН, профессор кафедры параллельных вычислений факультета информационных технологий Новосибирского государственного университета.
Его e-mail debelov@oapmg.sgcc.ru.

Кушнер Константин Григорьевич, студент кафедры систем информатики факультета информационных технологий Новосибирского государственного университета.
Его e-mail kushner.k.g@gmail.com.

Ускорение алгоритма излучательности на графических процессорах

А.С. Щербakov¹, В.А. Фролов^{1,2}

alex.shcherbakov@graphics.cs.msu.ru|vladimir.frolov@graphics.cs.msu.ru

¹Московский государственный университет имени М.В. Ломоносова;

²Институт прикладной математики имени М.В. Келдыша РАН

В данной работе предложен метод преобразования матрицы форм-факторов, позволяющий ускорить вычисление вторичного освещения методом излучательности. Рассмотрена адаптация этого метода для графических процессоров. В частности, предложено использование DXT-текстур для хранения матрицы форм-факторов и перепорядочивание столбцов и строк матрицы в целях уменьшения потерь при сжатии. Предложенные оптимизации позволяют повысить скорость работы алгоритма излучательности до 10 раз и уменьшить до 3 раз объем занимаемой памяти GPU.

Ключевые слова: излучательность, глобальное освещение, GPU.

Accelerating radiosity on GPUs

A.S. Shcherbakov¹, V.A. Frolov^{1,2}

alex.shcherbakov@graphics.cs.msu.ru|vladimir.frolov@graphics.cs.msu.ru

¹Lomonosov Moscow State University;

²Keldysh institute of applied mathematics

We propose a novel approach to implement radiosity on GPU with specific optimizations via form-factor matrix transformations. The proposed transformations enable to reduce the amount of computations for multiple-bounce global illumination and apply DXT compression (with subsequent hardware decompression when reading formfactors on GPU). Our implementation is 10 times faster running and requires 3 times less memory than the naive radiosity GPU implementation.

Keywords: radiosity, global illumination, GPU.

Глобальное освещение 3D-сцен складывается из первичного освещения – полученного из источника света, и вторичного – многократно отраженного от поверхностей сцены. Наибольшую сложность представляет вычисление вторичного освещения, так как размерность интеграла освещенности увеличивается с каждым отражением. Поэтому в приложениях реального времени используют различные методы приближенного вычисления.

1. Обзор существующих методов

1.1. Instant Radiosity

[1] является одним из самых популярных методов благодаря своей простоте. Для расчёта вторичного освещения используются «вторичные» источники света, создаваемые при помощи трассировки лучей из первичных источников. Вторичное освещение, таким образом может рассматриваться как первичное от «вторичных» источников. Развитием метода Instant Radiosity для GPU является алгоритм Reflective Shadow Maps (RSM) [3]. Вместо трассировки лучей в RSM для создания вторичных источников используются карты теней (shadow maps). Основным недостатком данного метода – низкая точность.

1.2. Light Propagation Volumes

[2] создаёт вторичные источники света так же, как это делает Instance Radiosity, но расчёт вторичного освещения производится при помощи распростране-

ния света по трёхмерной сетке. Основным недостатком данного метода – высокий расход памяти и низкая эффективность метода при расчёте распространения света через пустые пространства.

1.3. Voxel Cone Tracing

[5] производит сбор освещения для каждого пикселя путём трассировки нескольких конусов из заданной точки на поверхности, имитируя монте-карло трассировку лучей по полусфере. VCT так же как и LPV использует воксельную сетку для представления упрощённой геометрии, а сама трассировка конусов аналогична шаганию по лучу (ray marching). Отличие в том, что с увеличением расстояния выборка производится из более грубых мип-уровней воксельной сетки, за счёт чего и получается геометрическая аппроксимация конуса. Недостатком алгоритма является высокая вычислительная сложность и зависимость скорости от разрешения.

1.4. Spherical Harmonics

[8] основывается на разложении сложных функций освещенности в сумму более простых для вычисления величин. Для некоторых точек поверхности (как правило, вершин) вычисляются коэффициенты разложения их функций освещения по базису. Функции освещения из источников также раскладываются по базису. В итоге вычисление освещения в точке с разложенной в ней функцией освещенности сводится к скалярному произ-

ведению векторов состоящих из коэффициентов функции освещённости в данной точке и функции освещения из источника. Данный метод широко используется при визуализации открытых пространств, но уступает в точности на закрытых помещениях методу излучательности.

1.5. Излучательность

[6] позволяет получать качественные изображения для закрытых помещений с диффузными поверхностями, во многом не уступая более современным методам. Однако, время выполнения и требуемые ресурсы очень сильно зависят от сцены. На сценах содержащих сотни тысяч треугольников прямое применение излучательности затруднено из-за квадратичной сложности и затрат памяти в зависимости от количества примитивов. Поэтому на практике алгоритм излучательности выполняется для упрощённой сцены (содержащей меньшее количество площадок) и результат расчёта переносится на исходную сцену [10]. Следует подчеркнуть, что на ряду со сферическими гармониками, алгоритм излучательности переносит основную вычислительную сложность на этап предпросчёта, за счёт чего и достигается хороший баланс точность/скорость по сравнению с остальными методами.

2. Предложенный метод

В классическом алгоритме излучательности используется матрица форм-факторов F , размера $n \times n$, где n – количество площадок сцены. Для вычисления освещения после отражения данная матрица умножается на n -компонентный вектор $emission$, содержащий светимость площадок:

$$incident^{(1)} = F \cdot emission \quad (1)$$

Умножая полученный вектор на отражающую способность площадок ρ , на которые пришёл свет, вычисляют светимость площадок после отражения:

$$excident^{(1)} = incident^{(1)} \circ \rho \quad (2)$$

Элементами векторов в данных формулах являются 3-компонентные векторы, содержащие информацию по каждому цветовому каналу. Элементами матрицы форм-факторов являются вещественные числа от 0 до 1.

Приведённые вычисления можно повторять используя векторы $excident^{(i)}$ вместо изначальной светимости площадок, для получения света, пришедшего на поверхности сцены, после произвольного отражения:

$$incident^{(i)} = F \cdot excident^{(i-1)} \quad (3)$$

Полное освещение сцены после k отражений получается путём суммирования векторов $incident^{(i)}$:

$$indirect = \sum_{i=1}^k incident^{(i)} \quad (4)$$

2.1. Предпросчёт нескольких отражений

Предложенная модификация заключается в использовании преобразованной матрицы форм-факторов. Вначале введём «цветную» матрицу форм-факторов:

$$F_{ij}^C = F_{ij} \cdot \rho_j$$

Эта матрица содержит информацию о переносе света между площадками сцены по каждому каналу в отдельности. Таким образом, для её хранения требуется в 3 раза больше памяти. В данной работе рассматривается использование алгоритма излучательности для вторичного освещения сцены. Поэтому вычисление излучательности начинается не с получения вектора $incident^{(1)}$, а с вычисления светимости после первого отражения $excident^{(1)}$. Можно считать, что вектор $incident^{(1)}$ уже посчитан.

Мы можем преобразовать выражение (3) освещённости вектора используя формулу (2) для произвольного индекса.

$$\begin{aligned} incident^{(i)} &= F \cdot (\rho \circ incident^{(i-1)}) = \\ &= F^C \cdot incident^{(i-1)} = (F^C)^{i-1} \cdot incident^{(1)} \end{aligned} \quad (5)$$

Преобразуем формулу (4) с использованием формулы (5):

$$\begin{aligned} indirect &= \sum_{i=1}^k (F^C)^{i-1} \cdot incident^{(1)} = \\ &= \left(\sum_{i=1}^k (F^C)^{i-1} \right) \cdot incident^{(1)} \end{aligned} \quad (6)$$

Матричный полином в скобках не зависит от первичного освещения сцены, а зависит только от геометрии сцены и материалов поверхности. Поэтому он может быть вычислен на этапе предрасчёта.

$$S = \left(\sum_{i=1}^k (F^C)^{i-1} \right)$$

Таким образом вычисление вторичного освещения методом излучательности после k отражений сводится к одному умножению вектора на матрицу:

$$indirect = S \cdot incident^{(1)}$$

Использование матрицы такого вида позволяет ускорить вычисления в k раз, однако такая матрица требует в 3 раза больше памяти.

2.2. DXT-сжатие

Так как предложенная модификация алгоритма требует большего объёма памяти, был разработан метод хранения матрицы форм-факторов в сжатом виде. В качестве формата хранения был выбран формат DXT-текстур, поддерживаемый многими GPU (имеется ввиду аппаратная декомпрессия при чтении). Для этого

формата все значения в матрице необходимо привести к диапазону от 0 до 255. На рисунке 1 показано распределение значений в матрице форм-факторов. При масштабировании этих чисел в промежутке от 0 до 255 большая часть чисел станет 0. Поэтому большие числа, которые вносят основной вклад в вычисление излучательности сохраняются отдельно от остальной матрицы, а к остальным применяется следующее преобразование:

$$value'_i = \frac{\max \left(\log \left(\frac{value_i}{\max value_j} \right) + shift, 0 \right) \cdot 255}{shift}$$

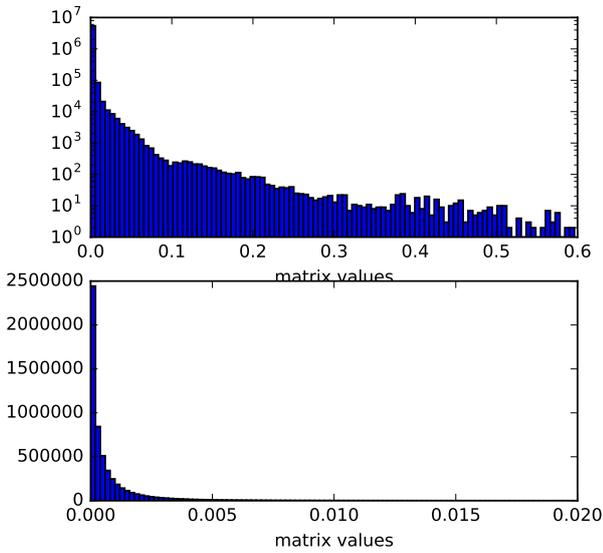


Рис. 1. Распределение значений в матрице форм-факторов (логарифмическая и линейная шкалы).

Текстура, которая получается в результате такого преобразования показана на рисунке 2 слева.

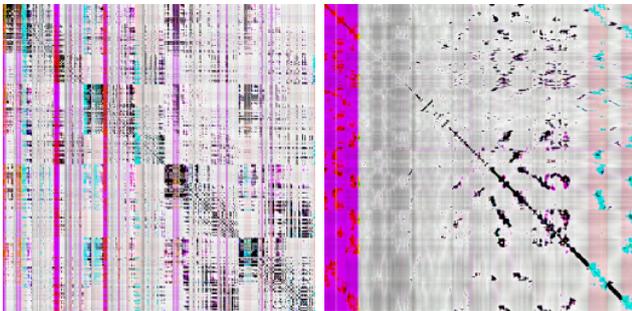


Рис. 2. Фрагмент матрицы форм-факторов до (слева) и после (справа) сортировки строк и столбцов.

Однако DXT-сжатие допускает потери. Чтобы минимизировать потери при сжатии строки и столбцы матрицы пересортируются так, чтобы уменьшить разницу между соседними значениями (рис. 2). Так как

матрица является отношением для пар площадок сцены, то строки и столбцы должны меняться местами одновременно (рис. 3). В результате среднеквадратичная ошибка при сжатии уменьшается до 5 раз (рис. 4).

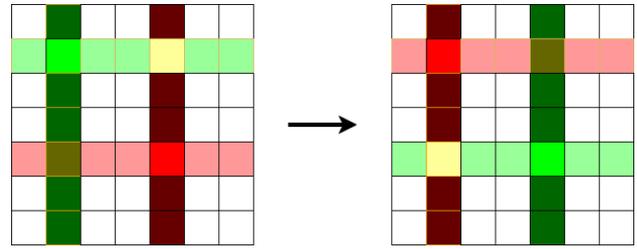


Рис. 3. Схема перестановки строк и столбцов.

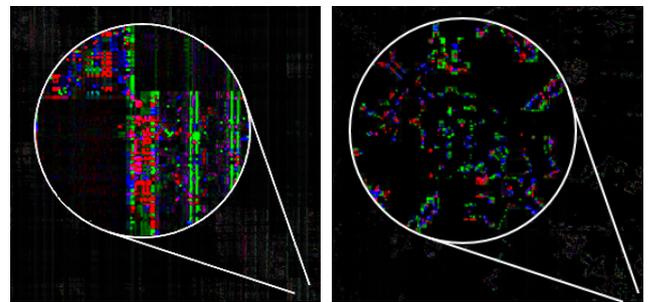


Рис. 4. Визуализированная разница между сжатой и не сжатой текстурами (увеличенный фрагмент) до переупорядочивания строк и столбцов (слева) и после (справа).

3. Детали реализации

Так как современные 3D-сцены содержат сотни тысяч треугольников и выполнение алгоритма излучательности является непроемлемым для таких порядков элементов сцены, для вычисления вторичного освещения была выбрана упрощённая версия той же самой сцены [11].

3.1. Предобработка сцены.

Предобработка сцены осуществляется по следующей схеме:

1. Строится упрощенный аналог сцены на основе вокселизации.
2. Вычисляются форм-факторы для площадок упрощенной сцены.
3. Вычисляется матрица форм-факторов учитывающая несколько отражений света.
4. Числа матрицы больше порогового значения сохраняются в отдельный файл, на их место ставятся 0.
5. Значения в матрице приводятся к диапазону от 0 до 255.
6. Осуществляется перестановка строк и столбцов.
7. Полученная матрица сохраняется в виде DXT-текстуры.

3.2. Визуализация сцены.

Визуализация происходит по следующему алгоритму:

1. Создаётся карта теней;
2. Вычисляется освещение площадок упрощённой сцены источником света.
3. Вычисляется вторичное освещение методом излучательности с помощью матрицы форм-факторов сохранённой в виде текстуры.
4. Вторичное освещение переносится с упрощённой сцены на исходную.

4. Сравнение результатов

Предложенные модификации алгоритма позволяют ускорить алгоритм излучательности в сумме до 10 раз (рис. 5, 8, 7). Использование матрицы форм-факторов учитывающей несколько отражений увеличивает размер файла с матрицей в 3 раза, однако, использование DXT-сжатия позволяет уменьшить требуемую память до 3 раз по сравнению с изначальной матрицей форм-факторов (и классическим алгоритмом излучательности, рис. 6). Мы провели сравнение с изображениями полученными методами Light Propagation Volumes из Unreal Engine 4, классической излучательности и трассировки путей (эталон). Предложенный метод показывает результат сравнимый по точности с классической излучательностью. При этом он ближе к эталону, чем изображение полученное методом LPV при одинаковой частоте кадров.

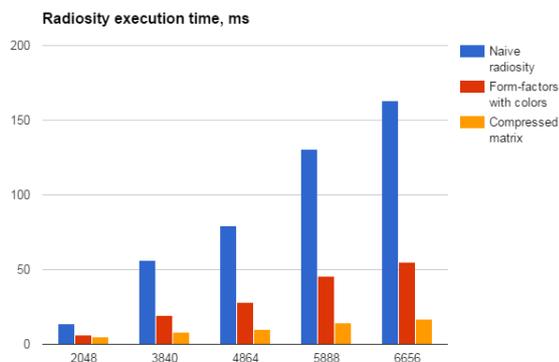


Рис. 5. Сравнение скорости вычисления излучательности. *Сжатие увеличивает скорость*, т.к. алгоритм ограничен памятью а не вычислениями.

5. Выводы и обсуждения

В отличие от других распространённых методов решения уравнения излучательности предложенный метод позволяет вычислить глобальное освещение за $n^2 + O(n)$ арифметических операций и чтений из памяти где n – число площадок, поскольку сводится к *единственному* умножению матрицы на вектор.

Аналогичного результата можно было бы добиться решая СЛАУ при помощи LU-разложения. Однако, эффективная реализация LU разложения на GPU

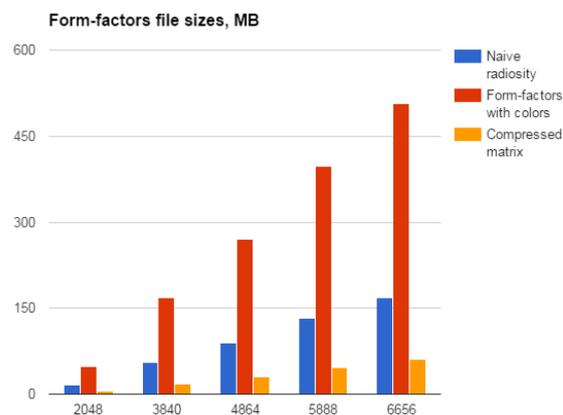


Рис. 6. Сравнение требуемой памяти.

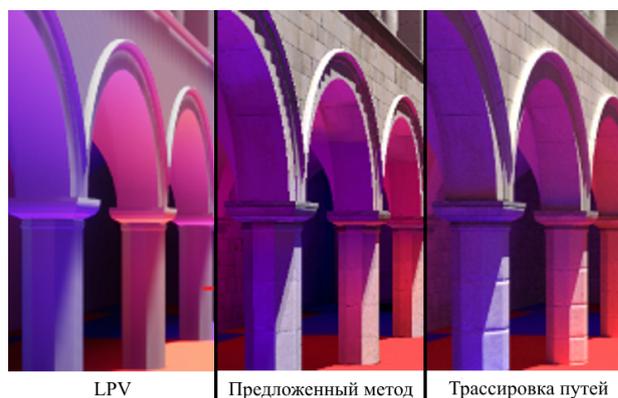


Рис. 7. Сравнение предложенного метода с LPV и трассировкой путей (эталон).

нетривиальна, а при использовании сторонних библиотек (например CUBLAS) отсутствует возможность использовать сжатие. Последнее, как уже было отмечено, критично для алгоритма излучательности (рис. 5, 6).

6. Литература

- [1] Alexander Keller Instant radiosity // In Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97). ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 49-56. DOI=<http://dx.doi.org/10.1145/258734.258769>
- [2] Anton Kaplanyan and Carsten Dachsbacher Cascaded light propagation volumes for real-time indirect illumination // In Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games (I3D '10). ACM, New York, NY, USA, 99-107. DOI=<http://dx.doi.org/10.1145/1730804.1730821>
- [3] Carsten Dachsbacher and Marc Stamminger Reflective shadow maps // In Proceedings of the 2005 symposium on Interactive 3D graphics and games (I3D '05). ACM, New York, NY, USA, 203-231. DOI=<http://dx.doi.org/10.1145/1053427.1053460>

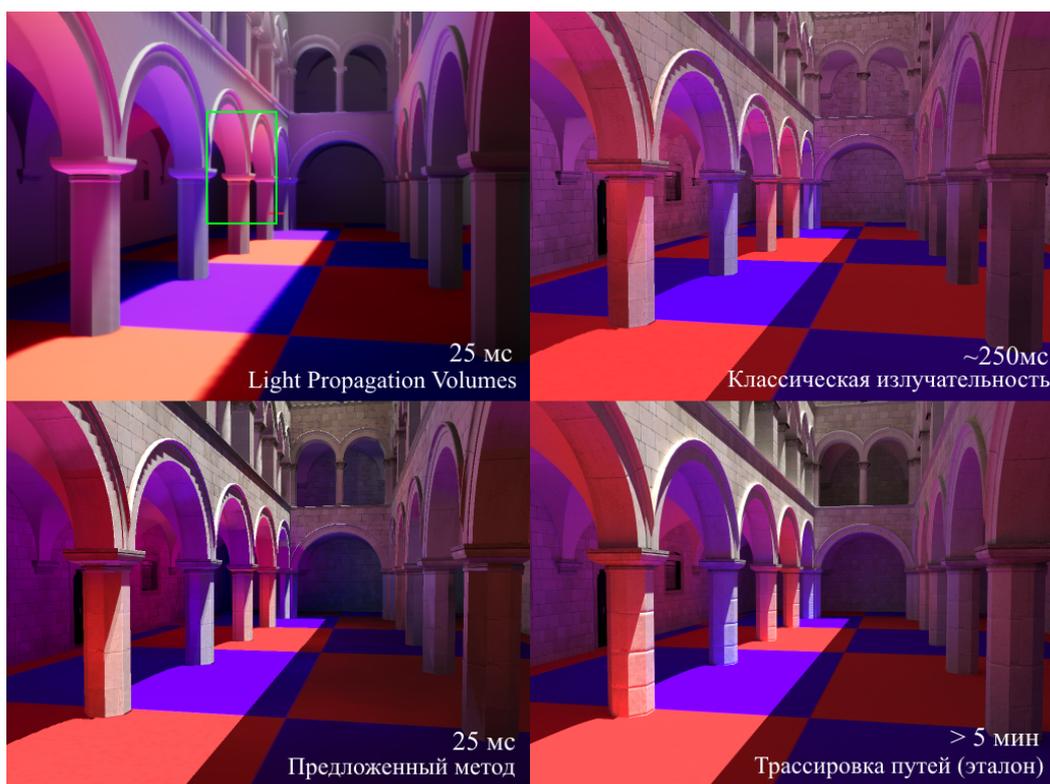


Рис. 8. Сравнение изображений, полученных различными методами и времени генерации изображений.

- [4] Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg, and Bennett Battaile Modeling the interaction of light between diffuse surfaces // In Proceedings of the 11th annual conference on Computer graphics and interactive techniques (SIGGRAPH '84), Hank Christiansen (Ed.). ACM, New York, NY, USA, 213-222. DOI=<http://dx.doi.org/10.1145/800031.808601>
- [5] Cyril Crassin, Fabrice Neyret, Miguel Sainz, Simon Green, and Elmar Eisemann Interactive indirect illumination using voxel-based cone tracing: an insight // In ACM SIGGRAPH 2011 Talks (SIGGRAPH '11). ACM, New York, NY, USA, , Article 20 , 1 pages. DOI=<http://dx.doi.org/10.1145/2037826.2037853>
- [6] Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg, and Bennett Battaile. Radiosity on graphics hardware Modeling the interaction of light between diffuse surfaces. // In Proceedings of the 11th annual conference on Computer graphics and interactive techniques (SIGGRAPH '84), Hank Christiansen (Ed.).
- [7] Greg Coombe, Mark J. Harris, and Anselmo Lastra Radiosity on graphics hardware // In Proceedings of Graphics Interface 2004 (GI '04). Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 161-168.
- [8] Ian G. Lisle and S.-L. Tracy Huang Algorithms for spherical harmonic lighting // In Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia (GRAPHITE '07). ACM, New York, NY, USA, 235-238. DOI=<http://dx.doi.org/10.1145/1321261.1321303>
- [9] Nathan A. Carr, Jesse D. Hall, and John C. Hart GPU algorithms for radiosity and subsurface scattering // In Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware (HWWS '03). Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 51-59.
- [10] SamMartin, Per Einarsson A Real Time Radiosity Architecture for Video Games // Siggraph 2010. [http://advances.realtimerendering.com/s2010/Martin-Einarsson-RadiosityArchitecture\(SIGGRAPH%202010%20Advanced%20RealTime%20Rendering%20Course\).pdf](http://advances.realtimerendering.com/s2010/Martin-Einarsson-RadiosityArchitecture(SIGGRAPH%202010%20Advanced%20RealTime%20Rendering%20Course).pdf)
- [11] Щербаков А., Фролов В. Автоматическое упрощение геометрии для расчёта вторичной освещенности методом излучательности // в сборнике Сборник трудов Графикон 2016, ННГАСУ, с. 34-38.

7. Благодарности

Работа поддержана грантом РФФИ 16-31-60048 «мол.а.дк».



Научная визуализация
и визуальная аналитика

Scientific Visualization
and Visual Analytics

GraphiCon 2017

24–28 сентября 2017
Пермь, Россия

The Brush Metaphor: an Example of Metaphor Adaptation

V.L. Averbukh^{1,2}, M.O. Bakhterev^{1,2}, Z.S. Chayka³, P.A. Vasev¹
 averbukh@imm.uran.ru|ziliyag@yandex.ru

¹N.N. Krasovskii Institute of Mathematics and Mechanics of the Russian Academy of Sciences,
 Ekaterinburg, Russia;

²Ural Federal University named after B.N. Eltsin, Ekaterinburg, Russia;

³Urals State Medical University, Ekaterinburg, Russia

In this paper, we describe the adaptation of popular 'brush metaphor' for usage in teeth masticatory surface analysis system.

Keywords: visualization metaphor, interface metaphor.

1. Introduction

Visualization and interface metaphor theory is still in early stages of grow. However, society can already benefit from it. In this paper we demonstrate one case of such benefit.

One project of our team is the specialized system for dentists for analyse of some properties of masticatory surfaces (described below). During its development we encounter a problem with user interface methods required for system operation. We examined various approaches and came to idea to use known Brush metaphor. We adapted that metaphor to project computational needs which fully solved the problem.

2. Sharpeye visualization builder

Specialized visualization systems created for concrete scientific project or user are extra useful in practice because it supply the researcher with adopted analysis tools. At the same time such systems require a lot of resources and time for development.

In Krasovskii Institute the SharpEye visualization system builder was developed. This tool allows to create a new specialized visualization systems relatively easy. Each visualization case is implemented as a plugin. Such plugin reads data from task custom format and manipulates 3d scene via API. More information available in [Vasev et al, 2012].

Thus the development of new specialized visualization system transforms from hard project to the process of adaptation, configuration and extension of already existing system.

Usually, SharpEye is being used to visualize the results of computer simulation of physical processes and optimal control. An abstract imagery is charactering to these areas. Such imagery is controlled by source domain, e.g. by examined objects and phenomenon. User interaction is used only to specify source data and to control visual output. Thus, usually new or complicated visualization and interface metaphors are not used.

3. Masticatory surface analysis

The task of tooth masticatory surface analysis was raised by [Z. S. Chayka et al, 2011] and is the following. We have a set of tooth, represented by triangles (produced by 3D scanner in STL file format). For arbitrary tooth, we need to select its masticatory surface and to calculate its area. Additionally, we want to see and to control these processes visually.

To visualize the teeth we used natural imagery.

The analysis was introduced by algorithm of automatic detection of masticatory surface. It was implemented as plugin for SharpEye. Masticatory surface detection is activated by mouse click on a tooth. See figure 1.

However in medicine there is no strict mathematical definition of masticatory surface. Therefore our automatic

detection algorithm finds only one probable case of such surface.

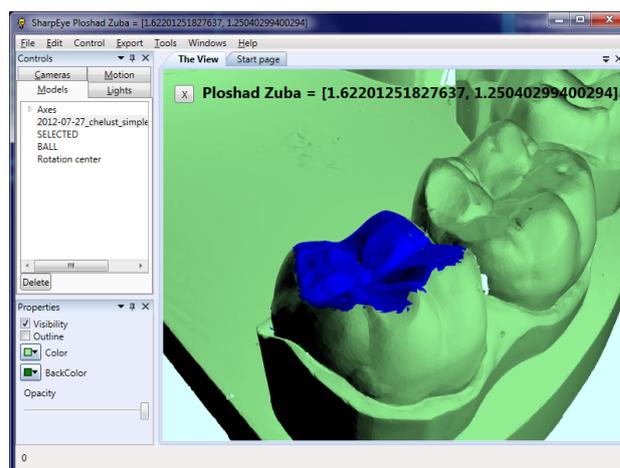


Fig. 1. Automatic masticatory surface detection.

Therefore physicians wanted to manually specify the desired surface too. This emerges the task to specify an arbitrary part of tooth surface.

4. Finding the brush metaphor

Originally, it was suggested to select surface by specifying border polyline. All triangles that fit inside constructed polygon form desired surface. This method was implemented, see figure 2.

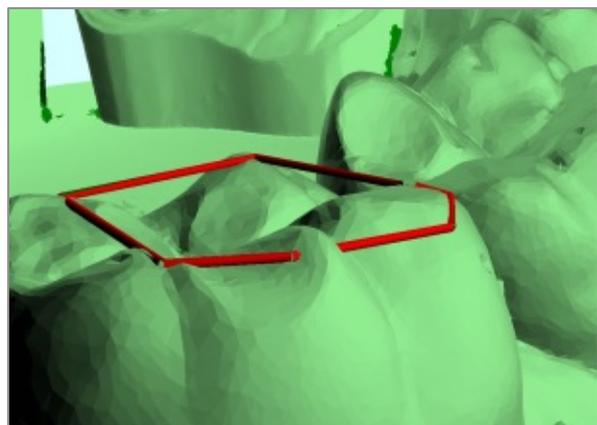


Fig. 2. Using polyline to define masticatory surface.

Besides its mathematical clear meaning, this method has some problems. To modify the selected surface, user has to adjust polyline, e.g. change node point coordinates, add and remove additional nodes, and so on.

The first idea was to provide some graphical tools to adjust the polyline, like in CAD systems. However, it is relatively hard to implement, and it makes user interface pretty complex.

Another idea was to use text area with polyline node points coordinates. User can change the numbers and so affect the polyline. He can also add or remove lines, therefore adding or removing node points. This approach was easy to implement, but it is comfortless for users (physicians).

In this situation, we came to idea of the Brush metaphor. It was inspired by operations we knew dentists do: they use brushes to paint by colorants on teeth shapes.

In our system, the brush is a tool looked like a ball and controlled by mouse. Dentist «paints» on tooth the surface he needs. Also he can erase some parts of his paintings. The painted surface considered as masticatory and its area is calculated.

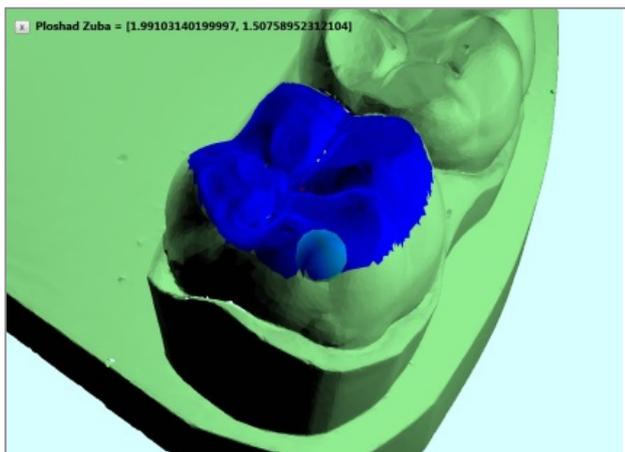


Fig. 3. Using brush to «paint» masticatory surface.

Implementation of the brush approach demonstrated that it is more natural than selection by polylines. There is no need to add, remove or adjust polyline nodes. Using brush a dentist naturally points out what part of a tooth is desired as masticatory surface.

The brush approach has another advantage: it keeps automatic surface detection algorithm still useful. Dentist can launch the algorithm, thus perform rough selection automatically, and then adjust selection using brush (figure 3).

We created a tool based on the Brush metaphor. Dentists confirmed that this tool is comfortable and can be used in their analysis. There is no need for other tools for surface selection in their research.

5. Observations and thoughts

First, we want to note the benefit from bringing new metaphor to the user interface. We had straightforward idea to use polyline to select the surface. It was difficult to implement and difficult to use. So we found and used another metaphor for the same task – the Brush. It happen to be easy to implement and easy to use.

Secondly, the Brush metaphor is being used in a wide range of graphical systems, for example in image painting. The brush tool in our software allows to select surface parts. At that, the tool is connected with computations, and produces input data for them. From the point of view of metaphor theory, we observe an adaptation of know Brush metaphor from one IT subdomain, digital painting, to another one: computer modelling.

At third, we guess that source domain for the Brush metaphor in our case could be not just dentist practice, but also mentioned above IT subdomain of digital painting. If this true, we observe a cultural progress. Early, IT domain was only the target for metaphorization, with metaphors coming from non-IT domains. Nowadays IT domain became the source of metaphors.

6. Conclusion

We observed the case of metaphor adaptation. On the one hand our variant of Brush metaphor is connected with our (inexact) impression about some aspects of dentist activity. On the other - one can consider the case of adaptation of Brush metaphor as an example of the metaphor transfer. The metaphor existing in one of IT subdomain is carried out to a new one.

The goal of metaphorization consists of the expansion of expressiveness for objects under researches. During process of metaphorization some objects of target domain are structured on an example of objects of source domain and there is a metaphorical mapping (projection) of one domain onto another.

Moreover, not all objects are selected (and not even all of their properties, or structure elements), but only those that are the most interesting for us. We note, the use of computer metaphors doesn't refer to exact matching of reality but conversely needs in additional "irreal" opportunities [Averbukh et al, 2008]. In current case one can consider the transfer of a weak metaphorized tool on a new subdomain.

That is the Brush metaphor was adapted on the visual environment of computer modelling. Natural imagery and natural interaction are retained. We suggest that stated variant of Brush metaphor may be applied in other specialized visual systems of computer modelling.

7. References

- [1] Averbukh V.L. et al., 2008. Searching and Analysis of Interface and Visualization Metaphors. // Human-Computer Interaction, New Developments, Chapter 3, Vienna, In-tech, pp. 49-84.
- [2] Chayka Z.S., Ron G.I., Using computer technologies in estimation of the state of ceramic inlays // Problems of stomatology, 2011, N3, pp. 8-11 (in Russian).
- [3] Vasev P., Kumkov S., Shmakov E. 2012. Extensible Scientific Visualization System // Scientific Visualization. Quarter: 2, Volume: 4, Number: 2. Pp. 64-77.

About authors

Vladimir L. Averbukh. Ph.D. Head of the researcher's section at Krasovskii Institute, associate professor at Ural Federal University. Research interests: computer visualization, human-computer interaction. Email: averbukh@imm.uran.ru.

Mikhail O. Bakhterev, researcher at Krasovskii Institute, associate professor at Ural Federal University. Research interests: parallel and distributed programming, AI. Email: mob@k.imm.uran.ru.

Ziliya S. Chayka, Ph.D of Medical Sciences, lecturer at the Department of General Practice Dentistry, Urals State Medical University. Research interests: restorative dentistry. Email: ziliyag@yandex.ru.

Pavel A. Vasev, researcher at Krasovskii Institute. Research interests: scientific visualization, information representation and user interaction. Email: vasev@imm.uran.ru.

A Visualization System for Binary Rooted Trees

M. Forghani¹, P. Vasev², V. Averbukh³
 majid.forqani@gmail.com|vasev@imm.uran.ru|averbukh@imm.uran.ru

¹Ural Federal University, Yekaterinburg, Russia;

²IMM UB RAS, Yekaterinburg, Russia;

³IMM UB RAS & Ural Federal University, Yekaterinburg, Russia.

A system VirTree has been designed for visualization of ultrametric phylogenetic trees, which creates a three dimensional view and also provides a virtual reality presentation. The system can validate clusters in tree and adding a new dimension gives advantage to visualize each cluster independently. In addition to customization of 2D & 3D displays, a virtual reality presentation is provided in system through Viewlang. Viewlang is a web environment of three-dimensional visualization that has a high development speed due to QML language. The virtual reality can dramatically increase the information content of visualization and provides a wide range of view to see the general picture of tree with details while due to limited size of monitor screen this is not possible for large trees.

Keywords: ultrametric phylogenetic tree, visualization, virtual reality, Viewlang.

1. Introduction

Sequence alignments and phylogenetic trees can provide information about the evolutionary history of the proteins that comprise a family [7]. The evolutionary history is achieved based upon similarity and differences in genetic characteristics. Phylogenetic trees represent a backbone for various other biological studies and it is essential to have state-of-the-art tools for their visualization, customization and interpretation [4]. Due to growth of genetics databases, the visualization and customization of phylogenetic tree becomes difficult especially for large trees. The aim of tree visualization is to display complex evolutionary relationships between species to graphically human-readable presentation.

An ultrametric tree is a weighted rooted tree that all of its leaves are at the same depth. Each leaf represents a species and is labeled with unique taxon while two nodes can be joined together in the tree by common ancestor. In fact, each branch defines a relationship between sequences of proteins. A rooted phylogenetic tree is a directed tree with a unique node that is most recent ancestor of all leaves [2]. The distance between a node and its ancestor is evolutionary distance and represents how distant the node (taxa) is from its most recent ancestor.

Traditionally, Phylogenetic trees have been represented as two dimensional (flat) diagrams and the difficulties of interpreting of such tree are growing in size and complexity [3]. Trees can be displayed in a number ways, e.g., circular radial, phylogram and dendrogram. There are several tools for drawing phylogenetic trees both online and standalone, e.g. TREEVIEW, PHYLO_WIN, TREECON, FIGTREE, SEEVIEW, iTOL.

The goal of visualization is to leverage existing scientific methods by providing new scientific insight through visual methods. To visualize in 3-dimensional, tree is displayed in a circular (radial) mode. This form of displaying is particularly useful for the visualization of mid-sized trees (up to several thousand leaves) [4]. A high-resolution 3D phylogenetic tree can be constructed using parameters such as isoelectric focusing, molecular weight and immuno-crossreactivity or any other relevant attribute of a character or trait that gives discrete phylogenetic relationships [5].

As an example of 3D visualization of phylogenetic tree, 3D Phylogeny Explorer (3DPE) is a novel phylogeny tree viewer that maps trees onto three spatial axes (species on the X-axis; paralogs on Z; evolutionary distance on Y), making interpretation of the tree visually intuitive [4]. In addition to 3DPE, Hughes et al. [1] had introduced the idea of visualizing phylogenetic trees in three dimensional hyperbolic space with the Walrus graph visualization tool. With Walrus, it becomes

possible to visualize and navigate phylogenetic trees with more than 100,000 nodes.

In current paper, a system for 3D visualization of ultrametric tree is presented. In this tree, all leaves have the same distance from root. Tree is displayed in circular (radial) mode and by default all arc lengths between each two neighbor leaves are equal. The program gives two views 2D and 3D. In general, using a relevant attribute of a character or trait of leaves such as isoelectric focusing, can led to a complex visualization that is not easy to interpret while VirTree defines Z distance between two leaves equal to arc length between them. This allows bettering visualization of circular tree and tree can be rotated and displayed freely.

Virtree system consists of two parts. The first one is written as a Matlab program that computes coordination of visualization. The second part is based on the virtual reality presentation of results by Viewlang system. Viewlang is the technology of creating interactive 3D-applications for the Internet that has a high development speed due to QML language and supports virtual reality and mobile devices [6, 8]. Virtual Reality (VR) environments are actively used to provide leaping into a new quality of cognitive visualization. Virtual environments are characterized by such features as egocentric point-of-views and user-centered often multi-sensory interactions. The VR environments are dynamic, rather than static. The user's experience of the virtual world may combine a visual channel with the auditory or kinesthetic feedbacks. The immersion and sense of presence (feeling of "being there") are factors which defines VR. The sense of presence distinguishes VR from "traditional" 3D Computer Graphics. Users "immersed" in VR control the graphics output. Thus presence and immersion effects are integral parts of VR. The VR loses its most remarkable ability without them. Similar to color and animation, presence and immersion open a new dimension in visualization. Color and animation directly extend informativness of visualization, but presence and immersion create conditions for another perception of information. In scientific visualization, as in applying VR in industrial applications, it must be looked for new ways to not only increase the quality of interpretation, but also to control the visualization processes.

2. VirTree System

As mentioned above, Virtree consists of two parts. The first part is a GUI Matlab program that allows user to customize the visualization and produce a file for Viewlang representation. A 2D plot displays two-dimensional (flat) circular tree while 3D one shows the same tree but with additional dimension Z (see Fig.1). Final 2D and/or 3D customized graphical representation of phylogenetic trees can be saved separately.

There is an option to validate clusters in phylogenetic tree based upon “cluster” command for phytree object in Matlab [9]. When clusters are declared, uniform color will be assigned to a cluster automatically. For better visualization of clusters, within-cluster leaves distance can be decreased by ratio and this allows to increase between-clusters distance (see Fig. 2.). The compaction ratio slider controls how within-cluster leaves be close together. Similarly, the compaction also affects the Z distance between and within clusters. There are also additional options to show trunk and labels, control the markers size and select color of markers.

Phylogenetic trees can be loaded from a file in the standard Newick format. The GUI program does not include alignment or tree reconstruction methods. It is a tool for visualization. It provides an output JSON file that has information of 3D plot presentation.

Virtual reality provides a wide range of view comparing with monitor screen. Using Viewlang web-based presentation (that is second part of system), tree can be visualized in VR by JSON file. This provides a new quality of visualization and enables the user to view a general picture while the details are simply visible. In fact, Viewlang allows a direct virtual reality presentation within Matlab.

As an example, an ultrametric tree has been visualized in VirTree (see Fig.3.). VR provides us with a wider view space that allows user change some visualization parameters such as the thickness of the tree, the scaling parameter of the tree, the size of the label leaf. In order to see an individual leaf or branch, the user just needs to go closer to it by passing through VR environment. That is much more convenient than zoom in /out in 2D format. In other words, VR visualization is more flexible than 2D presentation.

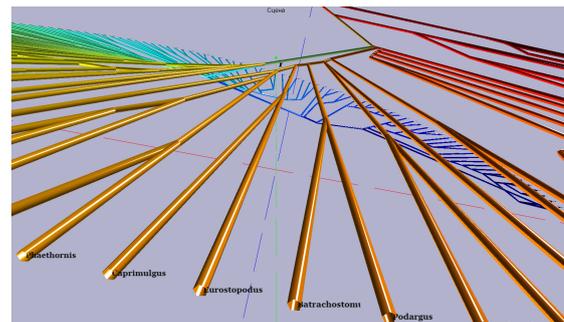
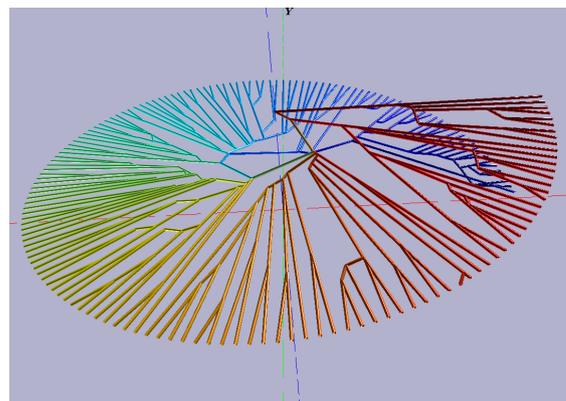


Figure 3. Virtual reality presentation of tree.

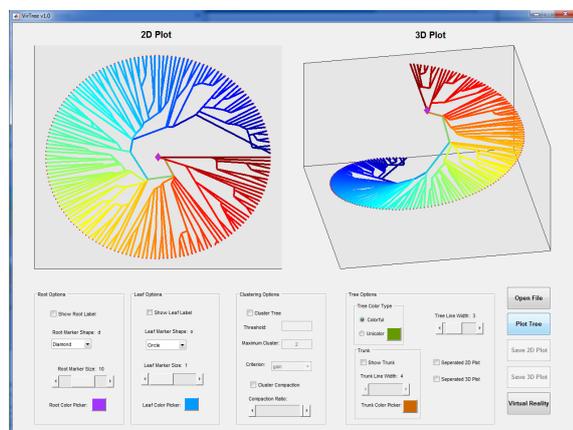


Figure 1. General representation of tree visualization.

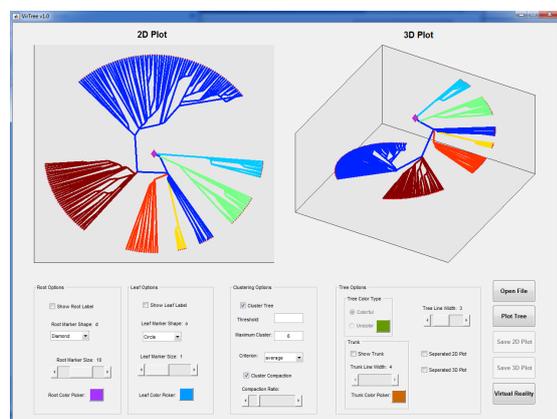


Figure 2. Cluster representation of tree from Fig.1.

3. Conclusion

In summary, Virtree allows user to customize the tree visualization. It supplies 2D and 3D circular tree presentation. Moreover, system can validate tree clusters and additional dimension Z provides better recognition of clusters in plot. The Z dimension also makes a distance between leaves labels for easy to read them. Addition virtual reality capabilities will be used to enable a wide range of view. Immersion into three-dimensional space and easy navigation in it should provide researchers with more opportunities for studying the structure of trees. In future work, the system will be improved by adding alignment and tree reconstruction modules that allows working with FASTA format and different types of tree (root and unrooted). In addition, the interface in virtual reality will be more interactive that provides tree customization in virtual space directly.

4. References

- [1] Hughes, Timothy, Young Hyun, and David A. Liberles. Visualising very large phylogenetic trees in three dimensional hyperbolic space. // BMC bioinformatics 5.1 (2004): 48.
- [2] Huson, Daniel H., and Celine Scornavacca. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. // Systematic biology 61.6 (2012): 1061-1067.
- [3] Kim, Namshin, and Christopher Lee. Three-Dimensional Phylogeny Explorer: Distinguishing paralogs, lateral transfer, and violation of " molecular clock" assumption with 3D visualization.// BMC bioinformatics 8.1 (2007): 213.
- [4] Letunic, Ivica, and Peer Bork. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.// Bioinformatics 23.1 (2006): 127-128.
- [5] Milner, M., et al. Constructing 3D phylogenetic trees. // Current Science (2003): 1471-1478.
- [6] V Vasev P. Three-dimensional visualization in a web based environment based on Qml declarative description // International (47th all-Russian) youth school-conference,

Yekaterinburg, January 31 - February 6, 2016.
<http://data2.lact.ru/fl/s/0/299/basic/1605/992/vasev-sopromat-2016-1.pdf?t=1482676145>

- [7] Zhai, Yufeng, Jason Tchieu, and Milton H. Saier Jr. A web-based Tree View (TV) program for the visualization of phylogenetic trees." *Journal of molecular microbiology and biotechnology* 4.1 (2002): 69-70.
- [8] <http://viewlang.ru>.
- [9] <https://www.mathworks.com/help/bioinfo/ref/clusterphytree.html>.

About the authors

Vladimir L. Averbukh, Ph.D., head of the researcher's section in the Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences, associate professor, Ural Federal University. E-mail averbukh@imm.uran.ru.

Pavel F. Vasev, researcher in the Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences. E-mail vasev@imm.uran.ru.

Majid A. Forghani, researcher in the Ural Federal University. E-mail majid.forqani@gmail.com.

Special Optical Fiber Preform Layers 3D-Reconstruction

Yu.A. Konstantinov², K.P. Latkin^{1,2}, F.L. Barkov^{1,2}, A.S. Smirnov^{1,2}, M.M. Poskrebyshev⁴,
V.P. Pervadchuk¹, D.B. Vladimirova¹, Yu.A. Konin^{2,3}, A.I. Garanin^{1,2,3}, V.V. Burdin^{1,2}
latkin.k.p@ya.ru

¹Perm National Research Polytechnic University, Perm, Russia;

²Photonics Lab of Perm Federal Research Centre of Urals Branch of Russian Academy of Science, Perm, Russia

³Inversion Sensor Company, Perm, Russia

In this paper, we propose a combination of different techniques to reconstruct the 3D-allocation of the different layers inside special optical fibre preform. These techniques allow for securing full information about the layers with pure silica, erbium-doped silica, as well as air channels geometry.

Keywords: 3D-reconstruction, optical fibre preforms, on-line measurements and visualization.

1. Introduction

The optical fibers are widely used in telecommunications, sensor systems and other areas of science and technology [4,6]. Usually silica glass optical fibers are manufactured from fiber preforms that may be produced using several well-known techniques (MCVD – Method of Chemical Vapor Deposition, VAD – Vapor Axial Deposition, OVD – Outside Vapor Deposition and their modifications) [7]. Then the preform usually passes the stage of mechanical processing and after that it should be drawn at the fiber draw tower [3]. The manufacturing technology of standard telecommunication fibers is well understood and controllable, but there is still significant scope for improved fabrication of non-conventional fibers designed for optoelectronic circuits. In the case of active fiber, the technological process has the following differences [1]. During the chemical vapor deposition the preform core is doped not only by the germanium oxide, but also by the erbium oxide to achieve the luminescence in the fiber at the wavelength of 1550 nm. In the case of polarization-maintaining (PM) fiber, after the deposition, the preform goes to the stage of air-channel drilling (air-channels are used for the insertion of the stressrods that will cause the birefringence in the fiber). The major parameters under control are: outer preform diameter at the stage of MCVD process during the deposition (including deflection control) and outer surface 3D-geometry; erbium dopant 3D-distribution along the preform length; air channels 3D-geometry. The glass refractive index depends mainly on concentration of germanium, therefore, erbium concentration cannot be measured using optical techniques. The spatial 3D-distribution of refractive index along the preform as well as in cross section is also important, but at present it is successfully controlled by the conventional instruments.

2. Preform outer diameter control

The first technological stage that we are going to discuss is outer diameter control and preform 3D-reconstruction.

The MCVD process requires the stability of the quartz tube temperature (the tube that will become a part of fiber optic preform) in the heating zone by changing the flow rate of hydrogen in the burner. This flow rate is adjusted according to the readings of the optical pyrometer. The level of doping of the precipitate varies according to the program given by adjusting the supply of reagent vapors to the reaction zone. Many years' experience of utilization of this process revealed the need for continuous monitoring of the outer diameter of the pipe during the deposition process and especially when the tube is collapsing. It should be noted that a considerable time of the deposition process (several hours) leads to the evaporation of quartz from the outer surface of the pipe and reduces its diameter.

Collapse of the tube (transformation into the preform) is usually carried out in several passes, during each of them the outer diameter of the pipe should be changed by a certain value. The need to impose strict limitations on the deviation of the tube from cylindrical shape (no more than 5% of the nominal diameter) also requires continuous monitoring of its geometry. Until now, the control of the tube dimensions has been performed visually by the operator with use of the 173 mm working distance microscope. Under the strong thermal radiation from the tube, this technique did not provide the required accuracy and stability of the technological process. Measuring the preform diameter by contact methods is also excluded since any measuring instrument can damage the heated rotating preform. Moreover, not all the MCVD lathe in the basic configuration are equipped with a system for contactless measurement of the preform outer diameter, therefore manufacturers of fiber-optic light-guides are forced to find their own technical solutions.

In line with the above, a technique for non-contact measurement of the geometric characteristics of cylindrical preforms was developed by usage of a megapixel CCTV camera with the interface USB 2.0 VEC 545 (LLC "EVS", St. Petersburg) and our own software (Fig. 1).

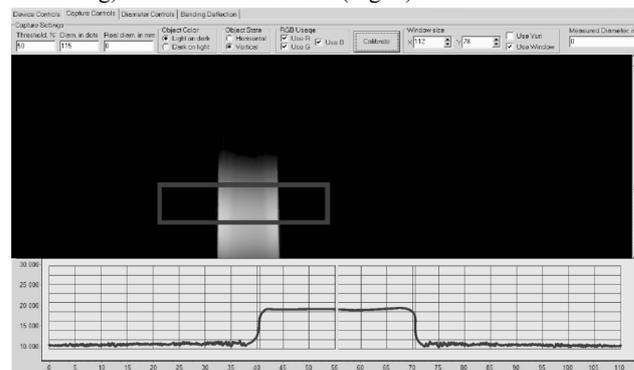


Fig. 1. Developed MCVD preform geometry control software [5].

To enlarge the image obtained from the camcorder, the Computar TG10Z0513FC-2 lens with a four-fold gray protective filter is used. The video camera is mounted on the bracket of the pyrometer of the MCVD lathe. The software supplied with the camera does not provide any information processing; for this purpose, a special own utility has been developed. The developed software has a user-friendly interface. The user has an access to the video image settings directly in the program, without having to run driver's user interface, and select the area under investigation, which greatly simplifies and speeds up the calculations. At the user's request, the spatial orientation of the sample can be changed, the

filtering on the RGB color components or the image inversion can also be turned on or off.

The software allows the use of threshold and correlation algorithms to determine the diameter of the preform. At a constant burner temperature and stable lighting conditions, it is appropriate to use the threshold algorithm:

$$G(j, k) < T_L, G(j, k) < T_H,$$

where $G(j, k)$ is the image data array describing the variation of brightness; T_L and T_H are the lowest and the highest threshold values. The differences in brightness in both horizontal and vertical dimensions can be written as

$$G_H(j, k) = F(j, k) - F(j, k + w),$$

$$G_V(j, k) = F(j, k) - F(j + w, k),$$

where $F(j, k)$ is a raw image data array (single frame) and w is a differential calculation step. In the case of dramatic temperature changes the correlation algorithm could be applied:

$$r = \frac{\sum_{j=1}^Z (G_H(j, k) - \bar{G})(R_{H(T)}(j, k) - \bar{R})}{\sqrt{\sum_{j=1}^Z (G_H(j, k) - \bar{G})^2} \sqrt{\sum_{j=1}^Z (R_{H(T)}(j, k) - \bar{R})^2}},$$

where r is a cross-correlation value of $G(j, k)$ image and $R_{H(T)}(j, k)$ reference image saved at preform temperature T ; \bar{G} and \bar{R} are their mean values over j . All these techniques enable on-line geometry control during the MCVD-process. The next stage is off-line 3D-reconstruction of preform geometry. Due to the high noise level during the MCVD process the received data should be filtered. We used the Discrete Wavelet Transform (DWT) package for this purpose.

At the first stage of the analysis, the initial processing is performed over the available time series (after the MCVD stage the preform raw geometry parameters are presented as Preform Center and Preform Diameter), the methods of which include the graphical representation of the initial series in the time domain, as well as the elimination of the trend of the series. In the simplest case, this is the elimination of the constant term (centering the series). Next, the rows are processed using a continuous wavelet transform with the parent Daubechies wavelet of orders 1-10. As a result, we obtain the scalings of the wavelet coefficients of expansion and the lines of the local maxima so-called skeleton transformations. It is known that if the data contain harmonic or quasi-harmonic components, the topographic map of the skeleton will consist of lines oriented along the vertical scale axis. In the case of a noise component, the skeleton lines are stretched in a perpendicular direction, i.e. parallel to the horizontal time axis. Thus, visually estimating the skeleton lines, even at the first stage of the study, it is possible to select highly noisy signals (Fig. 2). Batch data processing on the basis of the MatLab platform, WaveletToolbox, was used.

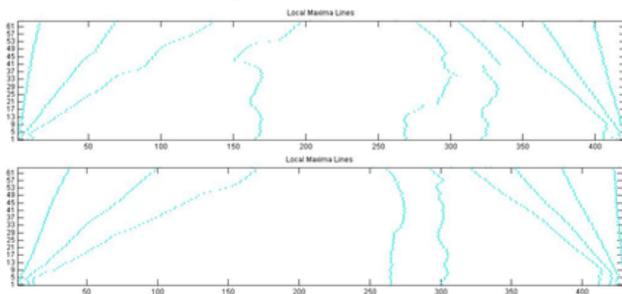


Fig. 2. Skeleton Image of the noised data (upper figure) and that of the data containing no noise (lower figure).

The second stage of the study is the identification of the noise components of the signal. Let the signal model have the form of an algebraic sum $s(t) = f(t) + \sigma e(t)$, where $f(t)$ – desired signal, $e(t)$ – noise, σ – noise amplitude, $s(t)$ – signal under test.

Standard SoftThresholding procedure corresponds to the wavelet coefficients transformation of the following form:

$$w \rightarrow w_t = \begin{cases} 0, & |w| \leq \lambda \\ \text{sign}(w)(|w| - \lambda), & |w| > \lambda, \end{cases}$$

where λ is a predefined threshold value. Thus, before performing the inverse wavelet transform, wavelet coefficients with absolute value below certain cut-off level are set to zero.

As a parent wavelet in the construction of a wavelet filter, wavelet functions with a compact support, such as Daubechies wavelets, Coiflets and Symlets are usually considered. Coiflets are special case of Daubechies wavelets with vanishing moments of scaling function.

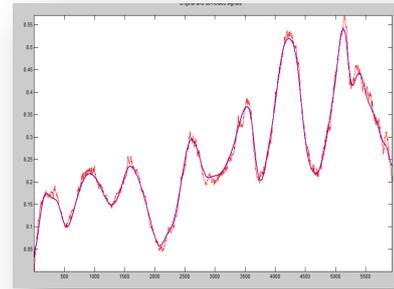


Fig. 3. Raw signal (red color) and denoised signal (blue one) using Coiflets.

The figure above demonstrates the effect of noise reduction. A filter with small scales corresponding to the lower part of the spectrum separates high-frequency components from the signal, which usually correspond to extraneous noise. Changing the scale, one can achieve complete suppression of high-frequency noise.

At the same time, however, there is a risk of distortion of the desired signal. In practice, it is more convenient to choose the scale after the transformation, by calculating the wavelet spectrum in a sufficiently wide range of scales. Then the necessary scales are selected on the spectrum map, and the inverse wavelet transform is produced. It should be noted that one can select two or more bands of the spectrum, not limited to only one area. This technique allows to calculate the frequencies at which the noise components appear and to identify the nature of noise. There is a certain relationship, though not obvious at first glance, between the concept of Fourier harmonics and the scale of the wavelet. The main point in this relationship is the inverse proportionality of the own frequency and scale, weighted by some conversion factor: $\omega = k/a$, where ω – own frequency, a – scaling coefficient.

At present it is clearly seen that it is possible to build the preform 3D-surface reconstruction after performing all steps (see figure 4). It is a very useful option in the case of the MCVD process automation, but it is not applicable as a quality test technique, because after the wavelet noise reduction the surface has non-uniformities tens or hundreds times bigger than they really are.

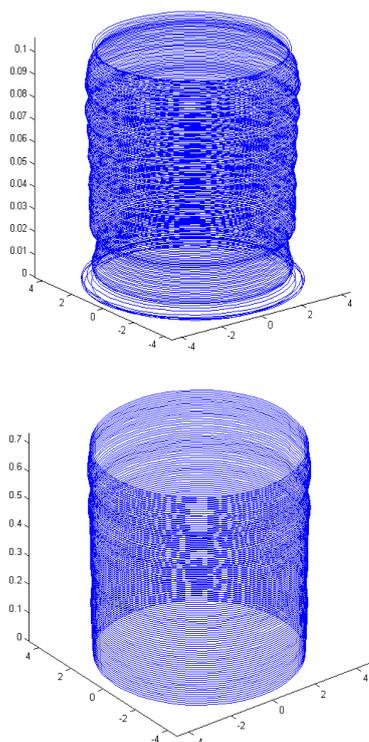


Fig. 4. 3D-surface reconstruction using raw signal (upper figure) and denoised signal using Coiflets (lower figure).

For the precise 3D-image (we shall talk about this in the final part of this article) we suggest to use the data from standard preform analyzer not related to the preform fabrication process.

The obvious drawback of the method of using a standard preform analyzer is a small sampling of the preform length coordinate. Long-term measurements are possible, but this is not always convenient for some applications; therefore, we recommend using both data from the lathe sensors and those obtained in the laboratory on standard equipment to construct a 3-D surface. Later in this article, we are also considering one more possible method – holography.

3. Erbium ions location data acquisition

The measurement is carried out using a pumping diode with a wavelength of 978 nm (near the erbium pumping wavelength), which illuminates the core of the preform from the side, and an integrated receiver with a quartz filter that passes only the luminescence band. The filter is located perpendicularly to the rays of the radiation source, which makes possible reducing the influence of the pump diode radiation on the results obtained. The erbium concentration, in this case, is directly proportional to the receiver's voltage. The results of a test preform measurement is presented below (Fig. 5) Thus, the basis of measurement is the luminescent properties of a material in the glass structure. Using this effect we have constructed an instrument ErGO-1 (Fig. 6), Erbium Grand Observer, model 1. It consists of precise optomechanics and electronic components that allow full radial scan of the preform. ErGO software is the managing application, that controls the positioning motor's and pump diode's drivers as well as ADC and other instrument units. It allows to obtain the comprehensive amount of data that could be used for the further visualization of erbium oxide distribution via every available method (from MathLAB and Excel to GL Scene and other libraries and programs).

It is clearly seen that by changing the angle of investigation we can obtain the 3D-reconstruction of erbium dopant location in the preform. It is useful in the case of Large Mode Area (LMA) fibers.

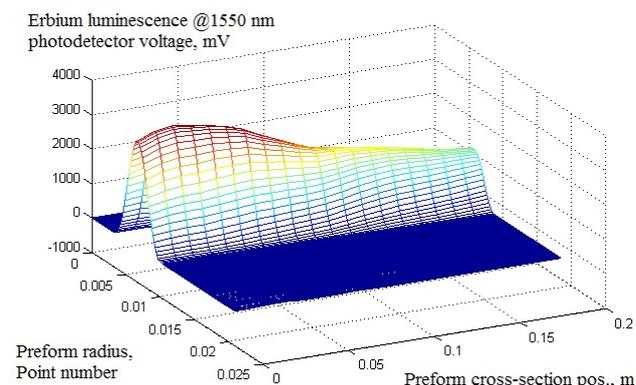


Fig. 5. Developed ErGO-1 analyzer data.

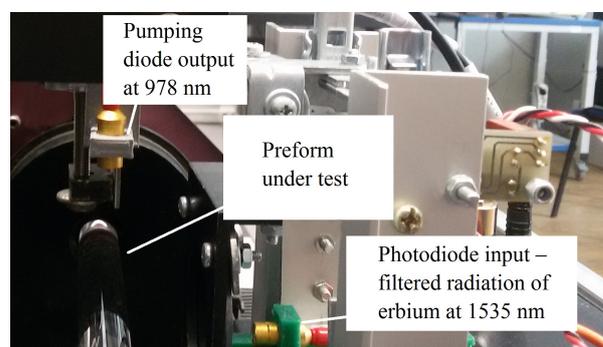


Fig. 6. Analyzer's opened case.

4. Air channels investigation

Air channels geometry is a very important thing which is needed to achieve the constant birefringence of the fiber along its length [2]. The main bottleneck is that contact measurements are not allowed as well as immersion of the preform into liquids. Here we present a non-contact technique that consists of the usage of the special fiber sensor inside the air channel. The fiber tip should be curved in a special way: its bending loss should not have strong influence on the signal attenuation, but it has to be placed perpendicularly to the air channel surface.

We have used a test sample to make a simple method demonstration. A short piece of preform with the drilled channels was mounted on nanopositioning stage, the sensor fiber which was connected with the IR-source and receiver via fiber coupler, is designed to measure the back-reflection optical power from the air channel surface.

After the first experiments we have found out that the amplitude back-reflection sensor does not have the expected performance. This is why we have transformed it into the interferometric scheme by adding the reference mirror with the adjustment microstage. We will report about the applicability of this setup as soon as comprehensive tests are performed.

5. The usage of holographic methods for preform investigation

The methods of holography are not yet widely spread in the field of optical fiber preform geometry properties study, but we think in future it can be applied for rapid analysis.

The process of digital reconstruction of holograms consists of calculating the backward propagation of light from the hologram to the reconstruction plane. The reconstruction

process could be described by a finite discrete form of the Fresnel approximation for the integration of diffraction:

$$b'(n\Delta x', m\Delta y') = A \sum_{j=0}^{N-1} \sum_{l=0}^{M-1} h(j\Delta\xi, l\Delta\eta) r^*(j\Delta\xi, l\Delta\eta) \times \exp\left\{\frac{i\pi}{d'\lambda} (j^2\Delta\xi^2 + l^2\Delta\eta^2)\right\} \exp\left\{2i\pi\left(\frac{jn}{N} + \frac{lm}{M}\right)\right\}.$$

The parameters used in this formula to calculate the complex number in the image plane are given by a CCD matrix having an $N \times M$ pixel size and a pixel pitch $\Delta\xi$ and $\Delta\eta$ in two orthogonal directions. The hologram is stored in array $h(j\Delta\xi, l\Delta\eta)$. The distance between the object and the CCD matrix is denoted by d , and as a rule $d' = d$. The complex coefficients do not depend on the hologram in question, they are contained in the data of the CCD-describing matrix.

Reconstructed image displays restoration of sharpness and borders of the smallest details.

As a test sample we have placed the preform with drilled air channels into the developed instrument. In the figure 7 we can observe the non-uniformities in silica as well as core and channels borders.

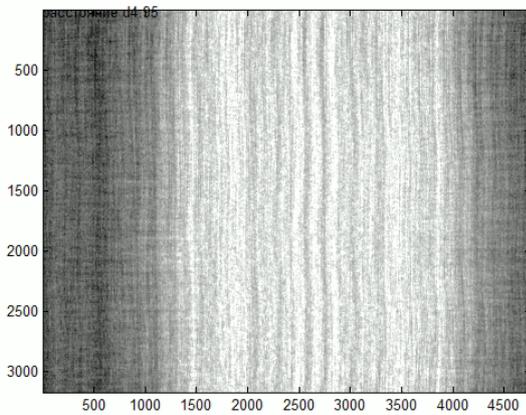


Fig. 7. Holographic image of the preform

The reason for use of the technique described is to distinguish the non-uniformities that conventional preform analyzer can not locate (for example, air bubble could be missed if its location does not coincide with the cross-section under test). After locating the non-uniformity the nearby area could be investigated more precisely using preform analyzer. Also we suppose it is possible to find some correlation with refractive index (RI) of the preform layers.

6. Conclusion

As a conclusion we would like to join all the data obtained from described stages and decide if they are enough to build the full 3D-distribution of all optical and geometrical parameters. As we mentioned above the data from MCVD stage is not applicable for this purpose, so this could be taken only as a draft, and at the final stage we need to substitute this by data received with preform profile analyzer. The only thing we have to take from the developed MCVD measurement system is preform deflection (at present this option is not included into conventional preform analyzers). So this method is rather useful as well as holographic method. Rare-Earth ions location setup has been developed and constructed and it gives the full information about erbium 3D-spatial distribution. The method of the air channel investigation is still being developed. It will be discussed in the future publications. The idea of joining all these data is shown below (Fig. 8).

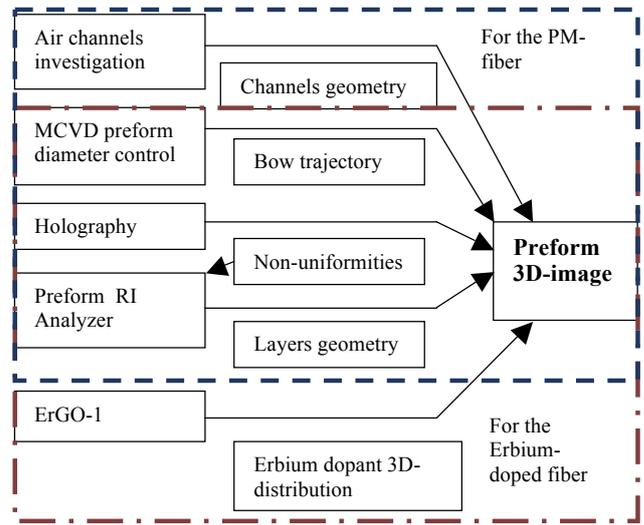


Fig. 8. The idea of preform 3D-image construction.

There are a lot of well-known methods for 3D-visualization. The choice depends on the current application. If it is designed for the lab test we suggest to use MatLAB built-in function. Also it is possible to use scaled bitmap sequence to show the parameters changes at every single cross-section (Fig. 9).

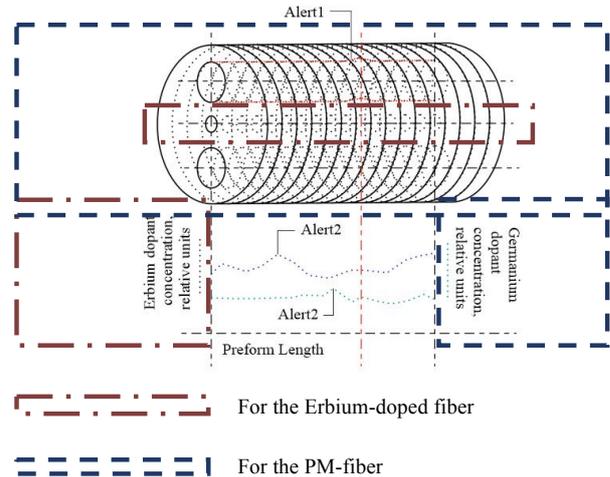


Fig. 9. An example of preform visualization.

At present the described approach is not universal, it has narrow application range in the field of special optical fibers (erbium doped fibers, panda-type fibers), and it is recommended to use this in combination with other techniques. Our set-ups allow to get information about the erbium oxide distribution inside the fiber preform, preform outer diameter during its fabrication process, the stress-rods geometry investigation set-up is now at the developing stage. The world-known firms like Photon Kinetics produce preform investigation instruments, but their comparison with our schemes is not reasonable due to the different kinds of data they allow to obtain. For the detailed studies we recommend to use them both. According to our information, we described in literature the feature of the erbium oxide distribution investigation in optical fiber preforms for the first time.

7. Acknowledgements

We thank Perm Scientific-Industrial Instrument Making Company for providing samples for tests as well as G.K. Shakirova and A.R. Khlopin for very helpful assistance.

8. References

- [1] A. Dhar et al., "The mechanism of rare earth incorporation in solution doping process", *Opt. Express* 16 (17), 12835 (2008).
- [2] A. Carter; B. Samson, "PANDA-style fibers move beyond telecom". *Laser Focus World* / August, 2004.
- [3] F. Just et. al., "The influence of the fiber drawing process on intrinsic stress and the resulting birefringence optimization of PM fibers", *Optical Materials*, vol. 42, April 2015, p. 345-350.
- [4] G. Keiser et. al., "Review of diverse optical fibers used in biomedical research and clinical practice", *J Biomed Opt.* 2014 Aug;19(8):080902.
- [5] Yu. A. Konstantinov, M.M. Poskrebyshv., N.A. Kharlamova, I.I. Kryukov, "Television measurement system of optical fibre preforms dimensions during the process of chemical vapour deposition" / *St. Petersburg Polytechnic University Journal of Engineering Science and Technology* / 2011, #6-2(138) – in Russian.
- [6] B. Lee, "Review of the present status of optical fiber sensors" / *Optical Fiber Technology*, vol. 9, Issue 2, April 2003, p. 57-79.
- [7] S. Nagel et al., "An overview of the modified chemical vapor deposition (MCVD) process and performance", *IEEE J. Quantum Electron.* 18 (4), 459 (1982).

About the authors

Yuri Alexandrovich Konstantinov, PhD in technical sciences, researcher at Photonics Lab of Perm Federal Research Centre of Urals Branch of Russian Academy of Science, Perm, Russia. E-mail: yuri.al.konstantinov@ro.ru.

Konstantin Pavlovich Latkin, PhD student at Perm National Research Polytechnic University, Perm, Russia; junior researcher at Photonics Lab of Perm Federal Research Centre of Urals Branch of Russian Academy of Science, Perm, Russia. E-mail: latkin.k.p@ya.ru.

Fedor Leonidovich Barkov, PhD in physical and mathematical sciences, senior researcher at Photonics Lab of Perm Federal Research Centre of Urals Branch of Russian Academy of Science, Perm, Russia; associate professor at General physics chair of Perm National Research Polytechnic University, Perm, Russia. E-mail: fbarkov@pstu.ru.

Alexandr Sergeevich Smirnov, PhD student at Perm National Research Polytechnic University, Perm, Russia; junior researcher at Photonics Lab of Perm Federal Research Centre of Urals Branch of Russian Academy of Science, Perm, Russia. E-mail: a.s.smrnv@gmail.com.

Maxim Mikhailovich Poskrebyshv, project supervisor at Inversion Sensor Company, Perm, Russia. Email: pmaximm@gmail.com.

Vladimir Pavlovich Pervadchuk, doctor of technical sciences, professor, head of Applied maths chair at Perm National Research Polytechnic University. E-mail: pervadchuk@mail.ru.

Daria Borisovna Vladimirova, PhD in physical and mathematical sciences, associate professor at Applied maths chair of Perm National Research Polytechnic University, Perm, Russia. E-mail: pervadchuk@mail.ru.

Yuri Alexandrovich Konin, development engineer at Perm Scientific-Industrial Instrument Making Company, Perm, Russia; junior researcher at Photonics Lab of Perm Federal Research Centre of Urals Branch of Russian Academy of Science. E-mail: yuri-konin@yandex.ru.

Andrey Ivanovich Garanin, PhD student at Perm National Research Polytechnic University, Perm, Russia; development engineer at Perm Scientific-Industrial Instrument Making Company, Perm, Russia; junior researcher at Photonics Lab of

Perm Federal Research Centre of Urals Branch of Russian Academy of Science. E-mail: kigaranin@yandex.ru.

Vladislav Viktorovich Burdin, PhD in physical and mathematical sciences, senior researcher at Photonics Lab of Perm Federal Research Centre of Urals Branch of Russian Academy of Science, Perm, Russia; associate professor at General physics chair of Perm National Research Polytechnic University, Perm, Russia. E-mail: vlaburdi@mail.ru.

Метод звуковых матриц в мультисенсорной аналитике для анализа многомерных скалярных полей

Е. Маликова¹, А. Пасько², В. Аджиев², В. Пилюгин³

malikova@smedx.com|pasko@mail.com|valchess@gmail.com|vvpilyugin@mephi.ru

¹ООО Медэкс, Самара, Россия;

²Национальный центр компьютерной анимации, Борнмутский Университет, Борнмут, Великобритания;

³Национальный исследовательский ядерный университет "МИФИ", Москва, Россия

В наши дни трудно представить современные научные исследования без использования инструментов визуальной аналитики. Исследователи столь же часто анализируют научные данные на базе анализа поставленных им в соответствие графических изображений, как и в повседневной жизни люди судят о предметах в основном на базе визуальной информации о них. Однако получать и анализировать информацию о мире вокруг можно не только с помощью глаз, но также и других органов чувств. Анализ научных данных мог бы быть существенно более эффективен при подключении дополнительных органов чувств. В данной статье представлен подход мультисенсорного анализа, обсуждаются теоретические вопросы, описаны основные концепции разработанных программных приложений, реализующих подход, представлен один из возможных видов аудио-визуального мэппинга для решения задачи анализа многомерных скалярных полей, так называемые «звуковые матрицы», и приводятся практические примеры анализа многомерных скалярных полей.

Ключевые слова: научная визуализация, многосенсорный анализ, FRep, сонификация.

Sound Matrices approach in multisensory analytics: for multidimensional scalar fields data analysis

E. Malikova¹, A. Pasko², V. Adzhiev², V. Pilyugin³

malikova@smedx.com|pasko@mail.com|valchess@gmail.com|vvpilyugin@mephi.ru

¹LLC SMEDX, Samara, Russian Federation;

²National Centre for Computer Animation, Bournemouth University, Bournemouth, United Kingdom;

³National Research Nuclear University "MEPhI", Moscow, Russian Federation

In our day it's hard to imagine modern research in sophisticated scientific area without use of visual analytics tools. It's become so common for scientists to obtain conclusions about scientific data on base of visual information as for people in their everyday life to make judgments about things on base of what they see. However, humans are able to receive and analyze information about world around them not only through their eyes, but with help of other sensory stimuli as well. Consequently, it would be not reasonable for researches to limit themselves to only one sensory stimuli in their research. In this work we introduce multisensory analytics approach, discuss theoretical issues, introduce main concepts of software tools implementing our approach, describe one of the possible auditory-visual mappings methods for multidimensional scalar fields analyses, that we called "sound matrices" and give practical examples of multidimensional scalar fields analyses.

Keywords: scientific visualization, multisensory analysis, FRep, sonification.

1. Multisensory visualization in modern research. Related works

Visual analysis of graphical representation of data has practically become a part of modern scientific research. It should be taken into consideration that researchers, as all people of creative professions, are predisposed to spatial creative thinking, so in the process of analysis of scientific data, they usually readily refer to various spatial and graphic images. This is one of the reasons, why scientific visualization and it's successor visual analytics as a method of data analysis has proven to be a very efficient tool often used by researchers. A quite general definition of visualization is "a binding (or mapping) of data to a representation that can be perceived" [1] is used more often nowadays and thus visual analysis is extended to become multisensory analysis. Among the sensory stimuli other than visual, the usage of sound has been widely

investigated since early 80-s [2], [3]. The human auditory perception is considered most quantitative because of its sensitivity to subtle changes in the sound characteristics. The technique of data representation using variable sound characteristics such as pitch, volume, note duration, rythm and others is called data sonification [4]. Let us look at sonification method characteristics more closely. Auditory perception has always been the human's early warning system, which may operate in the background mode. In [5] a small survey was made on the situations when using audio analysis may be even more effective than visual perception. The main classes of data that fall in this category are time-varying data and multidimensional data. The auditory perception brings the unique advantage to distinguish even small variations in the parameters of the single sound wave and to compare sound waves and thus can efficiently complement standard visual analyses, and help solving perceptual issues it often has to face

[6], [7]. As it was already noticed in [8] currently, it is considered that any person may be trained to develop an ear for music and ability to adequately judge on all its nuances allows one to take advantage of the most advanced extended sound analysis capabilities as well. In [9] the procedures of time-varying data representation in the graphical form using a musical accompaniment are considered. In the paper [10], there are examples of the presentation of scientific data in the form of musical fragments. This is largely a matter of sensory capabilities of a specific researcher, but in general in some cases just sound mapping and analyses can be successfully carried out. We will consider such possibility in this article, however, we are mostly concentrated on efficiently combining auditory and visual perception approach, that allows one to significantly enhance the ability to conduct analysis, taking advantages of two sensory organs that work differently, and to perceive the same information in different ways complementing each other. An extension of visualization through creating additional perceptual human inputs or more general a combination of several sensory stimuli for data representation is called data perceptualization [11], [12] (or data sensualization [13]). The typical combinations are between visual and auditory stimuli, visual and tactile/haptic stimuli [14], or three of these stimuli applied together [13]. As it was already mentioned in this article, we will concentrate on a visual-auditory data analysis practical case, although theoretical formalizations of establishing correspondences between the initial data and multiple sensory stimuli for multisensory analysis are assumed and were given in [15]. However, the problem of formalization of multiple sensory stimuli analysis and interpretation of analysis results in terms of initial data should be solved separately for each sensory stimuli.

2. An approach to multisensory data analysis pipeline. Visual-auditory analysis case

To obtain a multisensory representation we need to create a spatial scene [8], which is an assembly of spatial objects with their geometric, optical, sound and others descriptions. Then corresponding visual, sound and other stimuli can be generated using some specialized rendering procedures for further multicensory analysis. Based on the visual analytics process as presented in [16] and the idea of an intermediate multidimensional geometric representation of initial data [17], we proposed [8] the following interpretation of the basic multisensory analysis process 1.

In the diagram 1 perceptualization process is presented as a transformation (mapping) $M: D \rightarrow I$ from initial data D to insight I , which is the goal of the entire process. The mapping M is a superposition of mappings from one set to another in the diagram. Thus, the initial data undergo geometric interpretation and are mapped to the set G of multidimensional geometric models. The next step is to generate several sensory stimuli SS for human perception. The mappings from G to SS are facilitated by the introduction of a spatial scene, which is an assembly

of spatial objects with their geometric, optical, auditory, tactile and other properties (multimedia objects). Note that the geometric objects in the spatial scene can have their dimensionality reduced to 2D and 3D using geometric cross-sections and projections, which allows for applying well-known graphical rendering algorithms. When such a spatial scene is constructed, various sensory stimuli can be generated using corresponding rendering procedures: visual stimuli V (graphical images), auditory stimuli A (sounds), tactile and haptic stimuli T , and others. The final insight I can be either directly obtained from the generated sensory stimuli through human perception and analysis, or it is obtained in a combination with generating a hypothesis H and its analysis including automated methods. Note that the hypothesis H can be also represented with visual and other sensory stimuli, which can help to refine or redefine it in the process of analysis. The entire process has iterative character, which is shown by the feedback loop in the diagram. The user may tune or redefine not only the parameters of the data input, but also the introduced geometric models, the hypothesis, the selection of sensory stimuli and the type and parameters of rendering procedures. Applying the presented general approach the process of data analysis involving both human vision and hearing, we need to do the following:

1. To obtain a mapping of the given data onto its representation in the form of images and sound. To obtain a necessary model of a spatial scene, its geometric and optical models need to be extended by a sound model. Such a spatial scene augmented with sonification needs to be put in correspondence to the given data and then sound rendering can be applied with output to speakers or some other sound output device for further analysis.
2. To analyze the rendered images and sound and to interpret the results of this analysis in terms of the initial data.

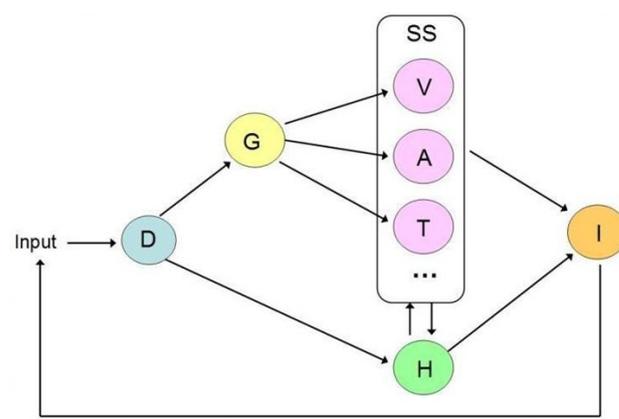


Fig. 1. Multisensory analysis process.

The definition of corresponding sound mappings that can be concretely analysed and easily interpreted by researchers is also a question that should be studied. Here we suggest that a researcher in common case should

be trained to interpret some not quite evident sound mappings similar to musicians training their ears for further music analysis in modern practice. In our work, we take advantage of musicians' approach adopting well-known concepts of music analysis and writing used by musicians from simple properties of sound analysis (pitch, volume, duration, etc.) to "music" properties analysis (tone, interval between tones, etc.). These concepts are taken as the base of sound mapping and accordingly of sound analysis. 2 presents some musical (sound) characteristics that musicians may distinguish auditorily and describe quantitatively: tone, note duration, interval between two notes are most often used ones.

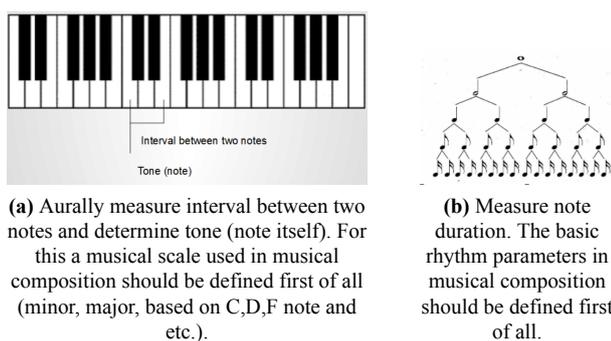


Fig. 2. Musical sound characteristics musicians may distinguish auditorily.

In this article, only simple cases of sound analysis, cases that for advanced analysis require some musical training (e.g., to determine interval and note) will be considered. From authors point of view, a camera, a sound receiver, a haptic cursor and other similar elements need to be explicitly placed in the spatial scene as spatial models of the human organs of perception. Thus, a spatial scene includes spatial objects representing data as well as other spatial objects representing their influence on human senses. Rendering of the spatial scene generates information for output devices provided for consideration by humans, namely a screen, speakers, a haptic device and others.

3. Multimedia coordinates based approach for visual-auditory analyses

Although some efforts have been made on the development of data perceptualization, a formal framework for establishing correspondences between data and multiple sensory stimuli has not been yet proposed. In [15] we introduced a framework based on the concept of multimedia coordinates introduced previously in [18] and applied in multidimensional shape modeling. The approach is formalization of mapping from a multidimensional geometric model to a multimedia object that can be treated as a multidimensional object with Cartesian, visual, audio, haptic and other types of multimedia coordinates. To operate with multimedia coordinates, one can introduce a system of normalized numerical coordinates (a unit cube) and its one-to-one

correspondence to the multimedia space. By selecting a real normalized value, one can use the corresponding value of the multimedia coordinate [15]. Each geometric coordinate variable takes values within a given interval and multimedia coordinates also have their own variation intervals as well. So to define the mapping, one has to establish a correspondence between these intervals through the normalized numerical coordinates. Thus a space mapping between geometric coordinates and multimedia coordinates establishes correspondence between the multidimensional shape and the multimedia object. In this way, a correspondence can be also established between the given scientific data and a multimedia object, because introducing a multidimensional geometric model is one of the steps in the visualization pipeline presented previously. As for multidimensional geometric model involved in multisensory visualization we used approach based on a constructive method to the creation of function evaluation procedures for geometric shapes, called the Function Representation (FRep) [19] and its extension HyperVolume [20], where the object is represented not by a single function, but by a vector-function with one component responsible for the object geometry and other components serving as point attribute functions representing such object properties as material, color, transparency, sound characteristics and others.

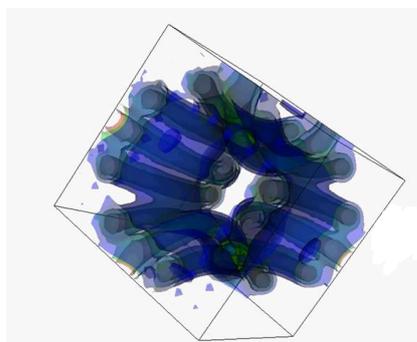
Thus both multimedia coordinates and hypervolume model for describing multimedia object are quite interconnected within presented approach used. To visualize a multidimensional geometry (or scalar field) special methods and geometric modelling techniques controlled quite well within multimedia coordinates concept are used. Usually they are:

1. Use of several semitransparent colored isosurfaces (see example 3(a)) in case when dimensionality of geometric object is not greater than 4D.
2. Projections on a subspace [PAS*96].
3. Cross-sections (see example 3(c)).

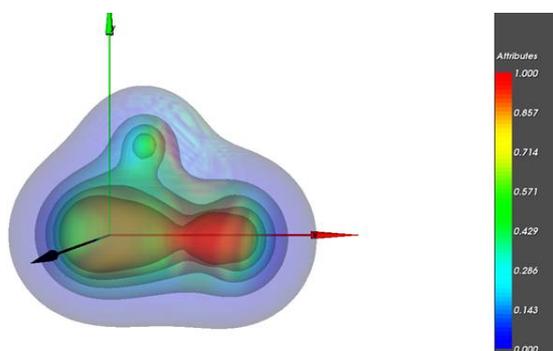
For more effective analysis these geometric operations may be defined interactively, through providing some specialized interactive widgets (plane, hypercube widgets to define cross-section and etc.). Also some special types of spatial scene may be introduced to provide special graphical representations called "matrices of cross-sections" or spreadsheets. An according type of multimedia coordinates was introduced in [18] and is called "spreadsheet coordinates". This type of coordinates allows for spreadsheet-like spatial organization of elementary images or shapes in the regularly or irregularly placed 1D, 2D or 3D nodes. In this work we will consider the case of a 1D node. Let us consider a simple case of 1D spreadsheet on Fig.3 (c) with the specific types of multimedia coordinates:

1. "x", "y" and "z" types correspond to world coordinates in the Cartesian coordinate system. They are used to describe a set of 3D isosurfaces $f(x,y,z)=c$
2. "c" type corresponds to a photometric coordinate, namely the color.

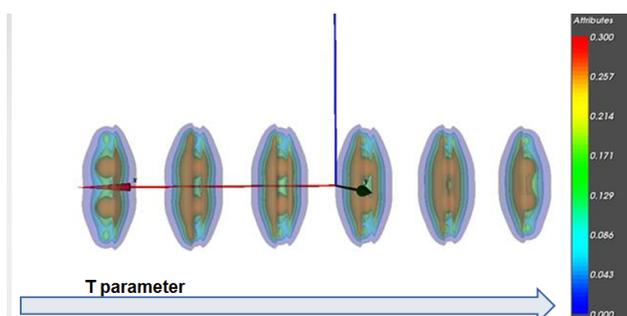
3. "v" type correspond to 1D spreadsheet coordinates. By assigning its discrete values we construct a horizontal 1D spreadsheet. Each section in Fig. 3 represents a 4D geometric object, displayed with 3D isosurfaces $f(x,y,z)=c$, where c is as well mapped to a photometric coordinate.



(a) Visualization of scalar order-parameter field distribution.



(b) Visualization of electron density and electrostatic potential field of NCH molecule.



(c) Visualization of dynamic electron density field of C₂H₂ molecule with use of spreadsheet technique.

Fig. 3. Examples of static and dynamic scalar fields visualization.

It should be noted that we have to reduce dimensionality not only to obtain graphical representation, but as well to map our data into sound and other type of sensory stimuli. So we have to introduce other type on spreadsheet coordinates, that will be called multisensory spreadsheet coordinates in general case. In[8] we demonstrated basic interactive techniques, when sound with predefined by data characteristics is produced in selected point. In this article we are concentrating on

more extended approach that we called Sound Matrices that in general may be used by researches without preliminary theoretical knowledge in music analyses, although these knowledge will make analyses more efficient and extended, and on base of them analyses may be carried out even without visual information. We'll give detailed examples and explain analyses process in details. As previously[8], we propose some generalizations on base multimedia coordinates approaches for specific type of multidimensional data multisensory analysis - scalar fields, bringing together some most well known interactive, photometric and geometrical techniques and demonstrating how they can be extended by other multisensory techniques on the example of sound.

4. Framework main features, cases examples

The presented below results of visual-auditory analyses were obtained with developed framework, based on HyperVolume object that influences various sensory stimuli (visual and sound in particular case) conception as base for computer scene model in multisensory analyses pipeline. The framework is designed on base of OpenSource projects VTK [23], Hyperfun [22] and OpenAL [24] in a form of API for writing standalone applications using C++ and scripting languages Hyperfun and Python. To summarize, we've tried two approaches to scalar fields data mapping to sound characteristics:

1. Direct sound characteristics mapping (frequency, volume) and generating wave with specified parameters. This an approach we've tried in our earlier works that helped us to make basic judgements on changes on specific mapped value (it's growing or falling and etc.) [15].
2. Mapping to abstract "musical terms" and operating them in entire analyses process. Notes, intervals, musical scale, musical instrument used, etc. As it will be demonstrated below this approach provides possibilities for more extended analyses.

As it was noted previously, in multimedia coordinates concept multimedia coordinates also have their own variation intervals, for example, color varies inside the color space (RGB cube). It's a well known that to judge about geometric value mapped into color an according color scale is used. Similarly, in music theory, there is an approach to judge about notes on base of their position on a specified scale (Cmaj for example), in other words, relative to the tonic, the first and main note of the scale from which each octave is assumed to begin. So in music terminology, that we adapt each note has a scale degree, that is it's position on scale. Example of C-major scale with according scale degrees C–D–E–F–G–A–B, in which C is the tonic). Usually such degrees are quite easy to learn to distinguish aurally after some special tuning for specified scale (a basic exercise in musical training). It is possible to assign a scale degree to the 12-tone chromatic scale, but this distinguishing such degrees and intervals variety is more difficult and demands training. In our framework we've chosen as base for sound analysis Cmajor scale

of 1-3 octaves. Notes sound generated on piano, guitar, flute or other instruments of user choice. As we mainly concentrate on a simple rendering cases, easy to perceive and analyze, the cases believed that practically anyone (without no damage of hearing system) can be capable to perceive and analyze after some simple harmonic ear training (musical school level) we introduce following data preprocessing steps in order to take into consideration perception issues of musical training and analyses as well: If we consider visual analyses of scalar fields, on base of visual information we mainly carry out two forms of analyses:

1. Shape analyses.
2. Feature analyses, in particular change in form.

Sound analyses also permits to judge about these and in particular efficient in all sort of changes detection as it was mentioned above. But we are limited by value of interval between notes that we may distinguish aurally to perform more extended analyses (not only to judge it value is growing or decreasing or by what value it changed). Usually we aurally quite easy distinguish changes by interval from “whole tone” (for some well-trained users it is “semi tone”). And it’s more easy to work with scale of 2 octaves or one. The obtained scalar values are mapped into degrees on Cmaj 1-2 octave scale. So, for example 1 octave, that has 8 degrees represent values within selected [xmin, xmax] range. So [xmin, xmax] range is spitted on 8 subranges [Ximin, Ximax] in order each degree represents scalar value within according subrange. Like this researcher may judge on approximate scalar value aurally by according not (degree on Cmaj scale). As well in order to improve analyses we suggest user may remove “unnecessary” information he may neglect in analyses process. So we suppose user may do following recommended image preprocessing steps, before mapping to sound:

1. ImageThreshold – to remove scalar values not falling into selected range [xmin,xmax] we are interested in or may perform first step analyses without. For example, if we deal with some CT of medical organ within body it has signal within given range and we want to analyze only shape boundaries and it’s internal structures.
2. Morphological operations on selected boundary– to remove small “island like” parts and smooth shape boundary, to reduce noise caused artifacts.

This preprocessing steps are conditional, however they may significantly increase analyses efficiency.

It should be noted that in general approach is quite similar to basic techniques used for organ segmentation in image processing as entire procedure of segmentation helps in particular to analyze organs better visually. The used sound mapping technique we implement is similar to one used in medicine as well, where 3D image scan is previewed and analyses mostly as a set of 2D image slices, where scalar value is defined at each pixel with color. We project in general case multidimensional geometry to 4D space x,y,z,f and build with defined step it’s crosssections $z_i=c_i$ that represent 2D sound images where instead of

pixel position we have sound source defined by x,y values and f is mapped according musical degree on Cmaj scale. We fix one by one some $y_j=c_j$ values within y range and moving sound automatically with defined step x_i along x axis with fixed y_i we sonify our 2D image line by line, thus obtaining what we call “sound matrix”. Below on 4 is presented 2D image slice of simple 4D scalar field, presented on 3a and notes of according played sound matrix 6x6, that “duplicates” the same information that we analyses visually. To simplify auditory-visual analyses the scalar field was preprocessed with image threshold technique in order to exclude from analyses all values greater than 0.5 (on image marked with wight color), so C3 (C note of first octave) is played in all this range.

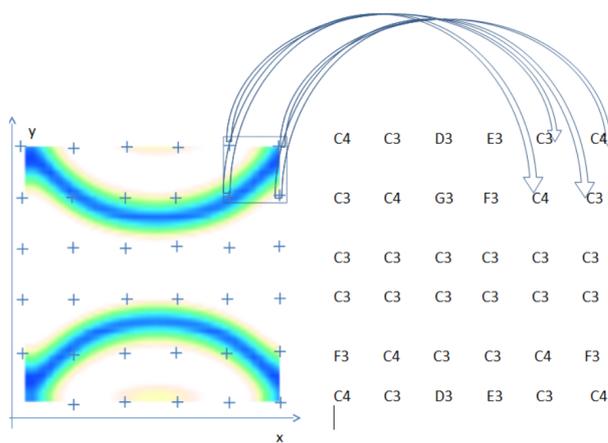


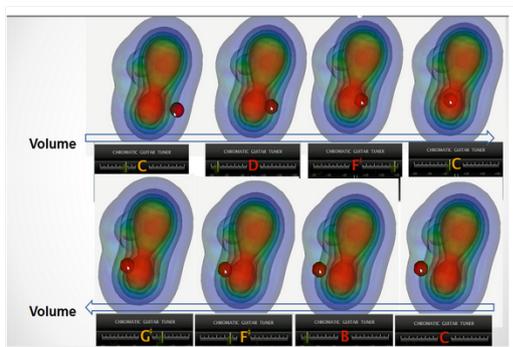
Fig. 4. Visualization of scalar order-parameter field distribution selected crosssection and according sound matrix with defined resolution 6x6.

It should be noted that different approaches depending on musical practical knowledge of researcher that analyses data with sound should be taken into consideration as well. So to move from basic analyses of scalar field main features (approximate form, symmetry) into details, we suggest the analyses process reiterativeness in general case, so it’s complexity may vary. This may suggest mapping to bigger interval, or readjusting or removing some preprocessing steps described above, or increasing sound matrix resolution to perform more detailed shape analyses. This may as well include additional interactive selection of image region of interest and it’s sonification, a technique that will be demonstrated further.

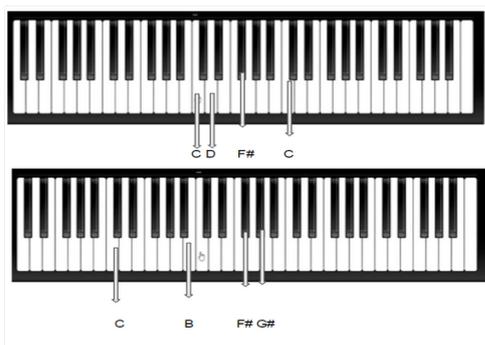
As in our case we mostly concentrate on use of the advantages and particularities of human hearing system (alarming system, that may operate is silent mode, those perception principles are quite different comparing to visual system) in analyses process in order to enhance it’s efficiency we suggest it’s most efficient combination with visual analyses. It should be noted that visual information content may duplicate those mapped in sound, or some extended techniques for image features extraction and visualization may be used as combination with sound analyses.

In example below, we'll concentrate on molecular fields as their form in crosssection in general case is quite simple (one or several spheres) and in case of such fields usually some general features are subject of study (symmetry, how fast is changing and etc.).

In[8] we've studied in detail a auditory-visual analyses of two molecular scalar fields, presented on 3(b) that are supposed to be analyzed together (a quite usual task in physical or medical research filed). In studied auditory-visual analyses case the sound wave was generated by a sound terminal with the frequency corresponding to the location of the point sound source and perceived by the user as a specific sound tone. Each sound tone generated at the location of the point source was defined on selected 2 octaves musical Cmajor scale 5. Here we receive following tones presented in 5(a) and can graphically present their place on musical scale 5(b). Here a basic guitar tuner was also used to illustrate the current note value 5(a).



(a) The use of an interactive “sphere” widget to define sound frequency w and volume v of the generated sound defined by the functions A_2 and A_3 values at fixed values of world coordinates x, y, z .



(b) Presentation of according notes on Cmajor scale (2 octaves) on piano. A researcher with well-trained musical ear and appropriate «auditory tuning» on Cmajor scale can easily aurally determine these notes and their place on piano musical scale and judge about how quantitatively sound was changed.

Fig. 5. Exploration of two scalar fields dependency and change with pitch and volume.

We may as well study both scalar fields slice by slice, obtaining crosssections with $z=z_i$ while moving along z axis with defined step 6.

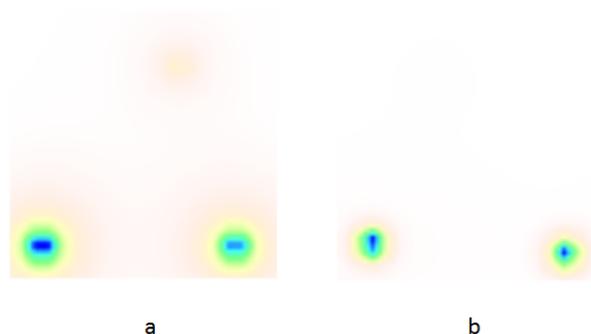


Fig. 6. Visualization of scalar fields selected crosssections $z=0$, electrostatic potential(a) and electron density(b) accordingly.

Such mapping will help us to define region where both scalar fields take biggest values and detect that their intensity spread can be characterized as caused by point source type (atoms position) for both fields, but those centers are a bit shifted in electron density field (a shift that is even hard to distinguish visually). The alternative to this approach may be selection of region of interest, where both fields take biggest value and simultaneous visualization of selected crosssections of one field (electron density) and sonification through sound matrix mapping of obtained crosssection of second field (electrostatic potential) as presented on 7. This will permit us to immediately and easily detect the main trend as sound matrix of second field will demonstrate practically symmetrical trend.

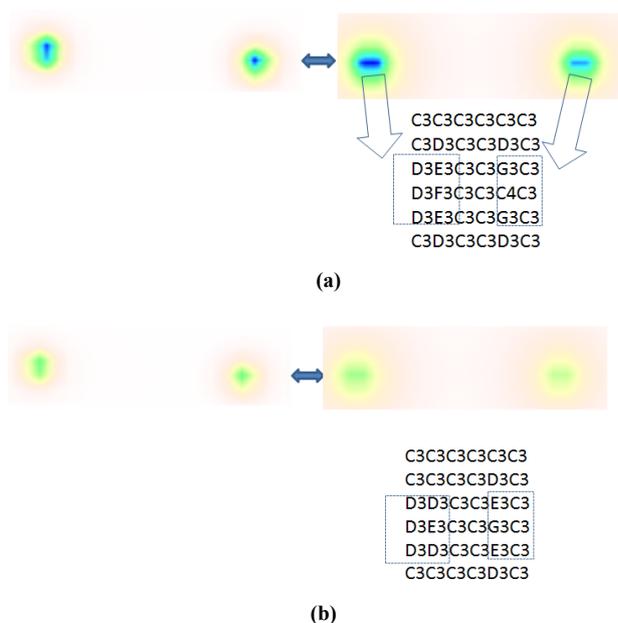


Fig. 7. Visual auditory analyses by selected crosssections (a,b) and sound matrixes of two scalar fields electron density and electrostatic potential accordingly of CNH molecule.

As well, to judge about difference in scalar field value change more efficiently we may compare generated sound matrices for two scalar fields, but for such analyses 8.

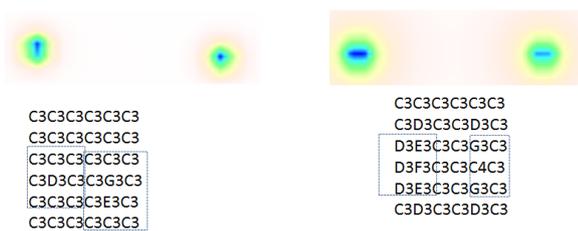


Fig. 8. Visual-auditory analyses by selected crosssections of two scalar fields. Sound matrix comparison.

5. Conclusions

In conclusion it may be said that the formalization of the mapping between the multidimensional geometric models and the spatial scene available for rendering multiple sensory stimuli is a big research question, still to be address. However, we have shown a possible solution in the case of the initial data represented by scalar fields (real functions of several variables) and illustrated this by the case study of the scalar fields analysis using interactive visual-auditory display on base of two possible techniques. Different types of interactive visual-auditory or auditory widgets, not only those based on combination of described two, as well as different types of sound mappings can still be studied as well as possible image preprocessing techniques that are used to simplify analyses oprocess. We are planning to continue to involve the concept of multimedia coordinates as a way to establish even more complex correspondences between initial data, the introduced multidimensional geometric models and multiple sensory stimuli.

6. References

- [1] Foley J., Ribarsky B. Next-generation data visualization tools, in *Scientific Visualization, Advances and Challenges* L. Rosenblum et al. (Eds.), Academic Press, 1994.
- [2] Yeung E. Pattern recognition by audio representation of multivariate analytical data *Analytical Chemistry*, vol. 52, No.7, 1980, pp. 1120–1123.
- [3] Bly S. Presenting information in sound *Proceedings of the CHI '82 Conference on Human Factors in Computer Systems*, ACM, 1982, pp. 371-375.
- [4] Kaper H., Wiebel E., Tipei S. Data sonification and sound visualization *Computing in Science and Engineering*, vol.1, No.4, 1999, pp.48-58.
- [5] Scaletti C., Craig Alan B. Using sound to extract meaning from complex data *In Proceedings SPIE*, 1459. 1991. pp. 207–219.
- [6] K. Gaither ; D. Ebert ; B. Geisler ; D. Laidlaw *In the Eye of the Beholder: The Role of Perception in Scientific Visualization* // *Visualization*, 2004. IEEE
- [7] C. Johnson *Top scientific visualization research problems* *IEEE Computer Graphics and Applications*, Volume: 24, Issue: 4, July-Aug. 2004 URL: https://www.cs.unc.edu/~taylorr/Comp715/papers/johnson_top_problems.pdf.
- [8] E. Malikova, V. Pilyugin, V. Adzhiev, G. Pasko, A. Pasko *Multisensory Analytics: Case of Visual-auditory Analysis of Scalar Fields* *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) - Volume 3: IVAPP*, pp. 322-329, ISBN: 978-989-758-228-8
- [9] Mezrich J. J., Frysiner S., Slivjanovski R. Dynamic representation of multivariate. time series data, *Journal of the American Statistical Association* 1984. Vol. 79, N. 385. pp. 34–40.
- [10] Lodha Suresh K., Beahan J., Heppe T. and etc. *MUSE: A Musical Data Sonification Toolkit* *In Proceedings of International Conference on Auditory Display (ICAD)*. 1997. pp. 36–40.
- [11] Grinstein G., Smith S *Perceptualization of scientific data* *Proc. SPIE 1259, Extracting Meaning from Complex Data: Processing, Display, Interaction*, 1990, pp. 190-199.
- [12] Ebert D. Extending visualization to perceptualization: the importance of perception in effective communication of information *in The Visualization Handbook*, C. Hansen and C. Johnson (Eds.), Academic Press, 2004, pp. 771-780.
- [13] Ogi T., Hirose M. Multisensory data sensualization based on human perception *VRAIS '96 Proceedings of the 1996 Virtual Reality Annual International Symposium*, 1996, pp. 66-71.
- [14] Maciejewski R., S. Choi, D. Ebert, H. Tan *Multi-modal perceptualization of volumetric data and its application to molecular docking*, *WHC '05 Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2005, pp. 511-514.
- [15] V. Adzhiev, E. Malikova, A. Pasko, G. Pasko, V. Pilyugin *Multidimensional scalar fields data analysis by means of multisensory visualization* *The 26-th International Conference GrahCon2016, September 19-23, 2016, NNGASU, Nizhny Novgorod, Russia*, pp. 142-146
- [16] Keim D., Mansmann F., Schneidewind J., Thomas J., Ziegler H. *Visual analytics: scope and challenges*, *Visual Data Mining* *Lecture Notes in Computer Science*, volume 4404,

- Springer, 2008, pp 76-90.
- [17] V. Pilyugin, E. Malikova, V. Adzhiev, A. Pasko
Some theoretical issues of scientific visualization
as a method of data analysis
Transactions on Computational Science XIX,
Lecture Notes in Computer Science, vol. 7870,
Springer-Verlag, 2013, pp. 131–142.
- [18] Adzhiev V., Ossipov A., Pasko A.
Multidimensional shape modeling in multimedia
applications, in MultiMedia Modeling: Modeling
Multimedia Information and Systems,
ed. A.Karmouch, World Scientific, 1999, pp. 39-60.
- [19] Pasko A., Adzhiev V., Sourin A., Savchenko V.
Function representation in geometric modeling:
concepts, implementation and applications
The Visual Computer, vol.11, No.8, 1995, pp.429-
446.
- [20] Pasko A., Adzhiev V., Schmitt B., Schlick C.
Constructive hypervolume modeling, Graphical
Models
vol. 63, No. 6, 2001, pp. 413-442.
- [21] Pasko A.A., Savchenko V.V “Projection operation
for multidimensional geometric modeling with
real functions”, Theory and Practice of Geometric
Modeling
Blaubeuren II conference, W. Strasser (Ed.),
October 14-18, 1996, Tuebingen, Germany.
- [22] URL: HyperFunproject. – HyperFun project.
- [23] URL: <http://vtk.org>. – VTK project.
- [24] URL: [http://connect.creativelabs.com/openal/
Documentation/OpenAL_Programmers_Guide.pdf](http://connect.creativelabs.com/openal/Documentation/OpenAL_Programmers_Guide.pdf).
– OpenAL Programmers Guide.

Исследование психологических состояний пользователей сред виртуальной реальности

В.Л. Авербух¹, Н.В. Авербух²

averbukh@imm.uran.ru|natalya_averbukh@mail.ru

¹ИММ УрО РАН, УрФУ, Екатеринбург, Российская Федерация;

²АНО ИНЭКС, Екатеринбург, Российская Федерация

Работа посвящена исследованию состояний, возникающих у пользователей систем визуализации на базе сред виртуальной реальности. Описываются феномены, связанные с виртуальной реальностью и ситуации, связанные с переживанием феномена присутствия. Рассказывается о проектировании и проведении экспериментов, связанных с изучением состояния присутствия у пользователей, решающих интеллектуальные задачи.

Ключевые слова: научная визуализация, виртуальная реальность, погружение, присутствие.

Evaluation of user's psychological states in virtual-reality environments

V.L. Averbukh¹, N.V. Averbukh²

averbukh@imm.uran.ru|natalya_averbukh@mail.ru

¹IMM UrB RAS, Ekaterinburg, Russia;

²ANO IEX, Ekaterinburg, Russia

The paper is devoted to evaluation of user's psychological states during interaction with virtual reality environments. The phenomena connected with virtual reality are described. Also the situations connected with presence state are considered. Experiments in the presence state and process of their design are described.

Keywords: scientific visualization, virtual reality, immersion, presence.

1. Введение

Изучение деятельности специалиста в процессе анализа и интерпретации данных при помощи визуализации – важная задача, позволяющая повысить эффективность всего компьютерного моделирования. Визуализация в рамках цикла компьютерного моделирования обеспечивает этапы анализа и интерпретации данных, проводимые пользователями систем визуализации. Как правило, пользователи систем научной визуализации являются также разработчиками компьютерных моделей и обладают большим объемом знаний по изучаемой проблематике. Визуализация для них – не набор занимательных картинок, а некоторый визуальный текст, несущий значимую информацию. “Чтение” такого текста можно рассматривать как последовательность восприятия визуальных объектов и уяснения их значения. Поэтому разработка специализированных систем научной визуализации требует учета, так называемого, человеческого фактора (human factor), что включает в себя исследование особенностей восприятия визуальной информации конкретными пользователями. Этому вопросу посвящено большое количество исследований, проведенных, прежде всего, специалистами в области компьютерной психологии.

Успех визуализации обеспечивается наличием связи между внутренней природой и структурой модельных объектов (и соответствующих им данных) и внутренними ментальными структурами пользователя, в которых представлена сложившаяся у него картина явления (так называемые “репрезентативные когнитивные структуры”). Проблема уяснения смысла визуализации может быть рассмотрена в связи с изучением инсайта. Особенно важными эти проблемы стали в связи с использованием сред виртуальной реальности, которые позволяют получить качественно новые возможности по

представлению, восприятию и анализу сложных объектов суперкомпьютерного моделирования.

Ниже в работе рассматриваются некоторые вопросы, связанные с изучением состояний пользователей систем визуализации на базе сред виртуальной реальности.

2. Развитие компьютерной визуализации

Визуализация, понимаемая как зримое представление моделей, существовала задолго до изобретения современных компьютеров. Появление в 50-ых годах XX века интерактивной компьютерной графики дало качественный скачок в визуализации как таковой. «Черно-белая» графика на базе векторных дисплеев была основой первого этапа в развитии алгоритмических и программных средств визуализации. Аппаратные возможности дисплеев позволяли вывод ограниченного количества фигур и кривых. На векторных дисплеях некоторых типов можно было реализовать простые анимационные изображения. Средства графического ввода давали достаточно богатые возможности взаимодействия с моделирующими программами.

Внедрение цветной растровой графики, резкое увеличение возможностей дисплеев, дало толчок к созданию методов и алгоритмов фотореализма, что, в свою очередь, обеспечило качественные изменения визуализации за счет увеличения количества и качества воспринимаемой визуальной информации. На экране стало отображаться резко большее количество объектов. Появились различные способы отображения модельных существей в рамках одного графического вывода, например, форма, цвет, текстура, движение. Визуальные объекты стало легче воспринимать и различать. К концу 80-ых годов компьютерная визуализация оформилась как самостоятельная дисциплина. В системах научной визуализации начали активно использоваться трехмерная графика, анимация, а затем и фотореализм. Опыт разработки и использования средств научной визуализации показывают, что развитие аппаратных и алгоритмических средств интерактивной графики обеспечило последовательное увеличение количества и качества

воспринимаемой визуальной информации. Интересно, что современное развитие аппаратных и алгоритмических средств компьютерной графики, в значительной мере связанное с потребностями компьютерных игр, привело к тому, что богатые возможности компьютерной графики, по сути, значительно превышают потребности большинства современных систем компьютерной визуализации.

К началу 90-ых годов в системах научной визуализации начали использовать «большие экраны» и среды виртуальной реальности, первоначально разработанные для авиатренажеров. В случае виртуальной реальности компьютером при помощи специальных устройств создается особая среда, воспринимаемая пользователем как реальный мир, в котором он действительно находится (а не наблюдает со стороны) и с которым взаимодействует напрямую, так же, как и с обычным миром [4]. Уже в 1992 году в журнале *Communications of the ACM* появилась публикация, подробно описывающая возможности сред виртуальной реальности типа CAVE [3]. Начиная с 1991 года, исследователи NASA описывали проект разработки «виртуальной аэродинамической трубы» [2]. (Соответствующая система была закончена в 1994 году.) В настоящее время средства виртуальной реальности стали широкодоступными. Системы типа CAVE используются во многих научных и проектных организациях за рубежом и в нашей стране.

3. Феномены, связанные с виртуальной реальностью

Общая задача наших исследований – получить набор рекомендаций для проектирования специализированных систем визуализации, использующих виртуальную реальность.

Среды виртуальной реальности имеют такие характеристики, как трехмерность и стереоскопичность; динамичность; мультисенсорность (подача информации на несколько органов чувств); контроль над видимой частью виртуальной среды; возможность обеспечить взгляд изнутри (пространства) [1]. Пользователь имеет возможность погружения (*immersion*) в виртуальный мир, в связи с чем возникает феномен присутствия (*presence*), связанного с ощущением нахождения “там”, в виртуальном мире, в противоположность наблюдению за картинкой на экране со стороны, характерного для традиционной компьютерной графики. При описании феномена присутствия можно говорить о субъективном восприятии, в котором человек “забывает” о роли техники в создании той среды, с которой он взаимодействует. По мнению специалистов, именно феномен присутствия обеспечивает большие возможности при визуальном анализе научных данных в средах виртуальной реальности. В тоже время наша практика проектирования специализированных систем научной визуализации и работы с их пользователями, также как и личный опыт, связанный с переживанием присутствия в экспериментальных средах виртуальной реальности, вызывала определенные опасения. По некоторым параметрам присутствие приближается к изменённым состояниям сознания (хотя и не является таковым) [9]. Научные вычисления предполагают анализ большого объема данных, полученных в результате работы компьютерной программы. Анализ научных данных, характерный для научной визуализации, включает ряд операций, в том числе, идентификацию, сравнение, ранжирование, локализацию, классификацию, фильтрацию, установление ассоциаций и корреляций,

кластеризацию [1]. Появился целый ряд вопросов, связанных с влиянием состояния присутствия на выполнение деятельности пользователем. Не отвлечет ли новое состояние от задачи? Было важно и интересно узнать достижимо ли присутствие, как таковое, при работе с абстрактными образами? Сможет ли пользователь взаимодействовать со средой, если у него не будет субъективного ощущения пребывания в ней? Далее возник вопрос о взаимовлиянии феномена присутствия и интеллектуальной деятельности по обработке представленной информации. Могут ли они протекать одновременно, будет ли одно усиливать другое или мешать, препятствуя решению задач? Важно также отделить воздействие психологических эффектов от воздействия чисто технических средств. По сути, было необходимо выяснить, сможет ли пользователь (например, специалист в области прикладной математики), погруженный в среду виртуальной реальности точно и адекватно анализировать сложные данные, представленные в виде абстрактных образов. Также было важно узнать не вызовет ли присутствие каких-либо неприятных ощущений в ходе сложного анализа. Таким образом, вместе с исследованием феномена присутствия необходимо изучение влияния виртуальной реальности на решение интеллектуальных задач. Поэтому было решено организовать изучение состояний потенциальных пользователей систем научной визуализации на базе виртуальной реальности. В этой связи возник вопрос о выборе методики исследования.

4. Переживание феномена присутствия

Феномену присутствия посвящено значительное количество исследований. В то же время немного работ связано с исследованием феномена присутствия у пользователей систем научной визуализации. Поэтому выбор методик исследования переживания феномена присутствия у пользователей систем научной визуализации на базе сред виртуальной реальности представлял собой новую задачу. Естественно найти актуальную проблематику визуализации и в ее рамках проводить исследования. В этом качестве выбрана задача разработки средств визуализации для задач математической физики, в частности, связанных с сеточными методами вычислений. Основные данные, получаемые в результате расчетов, представляют собой массивы значений параметров в узлах расчетной трехмерной сетки. Точность результатов моделирования зависит от количества элементов сетки, то есть от числа ее разбиений. Необходимо изучение отдельных фрагментов сетки и обеспечение навигации в трехмерном пространстве. Пользователь в данном случае имеет дело с большим объемом сложно структурированных данных. Использование сред виртуальной реальности может дать большой эффект при визуализации сеток за счет эгоцентричной точки обзора, “погружения” пользователя в данные и возможности изучения сетки изнутри. Возможна реализация “путешествия” внутри сетки и изучения особенностей ее внутренних объектов.

Первой идеей была идея проведения исследования на базе системы визуализации с использованием очков виртуальной реальности и сеточных данных, генерируемых используемыми на практике системами моделирования. Это, во-первых, обеспечивало сравнительно быстрое достижение поставленной цели, а, во-вторых, позволяло работать с настоящими данными и настоящими пользователями, выясняя на практике все поставленные вопросы. Однако именно из-за проблем с пользователями эта идея оказалась нерезализуемой. Оказалась, что количество исследователей, способных полноценно анализировать и интерпретировать конкретные графические выводы, ограничено специалистами, знающими и понимающими сущности математической и компьютерной моделей. Другие

исследователи, компетентные в принципе, но, не имея полной информации о модели, могут не увидеть важные особенности. Делать какие-либо выводы на основе весьма малой выборки нельзя. Да и заставить специалистов заниматься исследованиями собственных свойств и особенностей достаточно затруднительно.

Далее в качестве модельной задачи была предложена проверка скорости и точности ориентации по трехмерной карте местности, реализованной в среде виртуальной реальности. Задачи пространственного ориентирования носят универсальный характер и могут быть поставлены как в дву-, так и в трехмерной среде. Возможна реализация передвижения по традиционной карте и по ландшафту, представленному в среде виртуальной реальности. Это позволяет сравнить между собой визуализацию, предназначенную для обычного вывода информации на экран и визуализацию, ориентированную на среды виртуальной реальности. Задачи на пространственное ориентирование доступны широкому кругу пользователей и в какой-то мере могут быть увязаны с задачей навигации в абстрактном виртуальном пространстве. Однако в этом случае не предполагается моделирование основных операций анализа данных в виде, характерном для научной визуализации.

В результате поиска подходящей основы для моделирования работы пользователя со специализированной системой научной визуализации на базе среды виртуальной реальности была выбрана методика, в которой испытуемым предлагалась задача на пространственное преобразование объекта. Этой методикой является тест «Кубики Коса» – субтест интеллектуального теста Векслера. В ходе тестирования испытуемый должен собрать соответствующие предоставленному образцу узоры с помощью набора из разноцветных кубиков. «Кубики Коса» позволяют тестировать восприятие, моторику, зрительно-моторную координацию, пространственные представления и эвристические способности, оценивает способность к выполнению основных мыслительных операций – сравнение, анализ и синтез [10]. Выбор данного теста позволил использовать методики исследования, наработанные за долгие годы психологической и медицинской практики. В первоначальном исследовании в ходе экспериментов рассматривались два вопроса. Мешает или помогает использование виртуальной реальности пользователям систем визуализации в их работе. Влияет ли переживание феномена присутствия на способность решения интеллектуальных задач?

В эксперименте приняло участие несколько десятков человек, в основном студентов, аспирантов и профессионалов технических, естественнонаучных, математических и компьютерных специальностей. Задача испытуемых заключалась в выполнении десяти заданий на составление узора из верхних граней красно-белых кубиков.

Экспериментальная группа испытуемых выполняла тест в среде виртуальной реальности, выводимой на стереоочки. Контрольная группа выполняла тест с выводом визуализации на экран монитора, на котором было установлено разрешение такое же, как на стереоочках. При этом не менялось ни содержание заданий, ни способ, ни последовательность их предъявления. Кроме этого, испытуемым обеих групп было предложено посмотреть в стереоочках два демонстрационных ролика, иллюстрирующих возможности сред виртуальной реальности [7].

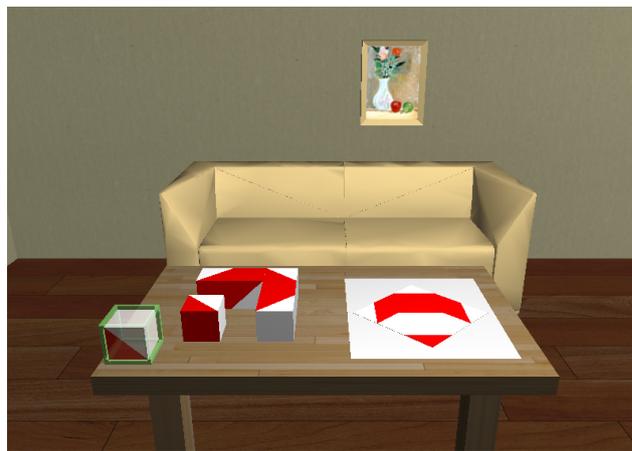


Рис. 1. Пример созданной для эксперимента виртуальной среды, позволяющей проходить тест «Кубики Коса» [7].

Данные показали, что виртуальная реальность не влияет на время решения заданий, однако влияет на правильность выполнения. В случае простых заданий влияние отрицательное. В экспериментальной группе чаще допускались ошибки. В случае же сложного задания, требующего обращения к пространственному воображению и выполнения операций мысленного вращения, виртуальная реальность даёт значительное преимущество. В экспериментальной группе не было допущено ошибок при выполнении усложнённого задания, тогда как в контрольной группе они были. Некоторые испытуемые в обеих группах не смогли даже ошибочно решить усложнённое задание. Большинство из них имели ошибочные стратегии решения, которые не смогли скорректировать и в итоге отказались от выполнения усложнённого задания. Однако только в экспериментальной группе некоторые испытуемые (10,3%) не смогли составить даже ошибочного плана решения и отказались после ряда хаотичных действий. В связи с этим была сделана гипотеза, что виртуальная реальность за счёт стереоскопического эффекта и создания объёмного изображения может способствовать операциям мысленного вращения, она также может и отрицательно влиять, скрадывая какие-то особенности изображения. Чтобы обеспечить производительность пользователей при проектировании визуальных систем необходимо учитывать особенности вывода и восприятия изображения в средах виртуальной реальности [7].

Переживание феномена присутствия определялось на основании субъективных самоотчетов, сделанных испытуемыми экспериментальной группы, и предложенного Витмером и Сингером опросника присутствия, «Presence Questionnaire» [5]. У части испытуемых в ходе решения выполнения теста было зафиксировано переживание феномена присутствия. Анализ показал, что феномен присутствия как таковой не оказал значимого влияния на время и правильность решения теста «Кубики Коса». Было выдвинуто предположение о том, что переживание феномена присутствия не будет оказывать влияние и на решение других интеллектуальных задач сходного типа и может не приниматься во внимание при проектировании виртуальных сред, предназначенных для сложной умственной деятельности [7].

На следующем этапе исследований тем испытуемым экспериментальной группы, которые указали на переживание феномена присутствия при выполнении теста, а также испытуемым контрольной группы, указавшим на переживание феномена присутствия при просмотре демонстрационных роликов, было предложено принять

участие в дополнительном эксперименте, цель которого состояла в изучении особенностей протекания переживания феномена присутствия. Испытуемым дали возможность «полетать» над виртуальным городом в условиях, максимально направленных на переживание участниками феномена присутствия. После этого они были опрошены с помощью структурированного интервью. В результате было описано протекание некоторых видов и уровней присутствия (средового, социального, личного присутствия). Отмечено различное влияние одних и тех же условий на разных людей, указавших на один и тот же уровень присутствия [7]. В ряде случаев у пользователей отмечались негативные эффекты, которые в литературе описываются, как «киберболезнь» (головокружение, тошнота, утомление и т.п.). (Интересно, что при экспериментальном использовании одной из моделей строчков, такие эффекты стали постоянными, но почти полностью исчезли при использовании очков очередной модификации.)

В настоящее время тест реализован на новой аппаратной и программной базе. Начат следующий этап исследований феномена присутствия в связи с решением интеллектуальных задач. Эти исследования становятся особенно актуальными в ходе разработки новых специализированных систем визуализации с использованием очков виртуальной реальности и с реализацией естественных (жестовых) интерфейсов. [6]. Средства виртуальной реальности применялись в ряде специализированных систем и примеров визуализации для представления вычислительных сеток, кристаллов, а также модели сердца, описанного в виде набора точек и сеток. При этом имели место естественная и абстрактная образности визуализации. Объем визуализируемых данных – 10×6 – 10×9 точек и выше. Появляется необходимость ориентироваться в сложно структурированных абстрактных пространствах. Для навигации в этих пространствах необходимо научиться находить нужное направление при перемещении. Наблюдения в целом подтвердили эффективность использования средств виртуальной реальности. Пользователи получают дополнительные возможности по изучению модельных сущностей и их особенностей при визуализации с помощью. Использование жестовых интерфейсов при работе в виртуальных средах значительно облегчает взаимодействие с объектами и навигацию в пространстве.

5. Изучение инсайта при визуализации

Получение инсайта рассматривается в качестве цели компьютерной визуализации, а его наличие, как один из основных критериев при оценке качества визуализации. Инсайт понимается как событие, при котором человек сразу же формулирует возникающую у него основную мысль. Рассматривается также немгновенный инсайт, при котором мысль формируется в течение некоторого времени как бы на наших глазах. [8] Задача получения инсайта определяет формы представления визуализации и временные ограничения при получении результата.

Наблюдения за пользователями систем визуализации позволяют говорить о нескольких типах поведения в ходе работы. Возможен быстрый просмотр изображения, позволяющий сделать те или иные выводы о сути представленной проблемы. В других случаях изучение «картинки» занимает много времени. Здесь также возможны различные варианты, например, отслеживание изменений в динамическом изображении, или своеобразное чтение визуального текста с анализом

деталей, в том числе, и в режиме взаимодействия с визуальными объектами. Кроме того, наблюдения показывают, что некоторыми пользователями визуализации результат был получен практически мгновенно. (В том числе решения задачи по кубикам Коса в рамках экспериментов со средами виртуальной реальности.) Можно заключить, что в этих случаях имел место «мгновенный» инсайт. Также ряд наблюдаемых случаев можно было описать в терминах «немгновенного» инсайта, когда пользователи формировали свои решения по мере вывода сложно структурированного визуального отображения.

По нашему мнению, инсайт не может рассматриваться как цель визуализации. Возможно его рассмотрение в качестве одного из средств получения результата – построения когнитивно-репрезентативных структур. Процесс визуализации можно рассматривать как построение визуального образа на основании абстрактных представлений об объекте. Эти абстрактные представления можно описать как модель исследуемого объекта, явления, или процесса, как-то связываемую с имеющимися у пользователя когнитивными структурами, описывающими данную сущность. Визуальные образы, представляющие моделируемую сущность, служат для того, чтобы создать или восстановить когнитивные структуры. Порождение когнитивных структур по визуальным образам и есть процесс интерпретации.

В тоже время нельзя не отметить важности изучения инсайта в связи с использованием сред виртуальной реальности и феноменом присутствия.

6. Феномен присутствия и эмоции

Как известно, среды виртуальной реальности очень активно используются в развлекательных комплексах для разнообразных демонстраций и игр. Полезно было бы обсудить ряд наблюдений, сделанных в связи с этим.

В случае неполного погружения, когда созданная компьютером среда окружает пользователя не со всех сторон, оставляя возможность наблюдать реальный мир, не создается даже частичного переживания присутствия до тех пор, пока наблюдатель не вовлекается в события среды. Однако те же люди указывают на полное переживание присутствия при вовлеченности. Возможна ситуация, когда при переживании полного зрительного погружения большая часть событий в виртуальной среде проходит вне сознания человека, не считая самого начала и самого конца интерактивной демонстрации [7].

В развлекательных комплексах предлагается набор игр с перемещениями в виртуальном пространстве и демонстраций, например, подводного мира. В играх важную роль играют хорошо реализованные виртуальные инструменты взаимодействия. Интересно, что даже имеющий достаточный опыт работы в средах виртуальной реальности и многократно наблюдавший случаи присутствия человек оказывается полностью или частично вовлеченным в виртуальный мир. Возможно возникновение чувства страха в связи с ситуациями и персонажами виртуального мира. При этом остается понимание нереальности мира, созданного виртуальной средой. То есть не было потери памяти, не было забывания, что перед глазами всего лишь разворачивается анимированная картинка. Человек может испытать сильное средовое присутствие и вести себя так, как будто эта среда реальна. Возможен страх из-за предметов, которые могли бы, нанести вред, будь ситуация реальна. Несмотря на самоконтроль, сила средового присутствия была весьма велика. Например, при «погружении на дно океана» напоминание самому себе, что ситуация не реальна, и человек не может дышать под водой, не помешало бояться проплывающих мимо медуз или кита.

Следует отметить, что инсайт зачастую тоже связан с эмоциональными переживаниями (в частности, определенный шок от открытия чего-то нового и важного, ощущение радости от успеха и т.п.).

7. Заключение

Таким образом, развитие компьютерных технологий порождает необходимость новых исследований, которые, в свою очередь, ставят новые вопросы перед исследователями. В настоящее время идет подготовка к очередному этапу экспериментальных исследований феномена присутствия на базе экспериментов с кубиками Коса. Новые исследования должны совместить методику исследования присутствия с помощью структурированного интервью и изучение интеллектуальной работы пользователей сред виртуальной реальности.

Отметим, что изучение феномена присутствия имеет значение не только при разработке систем научной визуализации, но и при создании тренажеров различного типа. В этом случае присутствие может играть как позитивную, так и негативную роль. Необходимо проектирование новой серии экспериментов для изучения феноменов виртуальной реальности в таких задачах.

8. Благодарности

Авторы выражают глубокую благодарность О.Е. Сурниной и А.А. Щербинину за активное участие в проектировании экспериментов и проведении исследований феномена присутствия.

9. Литература

- [1] Baker, M. P., Wickens, C. D. Human Factors in Virtual Environments for the Visual Analysis of Scientific Data // Technical Report, NCSA. 1995.
- [2] Bryson S., Levit C. The virtual windtunnel: an environment for the exploration of three-dimensional unsteady flows // VIS '91 Proceedings of the 2nd conference on Visualization '91. San Diego, California - October 22 - 25, 1991. Pp. 17-24.
- [3] Cruz-Neira C., Sandin D. J., DeFanti Th. A., Kenyon R. V., Hart J. C. The CAVE: audio visual experience automatic virtual environment // Communications of the ACM CACM, Volume 35 Issue 6, June 1992. Pp. 64-72.
- [4] Pettifer, S. An operating environment for large scale virtual reality. Unpublished doctoral dissertation, The University of Manchester. 1999.
- [5] <http://aig.cs.man.ac.uk/publications/papers/srp-phd.pdf>.
- [6] Witmer B.G., Singer M.J. Measuring Presence in Virtual Environments: A Presence Questionnaire // Presence. Vol. 7. No 3. 1998. P. 225–240.
- [7] Авербух В.Л., Анненкова О.Г., Артемова Н.А., Бахтерев М.О., Васёв П.А., Пестова М.С. Стародубцев И.С., Ушакова О.В., Разработка специализированной системы визуализации вычислительных сеток // Тезисы XVI Международной конференции Супервычисления и Математическое Моделирование». Тезисы. Саров. 2016, стр. 15-16.
- [8] Авербух Н.В., Щербинин А.А. Феномен присутствия и его влияние на эффективность решения интеллектуальных задач в средах виртуальной реальности // Психология Журнал Высшей школы экономики. Том 8, N 4, 2011, стр. 102-119.

- [9] Брушлинский, А.В. Субъект: мышление, учение, Воронеж. МОДЭК, Москва: Институт практической психологии, 1996.
- [10] Войскунский А. Е., Селисская М.А. Система реальностей: психология и технология. Вопросы философии 2005, N 11, стр. 119-130.
- [11] Филимонок Ю.И., Тимофеев В.И. Тест Векслера. Диагностика структуры интеллекта (взрослый вариант). Методическое руководство. СПб. «ИМАТОН», 2004.

Об авторах

Владимир Лазаревич Авербух – заведующий сектором компьютерной визуализации ИММ УрО РАН, доцент УрФУ. E-mail – averbukh@imm.uran.ru.

Наталья Владимировна Авербух – исследователь АНО ИНЭКС. E-mail – natalya_averbukh@mail.ru.

Приближенная оптимизация формы узла лопастей энергоустановки

Андреев С.В.¹, Бондарев А.Е.¹, Бондаренко А.В.², Визильтер Ю.В.², Галактионов В.А.¹, Гудков А.В.², Желтов С.Ю.², Жуков В.Т.¹, Иловайская Е.Б.², Князь В.А.², Мануковский К.В.¹, Новикова Н.Д.¹, Ососков М.В.², Силаев Н.Ж.², Феодоритова О.Б.¹

bond@keldysh.ru|vlgal@gin.keldysh.ru

¹Институт прикладной математики им. М.В.Келдыша РАН, Москва, Россия;

²ФГУП ГосНИИАС, Москва, Россия

Решается задача поиска оптимальной формы узла лопастей энергоустановки (ЭУ) с вертикальной осью. Рассматривается установка роторного типа с тремя изогнутыми лопастями. Ставится задача нахождения оптимальной формы лопастей с учетом ограничений на массу и момент инерции всего узла лопастей. Расчет целевой функции при заданной форме лопастей основан на численном решении трехмерной задачи обтекания в модели уравнений Навье-Стокса вязкого сжимаемого газа. В процессе таких расчетов определяется вращающий момент и силы, действующие на узел лопастей при обтекании его потоком воздуха с заданной скоростью. Для расчетов применялся пакет OpenFOAM. В работе приведен детальный анализ полученных результатов на основе разработанной оптимизационной технологии. В основу этой технологии положена параметризация конструкции энергоустановки тремя ключевыми параметрами. На дискретном множестве значений этих параметров решена задача поиска максимума двух целевых функций, определяющих аэродинамическое качество энергоустановки: модуля полной аэродинамической силы и модуля вращающего момента. В процессе работы создан виртуальный экспериментальный комплекс для моделирования процесса обтекания ЭУ на основе решения уравнений Навье-Стокса. Важной составной частью этого комплекса является визуализация представления результатов расчетов. Для визуального представления использовалась концепция прозрачной счетной области и рендеринга распределения газодинамических функций (давление, температура) по поверхности неподвижного узла лопастей энергоустановки в потоке воздуха. Для одновременной оценки процессов распределения газодинамических функций на поверхности лопастей и в счетной области использовалась комбинация рендеринга по поверхности лопасти и в поперечном сечении, рассекающем лопасть.

Ключевые слова: математическое моделирование, визуализация, оптимизационная задача, параметрический поиск.

Approximate Optimization of Shape of the Blades Assembly for Power Plant

Andreev S. V.¹, Bondarev A. E.¹, Bondarenko A. V.², Vizilter Y. V.², Galaktionov V. A.¹, Gudkov A. V.², Zheltov S. Y.², Zhukov V. T.¹, Ilovayskaya E. B.², Knyaz V. A.², Manukovskii K. V.¹, Novikova N. D.¹, Ososkov M. V.², Silaev N. Zh.², Feodoritova O. B.¹

bond@keldysh.ru|vlgal@gin.keldysh.ru

¹Keldysh Institute of Applied Mathematics RAS, Moscow, Russia;

²GOSNIAS, Moscow, Russia

The problem of finding the optimal shape of the blades assembly of a power plant with a vertical axis is solved. The power plant of a rotor type with three curved blades is considered. The task is to find the optimal shape of the blades, taking into account the constraints on the mass and inertia moment of the whole blade assembly. Calculation of the objective function for a given shape of the blades is based on the numerical solution of the three-dimensional flow problem in the model of the Navier-Stokes equations for a viscous compressible gas. In the process of such calculations, the torque and the forces acting on the blades assembly are determined when it is flowed by the airflow at a given speed. OpenFOAM software package was used for calculations. The paper provides a detailed analysis of the results obtained on the basis of the developed optimization technology. The basis of this technology is the parametrization of the design of the blades assembly by three key parameters. On a discrete set of values of these parameters, the problem of finding the maximum of two objective functions determining the aerodynamic quality of the power plant is solved, where the objective functions are the module of the total aerodynamic force and the torque module. In the process of work, a virtual experimental complex was created to simulate the process of flow around the blades assembly on the basis of solving the Navier-Stokes equations. An important integral part of this complex is the visualization of the results of calculations. For visual presentation the concept of a transparent calculation area and the rendering of the distribution of gasdynamic functions (pressure, temperature) over the surface of a stationary assembly of power plant blades in the air flow is used. For a simultaneous evaluation of the processes of distribution of gas dynamic functions on the surface of the blades and in the calculation area, a combination of the rendering along the blade surface and in the cross section cutting the blade is applied.

Keywords: mathematical modelling, visualization, optimization problem, parametric search.

1. Введение

Данная работа продолжает цикл исследований, посвященных разработке узла лопастей энергетической установки [1–4]. На предыдущих этапах работы была построена и протестирована вычислительная технология, представленная в работах [1–4] и включающая в себя построение CAD-модели узла лопастей, построение на основе этой модели расчетной сетки, расчеты обтекания,

визуализацию и анимацию результатов в моно- и стереорежимах.

В данной работе приведены результаты расчетных исследований по оптимизации формы узла лопастей энергетической установки (ЭУ) с вертикальной осью. Рассматривается ЭУ роторного типа [1–4] с тремя изогнутыми лопастями, форма которых характеризуется тремя параметрами. Ставится задача определения допустимого диапазона изменения этих параметров и нахождения приближенно оптимальной формы лопастей с

учетом ограничений на массу и момент инерции всего узла лопастей.

Строгое математическое решение оптимизационной задачи для нахождения экстремума заданной целевой функции, например, максимума вращающего момента узла лопастей – практически невозможно. Поэтому мы ограничились приближенным поиском экстремума (максимума) в дискретном пространстве трех ключевых параметров формы узла – двух угловых параметров лопасти и ее ширины. При этом возможное увеличение массы узла и его момента инерции не должно было значительно превышать аналогичные параметры для базового варианта формы узла. Расчет целевой функции при заданных ключевых параметрах основан на численном решении трехмерной задачи обтекания в модели уравнений Навье-Стокса вязкого сжимаемого газа. В процессе такого расчета определяется вращающий момент и силы, действующие на узел лопастей при обтекании его потоком воздуха с заданной скоростью. Для расчетов применялся пакет OpenFOAM (решатель rhoCentralFoam) [7]. Представленные расчеты проводились на вычислительном кластере К-100 в ИПМ им. М.В. Келдыша РАН [5] в режиме параллельных вычислений.

2. Технология расчетов

Построение численной технологии включает в себя ввод геометрической информации, ее обработку, построение сеточной модели, выбор алгоритмов расчета обтекания, расчет функционалов (сил и моментов), визуализацию, обработку результатов. Основой технологии является процесс математического моделирования узла лопастей ЭУ под заданной ветровой нагрузкой на основе уравнений Навье-Стокса. Следующим важным этапом является создание методики расчетов для выбора оптимальной формы лопастей, чему и посвящена данная работа. Для расчетов обтекания ЭУ строится геометрическая сетка, подготовка которой начинается с описания поверхности, ограничивающей трехмерное тело. Эта поверхность импортируется из CAD-пакета и может быть отредактирована, т.е. для получения объемной сетки нужного качества проводится дополнительная подготовка поверхности с использованием автоматических инструментов создания и редактирования поверхностных сеток, включая называемые в англоязычной литературе процедуры «surface wrapper» и «surface remesher». Первая процедура исправляет дефекты CAD-геометрии, создавая на выходе замкнутую поверхность желаемой детализации. Вторая процедура создает треугольную поверхностную сетку с контролем качества (гладкость в областях с большой кривизной, разрешение тонких мест, рост размера ячеек при удалении от областей с высокой детализацией, сохранение топологии объекта, локальное измельчение сетки и др.). На рис. 1 приведен пример исходной поверхностной «CAD-сетки с ошибками» (слева) и «сетки без ошибок» (справа).



Рис. 1. Пример поверхностной сетки: исходная CAD-поверхность с ошибками (слева) и исправленная треугольная сетка (справа).

На основе результирующей поверхностной сетки строится объемная расчетная сетка. В задачах обтекания

такая сетка, как правило, состоит из двух основных частей: призматической вблизи обтекаемых поверхностей и произвольной многогранной сетки на удалении от поверхностей. Для создания пристеночного призматического слоя с заданными параметрами используется последовательное наращивание призматических слоев. При этом допускается модификация текущего слоя, что обеспечивает создание призматического слоя, проникающего дальше от стенок вглубь расчетной области. Для заполнения внутреннего объема вне пристеночных слоев могут выбираться тетраэдры, гексагональные и многогранные ячейки. Пакет OpenFOAM в общем случае позволяет проводить расчеты на сетках из произвольных многогранных ячеек. В представленных расчетах применялись сетки с ячейками в виде треугольных призм в пристеночных слоях и тетраэдрами вне этих слоев.

3. Задача поиска оптимальной формы

Расчет обтекания ЭУ основан на решении уравнений Навье-Стокса [6]:

$$\frac{\partial p}{\partial t} + \nabla \cdot (\rho \bar{u}) = 0$$

$$\frac{\partial (\rho \bar{u})}{\partial t} + \nabla \cdot (\rho \bar{u} \otimes \bar{u}) = -\nabla p + \nabla \cdot \tau$$

$$\frac{\partial (\rho E)}{\partial t} + \nabla \cdot (\rho \bar{u} H) = \nabla \cdot [\bar{u} \cdot \tau + \bar{q}]$$

Здесь \bar{u} – вектор скорости осредненного течения с компонентами (u, v, w) ; ρ – плотность, p – давление, τ – тензор вязких напряжений; $E = e + |\bar{u}|^2 / 2$ – удельная полная энергия газа, e – удельная внутренняя энергия газа; $H = E + p / \rho$ – полная энтальпия; \bar{q} – вектор плотности теплового потока. Система уравнений дополняется уравнением состояния идеального газа: $p = (R / M) \rho T$, M – молекулярный вес, R – универсальная газовая постоянная, $e = C_v T$, C_v – удельная теплоемкость при постоянном объеме, T – температура. Тензор вязких напряжений и вектор теплового потока имеют вид $\tau = 2\mu(S - 1/3 I \nabla \cdot \bar{u})$, $\bar{q} = -\lambda \nabla T$, где $S = 0.5(\nabla \bar{u} + [\nabla \bar{u}]^T)$ – тензор скоростей деформации, $\mu(T)$, $\lambda(T)$ – коэффициенты динамической вязкости и теплопроводности. В данной работе влияние турбулентности не рассматривалось.

Работа энергетической установки с вертикальной осью вращения основана на возникновении вращающего момента за счет действия набегающего потока воздуха на лопасти. Благодаря высокому аэродинамическому качеству лопастей возникает подъемная сила, превышающая силу сопротивления, что в итоге приводит к появлению вращающего момента. Поэтому в качестве основных интегральных характеристик выступает полная аэродинамическая сила и вращающий момент.

Для визуального представления использовалась концепция прозрачной счетной области и рендеринга распределения газодинамических функций (давление, температура) по поверхности неподвижного узла лопастей энергоустановки в потоке воздуха. Для одновременной оценки процессов распределения газодинамических функций на поверхности лопастей и в счетной области использовалась комбинация рендеринга по поверхности лопасти и в поперечном сечении, рассекающем лопасть.

Анимированное представление реализовывалось с помощью подхода, названного «облет камеры» вокруг неподвижного узла лопастей в потоке. Для неподвижного узла лопастей подобный подход дает возможность контролировать распределение газодинамических функций во всех точках объекта, вовремя выявлять появление расчетных дефектов и отлаживать схему организации расчета. Кроме того, нашей задачей являлось создание стереофильма, позволяющего продемонстрировать результаты расчетов (создание последовательности стереопар, каждая из которых является стереокадром фильма), и метод «облета камеры» вокруг объекта является наиболее подходящим для этих целей.

Практически любая CAD-система позволяет визуализировать трехмерный объект с заданными координатами положения по отношению к пользователю. Поэтому задача создания стереофильма сводилась к построению последовательности изображений CAD модели с зафиксированной осью Z, проходящей через ее геометрический центр, и смещением на определенный угол. После каждого такого смещения проводился рендеринг модели, позволяющий достичь фотореалистичности, и сохранение результата в виде графического файла в высоком разрешении.

Одной из важнейших задач моделирования является поиск оптимальной формы узла лопастей ЭУ. Нас интересует такая форма изделия, которая обеспечивает максимальный вращающий момент при варьировании трех параметров узла – двух углов наклона лопастей и ширины каждой лопасти в предположении, что все три лопасти идентичны. На рис. 2 схематически показаны эти параметры: γ_1 – угол между направлением лопасти и осью вращения, γ_2 – угловой размер лопасти, L – ширина лопасти. Заметим, что всюду ниже результаты приводятся для тройки параметров (γ'_1, γ_2, L) , где $\gamma'_1 = 90^\circ - \gamma_1$. Параметры (γ'_1, γ_2, L) меняются в некотором диапазоне. В качестве базового набора выбран результат сканирования опытного образца: $(\gamma'_1, \gamma_2, L)_{\text{баз}} = (55^\circ, 120^\circ, 20 \text{ см})$. При вариациях γ'_1 принимает значения $45^\circ, 50^\circ, 55^\circ, 60^\circ, 65^\circ$; γ_2 – значения $100^\circ, 110^\circ, 120^\circ, 130^\circ$; ширина лопасти L выбирается из набора $0.8L_{\text{баз}}, 0.9L_{\text{баз}}, L_{\text{баз}}, 1.1L_{\text{баз}}, 1.2L_{\text{баз}}$.

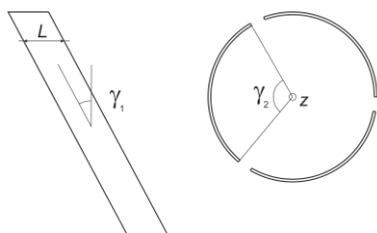


Рис. 2. Варьируемые параметры (γ_1, γ_2, L) .

Дополним схематическую картину, представленную на рис. 2, трехмерными изображениями, см. рис. 3–4. Рисунки представляют изменение формы узла лопастей при вариации вышеописанных параметров. Базовый вариант отмечен красным цветом.

Рассмотрим задачу условной оптимизации: найти значения параметров $x^* = ((\gamma'_1)^*, \gamma_2^*, L^*) = \text{argmax } M(\gamma'_1, \gamma_2, L)$ при выполнении ограничений на x^* . В качестве целевой функции M выбирается полная аэродинамическая сила \vec{F} или вращающий момент \vec{M} , а в качестве ограничений – условия $45^\circ \leq \gamma'_1 \leq 65^\circ, 100^\circ \leq \gamma_2 \leq 130^\circ, 0.8L_{\text{баз}} \leq L \leq 1.2L_{\text{баз}}$.

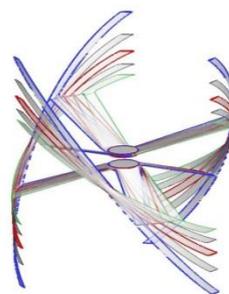


Рис. 3. Влияние γ'_1 на форму узла лопастей: 45° (зеленый), 50° (серый), 55° (красный), 60° (серый), 65° (синий).

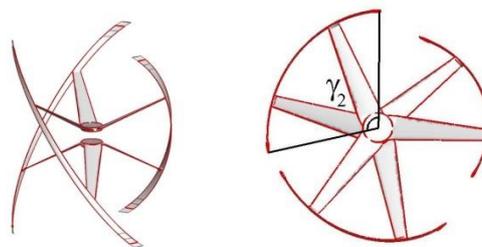


Рис. 4. Влияние γ_2 на форму узла лопастей.

Аналитической связи между целевой функцией и вектором независимых переменных x^* не существует, эту связь устанавливает трехмерная система уравнений Навье-Стокса. С формальной точки зрения задача поиска x^* сводится к задаче нелинейного программирования. Для приближенного решения этой задачи воспользуемся сеточным методом, определяя значений целевой функции на сетке узлов – дискретном наборе значений параметров γ'_1, γ_2, L в заданном диапазоне. Наибольшее значение целевой функции на данной сетке принимается за искомым экстремум. Предварим приближенную оптимизацию процедурой параметрического поиска для оценки направления изменения параметров. Для этого последовательно проварьируем отдельно каждый из параметров, т.е. исследуем три направления:

$$\begin{aligned} P_1 &= (\gamma'_1(N), (\gamma_2)_{\text{баз}}, L_{\text{баз}}); \\ P_2 &= ((\gamma'_1)_{\text{баз}}, \gamma_2(N), L_{\text{баз}}); \\ P_3 &= ((\gamma'_1)_{\text{баз}}, (\gamma_2)_{\text{баз}}, L(N)), \end{aligned}$$

где N – номер варианта из допустимого диапазона. Для каждого направления точки пронумерованы по возрастанию параметра. В результате проведения расчетов было установлено, что при увеличении параметров γ'_1, γ_2, L от значений базового варианта происходит увеличение, как модуля вращающего момента, так и модуля аэродинамической силы. Оценим возможное увеличение этих функций. Сначала зафиксируем ширину лопасти $L = L_{\text{баз}}$ и проварьируем γ'_1 и γ_2 в диапазонах их изменения. На рис. 6 показана зависимость модуля аэродинамической силы $F(\gamma'_1, \gamma_2)$ и модуля вращающего момента $M(\gamma'_1, \gamma_2)$ от вариации углов γ'_1 и γ_2 .

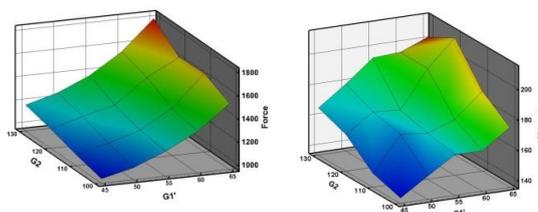


Рис. 6. $F(\gamma'_1, \gamma_2)$ (слева) и $M(\gamma'_1, \gamma_2)$ (справа); $L = L_{\text{баз}}$.

Анализ полученных результатов говорит, что в рассматриваемом диапазоне изменения γ_1' и γ_2 при $L = L_{\text{баз}}$ максимальный вращающий момент достигается при максимальных значениях γ_1' и γ_2 , что обеспечивает увеличение вращающего момента на 28%. При вариации ширины лопасти и фиксированных значениях параметров $(\gamma_1')_{\text{баз}}$ и $(\gamma_2)_{\text{баз}}$ наблюдается почти линейное увеличение вращающего момента. Так рост L на 20% приводит к росту вращающего момента на 16%. Одновременный рост всех трех параметров в принятых диапазонах приводит к росту момента инерции на 50% по сравнению с базовым вариантом геометрии. Считая постоянной плотность материала лопастей, учет ограничений по массе можно вести с помощью контроля объема узла лопастей. Все дальнейшие рассмотрения проводятся для слоя данных с максимальной шириной лопасти $L = 1.2L_{\text{баз}}$. На рис. 7 представлена зависимость модуля аэродинамической силы (слева) и модуля вращающего момента (справа) от γ_1' и γ_2 для $L = 1.2L_{\text{баз}}$.

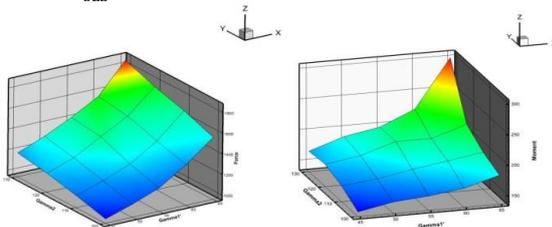


Рис. 7. $F(\gamma_1', \gamma_2)$ (слева) и $M(\gamma_1', \gamma_2)$ (справа); $L = 1.2L_{\text{баз}}$.

На рис. 8 (слева) показаны объемы узла лопастей $V(\gamma_1', \gamma_2)$ в зависимости от значений угловых параметров при фиксированном значении $L = 1.2L_{\text{баз}}$. Плоскость, ограниченная красным, соответствует значению объема для базового варианта $V = V_{\text{баз}}$. Соответственно, плоскость, ограниченная синим, отвечает значению объема $V = 1.1V_{\text{баз}}$. Линии пересечения обеих плоскостей с поверхностью значений объемов ограничивают вариацию формы изделия. На рис. 8 (справа) показана область допустимых значений при учете ограничений по объему на плоскости вариации угловых параметров: она заключена в области, ограниченной толстыми красной $V = V_{\text{баз}}$ и синей $V = 1.1V_{\text{баз}}$ линиями.

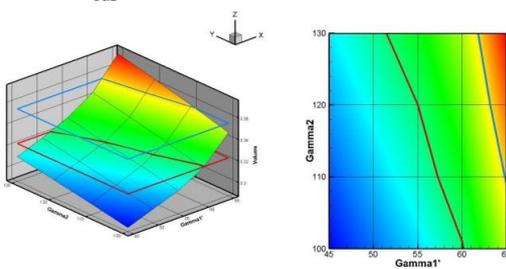


Рис. 8. Объем узла лопастей при $L = 1.2L_{\text{баз}}$ (слева) и область ограничения по объему (справа).

Аналогичным образом определяются ограничения для момента инерции. Они показаны на рис. 9; слева изображен момент инерции в виде трехмерной поверхности $M_i(\gamma_1', \gamma_2)$. Плоскость, ограниченная красным, соответствует значению $M = M_{\text{баз}}$, а ограниченная синим – $M = 1.1M_{\text{баз}}$. Линии пересечения обеих плоскостей с поверхностью значений моментов инерции ограничивают вариацию формы изделия. На рис. 9 (справа) диапазон допустимых значений представлен на плоскости вариации угловых параметров в виде области между толстыми красной и синей линиями.

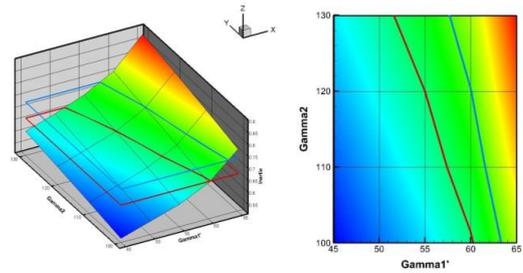


Рис. 9. Момент инерции (слева) и область ограничения по моменту инерции (справа) при вариации углов, $L = 1.2L_{\text{баз}}$.

Таким образом, для заданного слоя данных, соответствующих $L = 1.2L_{\text{баз}}$, мы получили ограничения по объему (рис. 8 справа) и моменту инерции (рис. 9 справа). Для совместного учета этих ограничений изобразим их на одном рисунке и выберем диапазон вариации углов, соответствующий наиболее жесткому ограничению, см. рис. 10, где на плоскости вариации угловых параметров совмещены ограничения по объему (красные линии) и моменту инерции (пунктирные голубые линии). Как видно на рисунке, нижние ограничения по объему и моменту инерции практически совпадают, а верхние существенно разнятся. Очевидно, что ограничения по моменту инерции являются более жесткими, чем ограничения по объему. В качестве наиболее жесткого ограничения при суммарном учете ограничений по объему и моменту инерции необходимо выбрать диапазон вариации угловых параметров, заключенный между пунктирными голубыми линиями.

Оценим влияние полученных при совместном учете ограничений на вращающий момент, который мы рассматриваем в настоящем исследовании как основную силовую характеристику узла лопастей.

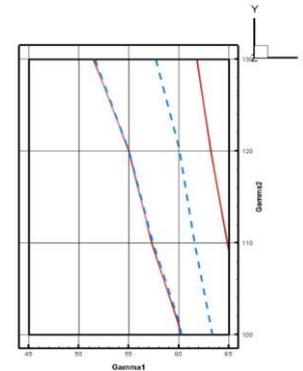


Рис. 10. Совместный учет ограничений по объему и моменту инерции при вариации углов, $L = 1.2L_{\text{баз}}$.

Для этой цели на рис. 11 показано совмещение поверхности вращающего момента (см. рис. 9, справа) при вариации углов и полученного диапазона ограничений (см. рис. 10). Распределение вращающего момента на рис. 11 (слева) представлено трехмерной поверхностью, а ограничения представлены вертикальными поверхностями, отсекающими нужный диапазон на поверхности момента.

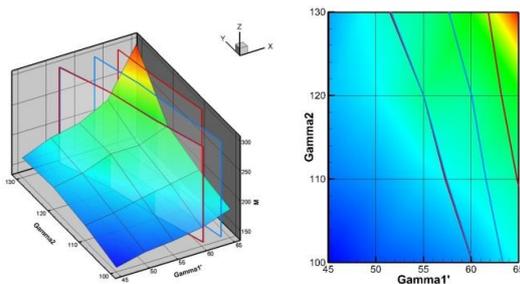


Рис. 11. Момент инерции с учетом ограничений при вариации угловых параметров в 3D (слева) и 2D (справа).

Для наглядности на рис. 11 (справа) представлены данные в двумерном виде. Это представление даст возможность более точно определить диапазон вариации угловых параметров, обеспечивающий максимум вращающего момента с учетом наложенных ограничений на изменение объема и момента инерции узла лопастей. Можно утверждать, что при учете выбранных ограничений по объему и моменту инерции нужный диапазон заключается в пределах вариации угла γ_1' от 55° до 60° , а угла γ_2 – от 120° до 125° . В итоге выделен диапазон геометрических параметров, обеспечивающих увеличение вращающего момента и найдены приближенно оптимальные параметры.

4. Заключение

Выполнены расчетные исследования по определению оптимальной формы узла лопастей с учетом возможных ограничений на конструкцию, проведен детальный анализ полученных результатов на основе разработанной оптимизационной технологии. В основу этой технологии положена параметризация конструкции тремя ключевыми параметрами. На дискретном множестве значений этих параметров решена задача поиска максимума двух целевых функций, определяющих аэродинамическое качество энергоустановки: модуля полной аэродинамической силы и модуля вращающего момента. В процессе работы создан виртуальный экспериментальный комплекс для моделирования процесса обтекания ЭУ на основе решения уравнений Навье-Стокса. Следует отметить, что визуализация представления результатов расчетов является важнейшей частью созданного комплекса, обеспечивая не только контроль и верификацию результатов на всех этапах работы вычислительной технологии, но и решение оптимизационной задачи с рядом ограничивающих условий.

5. Благодарности

Данная работа выполнена при поддержке грантов Российского фонда фундаментальных исследований (проекты 16-01-00553а и 17-01-00444а).

6. Литература

[1] Андреев С.В., Бондарев А.Е., Бондаренко А.В., Визильтер Ю.В., Галактионов В.А., Гудков А.В., Желтов С.Ю., Жуков В.Т., Иловайская Е.Б., Князь В.А., Мануковский К.В., Новикова Н.Д., Ососков М.В., Силаев Н.Ж., Феодоритова О.Б., Бондарева Н.А. *Моделирование и визуализация работы узла лопастей сложной формы в энергетической установке* // Журнал «Научная визуализация», 2015, т. 7, № 4, С. 1-12.

- [2] Андреев С.В., Бондарев А.Е., Бондаренко А.В., Визильтер Ю.В., Галактионов В.А., Гудков А.В., Желтов С.Ю., Жуков В.Т., Иловайская Е.Б., Князь В.А., Мануковский К.В., Новикова Н.Д., Ососков М.В., Силаев Н.Ж., Феодоритова О.Б., Бондарева Н.А. *Моделирование и визуализация работы энергетической установки сложной формы в режиме стереоанимации* // Труды 25-й Международной Конференции по Компьютерной Графике и Зрению ГрафиКон2015, Протвино, Россия, 22-25 сентября 2015 г., с. 183-187.
- [3] Андреев С.В., Бондарев А.Е., Бондаренко А.В., Желтов С.Ю., Жуков В.Т., Иловайская Е.Б., Мануковский К.В., Новикова Н.Д., Ососков М.В., Феодоритова О.Б. *Моделирование и визуализация работы энергетической установки сложной формы* // Математическое моделирование, 2016. 28:9, 125–136.
- [4] Бондарев А.Е., Жуков В.Т., Мануковский К.В., Новикова Н.Д., Феодоритова О.Б. *Разработка и организация математического моделирования обтекания неподвижной лопатки энергетической установки.* – Препринт № 60 ИПМ им. Келдыша, 2014
- [5] Гибридный вычислительный кластер К-100. URL: <http://www.kiam.ru/MVS/resources>
- [6] Ландау Л.Д., Лившиц Е.М. *Гидродинамика.* – Москва, Наука, 1986.
- [7] Программный комплекс OpenFOAM. URL: <http://www.openfoam.com>

Алгоритм обработки разнородных данных для мультисенсорной СТЗ на примере анализа температуры и концентрации газа

Р.А. Багутдинов¹, С.Г. Небаба¹, А.А. Захарова¹
ravil_bagutdinov@yahoo.com|stepanlfx@tpu.ru|zaa@tpu.ru
¹Томский Политехнический Университет, Томск, Россия

Решается прикладная задача определения тепловых потерь газа в газовой смеси в замкнутой области за счет теплопроводности и теплового излучения по данным датчиков температуры и концентрации. Предложен алгоритм комплексного анализа разнородных и разномасштабных данных, полученных в режиме реального времени при условии, что данные синхронизированы по времени на всех датчиках в составе мультисенсорной системы, позволяющий использовать их для последующего графического и статистического анализа. Проведен эксперимент с данными газоанализатора и соответствующими показаниями тепловизора внутри определенной кубической комнаты. Результаты эксперимента подтверждают эффективность предложенного алгоритма.

Ключевые слова: разнородные данные, разномасштабные данные, датчики, робототехнические комплексы, системы технического зрения, мультисенсорные системы.

Algorithm for processing heterogeneous data for multisensory vision system using the example of temperature and gas concentration analysis

R.A. Bagutdinov¹, S.G. Nebaba¹, A.A. Zakharova¹
ravil_bagutdinov@yahoo.com|stepanlfx@tpu.ru|zaa@tpu.ru
¹Tomsk Polytechnic University, Tomsk, Russia

An applied problem is solved for determining the thermal losses of gas in a gas mixture in a closed area due to thermal conductivity and thermal radiation from temperature and concentration sensors data. An algorithm for the complex analysis of heterogeneous and multiscale data obtained in real time is proposed, provided that the data are synchronized in time on all sensors in the multisensory system, which makes it possible to use them for subsequent graphical and statistical analysis. An experiment was performed with the data of the gas analyzer and the corresponding readings of the thermal imager inside a certain cubic room. The results of the experiment confirm the effectiveness of the proposed algorithm.

Keywords: heterogeneous data, multiscale data, sensors, robotic complexes, vision systems, multisensory systems.

1. Введение

В робототехнических комплексах, как и в физическом мире, основным видом осязания является зрение. Первое устойчивое применение в робототехнических комплексах системы технического зрения (СТЗ) получили в системах управления манипуляторами и мобильными роботами человеком-оператором. Следующим шагом в развитии таких систем стали системы супервизорного управления. Они оказались особенно эффективны для телеуправления роботами при значительном временном запаздывании в канале связи, когда становится затруднительным управление оператором в реальном времени, а также для работы в условиях плохой видимости [1]. Последний случай характерен, например, для подводных роботов, работающих вблизи дна, когда движение робота вызывает подъем осадков и замутнение воды; в чрезвычайных случаях в условиях задымленности, при загазованности рабочего пространства на производстве; других обстоятельствах, препятствующих видимости [3]. Поэтому применяемые в этих случаях супервизорное управление сводится к периодическому заданию оператором очередной выполняемой автоматически элементарной операции в моменты восстановления видимости. В подобных случаях целесообразно использовать мультисенсорные СТЗ, где акцентируется внимание на качестве получаемых данных. Успешное решение этой проблемы приведет к существенному прогрессу в широком спектре прикладных задач за счет повышения точности, эффективности и надежности таких систем.

2. Алгоритм оценки тепловых потерь газа по данным двух датчиков

Системы управления робототехническими средствами на основе технического зрения и сбор информации для последующего анализа передвижения роботов имеют особое значение в роботостроении [13]. Поскольку определение движения и создание картины структуры окружающей среды является неотъемлемой частью человеческого зрения, то одной из ключевых задач технического зрения будет реализация этой врожденной способности средствами компьютерных систем. Теоретические методы данного исследования основываются на методах цифровой обработки изображений, распознавания образов, дискретных преобразований и системного анализа [2].

Одной из актуальных задач, которые могут быть решены с привлечением методов СТЗ и использованием специализированных датчиков, является определение температуры и концентрации газов в закрытом помещении.

Определение источника утечки ряда газов имеет высокое значение при ликвидации аварий, в рассматриваемом случае поставлен эксперимент, в ходе которого проводились замеры концентрации диоксида азота и температуры газовой смеси. Диоксид азота применяется при производстве серной и азотной кислот, а также используется в качестве окислителя в жидком ракетном топливе на космодромах и смесевых взрывчатых веществах, на химических предприятиях. Экспериментальные данные по концентрации NO₂ и

температуре получены в условном ограниченном пространстве.

Разнородные и разномасштабные данные, полученные с газоанализатора, содержат информацию о составе газовой смеси различных газов, количество этих данных увеличивается с увеличением количества самих датчиков, в результате для рассмотренного эксперимента данные концентрации газов получены в каждой из десяти ключевых точек. В свою очередь данные, полученные с тепловизора, имеют графическую информацию и информацию о характере изменения температуры в ключевых точках.

Выдвигается следующая гипотеза: данные, полученные с газоанализатора, должны иметь линейную зависимость от распределения тепла, что следует из свойств газа, на физическое состояние которого прямо влияют его температурные показатели. В перспективе использование разнородных данных с различных датчиков предоставляет возможность выявлять более сложные зависимости между данными и за счет этого повышать качество и точность получаемой информации о температуре и составе газовой смеси.

На основе сравнения полученных экспериментальных данных предложен алгоритм определения тепловых потерь газа в газовой смеси произвольного состава в замкнутой области за счет теплопроводности и теплового излучения по данным датчиков температуры и концентрации.

Этапы алгоритма определения тепловых потерь газа по данным двух датчиков:

1. Получение двух массивов данных концентрации и температуры газа в замкнутой области $M*N$ через равные промежутки времени t .
2. Приведение данных к единой шкале соотношения концентрации газа и температуры за счет аппроксимации массива данных газоанализатора.
3. Вычисление функции корреляции от двух числовых последовательностей.
4. Определение среднего значения корреляции в единицу времени t_i .
5. Полученное значение корреляции имеет линейную зависимость от коэффициента теплопередачи газов в газовой смеси: в случае, если известный газ в начальный момент времени t_0 имеет большую температуру, чем газовая смесь, в которую он был помещен, то это значение характеризует тепловые потери; если его температура была ниже температуры газовой смеси – значение характеризует приобретенное за счет теплообмена и излучения газовой смеси тепло.

В областях с недостаточными данными от газоанализатора проводилась аппроксимация общепринятой центральной разностной схемой типа «крест», и решение в них определялось итерационным методом Гаусса–Зейделя. В данном случае аппроксимация используется для того, чтобы привести данные к единому общему масштабу. Ключевые значения данных с газоанализатора сравнивались с картинкой, полученной с тепловизора путем наложения изображений.

Для вычисления коэффициента корреляции (пп. 3-4 алгоритма) двух нормированных матриц, описывающих концентрацию и температуру, применима следующая формула:

$$R(x, y) = \frac{\sum_{x', y'} (C(x', y') \cdot T(x + x', y + y'))}{\sqrt{\sum_{x', y'} C(x', y')^2 \cdot \sum_{x', y'} T(x + x', y + y')^2}} \quad (1)$$

где C – матрица значений концентрации, T – матрица значений температуры. Оценка, получаемая по формуле (1), лежит в диапазоне от 0 до 1, где 0 – отсутствие корреляции данных, а 1 – полное совпадение данных. Граничное значение 0 в данном случае определяется характером обработки цифровых изображений, которые лежат в диапазоне от 0 до 255 для случая 1-байтового кодирования, а граничное значение 1 – выбором нормированной корреляции для наглядности получаемого результата.

Применяя операцию вычисления $R(x, y)$ по формуле (1) для двух разномасштабных матриц, можно получить изображение в градациях серого, в котором наиболее яркое пятно соответствует наиболее похожему участку двух матриц (рис. 1).

В случае матриц одинакового масштаба матрица $R(x, y)$ является матрицей, состоящей из одного элемента, который будет искомым коэффициентом корреляции концентрации и температуры, определяющим тепловые потери анализируемого газа.

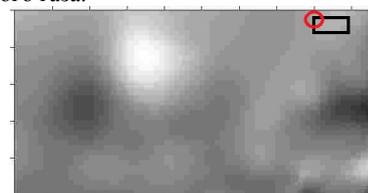


Рис. 1. Результат вычисления матрицы корреляции для двух нормированных массивов данных.

Существующие методы вычисления теплопередачи различных газов оперируют лишь информацией, полученной от одного датчика, и априорными знаниями (такими как состав газовой смеси, теплоемкость, долевое отношение компонентов). В отличие от них, разработанный алгоритм использует потоки разнородной информации с двух синхронизированных по времени датчиков и позволяет сделать заключение о характеристиках анализируемого газа уже на основании статистического анализа в режиме реального времени.

Такой алгоритм может быть использован также для определения неизвестного газа путем смешения с известным газом и мониторинга их температур и диффузии различными датчиками в составе мультисенсорной системы.

3. Эксперимент и анализ результатов

В рамках данной работы проведен эксперимент по определению зависимости между данными тепловизора и газоанализатора в режиме реального времени.

Ниже представлен план проведенного эксперимента.

1. Синхронизация и калибровка двух датчиков (газоанализатора и тепловизора).
2. Определение типа данных и объема выборки.
3. Получение данных с газоанализатора и тепловизора.
4. Выполнение 10 замеров в 50 интервалах времени для отслеживания динамики изменений состояния.
5. Группировка полученных данных по временным срезам.
6. Выведение полученных результатов эксперимента и сравнение с результатами тепловизора на основе предложенного алгоритма (разброс температуры от 25 °C до 140 °C, соответствующий газовому состоянию NO_2).

При расчетах использовались данные с газоанализатора для научных исследований «Тест 1-6» (ООО "БОНЭР") для измерения отдельных компонентов газовой смеси при

длительном непрерывном измерении относительно высоких концентраций. Устройство оснащено опτικο-абсорбционным инфракрасным и электрохимическим сенсорами и позволяет получать данные в режиме реального времени, архивировать данные и экспортировать их в стандартные программы.

Все датчики статические, один тепловизор и десять датчиков с газоанализатора. Тепловизор подвешен в центре комнаты под потолком, для максимального охвата комнаты. Датчики газоанализатора размещены равномерно по всей поверхности потолка.

Интервал измерений 15 сек. Диапазон измерений NO_2 - 1000 ppm. Приведенная погрешность измерения (кроме O_2) - 5%; по O_2 - 2%.

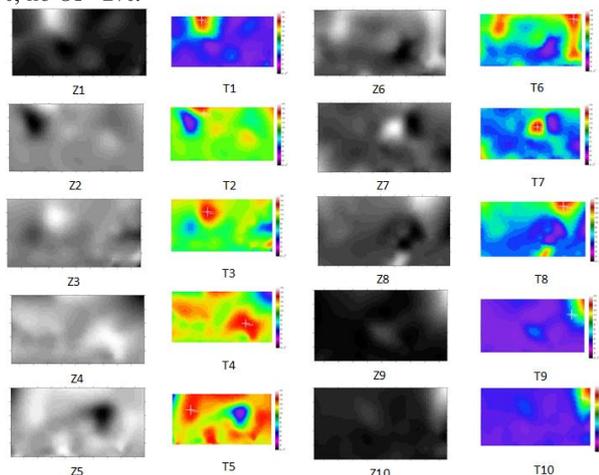


Рис 2. Сравнительный графический анализ (чёрно-белые изображения Z1-Z10 – по данным газоанализатора, цветные T1-T10 – по данным тепловизора).

На рисунке 2 можно проследить динамику изменения концентрации газа и сравнить с результатами, полученными с тепловизора. На изображениях Z1-Z10 представлены в графическом виде данные, полученные с газоанализатора, на изображениях T1-T10 – данные с тепловизора, номера соответствуют 10 интервалам времени, отражающие характерные особенности изменения концентрации NO_2 . На рисунке видно, что данные с тепловизора коррелируют с данными, полученными с газоанализатора, можно проследить динамику распространения газа в комнате. На рисунках Z2T2-Z5T5 видно, как распределение газа влияет на повышение температуры в комнате, на рисунках Z3T3-Z8T8 заметны турбулентные течения, вызванные смещением источника газа в направлении выхода из комнаты в правом верхнем углу, на рисунках Z6T6-Z10T10 прослеживается снижение концентрации газа и сопутствующая нормализация температуры.

Для статистического представления данных использовались стандартные методы ПО Surfer. По данным, полученным в ходе эксперимента выявлена прямая корреляция показаний датчиков температуры и концентрации, что свидетельствует о диффузионном процессе, близком к изотермическому. Согласно выдвинутой гипотезе, что коэффициент корреляции между показаниями датчиков концентрации и температуры отражает потери NO_2 путем теплопроводности и теплового излучения в газовой смеси, можно заключить, что эти тепловые потери равны 12% от исходного количества теплоты.

Полученные в ходе эксперимента теоретические результаты и формулируемые на их основе выводы подтверждаются строгостью математических выкладок,

базирующихся на аппарате интегрального и дифференциального исчисления, теории вероятностей и математической статистики. Справедливость выводов относительно эффективности предложенной системы подтверждена статистическим моделированием и опытно-методической обработкой реальных результатов.

Верификация имитационных моделей проводилась экспериментально.

4. Заключение

Предложен алгоритм обработки разнородных данных в решении прикладной задачи определения тепловых потерь газа в газовой смеси в замкнутой области за счет теплопроводности и теплового излучения по данным датчиков температуры и концентрации. Достоверность и точность результатов основываются на аппроксимации разреженных данных и анализе взаимной корреляции комплексных данных мультисенсорной системы.

Разработан алгоритм, использующий потоки разнородной информации с двух синхронизированных по времени устройств, который позволяет сделать заключение о характеристиках анализируемого газа уже на основании статистического анализа в режиме реального времени, в отличие от существующих традиционных методов, которые оперируют информацией, полученной от одного датчика, и априорными знаниями.

Выявлено, что коэффициент корреляции между показаниями датчиков концентрации и температуры отражает потери NO_2 путем теплопроводности и теплового излучения в газовой смеси, эти тепловые потери равны 12% от исходного количества теплоты для данных, полученных в ходе эксперимента. Это подтверждает предложенную гипотезу о возможности вычисления тепловых потерь по данным датчиков температуры и концентрации газовой смеси.

Если результаты анализа показывают слабую корреляцию на уровне областей, т.е. возникает аномалия соотношения температуры и концентрации газа, то можно говорить о том, что в этих точках, вероятно, произошла утечка газа, изменились внешние условия среды или граничные условия. В этом случае можно в режиме реального времени, не дожидаясь критического увеличения концентрации газа и его распространения в комнате, оперативно определить пространственную локализацию источника, принять решения по её ликвидации и поднять тревогу.

Результаты исследования применимы в сфере мониторинга и обработки разнородных данных для повышения эффективности выполняемых работ с помощью СТЗ. Данные исследования могут быть использованы при моделировании ситуаций, требующих быстрого реагирования, таких как: осуществление ликвидации аварий, моделирование эвакуации людей из зданий в чрезвычайных ситуациях, моделирование ситуаций при террористических атаках.

5. Благодарности

Работа выполнена в рамках госзадания № 2.1642.2017/4.6 на выполнение проекта по теме «Когнитивные методы визуализации и анализа многомерных данных при моделировании нелинейных динамических систем».

6. Литература

- [1] Адилев Р.М. Исследование и разработка методов анализа многоградационных растровых изображений в

- системах технического зрения: Дис. канд. техн. наук : 05.13.17/ Пенза, 2005.
- [2] Багутдинов Р.А. Подход к алгоритмизации адаптивных систем технического зрения, применяемых в робототехнике / В сборнике: Фундаментальные и прикладные научные исследования: актуальные вопросы, достижения и инновации. Сборник статей победителей IV Международной научно-практической конференции. 2017. С. 20-22.
- [3] Багутдинов Р.А. Гносеологические аспекты к определению назначения и состава стз в задачах проектирования и разработки робототехнических комплексов / Программные системы и вычислительные методы. 2017. № 1. С. 39-45.
- [4] Барский, А.Б. Нейронные сети: распознавание, управление, принятие решений [Текст] / А. Б. Барский. — М. : Финансы и статистика, 2004. — 176 с. — ISBN 5-279-02757-X.
- [5] Бородин А.В, Златкис В.М., Зуев Ю.П., Шишкирин В.И. Проблемы технического зрения // Разработка систем технического зрения и их применение в промышленности: Тез. докл. ВНТК.-Устинов, 1986.-С. 79-80.
- [6] Бурков А.П., Комин В.Г., Красильникьянц Е.В. Промышленная система технического зрения на базе интеллектуальной цифровой камеры IntCAM 285-1// Вестник ИГЭУ. Вып. 4. 2007.
- [7] Воротников С.А. Информационные устройства робототехнических систем. М.: изд. МГТУ им Н.Э. Баумана, 2005.
- [8] Гришин В.А. Системы технического зрения в решении задач управления беспилотными летательными аппаратами // Датчики и системы, №2, 2009.-С. 46-52.
- [9] Дворкович А.В., Мохин Г.Н., Нечпаев В.В. Об эффективности цифровой обработки статических изображений //Цифровая обработка сигналов и ее применение Докл. 1 Междун. конф.-М, 1998-Т 3-С 202-204.
- [10] Клевалин В.А., Поливанов А.Ю. Цифровые методы распознавания в системах технического зрения промышленных роботов// Мехатроника, автоматизация, управление, 2008, № 5.-С. 56-56.
- [11] Коробейников А.Г., Федосовский М.Е., Алексанин С.А. Разработка автоматизированной процедуры для решения задачи восстановления смазанных цифровых изображений // Кибернетика и программирование.- 2016.-1.-С. 270-291.
- [12] Куафе Ф. Взаимодействие робота с внешней средой : Пер. с франц.-М.: Мир, 1985.-285 с.
- [13] Мошкин В.И., Петров А.А., Титов В.С., Якушенков Ю.Г. Техническое зрение роботов – М.: Машиностроение, 1990. – 272 с.:ил.
- [14] Островский О.А. Криминалистическая портретная экспертиза по видеоизображениям, как форма определения личности // В сборнике: The Eleventh International Conference on Eurasian scientific development Proceedings of the Conference. Editor Jana Puna, Russia. 2016. С. 176-179
- [15] Островский О.А. Принцип объектной декомпозиции в систематизации идентификационных кодов, характеризующих преступления в сфере компьютерной информации / Полицияская деятельность. 2017. № 3. С. 10-18.
- [16] Попов А.В., Юревич Е.И. Роботы с силовым оучувствлением. СПб.: Астерион, 2008. . Каляев И.А., Мельник Э.В. Децентрализованные системы компьютерного управления. Ростов-на-Дону: изд. ЮНЦ РАН, 2011.
- [17] Попов Е.П., Писменный Г.В. Основы робототехники: Введение в специальность. Учеб. пособие для вузов по спец. «Робототехн. Системы и комплексы». М.: Высш. шк., 1990.-224с.
- [18] Поваркова А.В. Компьютерный анализ изображений: общие сведения, системы, примеры использования. // Вестник инфектологии. Электронный журнал. <http://www.infectology.ru/microscopy/today/analysis/index.aspx>
- [19] Семин, М.С. Прикладные задачи, решаемые с помощью систем технического зрения Текст. / М.С. Семин // Специальная техника. 2002.-№6. С 12-17.
- [20] Солопченко Г.Н. Измерительные информационные системы. СПб.: изд. Политехнического университета, 2010.-200 с.
- [21] Техническое зрение роботов / Под ред. А. Пью-М.: Машиностроение, 1987.-320 с.
- [22] Федоткин И.М. Математическое моделирование технологических процессов. Либроком, 2011.-416 с.
- [23] Фу К., Гонсалес Р., Ли К. Робототехника: Пер. с англ.-М.: Мир, 1989.-624 с.
- [24] Хорн Б. К. Зрение роботов: Пер. с англ. – М.: Мир, 1989. – 487 с., ил.
- [25] Юревич Е. И. Сенсорные системы в робототехнике : учеб. пособие / Е. И. Юревич. — СПб. : Изд-во Политехн. ун-та, 2013. — 100 с.
- [26] Bagutdinov R.A., Zakharova A.A. The task adaptation method for determining the optical flow problem of interactive objects recognition in real time // Journal of Physics: Conference Series. — 2017. — Vol. 803 : Information Technologies in Business and Industry (ITBI2016).
- [27] Bagutdinov R.A., Narimanov R.K. The calculating the fet based schottky hydrodynamic model // The First European Conference on Informational Technology and Computer Science 2015. С. 17-22.
- [28] Nebaba, S.G., Zaharova, A.A., Andreev, S.J. Automatic estimation of the face angle in the frame and transformation of images into zero rotation angles (2015) Scientific Visualization, 7 (4) .
- [29] Savitskiy, Yu.V., Nebaba, S.G., Spitsyn, V.G., Andreev, S.Yu., Makarov, M.A. Analysis of methods of features extraction from image of human face for identification (2016) Scientific Visualization, 8 (2), pp. 107-119.

Об авторах

Багутдинов Равиль Анатольевич – аспирант кафедры Инженерной графики и промышленного дизайна Института кибернетики Томского политехнического университета. E-mail: ravil_bagutdinov@yahoo.com.

Небаба Степан Геннадьевич – инженер кафедры Инженерной графики и промышленного дизайна Института кибернетики Томского политехнического университета. E-mail: stepanlfx@tpu.ru.

Захарова Алена Александровна – профессор, д.т.н., зав. кафедрой Инженерной графики и промышленного дизайна Института кибернетики Томского политехнического университета. E-mail: zaa@tpu.ru.

Модификация алгоритма анализа области частот для отпечатков пальцев

О.В. Бехтольд¹, В.Ю. Гудков²
olegbehtold@gmail.com|gudkovvy@mail.ru

¹Южно-Уральский государственный университет (НИУ), Челябинск, Россия;

² Южно-Уральский государственный университет (НИУ), Челябинск, Россия;

В данной работе рассмотрен модифицированный алгоритм построения поля качества изображения. Данное поле качества строится на основе частотного анализа локальных областей изображения. Модификация данного алгоритма заключается в добавлении фильтрации между этапом расчета частотной области локальных областей и этапом построения поля качества изображения. Поле качества изображения строится для оценки качества исходного изображения и принятия решения о модификации «плохих» участков, например, для компенсации морщин и порезов на отпечатках пальцев.

Ключевые слова: компьютерное зрение, дактилоскопия, отпечаток пальца, дискретное преобразование Фурье, частотная область, фильтрация.

Fingerprint frequency domain algorithm modification

O.V. Behtold¹, V.Y. Gudkov²
olegbehtold@gmai.com|gudkovvy@mail.ru

¹South Ural State University (NRU), Chelyabinsk, Russian Federation;

²South Ural State University (NRU), Chelyabinsk, Russian Federation;

Fingerprint image filtration is used to create quality field. This quality field based at frequency analysis of local fingerprint fields. We used additional filtration between frequency domain calculation step and quality field calculation step. Image quality field is used to estimate source image quality. This can help to find bad local effects, i.e. wrinkles and defects.

Keywords: computer vision, fingerprinting, fingerprint, DFT, domain frequency, filtration.

1. Введение

Поле качества является одной из качественных характеристик отпечатков пальцев в стандарте «NFIQ 2.0», разработанного Национальным институтом стандартов и технологий США [2]. Данный стандарт описывает методы оценки качества отпечатков пальцев и их эффективность.

В данном стандарте поле качества строится на основе частотного анализа локальных областей с помощью дискретного преобразования Фурье.

Актуальность исследования заключается в том, что данная модификация алгоритма, примененная к практической задаче мэтинга отпечатков пальцев, может уменьшить вероятности ошибок первого и второго рода, что позволит занять более высокие позиции в международных тестах и рейтингах среди компаний, занимающихся дактилоскопией.

2. Алгоритм

В алгоритме построения полей качества можно выделить основные этапы, такие как поворот локальной области изображения по направлению линий, дальнейшее преобразование этих областей с помощью дискретного преобразования Фурье, фильтрация и построение поля качества.

Рассмотрим подробнее основные этапы:

2.1 Поворот изображения

Поворот изображения необходим для корректной работы дискретного преобразования Фурье, чтобы мы могли получить корректные частоты в частотной области для построения поля качества.

Угол, на который производится поворот, определяется полем направлений, рассчитанным через тензорный метод Базена [1].

Поворот изображения выполняется вокруг его центра, и для каждой точки в новом изображении ищется позиция соответствующей точки в исходном фрагменте. Чаще всего эта позиция будет не совпадать с целочисленными координатами, поэтому будем применять интерполяцию.

Так как позиция X' имеет вещественные координаты, то находится между четырьмя другими координатами пикселей, как показано на рисунке 1, и значение рассчитывается по формуле 1, где L и M – дробные части значений координат, а A , B , C и D – значения яркостей пикселей.

$$X' = ((A*(1-L) + B*L)*(1-M) + (C*(1-L) + D*L)*M) \quad (1)$$

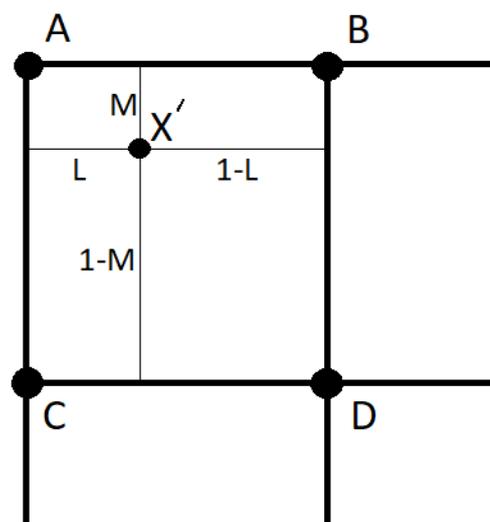


Рис.1. Позиция пикселя при переносе

2.2 Дискретное преобразование Фурье

В качестве исходных данных для дискретного преобразования Фурье используются средние значения яркостей локальных областей по вертикали. Таким образом, мы преобразуем наши локальные области в области частот по формуле 2 [3].

Для проверки расчетов используется визуализация области частот, пример такой визуализации показан на рисунке 2.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \dots, N-1 \quad (2)$$

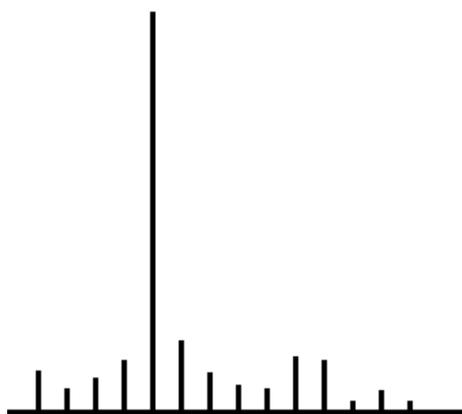


Рис.2. Дискретное преобразование Фурье

В области частот явно выделяются определенные значения, которые соответствуют частоте линий отпечатка пальца на выбранной локальной области.

2.3 Фильтрация

Модификация алгоритма заключается в фильтрации локальных областей частот. Для изображения порядка 256 на 256 пикселей локальные области выделим порядка 8 на 8 пикселей, и для каждой такой локальной области будем производить фильтрацию через плавающее окно 32 на 32 пикселя.

Для каждой локальной области нашего плавающего окна возьмем максимальную частоту от дискретного преобразования Фурье, и построим гистограмму из этих максимальных частот. Пик этой гистограммы будет соответствовать моде нашей выборки частот. В результате экспериментальных наблюдений предположим, что допустимый интервал вокруг моды равен $2/N$, где N – размер локальной области. Значение интервала выбрано на основе визуального анализа родных и чужих отпечатков пальцев таким образом, чтобы у родных отпечатков пальцев частота линий в локальных областях лежала в этом интервале.

Таким образом получаем, что все локальные области, максимальные частоты которых не входят в допустимый интервал, выбиваются из общей картины отпечатка пальца и помечаются как «плохие» участки на поле качества.

2.4 Поле качества

Поле качества строится на основе области частот по формуле 3, где F_{max} – максимальная частота, $A(F)$ – значение дискретного преобразования Фурье в определенной частоте, N – сторона квадратной локальной области изображения, $C=0,3$, в соответствии со стандартом ISO/IEC TR 29794-4:2010 [2].

$$Q_{FDA}^{local} = \frac{A(F_{max}) + C(A(F_{max} - 1) + A(F_{max} + 1))}{\sum_{F=1}^{N/2} A(F)} \quad (3)$$

По формуле 3 значение поле качества считается для каждой локальной области, кроме отмеченных ранее «плохих» участков. Для них качество прогнозируется как среднее значение качества близлежащих участков. На рисунке 3 приведен пример поля качества для всего изображения.



Рис.3. Пример поля качества отпечатка пальца

Поле качества является одним из критериев родственности отпечатков пальцев. Обычно у родных отпечатков плотность линий близка, а у чужих различается. Например, чем больше различие плотностей отпечатков, тем ниже степень похожести, т.е. больше вероятность что отпечатки чужие.

3. Заключение

Критериями, позволяющими оценить успешность исследования, являются визуализированные изображения полей качества отпечатков пальцев, а также программа на языке C++, реализующая стандартные алгоритмы и их усовершенствования.

Модификация алгоритма построения поля качества для отпечатков пальцев является частью более широкого исследования, целью которого является компенсация плохих участков отпечатков пальцев, морщин и порезов. Поле качества изображения дает информацию для принятия решений о дальнейшей модификации той или иной локальной области. Дальнейшее исследование улучшения полей качества отпечатков пальцев планируется провести с помощью добавления и анализа дополнительных характеристик, отвечающих за качество изображения.

4. Литература

- [1] Bazen, M. Fingerprint Identification - Feature Extraction, Matching, and Database Search / M. Bazen. – Enschede: Twente University Press, – 2002. – 187 p.
- [2] Elham Tabassi, NIST Fingerprint Image Quality 2.0, – 2016. – 95 p.
- [3] R. Gonzalez, R. Woods. Digital image processing, M.: Technosphere – 2012. – 961 p.

Об авторах

Бехтольд Олег Владимирович, аспирант кафедры ЭВМ высшей школы электроники и компьютерных наук ЮУрГУ. Его e-mail olegbehtold@gmail.com

Гудков Владимир Юльевич, д. ф.-м.н., профессор кафедры ЭВМ высшей школы электроники и компьютерных наук ЮУрГУ. Его e-mail gudkovvy@mail.ru

Визуальный анализ и обработка многомерных данных

А.Е. Бондарев, В.А. Галактионов, Л.З. Шапиро
 bond@keldysh.ru|vlgal@gin.keldysh.ru|pls@gin.keldysh.ru
 Институт прикладной математики им. М.В.Келдыша РАН, Москва, Россия

В работе рассматриваются проблемы визуального анализа многомерных наборов данных. Для визуального анализа применяется подход, основанный на построении упругих карт. Для анализа кластерных структур в исходном объеме данных упругие карты используются в качестве методов отображения исходных точек данных в замкнутые многообразия с меньшей размерностью. Уменьшая параметры упругости, можно проектировать поверхность карты, которая намного лучше аппроксимирует многомерный набор данных. Точки исследуемого объема данных проецируются на карту. Расширение построенной карты на плоскость вкупе с отображением в пространство первых трех главных компонент позволяет получить представление о кластерной структуре многомерного набора данных. Построение упругих карт не требует априорной информации о данных и не зависит от характера данных, происхождения данных и т. д. Применение метода «квази-зум» позволяет существенно улучшить результаты в области сгущения точек изучаемого многомерного пространства. В работе приведены результаты применения упругих карт для визуального анализа различных многомерных наборов данных.

Ключевые слова: многомерные данные, визуальный анализ, упругие карты, кластерные структуры.

Visual analysis and processing of multidimensional datasets

A.E. Bondarev, V.A. Galaktionov, L.Z. Shapiro
 bond@keldysh.ru|vlgal@gin.keldysh.ru|pls@gin.keldysh.ru
 Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

The article considers the problems of visual analysis of multidimensional datasets. For visual analysis an approach based on the construction of elastic maps is applied. To analyse clusters in original data volume the elastic maps are used as the methods of original data points mapping to enclosed manifolds having less dimensionality. Diminishing the elasticity parameters one can design map surface which approximates the multidimensional dataset in question much better. The points of dataset in question are projected onto the map. The extension of designed map to a flat plane with mapping into the space of the first three main components allows one to get an insight about the cluster structure of multidimensional dataset. The construction of elastic maps does not require a priori information about the data and does not depend on the nature of the data, the origin of the data, etc. The application of the "quasi-Zoom" method allows one to get the significantly improved results in the area of condensation of the points of the multidimensional space under study. The paper presents the results of applying elastic maps for visual analysis of various multidimensional datasets.

Keywords: multidimensional data, visual analysis, elastic maps, cluster structures.

1. Введение

В анализе многомерных данных особое место занимают задачи классификации. При классификации объема многомерных данных может решаться как задача разделения исследуемой совокупности явлений на классы, так и отнесения одного или нескольких явлений к уже существующим классам. Для решения подобных задач используются методы кластерного анализа. Методов и алгоритмов кластерного анализа на современном этапе существует очень много, они постоянно развиваются и отличаются большим разнообразием. Многообразие алгоритмов кластерного анализа обусловлено множеством различных критериев, выражающих те или иные аспекты качества автоматического группирования. При решении задач классификации весьма полезными оказываются подходы визуальной аналитики, являющиеся синтезом нескольких алгоритмов понижения размерности и визуального представления многомерных данных во вложенных в исходный объем многообразиях меньшей размерности.

К таким алгоритмам можно отнести отображение исходного многомерного объема в упругих картах (Elastic Maps) [5,6,8,9] с разными свойствами упругости или эластичности. Эти методы позволяют тем или иным образом выделить из исходного многомерного объема данных содержащуюся в нем кластерную структуру.

Следует заметить, что интерес к упругим картам появился у нас при разработке вычислительной технологии для построения, обработки, анализа и визуального

представления многомерных параметрических решений задач газовой динамики. Вычислительная технология реализована как единая технологическая цепочка алгоритмов производства, обработки, визуализации и анализа многомерных данных. Такая технологическая цепочка может рассматриваться как прототип обобщенного вычислительного эксперимента для нестационарных задач вычислительной газовой динамики. Схема реализации подобного обобщенного вычислительного эксперимента представлена на рисунке 1.

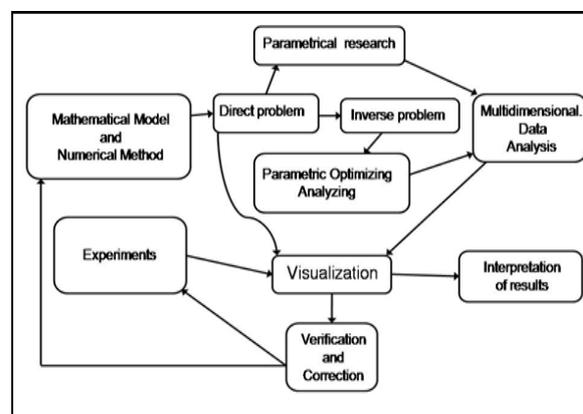


Рис. 1. Схема организации обобщенного вычислительного эксперимента.

Подобный обобщенный вычислительный эксперимент неявно предполагает наличие надежной математической

модели, численного метода для ее решения и набор экспериментальных результатов для верификации. В процессе вычислений необходимо реализовать организацию постоянного сравнения с экспериментальными данными при наличии такой возможности. Набор используемых методов должен включать в себя решение обратных и оптимизационных задач. Будучи реализованными с помощью описанных ранее параллельных интерфейсов, эти методы позволяют получать решения задач параметрического исследования и оптимизационного анализа в виде многомерных объемов данных.

Для обработки этих объемов и выявления скрытых взаимосвязей между изучаемыми в объеме параметрами необходимо интегрировать в общий набор алгоритмов методы анализа многомерных данных и их визуального представления. В итоге подобный обобщенный вычислительный эксперимент позволит получать решение не одной, отдельно взятой, задачи, а решение для целого класса задач, задаваемого диапазонами изменения определяющих параметров. Также следует отметить универсальность подобного обобщенного вычислительного эксперимента. Он может быть применен к широкому кругу задач математического моделирования нестационарных процессов. Практическая реализация подобного обобщенного эксперимента может обеспечивать организацию крупномасштабных промышленных расчетов. Описание элементов реализованной вычислительной технологии приведено в работах [2-4].

На практике упругие карты оказались полезным и достаточно универсальным инструментом, что позволило применять их к многомерным объемам данных разного типа. Например, данный подход был применен к задачам анализа текстовой информации, где в качестве числовых характеристик выступали частоты употребления слов [1].

2. Построение упругих карт

Идеология и алгоритмы реализации построения упругих карт подробно представлены в работах [5,8]. Подобная карта представляет собой систему упругих пружин, вложенную в многомерное пространство данных. Этот подход основывается на аналогии с задачами механики: главное многообразие, проходящее через «сердину» данных, может быть представлено как упругая мембрана или пластинка. Метод упругих карт формулируется как оптимизационная задача, предполагающая оптимизацию заданного функционала от взаимного расположения карты и данных.

Согласно [5,6] основой для построения упругой карты является двумерная прямоугольная сетка G , вложенная в многомерное пространство, которая аппроксимирует данные и обладает регулируемыми свойствами упругости по отношению к растяжению и изгибу. Расположение узлов сетки ищется в результате решения оптимизационной задачи на нахождение минимума функционала:

$$D = \frac{D_1}{|X|} + \lambda \frac{D_2}{m} + \mu \frac{D_3}{m} \rightarrow \min ,$$

где $|X|$ - число точек в многомерном объеме данных X ; m - число узлов сетки, λ, μ - коэффициенты упругости, отвечающие за растяжение и изогнутость сетки соответственно; D_1, D_2, D_3 - слагаемые, отвечающие за свойства сетки.

$$D_1 = \sum_{ij} \sum_{x \in K_{ij}} \|x - r^{ij}\|^2 .$$

D_1 является мерой близости расположения узлов сетки к данным.

Здесь K_{ij} - подмножества точек из X , для которых узел сетки r^{ij} является ближайшим:

$$x \rightarrow r^{ij}, \quad \|x - r^{ij}\|^2 \rightarrow \min, \quad K_{ij} = \{x \in X, x \rightarrow r^{ij}\},$$

Слагаемое D_2 представляет меру растянутости сетки:

$$D_2 = \sum_{ij} \|r^{ij} - r^{i,j+1}\|^2 + \sum_{ij} \|r^{ij} - r^{i+1,j}\|^2 .$$

Слагаемое D_3 представляет меру изогнутости (кривизны) сетки:

$$D_3 = \sum_{ij} \|2r^{ij} - r^{i,j-1} - r^{i,j+1}\|^2 + \sum_{ij} \|2r^{ij} - r^{i-1,j} - r^{i+1,j}\|^2 .$$

Варьирование параметров упругости заключается в построении упругих карт с последовательным уменьшением коэффициентов упругости, в силу чего карта становится более мягкой и гибкой, наиболее оптимальным образом подстраиваясь к точкам исходного многомерного объема данных. После построения упругую карту можно развернуть в плоскость для наблюдения кластерной структуры в изучаемом объеме данных. Применение упругих карт позволяет более точно и четко определять кластерную структуру изучаемых многомерных объемов данных.

Следует отметить, что при построении упругих карт в многомерном облаке данных, состоящем из сгущений и отдельных отдаленных точек, возникает проблема масштабируемости. Упругая карта будет пытаться подстроиться под рассматриваемый объем в целом – как к отдаленным точкам, так и к областям сгущения, что, естественно, не может получиться одинаково хорошо. Для того чтобы решить эту проблему и обеспечить четкое представление о данных в области сгущений в работе [1] был предложен подход, названный «квази-зум» (quasi-Zoom), заключающийся в вырезании области сгущения из рассматриваемого облака многомерных данных и построения для вырезанной области упругой карты заново.

3. Примеры построения упругих карт

Рассмотрим пример построения упругих карт для широко известного тестового объема многомерных данных IRIS [5]. Данный объем представляет собой набор данных, основанных на измерениях характеристик цветков ириса. Набор данных описывает три сорта ирисов и состоит из 150 точек в четырехмерном пространстве признаков. На рисунке 2 представлена «мягкая» упругая карта для этого набора данных с раскраской по плотности данных. Здесь и далее для построения и визуализации упругих карт применен программный комплекс ViDaExpert, подробно описанный в [8]. На рисунке 3 представлена та же самая карта, развернутая на плоскость. В таком виде карта дает достаточно четкое представление о разделении многомерного объема данных на три класса.

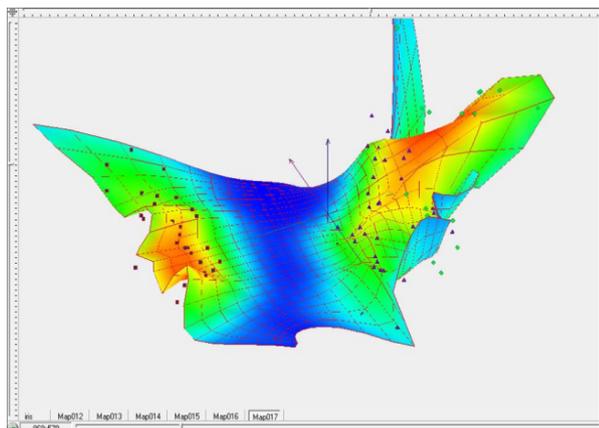


Рис. 2. «Мягкая» упругая карта в пространстве первых трех главных компонент с раскраской по плотности данных.

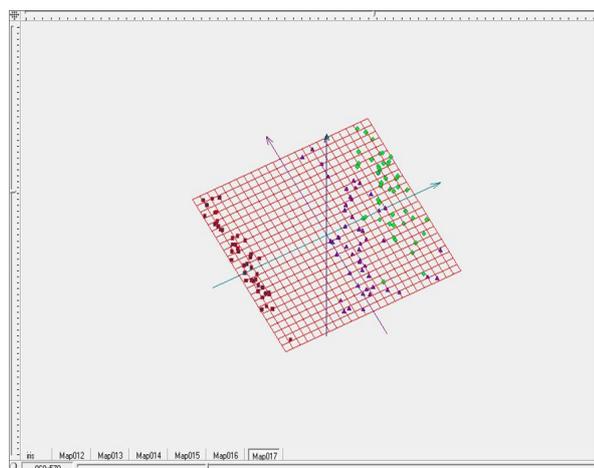


Рис. 3. Развертка «мягкой» упругой карты на плоскость.

Далее рассмотрим результаты применения подхода построения упругих карт к задаче анализа текстовой информации. С точки зрения построения упругих карт исходный многомерный объем является совершенно стандартным. Рассмотрим ниже результаты построения упругих карт для тестового объема [1]. Для первичных тестов было отобрано около 100 глаголов со 155 наиболее связанными с ними существительными. Полученные таким образом данные далее рассматриваются как многомерный объем данных, представляющий собой 100 точек в 155-мерном пространстве. Числовые значения получающейся в результате матрицы определяются как частоты совместного употребления.

Следует отметить, что среди отобранных глаголов содержался ряд пар, представляющих схожие глаголы совершенного и несовершенного вида. Это было сделано для дополнительного контроля в силу предположения, что точки, соответствующие подобным парам, должны находиться недалеко друг от друга на результирующем изображении. Пример «мягкой» упругой карты приведен на рисунке 4. На рисунке 5 представлена та же самая карта в развернутом виде. Здесь видно, что изучаемый объем данных, содержит область высокой плотности данных и точки, достаточно далеко отстоящие от этой области.

Именно в таких случаях возникает проблема масштабируемости, описанная в предыдущем разделе. Для решения этой проблемы был разработан подход «квази-зум» (quasi-Zoom), представленный в работе [1].

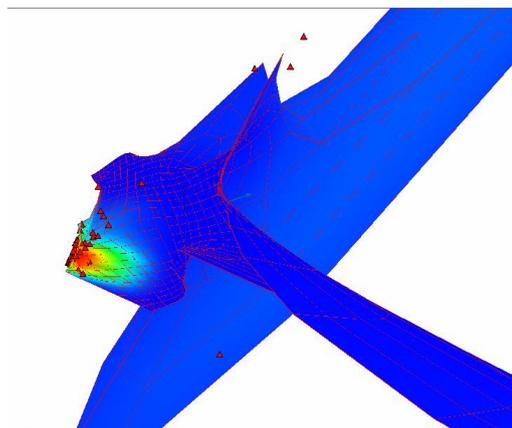


Рис. 4. Упругая карта с раскраской по плотности в применении к объему частот совместного употребления слов [1].

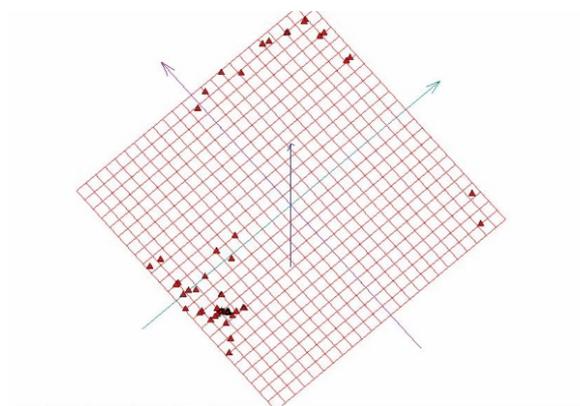


Рис. 5. Построение развертки предыдущей упругой карты на плоскость первых двух главных компонент [1].

При исследовании объема частот совместного употребления глаголов и существительных [1] практическая задача ставилась следующим образом. Нужно было максимально разделить «слипшиеся точки». Для этой цели был использован подход «квази-зум», который позволил решить эту задачу. Суть этого технологического приема заключается в том, что для более тонкой подстройки необходимо выделять большие кластеры в исследуемом объеме многомерных данных и проводить построение упругих карт для выделенных кластеров отдельно, организуя тем самым эффект, подобный функции «zoom» в современной фототехнике. Результаты применения подхода представлены на рисунке 6.

Применение технологий построения упругих карт для решения задач кластерного анализа не предполагает никакой априорной информации об изучаемых данных. Это дает возможность применять их к анализу данных самого различного типа вне зависимости от природы их происхождения. Подобное абстрагирование метода от типа и происхождения данных делает используемый подход построения упругих карт в достаточной степени универсальным инструментом анализа многомерных объемов данных.

Приведем пример применения построения упругих карт к анализу биомедицинских данных. Для этой цели были использованы данные работы [7]. Этот набор данных содержит значения шести биомеханических признаков, используемых для классификации ортопедических пациентов на 2 класса (нормальный «normal», или с отклонением от нормы «abnormal»). Каждый пациент

представлен в наборе данных шестью биомеханическими признаками, полученными из формы и ориентации таза и поясничного отдела позвоночника: тазовое опускание, тазовый наклон, угол поясничного лордоза, крестцовый наклон, радиус таза и степень спондилолистеза.

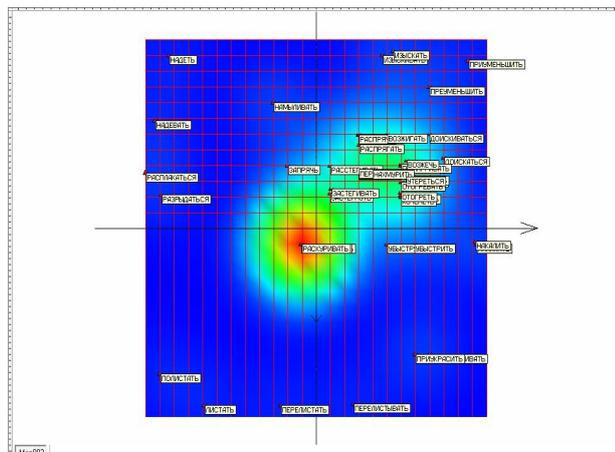


Рис. 6. Результаты применения подхода «quasi-Zoom» для разделения слипшихся точек (развертка упругой карты с раскраской по плотности данных).

Набор данных содержит 310 точек в 6-мерном пространстве. Рассматривая данный набор как многомерный объем в 6-мерном пространстве признаков, можно применять подход построения упругих карт, варьировать коэффициенты упругости карты для достижения наилучшего результата, далее проецировать точки многомерного объема на полученную карту и строить ее развертку в плоскости двух первых главных компонент.

На рисунке 7 представлена «мягкая» упругая карта для данного набора данных. Зеленые точки соответствуют категории «normal», красные точки соответствуют категории «abnormal». На рисунке 8 представлена развернутая карта с раскраской по плотности данных.

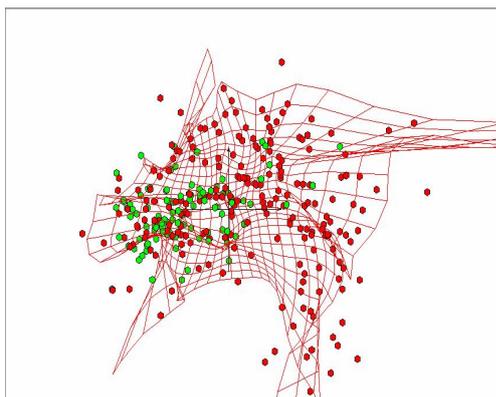


Рис. 7. Пример построения упругой карты для набора биомедицинских данных.

Полученные результаты дают возможность получить представление о взаимном расположении основных классов в изучаемом многомерном объеме. Однако на рисунке 8 видна область смещения данных из двух категорий. Для улучшения разделения необходимо в дальнейшем усовершенствовать алгоритм построения упругих карт за счет возможности сгущения исходной сетки карты в областях повышенной плотности данных. Подобная

возможность широко используется в задачах математического моделирования сплошных сред. Автоматическое сгущение сетки в области высоких градиентов в вычислительной механике жидкости и газа является отдельным направлением, где реализовано большое количество эффективных и апробированных на практике алгоритмов.

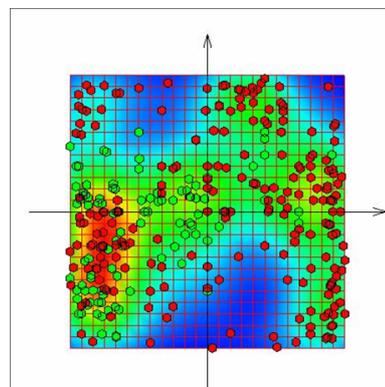


Рис. 8. Развертка предыдущей упругой карты на плоскость первых двух главных компонент с раскраской по плотности данных.

4. Заключение

Для анализа кластерных структур в многомерном объеме данных предлагается использовать технологии построения упругих карт, представляющие собой методы отображения точек исходного многомерного пространства на вложенные в это пространство многообразия меньшей размерности. Варьируя поверхность упругой карты за счет последовательного уменьшения коэффициентов упругости, можно добиться лучшего соответствия подстраивания карты под многомерное облако данных. После уменьшения коэффициентов изгиба и растянутости упругой карты, она становится более мягкой и гибкой, наиболее оптимальным образом подстраиваясь к точкам исходного многомерного объема данных. Применение технологий построения упругих карт для решения задач кластерного анализа не предполагает никакой априорной информации об изучаемых данных и не зависит от их природы, происхождения и т.п. Эти свойства позволяют применить технологии построения упругих карт для выявления кластерных структур. Для решения проблемы масштабируемости, когда упругая карта подстраивается как под область сгущения точек данных, так и к отдельно расположенным точкам облака данных, применяется подход «квази-зум» (quasi-Zoom). Суть подхода заключается в том, что для более тонкой подстройки в многомерном облаке данных выделяются большие кластеры, после чего проводится отдельное построение упругих карт для выделенных кластеров. Приведены примеры построения упругих карт для задач анализа текстов и для медицинских данных. Дальнейшее развитие используемого подхода представляется в данный момент в реализации такого алгоритма при построении упругих карт, где область высокой плотности данных рассматривалась бы аналогично области высоких градиентов в задачах математической физики, и предусматривала бы автоматическое сгущение сетки карты в подобных областях.

5. Благодарности

Данная работа выполнена при поддержке грантов Российского фонда фундаментальных исследований (проекты 16-01-00553а и 17-01-00444а).

6. Литература

- [1] Bondarev A.E., Bondarenko A.V., Galaktionov V.A., Klyshinsky E.S. Visual analysis of clusters for a multidimensional textual dataset / Scientific Visualization. V.8, № 3, pp.1-24, 2016, URL: <http://sv-journal.org/2016-3/index.php?lang=en>
- [2] Bondarev A.E., Galaktionov V.A. Analysis of Space-Time Structures Appearance for Non-Stationary CFD Problems // Proceedings of 15-th International Conference On Computational Science ICCS 2015 Rejkjavik, Iceland, June 01-03 2015, Procedia Computer Science, Volume 51, 2015, Pages 1801–1810.
- [3] Bondarev A.E., Galaktionov V.A. Multidimensional data analysis and visualization for time-dependent CFD problems // Programming and Computer Software, 2015, Vol. 41, No. 5, pp. 247–252, DOI: 10.1134/S0361768815050023.
- [4] Bondarev A.E. Visual analysis and processing of clusters structures in multidimensional datasets // Proceedings of the 2nd International ISPRS Workshop on PSBB, 15–17 May 2017, Moscow, Russia, ISPRS Archives, Volume XLII-2/W4, 2017, pp.151-154. <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W4/151/2017/>
- [5] Gorban A. et al. Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007.
- [6] Gorban A., Zinovyev A. Principal manifolds and graphs in practice: from molecular biology to dynamical systems International Journal of Neural Systems, Vol. 20, No. 3 (2010) 219–232.
- [7] Rocha Neto A., Barreto G., 2009. On the Application of Ensembles of Classifiers to the Diagnosis of Pathologies of the Vertebral Column: A Comparative Analysis, IEEE Latin America Transactions, 7(4):487-496.
- [8] Zinovyev A. Vizualizacija mnogomernyh dannyh [Visualization of multidimensional data]. Krasnoyarsk, publ. NGTU. 2000. 180 p. [In Russian]
- [9] Zinovyev A. Data visualization in political and social sciences, In: SAGE «International Encyclopedia of Political Science», Badie, B., Berg-Schlosser, D., Morlino, L. A. (Eds.), 2011.

Об авторах

Бондарев Александр Евгеньевич, к.ф.-м.н., старший научный сотрудник ИПМ им. М.В. Келдыша РАН. Его e-mail: bond@keldysh.ru.

Галактионов Владимир Александрович, д.ф.-м.н., профессор, заведующий отделом компьютерной графики и вычислительной оптики ИПМ им. М.В. Келдыша РАН. Его e-mail: vlgal@gin.keldysh.ru.

Шапиро Лев Залманович, к.ф.-м.н., старший научный сотрудник ИПМ им. М.В. Келдыша РАН. Его e-mail: pls@gin.keldysh.ru.

Стеганографическая защита изображений из PDF документов на основе конвертора PDF-SVG

В.Н. Горбачев¹, И.К. Метелёв¹, Е.М. Кайнарова¹, М.А. Полякова¹
helenkainarova@gmail.com|polyakovaqueen1998@mail.ru.

¹СПб Государственный университет промышленных технологий и дизайна, Санкт-Петербург, Россия

В работе рассмотрен вопрос о защите изображений, входящих в состав PDF документа. Основой предложенной стеганографической схемы служит конвертор PDF-SVG, который извлекает изображение из документа и возвращает его обратно после встраивания цифровых данных. Обсуждаются два варианта встраивания, использующие дискретное вейвлет преобразование и битовые плоскости полутонового контейнера.

Ключевые слова: стеганография, PDF, цифровые водяные знаки.

Steganographic protection of image from PDF document using a convertor PDF-SVG

V.N. Gorbachev¹, I.K. Meteleev¹, E.M. Kaynarova¹, M.A. Polyakova¹
helenkainarova@gmail.com|polyakovaqueen1998@mail.ru.

¹Saint-Petersburg State University of Industrial Technologies and Design, Saint-Petersburg, Russia

Protection of the images included in PDF document is considered. The basis of the proposed steganographic scheme is a convertor of PDF-SVG, that extracts images from PDF document and returns it back after embedding data. We discussed two embedding methods using a wavelet transform and bit planes of the grayscale cover.

Keywords: steganography, PDF, watermarks.

Будучи широко распространенным форматом представления электронных документов, PDF имеет разнообразную криптографическую защиту. Однако задача защиты информации не имеет единственного решения, что является источником новых предложений, в число которых входят стеганографические методы. Богатая структура PDF формата позволяет использовать морфологические и семантические характеристики текста, модификацию операторов визуализации текста, и многое другое [1].

При морфологическом подходе бит информации кодируется текстовой альтернативой: пассивный или активный залог, синонимы [2]. В технике кодирования пробелами, White Space Coding, используют обычный и неразрывный пробел, которые визуально неразличимы, но имеют разные коды [3]. В методе *TJ* (Text Justified) операторов, выравнивающих текст, данные встраиваются в интервалы между словами. [4]. При выравнивании длины интервалов принимают случайный характер, что дает возможность их модифицировать, используя LSB (Least Significant Bit) методы [5] или размытие (Dither Modulation) [6]. Сообщение можно закодировать перестановкой, которая встраивается в 1d или 2d массив. Встраивание в частотную область цифрового изображения специально выбранной перестановки, которая имеет циклы длиной не больше 2, ее называют инволюция, оказывается устойчивым к JPEG сжатию с потерей [7].

Встраивание цифровых водяных знаков (ЦВЗ) предусмотрено рядом приложений. К их числу относится Adobe Acrobat и редактор \TeX , который может создавать выходной файл в PDF. Эти приложения позволяют встроить видимые ЦВЗ, представляющие со-

бой полупрозрачный текст или изображения, которые затрудняют зрительное восприятие. Техника видимых ЦВЗ предполагает, что их может удалить законный пользователь за отдельную плату [8].

В отличие от рассмотренных методов мы сосредоточили внимание на защите изображений в PDF документах. Для этой цели предложен конвертор *PDF – SVG*. Он сохраняет изображения, извлеченные из PDF документа, в формате PNG или JPEG, в которые встраиваются как невидимые, так и видимые ЦВЗ. Защищенное изображение конвертор возвращает обратно в документ. Для случая видимых ЦВЗ целью является восстановление исходного изображения. В нашем методе величина PSNR (Peak Signal Noise Ratio) между оригиналом и восстановленным изображением может достигать порядка 55.8 дБ. Это значение свидетельствует о высоком качестве восстановленного изображения.

1. Встраивание с помощью конвертора

Стеганографическая схема содержит конвертор, который осуществляет преобразование между двумя форматами PDF и SVG, алгоритмы встраивания и детектирования.

1.1. Схема

Работа схемы содержит следующие шаги:

1. Конвертор извлекает изображения из PDF документа и сохраняет их в графическом формате PNG или JPEG. Эти или модифицированные изображения конвертор возвращает обратно в PDF документ.

2. Извлеченные изображения могут модифицироваться путем встраивания, детектирования или удаления ЦВЗ.

В качестве ЦВЗ используется бинарное изображение M , которое встраивается в контейнер C , полученный из графического файла PNG или JPEG. В случае невидимых ЦВЗ алгоритмы детектирования извлекают встроенные данные M' . В случае видимых ЦВЗ восстанавливается исходный контейнер CR . Из-за необратимости преобразований извлеченные данные и восстановленный контейнер будут отличаться от оригиналов. Однако в обоих случаях требуется выполнить условие неразличимости

$$\begin{aligned} M &\approx M', \\ C &\approx CR. \end{aligned}$$

Эти условия допускают две интерпретации. С одной стороны есть два изображения, которые выглядят одинаково. С другой стороны можно говорить о недетектируемости. Для невидимых ЦВЗ это означает невозможность определить с вероятностью единица наличие встроенных данных путем статистического анализа. Для видимых ЦВЗ это означает невозможность определить был ли удален встроенный ЦВЗ из изображения нелегитимным пользователем.

1.2. Конвертор

Важным свойством нашего конвертора является обратимость, что позволяет преобразовывать данные без потерь.

В сети Интернет можно найти большое число предложений как преобразовать изображение из PDF в графический формат PNG, JPEG и обратно. Были выбраны наугад два конвертора¹ и полутоновое изображение a.png с яркостью пиксела 0, 1, ..., 255. Мы рассмотрели преобразование

$$a.png \rightarrow \text{PDF} \rightarrow b.png.$$

Было найдено, что разность яркостей пикселей у изображений a.png и b.png достигала 129 единиц, что заметно при визуальном восприятии. Для нашего конвертора разность оказалась равно нулю, что означает обратимость преобразования.

Отметим точность, с которой получается этот результат. Для кодирования в графическом формате обычно используются целые неотрицательные числа, поэтому при вычислениях или записи в графический формат все числа округляются. Чтобы исключить случайное совпадение, было протестировано 200 различных pdf-документов. В результате наш конвертор можно считать обратимым по крайней мере в случае преобразования данных, представленных целыми неотрицательными числами. Это означает, что конвертирование не будет вносить ошибок в процессы извлечения ЦВЗ и восстановления контейнера.

¹<http://pdf-png-jpg.eu/>, <http://online2pdf.com/converto-r-png-to-pdf>

2. Частотное встраивание

Встраивание данных осуществлялось путем замены части коэффициентов дискретного вейвлет преобразования контейнера пикселями ЦВЗ.

2.1. Встраивание с помощью вейвлет преобразования

Алгоритмы встраивания и детектирования содержат следующие шаги:

1. Контейнер подвергается одноуровневому дискретному вейвлет преобразованию DWT (Discrete Wavelet Transform).
2. Часть коэффициентов из выбранного частотного блока заменяется на биты сообщения и осуществляется обратное преобразование IDWT.
3. Встроенные данные извлекаются из выбранного частотного блока.

Вейвлет преобразование извлеченного из PDF документа изображения C дает четыре блока коэффициентов

$$\text{DWT}(C) = \{cA, cH, cV, cD\},$$

где cA коэффициенты аппроксимации, cH , cV и cD коэффициенты горизонтальных, вертикальных и диагональных деталей. В литературе их также называют LL , LH , HL , HH частотными полосами от Low и High, высокий и низкий. Пусть сообщение M встраивается, например, в блок cD . Для этого M масштабируется путем изменения размера и яркости

$$M \rightarrow M_u \rightarrow aM_u,$$

где aM_u - изображение с новыми размерами и яркостью, которые определяются коэффициентами u и a . Эти коэффициенты играют важную роль, они подбираются экспериментально, исходя из поставленной задачи. Изменяя u или a , можно полностью или частично заполнить контейнер или сделать встроенные данные как видимыми, так и невидимыми. В результате часть коэффициентов cD заменяется на aM_u и блок со встроенными данными cD^* используется далее для формирования стегоконтейнера

$$S = \text{IDWT}\{cA, cH, cV, cD^*\}.$$

Данные извлекаются из стегоконтейнера S' , который получается из S путем некоторого преобразования T :

$$T : S \rightarrow S'.$$

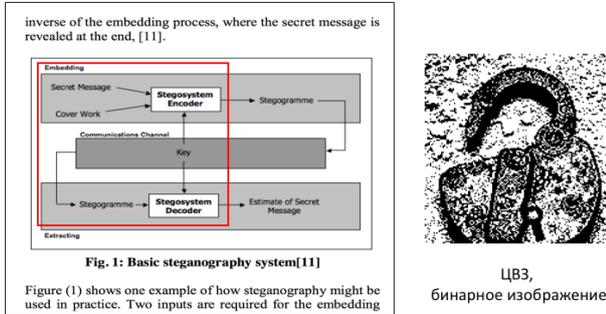
С помощью T описывается ряд процессов, в число которых входит запись изображения в графический формат, считывание, конвертирование, передача данных конвертору и др.

2.2. Эксперимент

Результат встраивания невидимого ЦВЗ приведен на рис. 1. Слева фрагмент PDF документа², он содержит

²Фрагмент из работы Y-C Lai, W-H Tsai, Covert communication via PDF files by new data hiding techniques, NSC project No, 97-2631-H-009-001

жит изображение Fig. 1: Basic steganographic system, которое расположено в тексте. Изображение защищено невидимым ЦВЗ, встроенным в область, отмеченную красным. Справа приведен ЦВЗ, представленный бинарным изображением. Для DWT был взят ортогональный вейвлет db6 из семейства Добеши.



Контейнер с ЦВЗ, a=20, u=0.5, вейвлет db2

Рис. 1. Фрагмент PDF документа с защищенным изображением.

Выбранные величины a=20 и u=0.5 означают, что яркость бинарного ЦВЗ увеличена в 20 раз и заменено 50% коэффициентов блока cD.

Мы рассмотрели встраивание с несколькими вейвлетами из семейства Добеши db1, db2, db6, db26, db41 для a=20 и u=0.5. При детектировании возникали ошибки, когда в извлеченных ЦВЗ от 2% до 15% пикселей воспроизводится неправильно. Было установлено, что основной причиной ошибки является потеря точности при округлении, которая возникает при записи цифрового изображения с ЦВЗ в графический формат.

3. Пространственное встраивание

Используя побитовое сложение, можно встроить бинарное изображение в битовую плоскость полутонного контейнера. Тогда ошибки округления при записи изображения в графический формат будут исключены.

3.1. Встраивание в битовые плоскости

В основе алгоритма лежит идея удвоения битовых плоскостей [9]. Благодаря избыточности, можно построить полутонное изображение с двумя одинаковыми битовыми плоскостями, причем визуально оно не будет отличаться от исходного. Тогда одна из плоскостей может быть использована для встраивания данных, а вторая для слепого детектирования ЦВЗ или восстановления контейнера.

Любое полутонное изображение с битовой глубиной k можно представить с помощью набора битовых плоскостей. Для k = 8

$$C = B_8 2^7 + \dots + B_1 2^0, \quad (1)$$

где $B_V = \text{bitget}(C, V)$ битовая плоскость, для вычисления которой введена функция bitget, $V = 1, 2, \dots, 8$. Плоскость B_1 называют LSB (Least

Significant Bit) плоскостью, она не содержит семантической информации. Используя (1), можно получить изображение с двумя битовыми плоскостями. Пусть B_V и B_U , где $V > U$ две битовые плоскости C. Тогда изображение C_D с двумя битовыми плоскостями B_V имеет вид

$$C_D = C - B_U 2^{U-1} + B_V 2^{U-1}. \quad (2)$$

Алгоритм встраивания, использующий (2), содержит следующие шаги.

1. Для выбранных V и U строится контейнер C_D с двумя битовыми плоскостями.
2. В плоскость V встраивается бинарное изображение путем побитового сложения

$$C \rightarrow S = C_D - B_V 2^{V-1} + (B_V \oplus M) 2^{V-1}. \quad (3)$$

Встроенное изображение будет видимым или невидимым в зависимости от выбора V. Мы рассмотрим случай видимых ЦВЗ, для которого в качестве V следует взять старшие битовые плоскости. Чтобы восстановить контейнер, старшая плоскость V, содержащая ЦВЗ, удаляется, а на ее место ставится U

$$S \rightarrow CR = S - \text{bitget}(S, V) 2^{V-1} + \text{bitget}(S, U) 2^{U-1} - \text{bitget}(S, U) 2^{U-1}.$$

Отличие контейнеров определяется младшей битовой плоскостью

$$C - CR = B_U 2^U. \quad (4)$$

Пример работы показан на рис. 2.

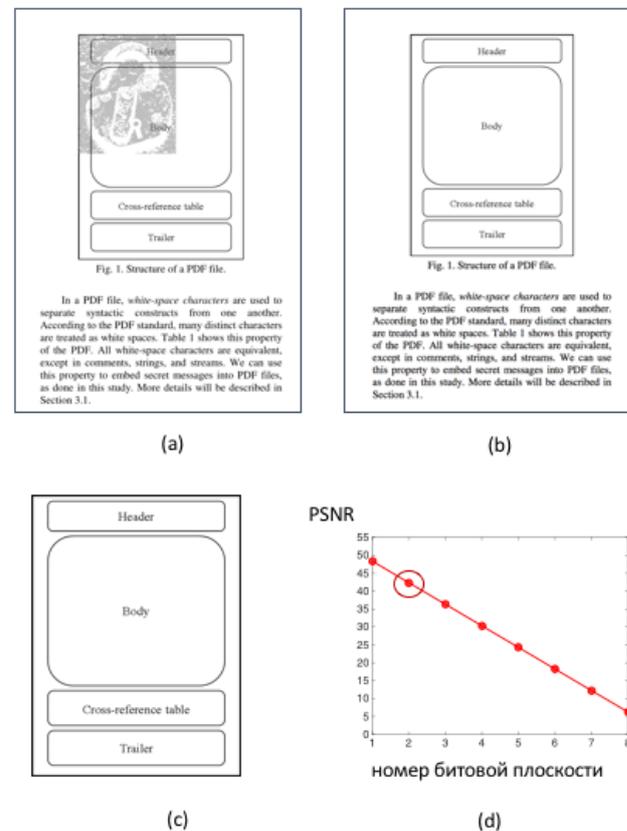


Рис. 2. Видимые и устранимые ЦВЗ.

На рис. 2 (а) приведен фрагмент PDF документа с изображением *Fig.1 Structure of PDF file*³, в которое встроен видимый ЦВЗ. Используются плоскости $V = 7$ и $U = 2$. Фрагмент PDF и изображение с удаленным ЦВЗ представлены на рис. 2 (b) и (c). После удаления ЦВЗ восстановленный документ визуально неотличим от исходного. На рис. 2 (d) приведена зависимость пикового отношения сигнал шум PSNR от номера удаленной битовой плоскости. Для рассматриваемого случая $U = 2$ величина PSNR=42.3417 Дб. Это значение совпадает с рассчитанным из эксперимента, который включает цепочку преобразований с конвертированием. Найденное совпадение подтверждает обратимость этих преобразований. Полученное значение PSNR свидетельствует о высоком качестве восстановленного изображения и совпадает с визуальной оценкой. В этом примере для дублирования использована вторая битовая плоскость $U = 2$. Если взять LSB плоскость $U = 1$, то можно получить более высокое PSNR: порядка 55.8 дБ, как это следует из графика на рис. 2 (d). Заметим, наши результаты получены для пространственного встраивания. В случае частотной техники, развитой в [10] для видимых ЦВЗ на основе DWT, значения PSNR оказываются несколько хуже: 37-40 дБ.

4. Выводы

1. Изображения, входящие в состав PDF документов, могут быть скопированы нелегитимным пользователем и потому нуждаются в своей собственной защите.
2. В стеганографической схеме может быть использован конвертор, который преобразует изображения из PDF в графический формат и обратно. Поэтому можно использовать разные методы защиты цифровых изображений, включающие пространственную и частотную технику.
3. Для встраивания в частотной области возникает характерная ошибка, обусловленная потерей точности из-за сохранения изображения в графическом формате. При использовании битовых плоскостей эта ошибка может быть исключена.

5. Литература

- [1] Ndongam R., Ekodeck S. G. PDF Steganography based on Chinese Remainder Theorem // arXiv:1506.01256v1 [cs.CR] 3 Jun 2015.
- [2] Meral H.M., Sankur B., Ozsoy A. S., Gungor T., Sevinc E. Natural language watermarking via morphosyntactic alterations, // Computer Speech & Language Journal, 2009. - Vol. 23, No 1, pp 107-125.
- [3] Wang J.T., Tsai W.H. Data hiding in PDF files and applications by imperceptible modifications of PDF object parameters. // Proc. of 2008 Conf. on Computer Vision, Graphics & Image Proc., Yilan, Taiwan, Aug. 2008, pp. 24-26.

- [4] Zhong S., Cheng X., Chen T. Data hiding in a kind of pdf texts for secret communication. International // Journal of Network Security, 2007. - Vol. 4(1), pp. 17-26.
- [5] Alizadeh-Fahimeh, F., Canceill-Nicolas, N., Dabkiewicz-Sebastian, S., Vandevenne-Diederik, D. Using Steganography to hide messages inside PDF files, // 2012, SSN Project Report.
- [6] Bitar A.W., Darazi R., Couchot J-F., Couturier R. Blind digital watermarking in PDF documents using Spread Transform Dither Modulation, // Multimes Tools Appl, DOI 10 1007 s11042-015-3034-2,
- [7] Chroni M, A. Fylakis A., Nikolopoulos S.D. Watermarking PDF Documents using Various Representations of Self-inverting Permutations. // arXiv:1501.02686 [cs.MM], 2015.
- [8] Mintzer F.C., Lotspiech J., Morimoto N. IBM, 1997. Safeguarding digital library contents and users: Digital watermarking. // D-Lib Magazin [Online]. Available: <http://www.dlib.org/dlib/december97/ibm/12lotspiech.html>.
- [9] Горбачев В.Н., Кайнарова Е.М., Денисов Л.А. Встраивание бинарного изображения в плоскости Грея. // Компьютерная оптика, 2013. - Т. 37, № 3, - С. 385-390.
- [10] Hu Y, S., Huang S.K.J An algorithm for removable Visible Watermarking.// IEEE. Transactions on circuits and systems video technology. 2006. - Vol. 16, No 1, pp.129-133.

³ Фрагмент страницы из [2]

Визуальные методы оценки и модели данных при проектировании разработки месторождений углеводородов

Д.А. Завьялов, А.А. Захарова, А.В. Шкляр
{zda, zaa, shklyarav}@tpu.ru
лаборатория 3D-моделирования Института кибернетики
Томский политехнический университет, Томск, Россия

В работе представлен новый подход к решению задачи поиска аналогий при проектировании разработки месторождений углеводородов. В основе подхода лежит использование трехмерной визуальной модели данных, позволяющей оценивать многомерный массив разномасштабной информации в комплексе и обосновывать принятие экспертного решения. В статье описано применение предложенного подхода при проектировании разработки нефтегазовых месторождений Томской области.

Ключевые слова: обработка данных, визуальная модель данных, научная визуализация, разработка месторождений.

Visual assessment methods and data models in the development of hydrocarbon fields

D.A. Zavyalov, A.A. Zakharova, A.V. Shklyar
{zda, zaa, shklyarav}@tpu.ru
lab of 3D Modelling of Institute of Cybernetics
Tomsk Polytechnic University, Tomsk, Russia

The paper represents a new approach to solving the problem of searching for similars in the design of hydrocarbon fields. The approach is based on the use of a three-dimensional visual data model that allows to estimate a multidimensional arrays of different-scale information in a complex and to justify the adoption of an expert decision. The article describes the application of the proposed approach in designing the development of oil and gas fields in the Tomsk region.

Keywords: data processing, visual data model, scientific visualization, field development.

1. Введение

При проектировании разработки нефтегазовых месторождений зачастую возникает задача реконструкции недостающих данных. В процессе разработки меняется энергетическое состояние залежи, вследствие чего изменяются значения параметров, принятые ранее. Кроме того, актуальность задачи обусловлена низким уровнем структурированности существующей информации, частым отсутствием данных для проведения прогнозного моделирования из-за высокой стоимости и трудоемкости их получения. Гипотезой решения задачи является предположение о возможности заимствования недостающих данных у объектов, характеризующихся схожими условиями залегания (аналогичных объектов) и, как следствие, близкими значениями ключевых параметров. Таким образом, целью визуального анализа становится быстрый поиск объектов, близких по ряду свойств, при условии, что решение о принципах оценки подобной близости остается за экспертом.

2. Задача поиска аналогий в проектировании разработки месторождений углеводородов

При проектировании разработки месторождений ситуация с отсутствием части собственных исследований не редкость, однако процедура поиска аналогий слабо формализована и во многом субъективна – решение всегда остается за проектировщиком или экспертом и часто базируется на личном опыте и не подкрепляется фактическим материалом. Особенностями рассматриваемой задачи поиска аналогий являются:

- 1) разнородность и разномасштабность анализируемых данных;
- 2) отсутствие строгой формализации и методики решения задачи;

- 3) субъективность получаемого решения (весовые коэффициенты параметров определяются экспертом);
- 4) нестрогость накладываемых ограничений (в случае отсутствия решения, доверительный интервал может быть расширен для увеличения объема обрабатываемой выборки, так как решение должно быть получено в любом случае).

Традиционно решение задачи выбора аналогий можно получить статистически, такой способ связан с обработкой табличных данных по имеющимся месторождениям и выявлением объектов, близких по определенным критериям. Визуальное решение задачи можно получить, например, используя гистограммы отклонений (по каждому параметру) или корреляционные диаграммы (по каждому месторождению), что затратно по времени, так как возникает необходимость анализировать большой объем информации. Примеры традиционного двухмерного решения задачи выбора аналогий показаны на рис. 1.

Кроме того, обозначенные способы решения задачи не позволяют выполнить комплексную оценку всех имеющихся данных в одной модели, что сократило бы время принятия решения. Главная причина этого – невозможность обеспечения достаточной информативности традиционными способами визуального анализа данных на основе плоских двухмерных моделей. В современных условиях постоянного роста объемов информации необходимо использовать многомерные визуальные модели данных [2].

3. Визуальные модели анализа данных

Для решения любой задачи требуется модель, соответствующая типу вырабатываемого визуального решения. Это говорит о необходимости активного привлечения возможностей пользователя при получении ответа на вопрос исследования. Кроме того, одним из

предварительных условий является сокращение времени анализа. Выполнение этого требования может быть обеспечено благодаря упрощению модели до типа, соответствующего обычному информированию, а также благодаря использованию выразительных визуальных средств, не требующих продолжительного знакомства и осмысления [5].

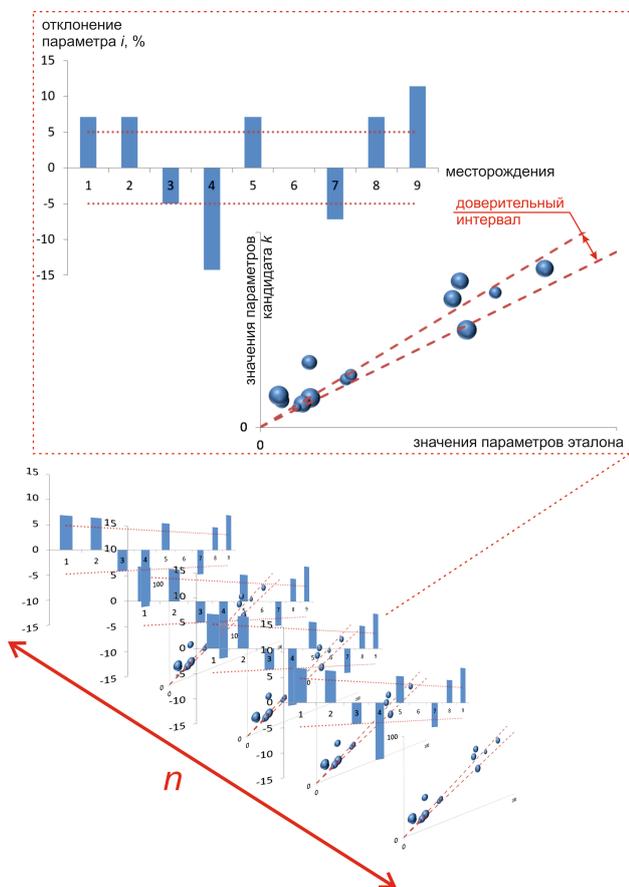


Рис. 1. Визуальное решение задачи выбора аналогий традиционным способом.

Для сокращения времени принятия решения при визуализации значительного числа параметров для визуальной модели выбирается образ, понятный пользователю. По этой причине, в создаваемой модели образ отдельного элемента, объединяющий заданное число параметров, представлен в виде трехмерного графа.

Одним из путей решения проблемы результативного использования времени является сокращение объемов избыточного информирования наблюдателя. С одной стороны, это достигается благодаря анализу информативности полученных образов и устранению элементов, смысл которых не связан с постановкой исследования. Но, кроме этого, возможно привлечение способности человека к мысленной интерполяции и экстраполяции данных, которая делает обоснованным воспроизведение в визуальной модели данных лишь части из них [1].

4. Визуальная модель данных для решения задачи поиска аналогий

В рамках задачи поиска аналогий происходит выбор наиболее близкого месторождения-аналога из выборки кандидатов с учетом весовых коэффициентов параметров и допустимых отклонений параметров кандидатов от эталона. Концептуальная схема предлагаемой визуальной модели данных для решения задачи поиска аналогий при

проектировании разработки месторождений представлена на рис. 2.

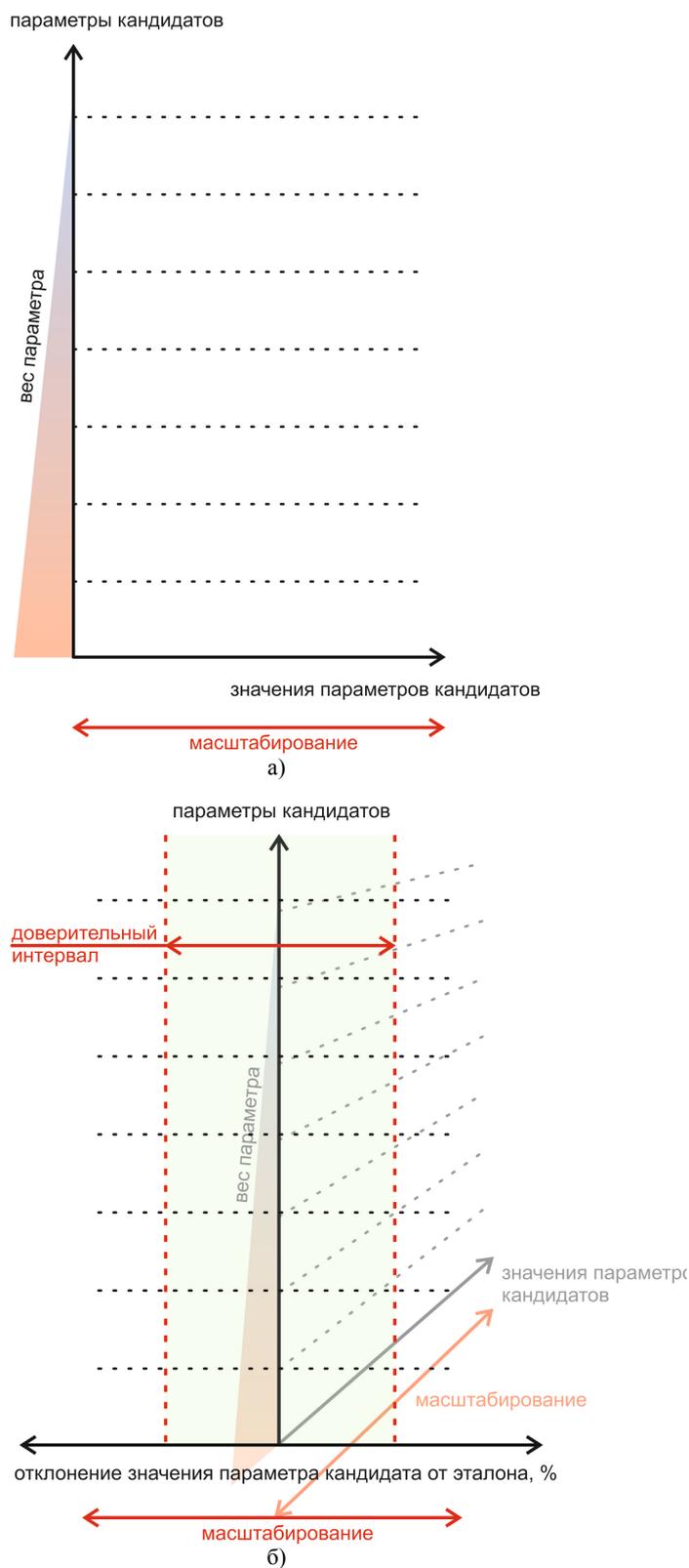


Рис. 2. Концептуальная схема визуальной модели данных.

Параметры месторождений-кандидатов в визуальной модели располагаются в соответствии со своими весовыми коэффициентами. В двухмерном режиме (рис. 2 (а)) модель позволяет выполнять анализ массива абсолютных значений с учетом весовых коэффициентов, при этом для лучшего визуального восприятия параметры масштабируются.

Третье измерение модели (рис. 2 (б)) реализует режим сравнения отклонений параметров месторождений-кандидатов от эталона, допустимое отклонение ограничивается доверительным интервалом. В таком режиме экспертом выполняется визуальная оценка и поиск наиболее подходящего месторождения-аналога, оптимальным решением будет кривая параметров, наиболее близкая (с точки зрения эксперта) к центральной линии (эталону).

В рассматриваемой задаче визуальная модель сопровождается максимально возможной степенью интерактивности системы управления, что позволяет формулировать мысленный вопрос на каждом шаге анализа таким образом, чтобы это соответствовало скорости и особенностям мышления пользователя.

Созданная процедура исследования данных подтверждает положение комплексного подхода к визуализации, в соответствии с которыми анализ имеет структуру последовательного продвижения от образа-вопроса к образу-ответу с привлечением визуальной верификации промежуточных результатов. Окончательное решение, предложенное для задачи анализа в этом случае, характеризуется возможностью быстрого выявления наилучшего ответа при минимальных усилиях. Это позволяет достигать цели исследования, соблюдая алгоритм трансформирования образа при помощи интерактивной системы управления и без привлечения дополнительных источников информации [4].

Помимо оценки массива разнородных данных в одной визуальной модели комплексность в данном случае заключается еще и в возможности выполнения одновременного визуального анализа разных аспектов данных: абсолютные значения, абсолютные и относительные отклонения от эталона, динамика данных.

Для опробования представленного подхода на реальных данных выбрано одно из месторождений Томской области, на котором отсутствуют собственные исследования керна, необходимые для прогнозирования разработки. Такие исследования могут быть использованы с одного из аналогичных месторождений, необходимо найти в исходной выборке кандидатов месторождение-аналог, наиболее близкое к эталону по ключевым параметрам. Исходная выборка включает 18 месторождений-кандидатов. Из всего набора данных (рис. 3) по месторождениям выделены 13 ключевых геолого-физических характеристик, которым экспертно назначены весовые коэффициенты и допустимые отклонения от эталона.

Параметры	Вес (1-5)	Допустимое отклонение, %	Кандидаты																					
			М1	М2	М3	М4	М5	М6	М7	М8	М9	М10	М11	М12	М13	М14	М15	М16	М17	М18	М19	М20		
Плотность воды в поверхностных условиях	1	+5, -5	1,02	1,001	1,001	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	1,021	
Средняя глубина залегания кровли	2	+5, -5	-2406,5	-2478	-2467	-2489	-2437	-2483	-2481	-2490	-2548	-2458	-2496	-2442	-2481	-2486	-2486	-2486	-2479	-2477	-2477	-2477	-2477	-2477
Коэффициент пористости	2	+15, -15	0,66	0,62	0,66	0,67	0,64	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	0,62	
Начальное пластовое давление	2	+15, -15	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	25,5	
Средняя общая толщина	3	+20, -20	13,8	12,6																				
Средняя эффективная нефтенасыщенная толщина	3	+15, -15	6	4,8																				
Вязкость нефти в пластовых условиях	4	+15, -15	0,56	0,61																				
Плотность нефти в поверхностных условиях	4	+8, -8	0,799	0,808																				
Объемный коэффициент нефти	4	+5, -5	1,178	1,229																				
Газосодержание	4	+10, -10	54	71,4																				
Коэффициент пористости	5	+10, -10	0,14	0,15																				
Коэффициент нефтенасыщенности пласта	5	+10, -10	0,58	0,65																				
Проницаемость	5	+10, -10	0,0051	0,005																				

Рис. 3. Исходная выборка данных для поиска аналогий.

Полученная трехмерная визуальная модель данных представлена на рис. 4. Двухмерный режим сравнения абсолютных значений (рис. 4 (а)) демонстрирует отсутствие связи между некоторыми данными (значения параметров откладываются от левой грани модели, горизонтальные оси соответствуют выбранным параметрам и располагаются в соответствии с их весовыми

коэффициентами) вследствие того, что в выборке присутствуют значения параметров различных пластов (пород-коллекторов, вмещающих углеводороды). Трехмерный режим поиска аналогий показан на рис. 4 (б) (центральная ось – эталон, линии – отклонения параметров кандидата от эталона в большую либо меньшую сторону, доверительный интервал обозначен параллелепипедом), здесь обозначена кривая параметров месторождения-аналога (линия с большей толщиной), выбранного проектировщиком в качестве решения, как наиболее близкого к эталону по всей совокупности параметров.

Представленный подход совмещает в одной трехмерной визуальной модели традиционный двухмерный анализ абсолютных значений параметров, анализ отклонений параметров от эталонного, ранжирование параметров по весовым коэффициентам, использование доверительного интервала, масштабирование значений параметров для лучшего восприятия пользователем. Все перечисленное, а также интерактивность трехмерной модели (возможность изменить ракурс и масштаб, как всей модели, так и отдельных ее параметров для оценки локальных эффектов) обеспечивает интуитивное восприятие пользователем больших массивов информации и позволяет сократить время анализа визуальной модели и принятия решений.

Реализованная комплексная трехмерная визуальная модель данных месторождений, представленная на рис. 4, позволила оперативно найти решение задачи поиска аналогий, благодаря чему были восполнены недостающие данные и построены корректные геологическая и гидродинамическая модели месторождения (в соответствии с действующими регламентными документами [6], [7]), а также выполнено прогнозное моделирование разработки, создан проектный документ и успешно пройдена государственная экспертиза.

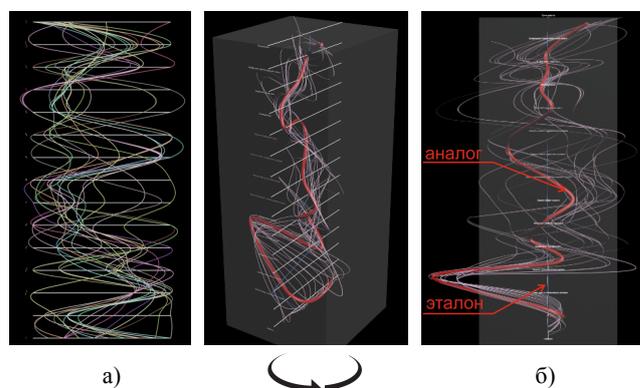


Рис. 4. Визуальная модель данных для задачи поиска аналогий.

Основной эффект от использования представленной комплексной трехмерной визуальной модели данных состоит в том, что одна такая модель заменяет массив двухмерных моделей, которые приходится анализировать эксперту при традиционном подходе (рис. 1), используя ряд универсальных и специализированных программных средств. Модель представляет собой инструмент для анализа данных, который позволяет не только проверить гипотезу, но также сформировать новое знание. По опыту проектирования разработки месторождений можно утверждать, что такой подход при анализе больших объемов данных быстрее и эффективнее традиционных.

Созданный инструмент визуального анализа и получения новой информации, является примером визуальной интерполяции данных благодаря реализации в модели функционального принципа визуальных аналогий. Использование разработанной визуальной модели

позволяет быстро обнаруживать в исследуемых данных объекты, близкие между собой, с точки зрения наблюдателя и признанных им принципов. Такой подход позволяет быстро и эффективно оценивать, анализировать и сравнивать между собой большие объемы данных, что затратно по времени при использовании традиционных способов анализа.

Представленная визуальная модель позволяет решать широкий спектр задач, связанных с анализом больших объемов разнородных данных. К примеру, модель позволяет производить верификацию данных или прогноза разработки месторождений, а также выполнять экспресс-оценку полученного проектного решения с учетом имеющегося опыта разработки аналогичных объектов.

При проектировании разработки месторождения одной из основных задач является оптимальное размещение проектных скважин (динамика и объем добычи определяются геолого-физическими характеристиками пласта в районе размещения скважины) и выбор их режимов работы. На рис. 5 показана визуальная модель, отображающая прогнозные объемы добычи нефти по скважинам одного из месторождений в динамике. В указанной модели горизонтальные оси соответствуют годам разработки месторождения, а в качестве эталона выбрана одна из действующих на месторождении скважин, режим работы которой признан экспертом удовлетворительным. Модель показывает отклонение в объемах добычи нефти проектных скважин от эталонной (действующей). При этом для удобства сравнения произведен сдвиг начала работы скважин к одной точке (при разработке месторождения бурение скважин происходит не одновременно), а доверительный интервал обозначает допустимый режим работы проектных скважин.

Модель позволяет сделать вывод о том, что режим работы части прогнозных скважин является неудовлетворительным, а именно: объемы добычи некоторых скважин (рис. 5 (а)) значительно превышают эталонные значения, что может свидетельствовать о некорректной настройке их режимов работы; динамика добычи других скважин (рис. 5 (б)) говорит о нецелесообразности их бурения – такие скважины не окупят себя, что может привести к отрицательной экономике проекта. Такое проектное решение необходимо доработать [3], [6].

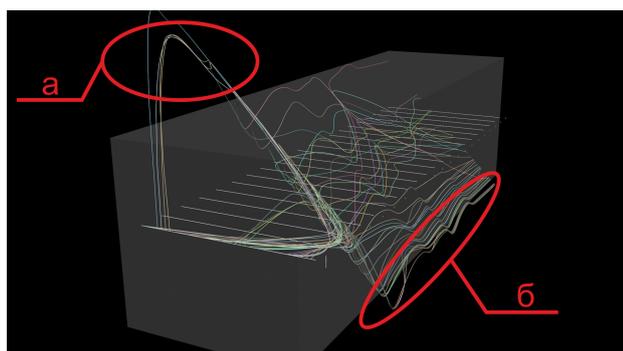


Рис. 5. Оценка режимов работы прогнозных скважин.

5. Заключение

В статье показана актуальность задачи поиска аналогий в процессе проектирования разработки месторождений, предложен новый подход к решению задачи на основе трехмерной визуальной модели данных, которая становится инструментом оперативного принятия решений. Апробация предложенного подхода проведена на реальных данных по месторождениям Томской области в процессе выполнения ряда НИР по геологическому и прогнозному

гидродинамическому моделированию для проектов подсчета запасов и разработки. Предложенный подход позволил оперативно найти решение задачи поиска аналогий, восполнить недостающие данные, получить корректные геологические и гидродинамические модели месторождений, выполнить прогнозное моделирование разработки и успешно пройти государственную экспертизу.

6. Благодарности

Работа выполнена при поддержке гранта Министерства образования и науки РФ № 2.1642.2017/4.6.

7. Литература

- [1] Mehrnoosh Oghbaie, Michael J. Pennock, William Rouse. Understanding the efficacy of interactive visualization for decision making for complex systems // 2016 Annual IEEE Systems Conference (SysCon).
- [2] Бондарев А.Е., Галактионов В.А., Четкин В.М. Анализ развития концепций методов визуального представления данных в задачах вычислительной физики // Журнал вычислительной математики и математической физики. 2011. Т. 51. № 4. С. 669-683.
- [3] Завьялов Д.А. Применение методики оценки рисков разработки месторождения // Молодежь и современные информационные технологии : сборник трудов XIII Международной научно-практической конференции студентов, аспирантов и молодых ученых, г. Томск, 9-13 ноября 2015 г. : в 2 т. — Томск : Изд-во ТПУ, 2016. — Т. 1. — [С. 326-327].
- [4] Захарова А.А., Шкляр А.В. Основные принципы построения визуальных моделей данных на примере интерактивных систем трехмерной визуализации // Научная визуализация. - 2014. - №2. - С. 62-73.
- [5] Захарова А.А., Шкляр А.В., Ризен Ю.С. Измеряемые характеристики задач визуализации // Научная визуализация. - 2016. - №1. - С. 95-107.
- [6] Методические рекомендации по подготовке технических проектов разработки месторождений углеводородного сырья // Минприроды России. Утв. 18.05.2016.
- [7] Методические рекомендации по применению классификации запасов и ресурсов нефти и горючих газов // Минприроды России. Утв. 01.02.2016.

Об авторах

Завьялов Дмитрий – аспирант каф. ИГПД ИК ТПУ. Его e-mail: zda@tpu.ru.

Захарова Алёна Александровна – зав. каф. ИГПД ИК ТПУ, д.т.н. Её e-mail: zaa@tpu.ru.

Шкляр Алексей Викторович – ст. преподаватель ИГПД ИК ТПУ. Его e-mail: shklyarav@tpu.ru.

Решение задач анализа данных с помощью аналитических визуальных моделей

А.А. Захарова, Е.В. Вехтер, А.В. Шкляр
zaa@tpu.ru|vehter@tpu.ru|shklyarav@tpu.ru

Институт кибернетики Томского политехнического университета, Томск, Россия

В работе показаны возможности и преимущества использования авторского метода комплексного визуального анализа, возникающие при решении практических задач исследования данных. Показан результативный подход к построению визуальных моделей, являющихся инструментами визуальной аналитики и обладающих предсказуемой логичностью и контролируемой результативностью. Успешность решения задач анализа обеспечена обоснованным использованием свойств визуальных моделей. Предложена базовая методика проведения аналитического исследования произвольных данных, позволяющая ввести понятие формальной визуальной модели.

Ключевые слова: визуальная аналитика, визуальная модель, анализ данных, интерпретация, визуальное восприятие.

Data analysis problems' solving with a use of analytical visual models

A.A. Zakharova, E.V. Vekhter, A.V. Shklyar
zaa@tpu.ru|vehter@tpu.ru|shklyarav@tpu.ru

Institute of Cybernetics, Tomsk Polytechnic University, Tomsk, Russian Federation

The work shows the features and benefits of using copyright method integrated Visual analysis that arise when solving practical tasks of research data. Shows a productive approach to building Visual models are tools for Visual Analytics and with predictable consistency and controlled performance. The success of decision analysis problems is provided justified using the properties of Visual models. Proposed basic methodology of analytical study of arbitrary data, allowing to introduce the notion of a formal Visual model.

Keywords: visual analytics, visual model, data analysis, visual interpretation, visual perception.

1. Введение

Визуализация является одним из способов моделирования и выполняет функции инструмента познания [1]. Полученный наблюдателем зрительный образ является искусственной формой, находящейся между чувственным восприятием и мышлением. Это позволяет объединить зрительную обработку информации и мысленный анализ данных, хранящихся в памяти исследователя.

Результатом визуального анализа является ответ на вопрос исследования, полученный с помощью визуальной модели в пределах поставленных ограничений [3]. Результативность визуального анализа возрастает в случае реализации комплексного подхода в функции визуального моделирования. Комплексный подход к визуальному анализу данных подразумевает сбалансированное и целенаправленное участие в функции визуального моделирования всех значимых для результата анализа составляющих.

В рамках комплексного подхода, в визуальном анализе выделяют три основные составляющие функции моделирования – функциональную, семиотическую и психоэмоциональную. Построение визуальной модели, обладающей предварительно заданными возможностями, соответствующими особенностям решаемой задачи, на основании комплексного подхода к визуализации данных эквивалентно определению способа и степени активности участия каждой из указанных составляющих.

Структура визуальной модели, предназначенной для проведения анализа данных, в т.ч. для задач с низким уровнем формализации, определяет последовательность и логику рассуждений исследователя. Таким образом, свойства визуальной модели оказывают прямое влияние на получаемое решение [2]. С другой стороны, визуальная модель, позволяющая получить решение таких задач,

становится полноценным аналитическим инструментом, реализующим логику рассуждений исследователя и его понимание изучаемой предметной области. Кроме того, визуальная модель становится воспринимаемой репрезентацией нового знания, возникающего в ходе проводимого анализа.

2. Структура визуального анализа

Анализ, в общем случае, является целенаправленным действием, которое может рассматриваться как последовательность, начинающаяся с постановки вопроса и продолжающаяся поиском гипотез, содержащих ответ на этот вопрос. В таком случае, правильность результата анализа зависит от двух факторов: сути вопроса, которая определяется степенью предварительного понимания анализируемой информации, а также от формы этого вопроса. Форма определяется языком, на котором поставлен вопрос, внутренней структурой, логикой и необходимостью взаимодействия с новыми источниками информации [5, 8].

Анализ, осуществляемый с помощью визуального образа исследуемых данных, можно представить аналогичным образом. Сам образ в момент начала анализа является моделью исходных данных. Вместе с моделью или на основании ее изучения у пользователя формируется вопрос дальнейшего исследования. Возникает циклически повторяемая ситуация, для которой характерно наличие поисковой и верифицирующей процедур. Все следующие действия наблюдателя направлены на построение и оценку гипотезы промежуточного ответа на основании информации, представленной в визуальной модели и, вероятно, с использованием знаний пользователя [14]. Следовательно, многошаговый последовательный процесс визуального анализа состоит в переходе от визуального образа-вопроса, являющегося формулировкой решаемой задачи, к мысленному образу-ответу, интерпретируемому

пользователем как обнаруженная закономерность в исследуемых данных (Рис. 1).

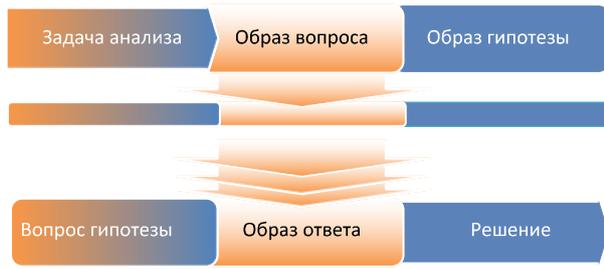


Рис. 1. Структура визуального анализа.

3. Структурная единица визуального анализа

Введем определение структурной единицы визуального анализа – элемента визуальной модели. Структурной единицей визуализации (визуальным элементом) будем считать состояние визуальной модели, интерпретация которого предоставляет наблюдателю объем информации, необходимый для последующего получения общего результата проводимого анализа. Согласно приведенному описанию структуры визуального анализа, такое состояние является визуально воспринимаемым образом, интерпретируемым как однозначный ответ на промежуточный вопрос. Сложность такого вопроса определяется ограниченностью времени, предоставляемого процедурой анализа, а также ресурсоемкостью процессов создания структурной единицы и верификации сведений, полученных пользователем, благодаря ее изучению.

Структурная единица визуального анализа представляет собой управляемую систему S с обратной связью $R(t)$. Объем исследуемых данных поступает на неуправляемый вход $V(t)$ структурной единицы (Рис. 2). Особенностью такой системы является возможность изменения состояния управляемого входа системы $U(t)$ в зависимости от полученных результатов $Y(t)$. Пользователь P является, с точки зрения комплексного подхода к визуальному анализу, обязательным участником структурной единицы.

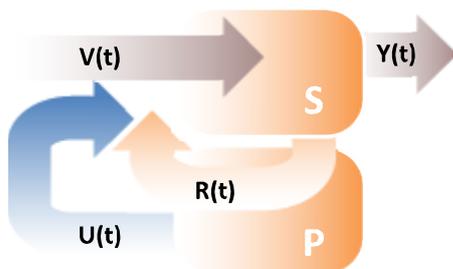


Рис. 2. Структурная единица визуального анализа.

4. Комбинированная модель

В ряде случаев, например, при исследовании сложных и объемных данных [7], также как при изучении особенностей изменяющихся состояний, может оказаться недостаточным использование одной визуальной модели. Это происходит в тех случаях, когда визуальный образ не может, по любым причинам, содержать в себе всего объема необходимых сведений, либо, если технически это возможно, но адекватная интерпретация вызывает у пользователя затруднения. Решением, позволяющим найти выход из этого затруднения, может быть построение набора визуальных моделей, каждая из которых соответствует некоторому фрагменту исходного объема данных. В результате, происходит построение сложной визуальной модели, имеющей более широкие возможности.

Визуальные модели различаются по своим возможностям, назначению и применимости для решения различных видов задач анализа. На основании предложенного описания структурной единицы визуального анализа, тип визуальной модели характеризуется параметрами активности использования информационных входов и выходов единиц, составляющих модель. В одном из простейших случаев, визуальный образ используется как индикатор состояния наблюдаемой системы. Целью подобной модели данных является оповещение наблюдателя об изменении состояния или, точнее, о наступлении ожидаемого события.

Визуальные модели данных получили широкое применение в системах поддержки принятия решений, обладающих различным уровнем сложности. В этом случае, задача анализа требует быстрого осмысления поступающих данных, формирования одной или нескольких гипотез ответа на вопрос исследования и выбор наилучшей из них. Для этого типа визуального анализа характерно активное использование дополнительных сведений и знаний пользователя при формулировании промежуточных вопросов и получении выводов, оказывающих влияние на дальнейшее существование исследуемой системы.

Объединение структурных единиц в функционирующую систему, приводит к построению визуальной модели, когнитивное значение которой превосходит результативность исследования отдельных образов. Необходимо наличие правила, объединяющего отдельные элементы в единую систему и обеспечивающего получение ожидаемого результата анализа. Для изменяющихся данных, таким правилом может быть построение хронологической последовательности, позволяющей наблюдать и сравнивать динамические характеристики системы. Однако, использование в качестве объединяющего правила естественных закономерностей, связанных с происхождением данных, не является обязательным.

Эмерджентность визуальной модели обеспечивается связями составляющих ее элементов, поэтому особое значение приобретает выделение и описание именно этих свойств структурных единиц. Условно, они могут быть разделены на две функциональные группы: информативные и управляющие связи.

Управляющие связи структурных единиц, ответственные за верификацию появляющихся гипотез в визуальных моделях, содержащих совокупность анализируемых образов, могут, а иногда и обязаны учитывать при выполнении своих функций всю информацию, полученную пользователем на уже пройденных этапах анализа. С точки зрения повышения достоверности результатов проводимого анализа, это обстоятельство является положительным достижением. При этом, более сложная процедура проверки может быть связана с дополнительными затратами времени, являющегося одним из основных критериев результативности визуальной модели, и это потребует уточнения разработанной системы образов для достижения сбалансированного решения.

В описанной таким образом логической структуре визуальной модели наряду с последовательностью наблюдаемых пользователем образов данных формируются еще два функциональных объекта, участвующих в визуальном анализе. Первым является цепочка промежуточных вопросов, руководящих процессом исследования данных и ответственных за его логичность и обоснованность. Вторым объектом становится управляющая последовательность, состоящая из процедур верификации, активно использующих возможности

визуального восприятия и эмоциональной оценки наблюдаемых исследователем образов. Успешность применения модели для решения задачи анализа достигается сбалансированным и полноценным использованием всех логических элементов, образующих структуру функционирования визуальной модели.

5. Аналитическая модель

Структура визуальной модели, предназначенной для проведения анализа данных в задачах с низким уровнем формализации, определяет последовательность и логику рассуждений исследователя (Рис. 3). Таким образом, свойства визуальной модели оказывают прямое влияние на получаемое решение. Следует обратить внимание на то обстоятельство, что результативность составляющих визуальной модели (функциональной, информативной и управляющей) зависит от использования определенных возможностей пользователя [4].

В случае сложной модели, последовательное приближение к решению задачи приводит к непрерывному и обязательному присутствию пользователя в процессе анализа. Таким образом, полнота логической схемы функционирования визуальной модели не может быть достигнута без включения в нее четвертого обязательного элемента – возможностей конкретного пользователя, влияющих на результат анализа на протяжении прохождения всей последовательности образов.

Наличие интерактивного управления визуальным образом обеспечивает прямое участие пользователя в манипулировании образом и является основанием для глубокого анализа данных [9]. Обоснованная система управления создает условия для постановки новых вопросов исследования и быстрое получение ответов, ускоряя достижение цели анализа. Таким образом, дополнение визуальной модели управляющим интерфейсом, соответствует предложенной структуре визуального анализа и позволяет сделать его инструментом аналитического решения любой поставленной задачи.

Основной идеей, обеспечивающей увеличение когнитивной результативности визуального анализа, становится направленное использование возможностей пользователя, реализуемое благодаря свойствам визуальной модели. Это создает возможность привлечения дополнительного ресурса для решения поставленной задачи анализа [10]. Перспективными направлениями в этом случае становятся привлечение не только информированности пользователя, но и особенностей его мышления, как для формирования гипотезы решения, так и для принятия решения о ее достоверности.

В случае исследования данных, относящихся к вопросам, не имеющим полноценного формального описания структуры соответствующих знаний, использование визуальных моделей становится не только способом обнаружения новых закономерностей, но и формирует понимание их смысла. Следовательно, приобретение визуальной моделью статуса аналитического инструмента возможно лишь в том случае, когда созданы

условия для интерактивного управления способом представления данных с ее помощью. Визуальная модель, предлагающая наблюдателю на первоначальном этапе исследования образ не только исходных данных, но и обнаруженных закономерностей, может использоваться в качестве *формальной визуальной модели* области знаний.

Анализ логической структуры визуальной модели, полученной в результате приведенных выше рассуждений, позволяет дать ответы на вопросы, относящиеся к увеличению результативности использования визуального анализа [13]. Основными критериями эффективности визуальной модели выступают два параметра: достоверность полученного результата и время, затраченное на его достижение. Наличие в полной схеме, как минимум, четырех функциональных элементов и их взаимное влияние приводит к необходимости определения правил, соблюдение которых необходимо для получения наиболее эффективной комбинации. С учетом динамичного характера проведения процедуры анализа, соблюдение и адаптация этих правил к конкретным условиям становятся еще одной целью интерактивного управления визуальной моделью.

Система интерактивного управления моделью является одним из способов верификации полученных решений и потому обеспечивает ускоренное достижение цели анализа в случае, когда этот способ удобен для пользователя. Интерактивность визуальной модели становится условием ее высокой результативности [6, 11].

Комбинирование единичных элементов позволяет получать информативные, визуально воспринимаемые объекты, которые могут быть использованы для решения самых разных задач. В их число входят визуальное представление информации, целенаправленное использование когнитивного потенциала исследователя, активизируемое воздействием образного восприятия, исследование данных, не имеющих формального описания или объяснения. Особый интерес представляет задача построения визуальных моделей, содержащих данные целых предметных областей. Это позволяет проводить визуальный анализ данных, имеющих различное происхождение, достоверность, формат и смысл, с целью их обобщения, верификации и поиска объяснения.

Таким образом, аналитический потенциал визуальной модели определяется выбором способов визуального моделирования изучаемых данных, максимально возможной скоростью построения зрительного образа при сохранении всех необходимых, с точки зрения функциональности, нюансов визуального представления [12]. Кроме этого, возможно, наиболее значимым элементом модели становится ее интерфейс, необходимый для использования информативного и когнитивного потенциала самого пользователя.

6. Исследование предметной области

В качестве практической задачи, для исследования результативности использования визуальных моделей при изучении многомерных данных, была решена задача



Рис. 3. Структура аналитической модели.

представления общей совокупности знаний в предметной области, содержащей экспериментальные сведения. Осложняющими обстоятельствами для подобной задачи стали большое число разнородных источников информации (публикации) и различающиеся структуры описаний сравниваемых данных. Источником исходной информации в базе исследуемых данных служили статьи, опубликованные в мировых рецензируемых изданиях.

В результате, построена *общая визуальная модель* эмпирических данных о современном состоянии изучения процессов получения карбида кремния электродуговым методом (Рис. 4). Модель представляет собой набор горизонтальных плоскостей, каждая из которых соответствует одной из характеристик исследуемых объектов. Линия, проведенная через точки, соответствующие одному объекту становится визуальным образом, характеризующим этот объект. Множество таких визуальных объектов (трек-образов) составляет визуальную модель, позволяющую проводить сравнение объектов между собой и анализировать весь объем исследуемых данных. Одним из значимых результатов проведенного анализа стало определение стратегии развития предметной области. Кроме того, проведена оценка достоверности опубликованных данных и изменений областей концентрации внимания исследователей во времени.

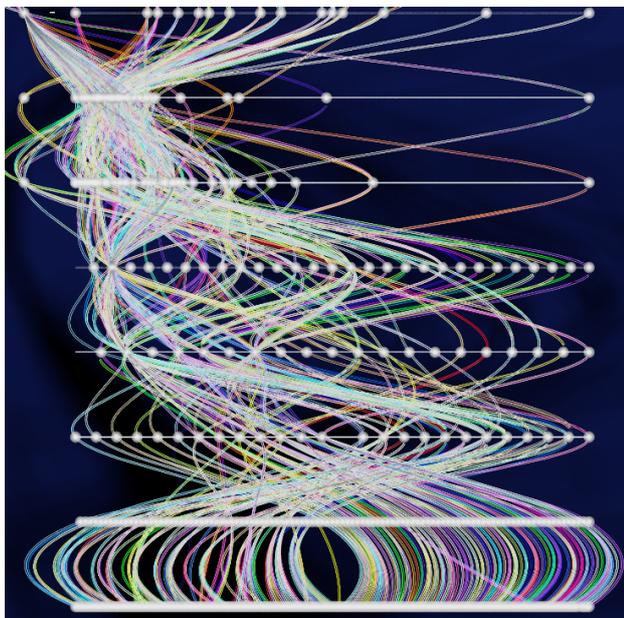


Рис. 4. Визуальная модель эмпирических данных.

С помощью построенной визуальной модели проанализированы сведения о 260 экспериментах, посвященных получению углеродных ультрадисперсных материалов в плазме электрической дуги постоянного тока. Получены выводы о росте объема экспериментальных данных преимущественно в отношении получения четырех наиболее популярных продуктов. При сравнении групп отдельных параметров (показатели эксперимента – получаемый продукт) обнаружено, что производительность рассматриваемых экспериментальных систем практически не оценивается за исключением обрывочных несистематизированных данных, что говорит о неготовности исследуемого метода к внедрению в промышленное производство [14].

Значительный интерес представляет использование фрагментированных моделей, представляющих образы комбинаций отдельных параметров изучаемых данных [13]. Создание таких моделей соответствует введенному

понятию декомпозиции исходной задачи анализа. Кроме того, построение подобных «частичных» моделей обеспечивает пользователя ответами на самостоятельно сформулированные промежуточные вопросы. Полученная исследователем новая информация позволяет формировать вероятные гипотезы ответов на основной вопрос исследования, принимать решение о переходе к следующему шагу анализа или проводить поиск внутренних закономерностей, неявно присутствующих в исходных данных.

Модель соответствия между годами исследований и напряжением, характеризующим отдельный эксперимент (Рис. 5), позволяет провести анализ изменений мировых тенденций и перейти к формулировке вопроса об их причинах. Преимуществом подобной модели становится быстрое определение особенностей данных (в приведенном примере, выделение значения напряжений, при которых проведено наибольшее число исследований на протяжении длительного времени). Одним из следствий подобного алгоритма исследований становится обоснованное исключение из анализа части данных, не представляющих интерес на основании промежуточных выводов. Этот шаг увеличивает результативность анализа при проведении исследований больших объемов данных.

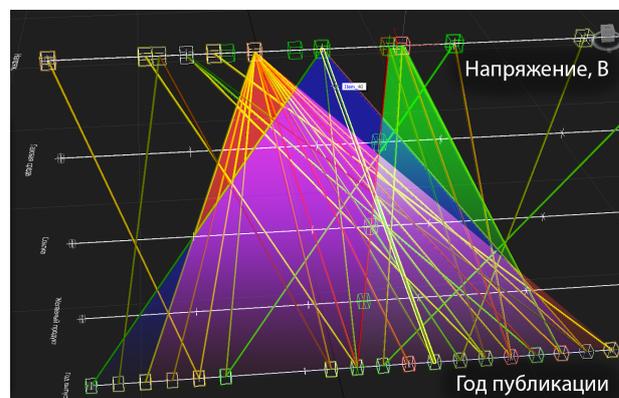


Рис. 5. Фрагментированная визуальная модель.

В результате проведенного исследования были подтверждены основные предположения, описывающие структуру визуального анализа. Благодаря использованию интерфейса управления визуальной моделью, достигнута высокая результативность анализа. Это связано с сокращением времени анализа, по сравнению с традиционными способами изучения подобных данных, для которых время исследования измеряется несколькими днями, даже с учетом опыта и квалификации специалистов.

По совокупности представленных сведений в отношении анализа диаграммы энергетических характеристик получен вывод о том, что полем для дальнейших исследований метода синтеза углеродных материалов в плазме дуги постоянного тока должна стать взаимосвязь уровня напряжения и электрической мощности с другими параметрами экспериментальных установок для получения материала с заданным фазовым составом и свойствами в контексте требований к электрической мощности системы и объема потребления электроэнергии. Основаниями для подобных выводов являются точки концентрации данных на определенных величинах выбранных параметров (Рис. 6), а также наблюдаемые в модели явные области недостаточной изученности.

7. Заключение

Предложенный в работе подход к решению задач анализа с помощью визуальных моделей данных позволяет

оперативно получать ответы на вопросы исследования, в том числе, высокой степени сложности. Построение модели, предоставляющей возможность визуализировать исходные данные и формулировать промежуточные вопросы с целью поиска неизвестных ранее внутренних закономерностей, дает пользователю аналитический инструмент для исследования поставленной перед ним задачи. Помимо возможности достичь цели анализа с помощью этого инструмента, пользователь получает возможность существенно сократить время, затрачиваемое им на поиск ответа или принятие решения в ситуации выбора наилучшего варианта. Сделанные утверждения проверены при анализе опубликованных данных, посвященных получению углеродных ультрадисперсных материалов.

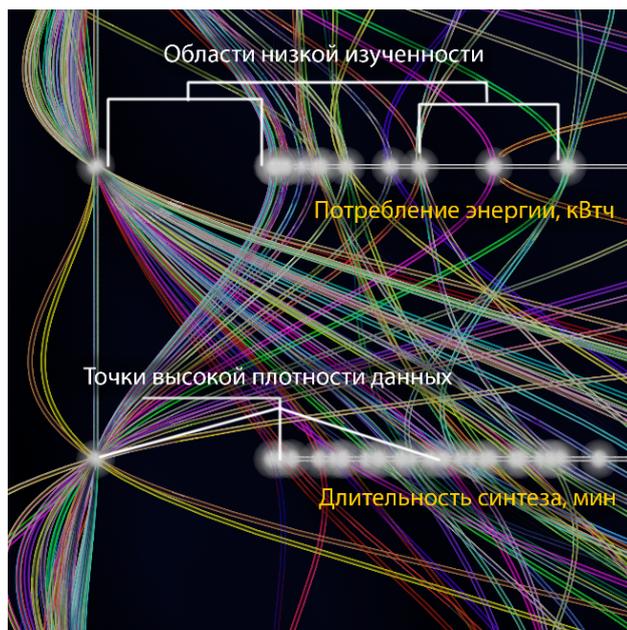


Рис. 6 Визуальный поиск закономерностей.

Работа выполнена в рамках госзадания № 2.1642.2017/ПЧ на выполнение проекта по теме «Когнитивные методы визуализации и анализа многомерных данных при моделировании нелинейных динамических систем».

8. Литература

- [1] Bondarev A.E., Galaktionov V.A.: Multidimensional data analysis and visualization for time-dependent CFD problems. *Programming and Computer Software* 41(5), 247–252 (2015). DOI: 10.1134/S0361768815050023.
- [2] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, Hanspeter Pfister, What Makes a Visualization Memorable?, *IEEE Transactions on Visualization and Computer Graphics*, v.19 n.12, p.2306-2315, December 2013 [DOI: 10.1109/TVCG.2013.234]
- [3] Chen, C.: *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. 2nd ed. London: Springer, (2013).
- [4] Chen C.: Top 10 unsolved information visualization problems. *IEEE Computer Graphics and Applications* 25(4), 12–16 (2005).
- [5] Eppler, M., Burkhard, R.A.: Visual representations in knowledge management: Framework and cases. *Journal of Knowledge Management*, vol. 11(4), pp. 112-122.

QEmerald Group Publishing Limited (2007). ISSN 1367-3270 DOI 10.1108/1367327071076275.

- [6] Huang, D., Tory, M., Adriel Aseniero, B., Bartram, L., Bateman, S., Carpendale, S., Tang, A., Woodbury, R.: Personal visualization and personal visual analytics. *IEEE Trans. Vis. Comput. Graph.* 21, 420–433 (2015).
- [7] Klimentov, A.; Buncic, P.; De, K.; Jha, S.; Maeno, T.; Mount, R.; Nilsson, P.; Oleynik, D.; Panitkin, S.; Petrosyan, A.; Porter, R. J.; Read, K. F.; Vaniachine, A.; Wells, J. C.; Wenaus, T.; Iop, Next Generation Workload Management System For Big Data on Heterogeneous Distributed Computing. *Proceedings from the 16th International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, Prague, Czech Republic, September 1–5, 2014; Iop Publishing Ltd: Bristol, U.K., 2015; Vol. 608.
- [8] Korobkin D., Fomenkov S., Kravets A., Kolesnikov S., Dykov M.: Three-Steps Methodology for Patents Prior-Art Retrieval and Structured Physical Knowledge Extracting. In: *Creativity in Intelligent Technologies and Data Science First Conference, CIT&DS 2015*, pp.124-138. Volgograd, Russia (2015).
- [9] Ozpeynirci, O., Ozpeynirci, S., Kaya, A.: An interactive approach for multiple criteria selection problem. *Computers and Operations Research* 78, 154-162 (2016).
- [10] Sedig Kamran, Parsons Paul, Hai-Ning Liang, Jim Morey: Supporting Sensemaking of Complex Objects with Visualizations. *Visibility and Complementarity of Interactions*, *Informatics* 3(4), 20 (2016). doi:10.3390/informatics3040020.
- [11] Shneiderman, B., Plaisant, C., Cohen, M.S., Jacobs, S.M., Elmqvist, N., Diakopoulos, N.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 6th ed. Pearson: Upper Saddle River, NJ, USA (2016).
- [12] Zakharova, A.A., Shklyar, A.V.: Visual presentation of different types of data by dynamic sign structures. *Scientific Visualization* 8(4), 28-37 (2016).
- [13] Zaripova V.M., Petrova I. Yu., Kravets A., Evdoshenko O.: Knowledge Bases of Physical Effects and Phenomena for Method of Energy-Informational Models by Means of Ontologies. In: *Creativity in Intelligent Technologies and Data Science First Conference, CIT&DS 2015*, pp. 224-237. Volgograd, Russia (2015).
- [14] Zavyalov, D.A., Zakharova, A.A., Shklyar, A.V., Bagutdinov, R.A.: An integrated approach to modeling by an example of a landfill of disposal of liquid oil waste. *Software systems and computational methods [Programmnye sistemy i vychislitel'nye metody]* 1, 22-30 (2017).

Об авторах

Захарова Алена Александровна, Институт кибернетики, Томский политехнический университет, Томск, Россия, zaa@tpu.ru

Вехтер Евгения Викторовна, Томский политехнический университет, Томск, Россия, vechter@tpu.ru

Шкляр Алексей Викторович, Институт кибернетики, Томский политехнический университет, Томск, Россия, shklyarav@mail.ru

Анализ видеопоследовательностей и термограмм лица для дистантного съема физиологических индикаторов психоэмоциональных состояний человека

И.А. Знаменская¹, Е.Ю. Коротева¹, В.В. Шишаков¹, А.В. Хахалин¹, Е.А. Кузьмичева¹, С.А. Исайчев²,
А.М. Черноризов²

znamen@phys.msu.ru|koroteeva@physics.msu.ru|shift@physics.msu.ru

¹Московский Государственный Университет им. М.В. Ломоносова, физический факультет, Москва, Россия;

²Московский Государственный Университет им. М.В. Ломоносова, факультет психологии, Москва, Россия

Данная работа направлена на установление взаимосвязи между динамикой теплового потока от лица человека, регистрируемого тепловизионной камерой, и физиологической оценкой реакций вегетативной нервной системы, измеренной контактными методами. Решается задача идентификации эмоционального возбуждения по активности мимических мышц, поверхностно залегающих кровеносных сосудов и дыхательных потоков на основе съемки лица человека в оптическом и инфракрасном диапазонах. Разработаны алгоритмы компенсации движений лица и сопоставления изображений с видеокамеры и тепловизионной камеры, а также составлено программное обеспечение, реализующее предложенные алгоритмы. Предложен метод выделения ключевых точек на лице человека на основе машинного обучения.

Ключевые слова: инфракрасная термография, обработка изображений, распознавание образов.

Analysis of video sequences and thermal images of human faces for remote detection of psychoemotional states

I.A. Znamenskaya¹, E.Yu. Koroteeva¹, V.V. Shishakov¹, A.V. Khakhalin¹, E.A. Kuzmicheva¹, S.A. Isaichev², A.M. Chernorizov²

znamen@phys.msu.ru|koroteeva@physics.msu.ru|shift@physics.msu.ru

¹Lomonosov Moscow State University, Faculty of Physics, Moscow, Russia;

²Lomonosov Moscow State University, Faculty of Psychology, Moscow, Russia

This work is focused on estimating the relationship between the dynamics of heat fluxes from a human face captured by a thermal camera, and the physiological response of the autonomic nervous system registered by contact methods. The emotional arousal is identified based on the activity of facial muscles, cutaneous blood vessels and breathing using optical and infrared imaging of human faces. The algorithms are developed for motion compensation and juxtaposition of video and thermal images, together with the software implementing these algorithms. The machine-learning based method is proposed to allocate the key points on human faces.

Keywords: infrared thermography, image processing, pattern recognition.

1. Введение

Проблема объективной диагностики функциональных состояний (ФС) человека в видах деятельности, связанных с высоким риском возникновения техногенных катастроф, становится не только актуальной, но и социально значимой проблемой. Роль надежности «человеческого фактора» значительно возрастает, если периоды активной деятельности, требующей быстрого принятия решения, сменяются длительными периодами монотонной работы. Такой тип деятельности характерен для большинства операторских профессий, водителей транспортных средств, военнослужащих, работников охранных и силовых структур, т.е. там, где велика вероятность возникновения экстремальных и чрезвычайных ситуаций.

Особый интерес вызывают научно-прикладные исследования, связанные с разработкой неинвазивных методов объективной идентификации ФС в режиме реального времени. Разработка методов, оптико-электронной аппаратуры и программных средств идентификации психо-эмоциональных состояний по индивидуальному особенностям поведения и внешнего облика человека является актуальной задачей для ученых и практиков в России и за рубежом.

Целью данной работы является разработка технологии дистантного съема физиологической информации о психо-эмоциональных состояниях человека (спокойное бодрствование, эмоции, стресс) на основании исследований

активности мимических мышц, кожных покровов, поверхностно залегающих кровеносных сосудов лица, а также дыхательных (газовых) потоков. Ставится задача идентификации эмоционального возбуждения по активности мимических мышц, регистрируемой в оптическом и инфракрасном (ИК) диапазонах. Задача осложнена необходимостью регистрации в условиях непрерывного смещения поверхности лица относительно сканирующих устройств – видеокамеры и тепловизора. В работе данная проблема решается путем комбинации методов регистрации в видимом и ИК диапазонах и за счет разработки и адаптации программы анализа динамически меняющихся зрительных сцен.

2. Термография в психофизиологии

Термография представляет собой способ измерения и визуализации теплового, инфракрасного излучения, испускаемого всеми нагретыми телами. Основным условием для формирования ИК изображения является наличие температурного контраста между объектом и фоном. Особенности излучения кожи человека определили широкое использование термографии в физиологической и биомедицинской диагностике [9]. Кожа имеет высокий коэффициент излучения, близкий по значению к коэффициенту излучения абсолютно черного тела, поэтому изменение ее температуры приводит к значительному изменению мощности ИК излучения. Кроме того, низкий

коэффициент отражения кожи минимизирует влияние окружающей среды на определение ее температуры.

Наиболее доступным для тепловизионных измерений является область лица. При неизменных внешних условиях поверхность лица физически здорового человека имеет неоднородное распределение температуры, определяемое физиологическими особенностями организма. Локальное изменение температурных полей в области лица может быть связано:

- 1) с активностью мимических мышц;
- 2) с расширением кровеносных сосудов лица;
- 3) с динамикой дыхательных потоков (как показали эксперименты [8]);
- 4) с усилением потоотделения в определенных зонах лица.

Отдельные области на лице человека обладают наибольшей температурной чувствительностью к различным внутренним и внешним раздражителям [3]. Такие области на термограмме называют областями интереса (Рис. 1).

3. Методика экспериментов

В работе проводились эксперименты по одновременной бесконтактной регистрации динамики теплового излучения от лица человека и контактной регистрации активности вегетативной нервной системы.

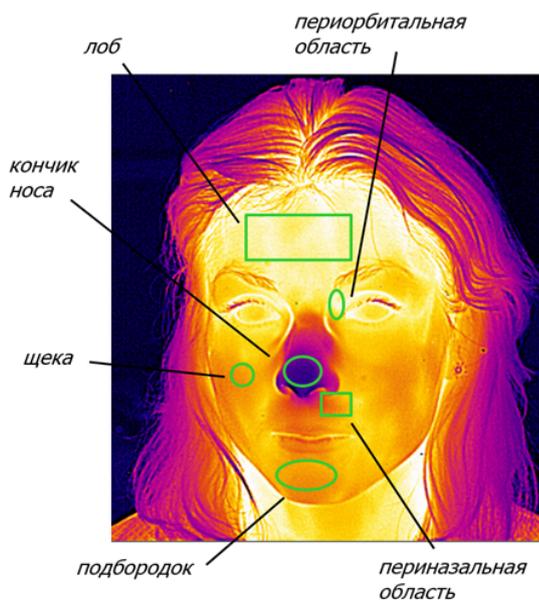


Рис. 1. Термограмма лица человека с выделенными областями интереса.

Для проведения инфракрасной съемки в экспериментах были использованы (по отдельности) две тепловизионные камеры. Первая камера марки FLIR SC7700 работает в диапазоне длин волн 3.7–4.8 мкм (MWIR) и позволяет получать тепловизионные изображения с частотой до 115 Гц с пространственным разрешением 640x512 пикселей и до 400 Гц с ограниченным разрешением. Управление камерой и обработка термограмм проводятся при помощи программного обеспечения Altair и FLIR ResearchIRMax. Вторая камера модели COX CX640 работает в диапазоне длин волн 8–14 мкм (LWIR) и позволяет получать тепловизионные изображения с частотой до 50 Гц с пространственным разрешением 640x480 пикселей. Для получения и анализа термограмм используется программное обеспечение Thermal Imaging Analyzer. Одновременно проводилась съемка в видимом диапазоне цифровой камерой Olympus OM-D E5.

В экспериментах принимали участие порядка 25 здоровых испытуемых (мужчин и женщин) в возрасте от 18 до 55 лет. Все измерения проводились в положении испытуемых сидя, после их адаптации к лабораторным условиям. Тепловизор и видеокамера располагались на расстоянии порядка 1 м от испытуемого. Время съемки варьировалось от 20 с до 15 мин. Частота съемки составляла 5–25 Гц. В помещении поддерживалась фиксированная комнатная температура (20–22°C).

Для анализа корреляции термографических измерений с оценкой психоэмоционального состояния человека синхронно с тепловизионной и видеосъемкой осуществлялась регистрация физиологических реакций человека при помощи многоканального энцефалограф-анализатора Энцефалан. Ход проведения эксперимента представлен на Рис. 2.

4. Алгоритм анализа термограмм

Для корректного сопоставления тепловизионной и видеосъемки необходимо приведение теплового поля на лице человека к стандартному представлению. В ходе решения данной задачи были разработаны алгоритмы компенсации движений лица и сопоставления изображений с видеокамеры и тепловизионной камеры, а также составлено программное обеспечение, реализующее предложенные алгоритмы. Кроме того, предложен метод выделения ключевых точек на лице человека на основе машинного обучения.

В использованном диапазоне длин волн контур человеческой головы или лица имеет весьма четкие границы из-за высокой разницы температуры фона и тела человека, однако различить отдельные черты лица может быть затруднительно из-за низкого контраста такого изображения в пределах самого лица. Изображение лица в видимом диапазоне имеет более выраженную детализацию, позволяющую точно определить расположение на изображении областей интереса (Рис. 1), компенсируя возможные искажения, такие как поворот головы, мимические искажения и т.п., а также точно оценить параметры этих искажений. Для этого требуется решение двух задач: 1) поиск лица на изображении по базовым контурам и точкам (глаза, нос); 2) детализированное описание параметров лица (контуров глаз, век, бровей, губ, крыльев носа).



Рис. 2. Часть экспериментальной установки (вид сверху).

Обе функциональности реализованы в широком спектре биометрического программного обеспечения, а также в программных инструментариях разработчика (SDK), таких как OpenCV [5]. Обнаружение лица на изображении выполняется с помощью каскадных фильтров Хаара. Более

детальная оценка поворота лица, обнаружение биометрических (мимических) контрольных точек на лице осуществляется с помощью активных моделей внешнего вида (ActiveAppearanceModel, AAM) и методом Активных Моделей Фигур (ActiveShapeModels, ASM).

Совмещение данных моделей AAM/ASM с тепловизионным изображением позволяет более точно определить расположение нужных точек контроля температуры на тепловизионном изображении. Эти методы позволяют также улавливать мимику человека, и использовать эту информацию для оценки его эмоционального состояния. Применение данных методов на видеоряде позволяет отслеживать траектории контрольных точек в пространстве и анализировать динамические характеристики мимики. Использование стереокамеры позволяет в ближней зоне более детально оценить рельеф поверхности лица, а в дальней зоне – оценить расстояние до исследуемого объекта. Алгоритмы восстановления карты глубины сцены реализованы во множестве инструментариев для разработчика (в т.ч. OpenCV).

Для анализа термографических изображений в данной работе использовалось программное обеспечение на языке Python с использованием библиотек OpenCV, Dlib и OpenFace [2,4,5]. Данные библиотеки, используя метод Виолы-Джонса [6,7], позволяют выделить лицо на изображении, а также найти характерные ключевые точки лица (глаза, губы, нос и т.д.). Далее, методом CLNF [1] на изображении уточняется расположение, форма и пространственная ориентация всех черт лица, а также положение головы относительно камеры. Это позволяет компенсировать движение головы относительно камеры, выделить на изображении лица конкретную точку, следить за изменением температуры в этой конкретной точке, а также оценить, видны ли эти участки на изображении, или нет. Данные алгоритмы применялись как к тепловизионным видеосъемкам, так и к визуальным

(Рис. 3). При одновременной съемке тепловизионной и обычной цифровой камерой для синхронизации видеоряда использовалось моргание глаз.

5. Предварительные результаты

Была проведена серия экспериментов по заранее подготовленным сценариям, содержащим интервалы, в которых испытуемые находились в спокойном состоянии, и интервалы, в которых моделировались стрессовые условия. На Рис. 4 представлен пример сопоставления динамики теплового потока от области интереса (кончик носа) на лице человека с результатами записи кожно-гальванической реакции (КГР). КГР – биоэлектрическая реакция, которая широко применяется в психофизиологических исследованиях в качестве высокочувствительного показателя уровня активности симпатической нервной системы, а также для оценки нейropsychического напряжения человека. На Рис. 4 видно, что стрессовая реакция на резкий звук приводит к локальному повышению, а интенсивная ментальная нагрузка – к локальному снижению температуры в области интереса (наиболее вероятно, связанному с ее охлаждением за счет усиления активности потовых желез).

Изначально планировалось, что использование дополнительного синхронного видеоряда в визуальном диапазоне поможет более точно оценивать положение контрольных точек лица на изображении и лучше компенсировать движение и повороты лица относительно камеры. Однако в дальнейшем, после накопления определенного количества синхронных записей с разными испытуемыми, удалось составить обучающую выборку из ИК-изображений и информации о распознанных лицах с визуального видеоряда, что позволило «натренировать» библиотеку OpenFace работать с ИК-изображениями отдельно, искать положение черт лица и оценивать ориентацию головы без подсказки с визуального канала.

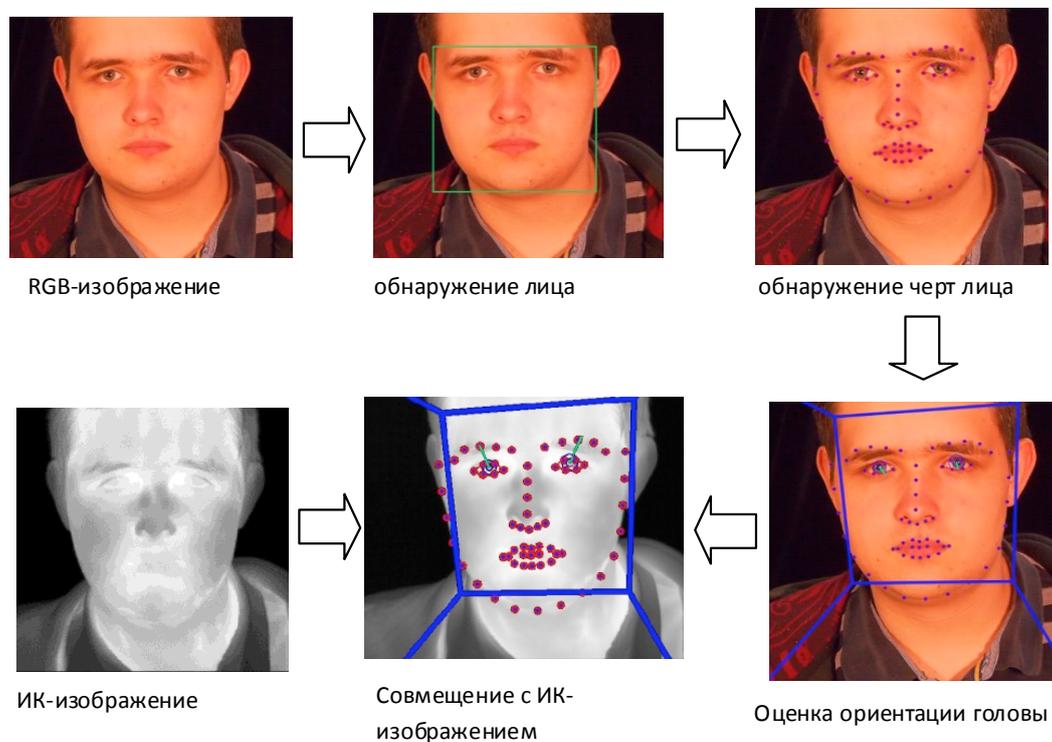


Рис. 3. Схема работы алгоритма анализа изображения лица, выделение черт и оценка поворота головы совместно на ИК и визуальном видеоряде.

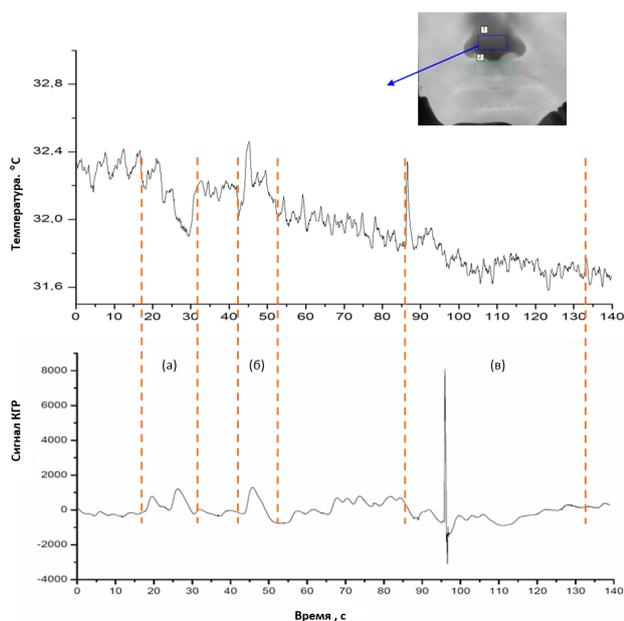


Рис. 4. Сопоставление динамики среднего температурного сигнала в области интереса (выделена на термограмме) и сигнала КГР во время прохождения теста: (а) глубокий вдох; (б) резкий звук; (в) ментальная нагрузка.

6. Заключение

В результате работы создан программный модуль для сопоставления изображения с видеокamеры и тепловизионной камеры. Исследована активность поверхности кожи лица человека в оптическом и инфракрасном диапазонах. Полученные результаты позволяют говорить о том, что динамика тепловых полей в области лица человека может служить одним из индикаторов изменения его психо-эмоционального состояния.

Проведен анализ выявленной психофизической реакции в форме локального изменения температурных полей в области лица для различных тестов. Сделаны выводы о возрастном и гендерном распределении выявленных реакций.

7. Благодарности

Работа выполнена с использованием оборудования, приобретенного за счет средств Программы развития Московского Университета, и при финансовой поддержке РФФИ (грант 16-18-00080).

8. Литература

- [1] T. Baltrušaitis, P. Robinson, and L-P. Morency, "OpenFace: an open source facial behavior analysis toolkit", IEEE Winter Conference on Applications of Computer Vision, Lacc Placid, NY, 2016.
- [2] Dlib: <http://dlib.net/>
- [3] Ioannou S., Merla A. "Thermal infrared imaging in psychophysiology: Potentialities and limits", Psychophysiology, 51, 2014, pp. 951–963.
- [4] OpenCV: <http://opencv.org/>
- [5] OpenFace: <http://www.cl.cam.ac.uk/research/rainbow/projects/openface/wacv2016.pdf>
- [6] P. Viola and M.J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), 2001.

- [7] P. Viola and M.J. Jones, "Robust real-time face detection", International Journal of Computer Vision, vol. 57, no. 2, 2004., pp.137–154.
- [8] И. А. Знаменская, Е. Ю. Коротева, А. В. Хахалин, В. В. Шишаков. "Термографическая визуализация и дистанционный анализ динамических процессов в области лица", Научная визуализация, 8(5):1–8, 2016.
- [9] Скрипаль А.В., Сагайдачный А.А., Усанов Д.А. Тепловизионная биомедицинская диагностика: Учеб. пособие. Саратов, 2009.

Об авторах

Знаменская Ирина Александровна, д.ф.-м.н., профессор кафедры молекулярных процессов и экстремальных состояний вещества физического факультета Московского государственного университета. Ее e-mail znamen@phys.msu.ru.

Коротева Екатерина Юрьевна, к.ф.-м.н., с.н.с. кафедры молекулярных процессов и экстремальных состояний вещества физического факультета Московского государственного университета. Ее e-mail koroteeva@physic.msu.ru.

Шишаков Виталий Владимирович, к.ф.-м.н., м.н.с. кафедры математического моделирования и информатики физического факультета Московского государственного университета. Его e-mail shift@physics.msu.ru.

Хахалин Андрей Владимирович, к.ф.-м.н., к.соц.н., с.н.с. кафедры молекулярных процессов и экстремальных состояний вещества физического факультета Московского государственного университета. Его e-mail avkhakhalin@mail.ru.

Кузьмичева Екатерина Алексеевна, студент кафедры молекулярных процессов и экстремальных состояний вещества физического факультета Московского государственного университета. Ее e-mail kakuzmicheva@gmail.com.

Исайчев Сергей Александрович, к.п.н., доцент кафедры психофизиологии факультета психологии Московского государственного университета.

Черноризов Александр Михайлович, д.п.н., профессор кафедры психофизиологии факультета психологии Московского государственного университета. Его e-mail amchern53@mail.ru.

Применение глубоких свёрточных нейронных сетей для визуализации препятствий на взлётно-посадочной полосе

В.В. Князь^{1,2}, О.В. Выголов¹, В.В. Федоренко¹, В.С. Горбачевич¹, В.А. Мизгинов¹, Д.Г. Клевцов¹
 vl.kniaz@gosniias.ru | o.vygovol@gosniias.ru | v.fedorenko@gosniias.ru | gvs@gosniias.ru
 vl.mizginov@gosniias.ru | k_dm@gosniias.ru

¹125319, Москва, ул. Викторенко, 7;

²Московский физико-технический институт (государственный университет), Москва, Россия

Контроль чистоты поверхности взлётно-посадочной полосы является важным фактором обеспечения безопасности воздушного движения. Современный уровень развития датчиков и методов обработки поступающих от них данных обеспечивает необходимые предпосылки для решения задачи обнаружения и активной визуализации препятствий по данным тепловизионных камер в автоматическом режиме, что позволяет повысить эффективность уборки взлётно-посадочной полосы. Перспективным средством анализа мультиспектральных видеопоследовательностей являются глубокие свёрточные нейронные сети. В данной статье рассматривается модифицированная архитектура глубокой свёрточной сети SqueezeNet для задач автоматического выделения препятствий по видеопоследовательностям видимого и дальнего инфракрасного диапазонов.

Ключевые слова: глубокие свёрточные нейронные сети, тепловизионные изображения, обнаружение препятствий, взлётно-посадочная полоса, семантическая сегментация

Deep convolutional neural network for visualisation of foreign objects on a runway

V.V. Kniaz^{1,2}, O.V. Vygovolov¹, V.V. Fedorenko¹, V.S. Gorbatsevich¹, V.A. Mizginov¹, D.G. Klevtsov¹
 vl.kniaz@gosniias.ru | o.vygovolov@gosniias.ru | v.fedorenko@gosniias.ru | gvs@gosniias.ru
 vl.mizginov@gosniias.ru | k_dm@gosniias.ru

¹State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia;

²Moscow Institute of Physics and Technology (MIPT), Russia

The presence of foreign objects on airport runways poses a significant threat to the safety of air travel. Infrared camera based runway monitoring systems for automatic detection and visualisation of foreign objects are highly demanded nowadays. Deep neural networks have recently become a powerful instrument for analysis of multispectral image sequences. This paper is focused on the development of a new deep neural network architecture for automatic detection of foreign objects on a runway. The architecture is based on the SqueezeNet network. The new network performs detection using a pair of images captured in visible and far infrared ranges.

Keywords: deep convolutional neural networks, thermal vision, obstacle detection, runway, semantic segmentation

1. Введение

Перспективным способом уборки взлётно-посадочной полосы (ВПП) являются колёсные роботы, оснащённые тепловизионными системами технического зрения. Управление роботом осуществляется удалённо из диспетчерского пункта. Оператор периодически производит осмотр поверхности полосы с использованием системы технического зрения мобильного робота.

Восприятие тепловизионных изображений требует навыков в силу значительных отличий от изображений видимого диапазона. Это обуславливает актуальность предварительной обработки тепловизионных изображений для акцентированной визуализации препятствий. Данная задача может быть решена с помощью методов семантической сегментации. Анализ последних работ в данной области [5–7] показывает, что наиболее перспективным подходом для решения задач семантической сегментации являются глубокие свёрточные нейронные сети.

В данной работе рассматривается разработка новой архитектуры глубокой свёрточной нейронной сети для выделения препятствий на ВПП. На вход сети поступает многоспектральное изображение, включающее каналы в видимом и инфракрасном диапазонах. На выходе сеть выдаёт меточное изображение, на котором визуализируются два класса объектов: фон и препятствие. Разработанная архитектура основывается на архитектуре сети SqueezeNet [8]. В отличие от задачи классификации объекта на изображении, при решении задачи семантической сегментации требуется предсказать тип объекта в каждом пикселе исходного изображения. Для восстановления исходного разрешения изображения к архитектуре сети SqueezeNet были добавлены два слоя деконволюции.

По сравнению с альтернативными методами выделения препятствий на ВПП предлагаемый подход обладает рядом преимуществ. Во-первых, свёрточные нейронные сети позволяют производить обучение без насыщения: при поступлении новых изображений в обучающую выборку качество распознавания препятствий постоянно растёт. Во-вторых, в отличие от систем, ос-

нованных на выделении препятствий на основе модели движения камеры [11, 12], предлагаемый подход обеспечивает выделение препятствий при произвольной траектории движения.

2. Постановка задачи

Рассматривается задача в следующей постановке: колёсный робот, оснащённый многоспектральной системой технического зрения (СТЗ) FLIR ONE (рис. 1) производит обзор поверхности ВПП. Требуется осуществить обработку изображений, поступающих от СТЗ, для акцентированной визуализации типовых препятствий (выделение контуров, маркирование цветом и т.п.). Нахождение визуальных контуров препятствия производится с помощью глубокой свёрточной нейронной сети. Таким образом, постановка задачи сводится к семантической сегментации многоспектральных изображений на два класса: фон и препятствие.



Рис. 1. Мобильный робот с многоспектральной камерой FLIR ONE

Целевая функция. Пусть на вход сети поступает многоспектральное изображение $\mathbf{X} \in \mathbb{R}^{H \times W \times 3}$. Требуется обучить целевую функцию $\hat{\mathbf{Y}} = \mathcal{F}(\mathbf{X})$, которая производит отображение комплексированного изображения в меточное изображение $\mathbf{Y} \in \{0, 1\}^{H \times W}$, где H, W – размеры изображения, а метки 0 и 1 обозначают, соответственно, фон и препятствие. Для достижения равномерной сходимости обучения при обучении используется многомодальная функция потерь L_{cl} [10], которая задаётся уравнением

$$L_{cl}(\hat{\mathbf{Y}}, \mathbf{Y}) = - \sum_{h,w} v(\mathbf{Y}_{h,w}) \sum_q \mathbf{Y}_{h,w,q} \log(\hat{\mathbf{Y}}_{h,w,q}) \quad (1)$$

3. Архитектура сети

Анализ последних работ в области семантической сегментации [3, 5–7] показывает, что наиболее высокой точности классификации позволяют достичь алгоритмы на основе глубоких свёрточных нейронных сетей.

В отличие от традиционных архитектур сетей, используемых для классификации изображений, в архитектурах сетей для семантической сегментации, как правило, отсутствуют полносвязные слои. Также в данных сетях широко используются слои деконволюции.

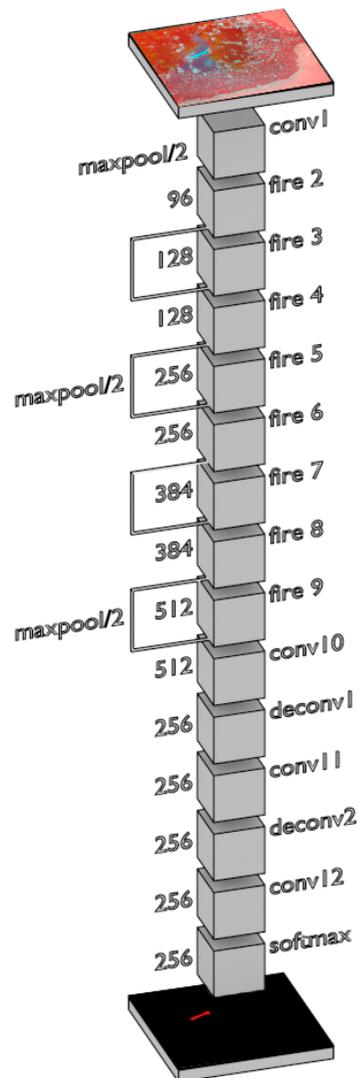


Рис. 2. Архитектура сети

Для обучения целевой функции $\hat{\mathbf{Y}}$ требовалось разработать архитектуру глубокой свёрточной нейронной сети для семантической сегментации. Для достижения вычислительной эффективности за основу архитектуры была взята сеть SqueezeNet [8]. Основным преимуществом сети SqueezeNet является выигрыш в производительности в 50 раз по сравнению с сетью AlexNet [4], при сохранении сопоставимой точности классификации. Для обеспечения вычислительной эффективности произведено уменьшение размеров фильтров свёртки с 3×3 на 1×1 . Благодаря этому число обучаемых параметров сократилось в 9 раз. Таким образом, сеть строится из однотипных модулей, называемых «fire module». Для обеспечения устойчивого выделения препятствий в архитектуру разрабатываемой сети (рис. 2)

был внесён ряд изменений. Во-первых, для восстановления пространственного разрешения входного изображения были добавлены два слоя деконволюции. Во-вторых, для уменьшения количества параметров удалён слой субдискретизации (global avgpool).

4. Построение обучающей выборки

Обучающая выборка для семантической сегментации состоит из пар геометрически выровненных многоспектральных изображений X и меточных изображений Y . Комплексование видимого и инфракрасного изображений осуществлялось путём замены красного канала исходного изображения видимого диапазона на изображение инфракрасного диапазона.

Для построения истинных меточных изображений для обучающей выборки использовался метод оценки траектории камеры, предложенный в [13]. Примеры исходных изображений, а также компенсированного и меточного изображений приведены на рисунке 3. Съёмка исходных изображений осуществлялась с помощью прототипа мобильного колёсного робота для уборки ВПП [1, 2].

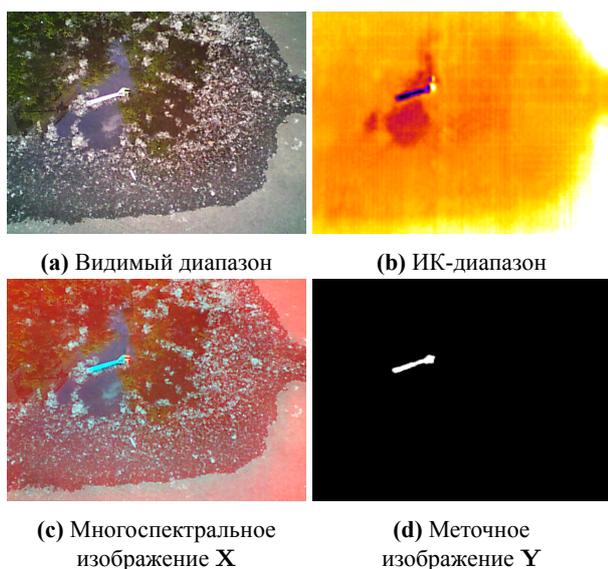


Рис. 3. Примеры изображений из обучающей выборки. Условное препятствие (разводной ключ) расположено в центре кадра

Прототип мобильного робота для уборки ВПП.

Для подготовки тестовой выборки использовался колёсный робот, оснащённый многоспектральной системой технического зрения. Робот построен на базе колёсной платформы «Hercules» и одноплатного компьютера «Raspberry Pi» (рисунок 1). В качестве многоспектральной системы технического зрения использовалась многоспектральная камера «FLIR ONE», подключённая к смартфону. Разрешение камеры в видимом диапазоне составляет 640x480 пиксел, в инфракрасном - 160x120 пиксел. Параметры камеры FLIR ONE приведены в таблице 1.

Характеристики FLIR ONE	
Диапазон температур	-20° C to 120° C
Вес	29 грамм
Размеры (Д x Ш x В)	72 x 26 x 18 мм
Чувствительность	0.1° C
ТВ камера	640x480 пикс.
ИК камера	160x120 пикс.
Поле зрения	46°x35°

Таблица 1. Характеристики камеры FLIR ONE

Управлением роботом осуществляется с персонального компьютера по каналу сети WiFi. Для управления роботом был разработан специальный протокол и приложения клиент/сервер для «Raspberry Pi» и персонального компьютера.

5. Обучение сети и анализ результатов

Обучение разработанной архитектуры сети производилось в среде NVIDIA DIGITS с использованием графического процессор Titan X PASCAL.

Тестирование обученной сети проводилось на тестовой выборке, объёмом в 300 пар изображений. Для определения качества работы сети использовалась бинарная мера сходства Intersection over Union (IU), определяемая соотношением:

$$IU = \frac{A_I}{A_U}, \quad (2)$$

где A_I – площадь пересечения области препятствия на построенном меточном изображении и на эталонном меточном изображении, A_U – площадь объединения данных областей. Как правило, значение $IU > 50.0\%$ считается показателем приемлемого качества сегментации [9]. Примеры полученных меточных изображений представлены на рисунке 4. Значения IU для различных условий приведены в таблице 2.

Условия	Intersection over Union
Тень	62.6
Солнце и тень	50.1
Мокрый асфальт	49.3

Таблица 2. Значения IU для различных условий

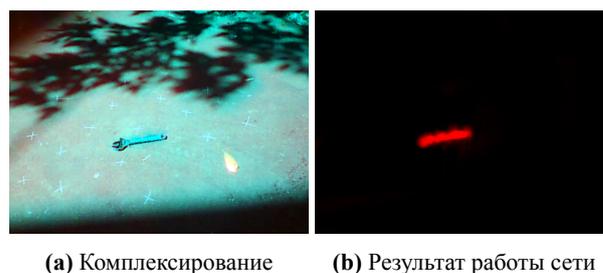


Рис. 4. Пример работы сети

6. Выводы

В работе предложен метод визуализации препятствий на поверхности ВПП с использованием глубокой свёрточной нейронной сети. Метод предназначен для решения задачи помощи оператору мобильного колёсного робота в обнаружении посторонних препятствий на поверхности ВПП. Он позволяет повысить эффективность обнаружения препятствий в ночное время и в условиях ограниченной видимости.

Для реализации метода разработана архитектура сети для выделения визуальных контуров препятствий. Произведено обучение и тестирование сети. Для обучения сети использовалась обучающая выборка объёмом 4000 пар комплексированных и меточных изображений. Съёмка изображений для обучающей и тестовой выборок производилась с помощью колёсного робота, оснащённого многоспектральной камерой FLIR ONE. Выборка включает снимки условных препятствий (разводной ключ, детали конструкции, предметы багажа и т.п.), полученные в различных условиях: прямой солнечный свет, граница света и тени, мокрый асфальт. Истинные меточные изображения были построены с использованием программы трёхмерного моделирования. Обучение сети производилось в среде NVIDIA DIGITS.

Тестирование разработанной сети производилось на независимой выборке объёмом в 300 пар изображений. Для определения точности выделения препятствий использовалась метрика Intersection over Union. Анализ полученных результатов показал, что сеть успешно производит выделение препятствия во всех рассмотренных условиях. Средняя точность выделения контуров препятствия по метрике Intersection over Union составила 54%.

7. Литература

- [1] Kniaz V V. Fast instantaneous center of rotation estimation algorithm for a skied-steered robot // SPIE Optical Metrology. — 2015. — Jun. — Vol. 9528. — P. 95280L–95280L–11.
- [2] Kniaz V V. Real-time optical flow estimation on a GPU for a skied-steered mobile robot. — 2016. — Apr. — P. 989706–989706–12.
- [3] Kniaz V. V., Gorbatshevich V. S., Mizginov V. A. THERMALNET: A DEEP CONVOLUTIONAL NETWORK FOR SYNTHETIC THERMAL IMAGE GENERATION // ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2017. — Vol. XLII-2/W4. — P. 41–45. — URL: <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W4/41/2017/>.
- [4] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. — 2012.
- [5] Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2015.
- [6] Long J., Shelhamer E., Darrell T. Fully Convolutional Models for Semantic Segmentation // CVPR 2015, and PAMI 2016. — 2016.
- [7] Semantic image segmentation for information presentation in enhanced vision / Oleg V. Vygolov, Vladimir S. Gorbatshevich, Nikita A. Kostromov et al. // Proc. SPIE. — 2017. — Vol. 10197. — P. 101970H–101970H–8. — URL: <http://dx.doi.org/10.1117/12.2262507>.
- [8] SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size / Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf et al. // CoRR. — 2016. — Vol. abs/1602.07360. — URL: <http://arxiv.org/abs/1602.07360>.
- [9] The Pascal Visual Object Classes (VOC) Challenge / Mark Everingham, Luc Van Gool, Christopher K I Williams et al. // International Journal of Computer Vision. — 2009. — Vol. 88, no. 2. — P. 303–338.
- [10] Zhang Richard, Isola Phillip, Efros Alexei A. Colorful Image Colorization. // ECCV. — 2016. — Vol. 9907, no. Chapter 40. — P. 649–666.
- [11] Zheltov S, Sibiryakov A. Method of 3d-object detection based on orthophoto difference analysis // International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B3. Amsterdam 2000. — 2000. — P. 1041–1047.
- [12] Бусурин В. И., Князь В. В. Автоматическое обнаружение препятствий на взлётно-посадочной полосе средствами технического зрения. — Труды МАИ, 2014. — Т. 81.
- [13] Князь В. В., Мизгинов В. А., Горбацевич В. С. Создание модельных изображений инфракрасного диапазона и оценка их визуального качества с использованием свёрточных нейронных сетей. — Научная визуализация, 2016. — Т. 8. — С. 67–79.

Об авторах

Князь Владимир Владимирович, инженер ГосНИИАС, магистрант ФУПМ МФТИ. Его e-mail vl.kniaz@gosniias.ru.

Выголов Олег Вячеславович, к.т.н., начальник лаборатории ГосНИИАС. Его e-mail o.vygodov@gosniias.ru.

Федоренко Вадим Вадимович, инженер ГосНИИАС. Его e-mail vfedorenko@gosniias.ru.

Горбацевич Владимир Сергеевич, начальник лаборатории ГосНИИАС. Его e-mail gvs@gosniias.ru.

Мизгинов Владимир Андреевич, инженер ГосНИИАС. Его e-mail vl.mizginov@gosniias.ru.

Клевцов Дмитрий Геннадьевич, инженер ГосНИИАС. Его e-mail k.dm@gosniias.ru.

Модели восприятия визуальной информации

Д.В. Манаков

manakov@imm.uran.ru

Институт математики и механики им. Н.Н. Красовского Уральского отделения РАН, Екатеринбург, Россия

Работа посвящена рассмотрению математических моделей зрительного восприятия. Целью изучения данного вопроса является не только повышение алгоритмического качества систем и устройств визуализации, но и возможность разработки самообучающихся тренажеров, основанных на генерации отрицательной обратной связи, например, с позиций диссипативных или функциональных систем.

Ключевые слова, восприятие, диссипативные системы, методы передачи объема и глубины пространства.

The visual information perception models

D.V. Manakov

manakov@imm.uran.ru

Institute of Mathematics and Mechanics im. N.N. Krasovskogo Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia

The paper is devoted to the mathematical models of visual perception. The purpose of this research is both to improve the algorithmic quality of visualization systems and devices, and to develop self-learning simulators based on the generation of negative feedback, for example, from the position of dissipative or functional systems.

Key words: perception, dissipative systems, methods of transferring the volume and the space depth.

1. ВВЕДЕНИЕ

Оценка эффективности визуализации должна рассматривать не только качество визуального дизайна, но и обеспечение соответствующего понимания, что особенно важно в случае когнитивной визуализации. Одним из параметров эффективности теории принятия решения является уровень доверия. Результаты тестирования показывают, что текстовое представление информации вызывает большее доверие, чем визуальное. Это можно объяснить тем, что в тексте меньше многозначности и неопределенности по сравнению с визуализацией. С целью минимизации этих факторов программисты должны максимально доверять математической модели, поэтому верификации программ и визуализации уделяют значительное внимание.

Рассмотрение моделей восприятия-понимания является частным случаем верификации программ визуализации. Разделение человеческого мышления на восприятие и понимание достаточно условно и не всегда оправданно. Принято выделять четыре основных (и связанных с математической формализацией) направления исследования процесса мышления:

1. Кибернетика (например, восприятие как параллельный процесс).
2. Бионика.
3. Нейронные сети.
4. Диссипативные системы (частный случай рассмотрения гомеостаза – функциональные системы).

Последнее направление связано с самоорганизацией, и мы остановимся на нем более подробно.

Целью данной работы является рассмотрение возможных постановок задач основных тематических направлений изучения зрительного восприятия в области компьютерной визуализации, прежде всего с позиций математического моделирования. Несмотря на высокую степень полимодальности человеческого зрения и восприятия в целом тематический круг подобных задач достаточно узок:

1. Сравнение текстовой и визуальной информации.
2. Зрительная фильтрация потока информации.

3. Визуальный анализ и визуализация неопределенности.
4. Оценка качества (эргономичности) дисплеев (реалистичности изображения).
5. Расширенная визуализация (в основном сонификация).
6. Сравнение стереоизображения и других методов передачи объема и глубины пространства (например, параллакс движения) и их интегрирование.
7. Виртуальная и расширенная (дополненная) реальность.
8. Тестирование и тренажеры.

Конечной целью изучения зрительного восприятия для компьютерной визуализации является не только повышение алгоритмического качества систем и устройств визуализации, но и возможность разработки самообучающихся тренажеров, основанных на генерации отрицательной обратной связи, например, с позиций диссипативных или функциональных систем.

2. ДИССИПАТИВНЫЕ СИСТЕМЫ

Диссипативная система - квазистационарная открытая система, характерной особенностью которой является процесс самоорганизации, происходящий в результате действия отрицательного вектора, например, силы трения.

Поясним отличие диссипативной системы от задачи оптимального управления в случае линейных систем:

$$\dot{x} = Ax - Bu,$$

где u – оптимальное управление, которое надо найти. Поскольку диссипативная система является открытой системой управление приходит извне, оно изначально задано. Введение в рассмотрение отрицательного вектора или отрицательной обратной связи сужает круг задач до моделей с насыщением.

Область применения диссипативных систем достаточно широка, начиная от микроуровня - рассмотрения клетки организма, до макроуровня – моделирование устройства вселенной. Теорию хаоса и теорию эволюции можно считать частными случаями диссипативных систем. Близкими по духу являются в области экономики наука синергетика, а в области программирования автономные вычисления. Поскольку данная работа посвящена восприятию визуальной информации, более подробно

остановимся на теории функциональных систем П.К. Анохина.

Человек находится в состоянии гомеостаза (почти статического равновесия). Выделяются два типа функциональных систем. Системы первого типа обеспечивают гомеостаз за счёт внутренних (уже имеющихся) ресурсов организма, не выходя за его пределы – замкнутые системы. Системы второго типа поддерживают гомеостаз за счёт изменения поведения, которое формируется в результате отрицательной обратной связи – открытые (диссипативные) системы.

Функциональные системы второго типа представляют интерес для создания самообучающихся тренажеров, основанных на генерации отрицательной обратной связи, но в этом случае, скорее всего, надо рассматривать задачу оптимального управления. Перспективным направлением является изучение энцефалограммы мозга с целью распознавания обратной связи и валидации моделей восприятия.

Последние исследования в области диссипативных структур позволяют делать вывод о том, что процесс самоорганизации происходит гораздо быстрее при наличии в системе внешних и внутренних шумов. Таким образом, шумовые эффекты приводят к ускорению процесса самоорганизации. Тремор глаз и тремор рук – колебания малой амплитуды, фактически это шум, и их необходимо учитывать, как в процессе моделирования, так и с целью ускорения обучения на тренажерах.

Ключевым понятием диссипативных систем является энтропия. Например, термодинамическая энтропия - мера необратимой диссипации энергии. Поскольку понятия термодинамической энтропии и информационной энтропии (информативности) близки или даже эквивалентны была поставлена цель найти иллюстративный пример динамики диссипативных систем в области информационных технологий. Оказалось проще взять известный пример [5] и рассмотреть его с точки зрения диссипативных систем. Рис. 1. представляет визуальный анализ деятельности кредитных организаций. Он позволяет строить суждения об образовании кластеров (диссипативных структур) и находить объекты, стремящиеся оказаться в кластере (аналогично, модели движения амёб в сторону вещества с наибольшей концентрацией). С точки зрения диссипативных систем на следующем шаге кластер должен разрушиться (обычно по превышению некоторого порогового значения) и на его периферии должен сформироваться новый кластер.

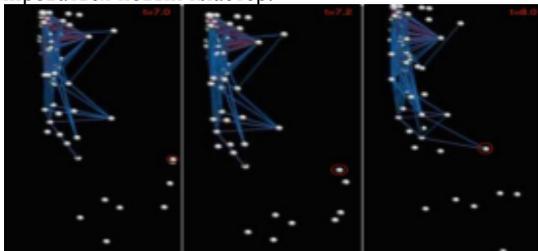


Рис. 1. Кластер и объект, стремящийся оказаться в кластере [5].

Как уже отмечалось, диссипативные системы являются не единственным направлением в области изучения зрительного восприятия и человеческого мышления, но имеют достаточно широкую область применения. Например, они применяются в области лингвистики, а частный случай - функциональные системы, может рассматриваться не только как одна из методик создания тренажеров, но и изучения зрительной фильтрации потока информации – одного из направлений изучения зрительного восприятия.

3. ОСНОВНЫЕ НАПРАВЛЕНИЯ ИЗУЧЕНИЯ ЗРИТЕЛЬНОГО ВОСПРИЯТИЯ

Эти направления были перечислены во введении работы. Конечно, они частично пересекаются между собой, а восприятие пересекается с пониманием, но имеют определенное самостоятельное значение, и поэтому все это можно считать базисом или семантическими единицами ментального пространства визуализации. Для того, чтобы рассматривать компьютерную визуализацию как самостоятельную дисциплину, необходимо сформировать ментальное пространство визуализации со своей семантикой, прагматикой и базисом. Визуальные тексты можно рассматривать в качестве базиса универсального пространства визуализации [4], то есть они имеют основополагающее истинностное значение. Верификация визуализации - один из основных инструментов создания данного пространства, и поэтому тематические направления изучения зрительного восприятия будут представлены не столько в качестве обзора литературы, а и в качестве математических постановок задач, требующих решения.

Кибернетическое направление рассмотрения визуального текста как непрерывного отображения является перспективным, но нерешенных вопросов слишком много. В том числе по модели потока данных, которая имеет частичное пересечение с методикой зрительной фильтрации потока информации [6]. Далее остановимся на этой методике изучения зрительного восприятия и на технологии фильтрации данных, которая имеет важное (близкое к единице) значение в случае обработки данных большого объема, включая результаты параллельных вычислений.

3.1 Зрительная фильтрация потока информации

Данная методика [6] формально связана с функциональными системами второго типа и теорий управления. С целью изучения зрительного восприятия применяется кибернетический подход манипуляции входным потоком информации (изменение двух параметров объектов: цвета и формы). Например, применение светофильтра в рамках натурального испытательного стенда позволят виртуальную блокировку зрительного нерва. Для данной методики напрашивается создание виртуального испытательного стенда: применение очков виртуальной реальности и программное управление потоком информации, включая фильтрацию данных.

Как отмечается [6], у методики хорошая валидность, но на уровне восприятия постоянно возникают труднообъяснимые визуальные артефакты. Например, если в очках виртуальной реальности рассматривать горизонтальную белую линию, то на периферии зрения можно заметить гладкое разложение этой линии на три линии разного цвета. Конечно, причины возникновения подобных визуальных артефактов можно формально объяснить, но не факт, что именно эта модель срабатывает на уровне зрительного восприятия человека. Например, можно рассмотреть взвешенную сумму стереоизображения, основанного на бинокулярной диспаратности, и других видов объемного зрения.

Модели фильтрации данных [3] также кибернетически переносятся в область изучения зрительного восприятия, поскольку стандартный графический конвейер, в том числе модель потока данных, представляет собой последовательность процессов фильтрации, геометрической обработки, рендеринга.

Задача визуализации данных большого объема не имеет решения без постановки задачи фильтрации данных или

семантически близких подходов, например, таких, как агрегация, редукция, кластеризация. Фильтрация данных (фильтр) – любая операция над данными, изменяющая их количество (“не хеширование”). В этом определении “не хеширование” рассматривается, как необходимое условие установление предпорядка. Формально можно говорить о построении идеала.

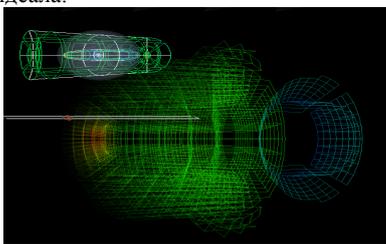


Рис. 2. Пространственная фильтрация с помощью метафоры альфа-сферы.

В качестве иллюстрации можно привести пример пространственной фильтрации данных, реализованный с помощью метафоры альфа-сферы (рис. 2). Если альфа-сферу или фильтр сечения плоскостью рассматривать как эпсилон-покрытия, то возможно применить модели дискретной оптимизации. Построение фильтров для конкретной задачи визуализации или данных можно также рассматривать как задачу оптимального управления [3].

Эпсилон-покрытия и визуальные артефакты можно рассматривать как неопределенности. Существует большое количество англоязычных обзоров по визуальному анализу, визуализации и пространственно неопределенности, часто содержащие противоречивую и недостоверную информацию, и поэтому требующие определенной ревизии. Одной из причин противоречий является то, что неопределенность может быть понятием различных математических дисциплин, например: логики (трехзначная, нечеткая логика), теории принятия решений. Мы считаем, что достаточно рассмотрения неопределенности с позиций математического анализа: предельной неопределенности - неопределенности, которая имеет конечный предел в конкретной метризуемой топологии.

3.2 Визуальный анализ, визуализация и распространение неопределенности

В работе [9] вводится понятие верифицируемой (верифицирующей) визуализации, которая отслеживает, как распространяется погрешность (неопределенность) на всем этапе вычислительного конвейера, включая визуализацию. Подобный подход в общем случае принято называть моделью с неопределенностью, а в частном - визуализацией с неопределенностью. С другой стороны в отображение интервала на интервал (компакта на компакт) нет никакой неопределенности ни с точки зрения восприятия, ни с точки зрения математики. Распространение неопределенности не так интересно, если оно не противоречит оценкам вычислительного метода, необходимо находить визуальные артефакты и объяснять их причины возникновения. Например, визуальные артефакты как следствие: зрительного восприятия, изменения ориентации нормалей (вырожденные ячейки), вычислительной погрешности. Если размер визуального артефакта уменьшился одновременно с уменьшением шага вычислений (аналогия с методом Рунге), то, скорее всего, это следствие вычислительной погрешности.

Неопределенность может быть, если отображение не является взаимно однозначным, тогда надо отображать раскрашенные интервалы. В визуальной семиотике используется понятие визуальной переменной (например, форма объекта) и ставится задача показать с позиций интуитивной логики, что знак этой переменной содержит

неопределенность. Постановка не лишена смысла, но достаточно рассмотрения стандартных математических задач: определение информативных признаков, визуальный анализ чувствительности решения в зависимости от параметров прикладной задачи, например, в области газовой динамики [2]. И ориентироваться надо не на интуитивную логику и семиотику, а на физический или математический смысл, например, в этом случае ламинарное течение, а в этом турбулентное, параметры такие. Разрешение неопределенности и отображение точек бифуркации в виде рогок, воспринимается в рамках трехзначной логики, но математически может быть связано с недифференцируемостью по параметру.

Многопараметрическая модель визуального текста является достаточно общей, и применима для визуального анализа и визуализации неопределенности, поскольку в модели, кроме пространственно-временных характеристик учитываются частные производные визуального текста по его информативным признакам [4]. Двухиндексная нотация неопределенности, в которой нижний индекс обозначает размерность независимых данных, а верхний количество зависимых переменных, пригодна для классификации неопределенности, и заимствована из классификации видов отображения (сущностей) [7], но не учитывает ни происхождения ни способа представления неопределенности.

В работе [10] рассмотрена применимость трех наиболее цитируемых теорий восприятия-понимания для анализа и визуализации неопределенности, основанных на использовании:

1. Визуальных переменных, которые начинают распознаваться на уровне сетчатки глаз.
2. Принципа графического совершенства - это то, что дает зрителю наибольшее количество идей за кратчайшее время с наименьшим количеством чернил в самом маленьком пространстве. (Возможны и другие минимаксные постановки, например, сложность интерпретации, фильтрация данных [3]).
3. Физиологических и когнитивно-психологических (в том числе гештальт-теории) исследованиях, объединяющих два предыдущих направления.

Изучение и моделирование зрительного восприятия невозможно без минимального знания нейрофизиологии особенно для объемного зрения. Структура основного зрительного пути, например, представлена в книге [6].

3.3 Стереозображение и другие методы передачи объема и глубины пространства

Чтобы не путаться будем разделять стереозображение, основанное на диспаратности зрения, как наиболее часто используемое и объемное изображение (зрение) – все остальные методы передачи объема и глубины пространства. Основные механизмы оценки объема и глубины пространства:

Нейрофизиологические: диспаратность зрения, **тремор** (движение глаз с частой 150 Гц и малой амплитудой), аккомодация хрусталика, конвергенция и дивергенция осей глаз, полимодальность зрения и восприятия;

Свойства пространства: параллакс движения, перспектива, тени, цветовые эффекты атмосферы (прозрачность).

Стоит отметить, что компенсаторная функция зрительного восприятия человека очень развита. Она напрямую связана с выживаемостью вида. Стереозрение полностью компенсируется другими способами объемного зрения, среди которых наиболее важную роль играет тремор глаз. Как уже отмечалось, компенсаторная функция может рассматриваться как взвешенная сумма. Кибернетическая модель стереозрения достаточно простая

– разложение в ряд (Фурье, Эйлер, Лапласиан) по двум базисным функциям, в частности по двум (криволинейным) координатам.

Считается доказанным, что во время восприятия цвета работает разностная схема по двум временным точкам, поскольку воспринимается не длина волны, а разность (изменение длины волны). Естественно предположить, что кибернетическая модель тремора глаз также разностная схема. Разностная схема типа крест, с начала выполняется дифференцирование (отображение тернарных треморных лучей в криволинейные координаты колечка), а затем обратное интегрирование криволинейных координат в мозолистое тело, соединяющее два полушария головного мозга. Косвенным подтверждением того, что работает разностная схема по трем временным точкам, являются эмпирические данные: частота тремора глаз 150 Гц, а время задержки восприятия одна пятидесятая секунды, следовательно, задержка составляет три такта тремора. Оценка производительности треморной модели зрительного восприятия: количество нейронов глаза 150000000 надо умножить на частоту 150 и результат возвести в квадрат (сложность вычислительного метода N^2), получается примерно 10^{20} . Для сравнения порядок экзафлопных вычислений 10^{18} . Если мы не боги, то производительность параллельных вычислений ограничена сверху зрительным восприятием.

Поскольку производительность зрительного восприятия чрезмерна, имеет право на существование гипотеза о том, что человеческое мышление является артефактом зрительного восприятия с точки зрения искусственных нейронных сетей. Слух и осязание являются врожденными чувствами, а зрение приобретенным. Зрение формируется в процессе обучения нейронной сети, повторяя механизмы этих органов чувств: стереозрению соответствует диспаратность слуха и осязания (левша - правша), тремору зрения - тремор рук. После обучения в нейронной сети осталось свободное место, природа не терпит пустоты и свободное место заняло мышление. Кроме физиологических оснований существуют косвенные психологические подтверждения этой гипотезы, например, решая трудную задачу, люди склонны отключаться от восприятия, тем самым освобождая дополнительное пространство. Данная гипотеза рассматривается всего лишь как модель, в любом случае полимодальность зрения и других органов чувств существенна при изучении зрительного восприятия.

Существуют работы, сопоставляющие стереозрение и свойства пространства: метафора восхода солнца оценивает влияние изменения освещенности Рис. 3. [4], в работе [8] рассматривается модель обработки стереоизображений, которая сначала определяет глубину пространства, полученную из параллакса движения, а затем корректируют бинокулярный дисбаланс.

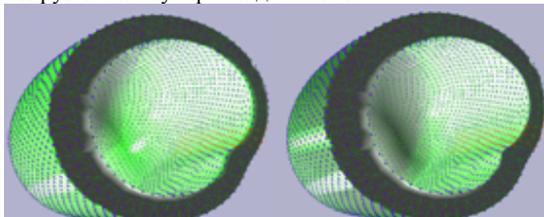


Рис. 3. Вращение источника света вокруг границы сетки.

По ряду причин не все перечисленные направления изучения зрительного восприятия были рассмотрены. В частности, потому что они частично пересекаются между собой, прежде всего на уровне моделей. Например, наивный байесовский подход применяется как в тестирование, так и в разработке тренажеров [1].

4. ЗАКЛЮЧЕНИЕ

Изучение зрительного восприятия является актуальной задачей для компьютерной визуализации. В данной работе предлагается структурирование материалов по тематическим направлениям (узконаправленным видам деятельности) и моделям формализации. Рассмотрение моделей восприятия-понимания является частным случаем верификации (математического доказательства правильности) визуализации. С целью валидации моделей в секторе компьютерной визуализации ИММ УрО РАН существует намерение начать разработку он-лайн сервисов параллельных вычислений. Например, на базе конструктора веб-визуализации реализовать автономную оценку эффективности для семейств алгоритмов с внешней памятью с позиции дискретной оптимизации (транспортная задача).

5. ЛИТЕРАТУРА

- [1] Алешин В.П. Технология виртуальной 3D среды в обратных задачах анализа визуального восприятия и интерпретации изображений // Труды 26-ой Международной конференции по компьютерной графике и зрению 19-23 сентября 2016, Нижний Новгород, с. 9-13
- [2] Бондарев А.Е., Галактионов В.А., Шапиро Л.З. Построение и визуальный анализ параллельных решений для оптимизационных и параметрических исследований в вычислительной газовой динамике // Труды 26-ой Международной конференции по компьютерной графике и зрению 19-23 сентября 2016, Нижний Новгород, с.211-215.
- [3] Манаков Д., Авербух В. Верификация визуализации // Научная визуализация 2016. Кв.1. Том 8. N: 1. Стр. 58 – 94.
- [4] Д. В. Манаков, В. Л. Авербух, П. А. Васёв. Визуальный текст как истинностное подмножество универсального пространства // Научная визуализация 2016. Кв.4. Том 8. N: 1. Стр. 58 - 94.
- [5] Д.Д. Попов, И.Е. Мильман, В.В. Пилюгин, А.А. Пасько. Решение задачи анализа многомерных динамических данных методом визуализации // Научная визуализация 2016. Кв.1. Том 8. N: 1. Стр. 55 – 57.
- [6] Е.А. Трофимов Эргономика зрительного восприятия. М. ООО “АИР” 2013.
- [7] Brodli, K.; Osorio, R.A.; Lopes, A. A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 81–109.
- [8] Petr Kellnhofer, Piotr Didyk, Tobias Ritschel, Belen Masia, Karol Myszkowski, Hans-Peter Seidel Motion Parallax in Stereo 3D: Model and Applications ACM Transactions on Graphics 35(6) (Proc. SIGGRAPH Asia 2016, The Venetian Macao, Macao).
- [9] Kirby R., Silva C. The need for verifiable visualization // IEEE Computer Graphics and Applications, 28(5) Sep 2008. Pp. 78–83.
- [10] T. Zuk and S. Carpendale. Theoretical analysis of uncertainty visualizations. In *Visualization and Data Analysis*, 2006.

Об авторах

Манаков Дмитрий Валерьянович – ведущий программист ИММ УрО РАН. Его e-mail manakov@imm.uran.ru.

Лазерная визуализация тепловых процессов в прозрачных твердых телах

Б.С. Ринкевичюс, И.Н. Павлов, И.Л. Расковская, А.В. Толкачев
rinkevbs@mail.ru|inpavlov@bk.ru|raskovskail@mail.ru|avtolk2008@yandex.ru
ФГБОУ ВО «НИУ «МЭИ», Москва, Россия

Экспериментально визуализированы поля температуры в слое полиметилметакрилата и исследована их динамика при неоднородном нагреве и остывании образца.

Ключевые слова: рефракция, визуализация температурных полей, оптически неоднородная среда, структурированные лазерные пучки, неоднородный нагрев твердого тела.

Laser visualization of thermal processes in transparent solid bodies

B.S. Rinkevichyus, I.N. Pavlov, I.L. Raskovskaya, A.V. Tolkachev
rinkevbs@mail.ru|inpavlov@bk.ru|raskovskail@mail.ru|avtolk2008@yandex.ru
National Research University "MPEI", Moscow, Russia

The fields of temperature in a piece of polymethylmethacrylate are experimentally visualized and its dynamics is shown upon inhomogeneous heating and cooling of a sample.

Keywords: refraction, visualization of temperature fields, optically inhomogeneous medium, structured laser radiation, inhomogeneous heating of a solid body.

1. Введение

В работах [1–6, 8–14] были представлены методики экспериментальной визуализации и восстановления полей температуры и других физических параметров в жидкости, основанные на регистрации рефрактограмм лазерных пучков. В статье [7] указанные методики были адаптированы для восстановления температуры в неоднородно нагретом слое прозрачного твердого тела и обобщены для исследования двумерной неоднородности при использовании для зондирования точечно-структурированного лазерного излучения.

Как хорошо известно, в настоящее время широко распространены компьютерные методы визуализации различных физических процессов, как в жидкостях, так и в твердых телах. Однако следует обратить внимание, что при компьютерном моделировании не всегда могут быть учтены все факторы, влияющие на физический процесс. Например, в (9) авторами было показано, что даже простейшая задача обтекания металлического нагретого шара, помещенного в холодную воду не позволяет учесть появление термиков сверху шара, которые хорошо визуализируются с помощью методов лазерной рефрактографии (1).

В данной работе экспериментально визуализированы поля температуры в слое полиметилметакрилата (оргстекла) и исследована их динамика при неоднородном нагреве и остывании.

2. Экспериментальная установка и методика измерений

Объектом исследований являлся прямоугольный элемент из органического стекла с размерами: ширина 153 мм, высота 140 мм и толщина 25 мм. На верхней грани элемента по центру располагался прямоугольный сосуд из тонкостенного алюминия толщиной 0,5 мм с внутренними размерами 26 × 38 × 74,6 мм. Ребра нижней грани имели радиус закругления 1 мм. В данный сосуд при проведении эксперимента заливалось 52 мл горячей воды. На рис. 1 показан внешний вид элемента из оргстекла и алюминиевого сосуда, расположенного на верхней грани

элемента. Фотографирование проводилось сверху под небольшим углом. Для обеспечения надежного теплового контакта между дном сосуда и поверхностью элемента из оргстекла была нанесена теплопроводная паста (кремнийорганическая) марки КПТ-8.



Рис.1. Внешний вид элемента из оргстекла и алюминиевого сосуда, расположенного на верхней грани элемента.

В качестве источника излучения был применен маломощный (5 мВт) твердотельный лазер с диодной накачкой, работающий на длине волны $\lambda = 532$ нм (зеленый цвет). Лазерный пучок направлялся на дифракционный оптический элемент (ДОО), на выходе которого формировалась система расходящихся лазерных пучков. В поперечном сечении данной системы лазерных пучков получалась сетка с числом пучков 20×20 , расположенных равномерно по сечению. Для того, чтобы сформировать систему параллельных лазерных пучков, использовалась длиннофокусная сферическая линза с фокусным расстоянием 260 мм и световым диаметром 75 мм. За счет подбора расстояния между ДОО и этой линзой обеспечивалась параллельность системы лазерных пучков после данной линзы. На рис.2 показаны элементы экспериментальной установки.

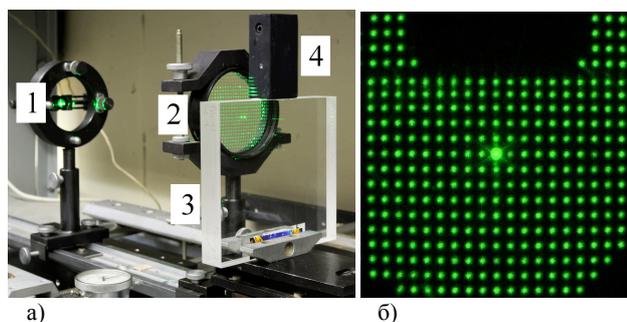


Рис. 2. а) – элементы экспериментальной установки:
1 – источник структурированного лазерного излучения,
2 – объектив, 3 – пластина из оргстекла, 4 –
алюминиевый сосуд,
б) – вид матового экрана для холодной пластины.

Система параллельных лазерных пучков использовалась для зондирования области в органическом стекле под алюминиевым сосудом с горячей водой.

На расстоянии 317 мм от задней поверхности элемента из оргстекла располагался матовый экран, на котором наблюдалось изменение во времени соответствующих рефрактограмм (по мере нагрева и остывания пограничного слоя оргстекла). За экраном была установлена цветная цифровая видеокамера для регистрации рефрактограмм.

Методика проведения эксперимента состояла в следующем:

- осуществлялась юстировка оптической системы экспериментальной установки;
- на экране видеокамеры (на мониторе, сопряженного с ней персонального компьютера) наблюдалась неискаженная точечная рефрактограмма;
- видеокамера включалась на «ЗАПИСЬ»;
- в алюминиевый сосуд заливалась горячая вода;
- видеокамера регистрировала изменение рефрактограмм во времени по мере остывания воды в алюминиевом сосуде (запись велась примерно 9 минут).

На рис. 2б) показан внешний вид отдельных элементов экспериментальной установки. На рис. 2а) показаны: дифракционный оптический элемент (ДОО), длиннофокусная линза, преобразующая расходящуюся после ДОО систему лазерных пучков в параллельную. За линзой расположен элемент из органического стекла с закрепленным на верхней его грани алюминиевым сосудом, наполненным холодной горячей водой. На рис. 2б) показан матовый экран с рефрактограммой, до проведения эксперимента (неискаженной нагревом верхней поверхности образца).

На рис. 3 приведена экспериментально полученная зависимость положения тонких пучков на экране, характеризующая остывание воды в алюминиевом сосуде, установленном на верхней поверхности элемента из органического стекла.

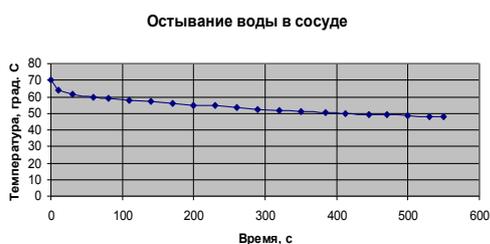


Рис. 3. Изменение температуры воды в алюминиевом сосуде, установленном на верхней поверхности элемента из органического стекла.

Перед проведением измерения температуры с помощью цифрового мультиметра с термопарой было проведено сравнение показаний ртутного термометра и системы мультиметр-термопара. Получено, что данные, измеренные мультиметром с термопарой, отличаются от показаний ртутного термометра не более, чем на $\pm 2\%$.

Для получения более полной информации о динамике прогрева верхней поверхности элемента из органического стекла этот процесс визуализировался прямым теневым методом при освещении исследуемой части элемента широким лазерным пучком.

С этой целью была смонтирована экспериментальная установка, в которой источником излучения являлся лазер с параметрами: мощность излучения 100 мВт; длина волны излучения 532 нм. С помощью микрообъектива с увеличением $\times 20$ формировался расходящийся в пространстве лазерный пучок.

Прямые теневые картины на матовом экране регистрировались цветной цифровой видеокамерой и их примеры показаны на рис. 4.

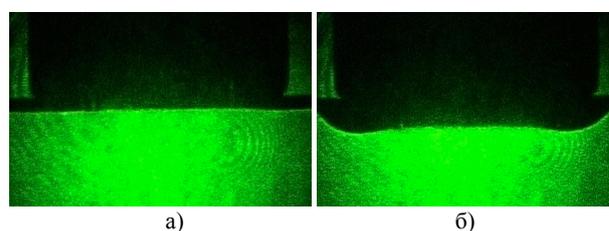


Рис. 4. Прямотеневые изображения зоны нагрева оргстекла: а) вид образца без нагрева в начальный момент времени $t = 0$ с; б) вид образца при температуре поверхности 58°C и времени нагрева $t = 120$ с.

Как следует из рис. 4, прямотеневые изображения являются малоинформативными с точки зрения количественной диагностики градиента температуры и зависящего от него градиента показателя преломления. Для получения количественной информации в этом случае требуется измерение распределения интенсивности в прошедшем пучке.

Использование структурированного лазерного излучения позволяет проводить количественную диагностику оптических неоднородностей на основе измерения смещения структурных элементов пучка, в данном случае «точек», отображающихся в сечении точно-структурированного пучка. На рис. 5а и 5б крупным планом показаны области рефрактограмм, соответствующая максимальным градиентам температуры. Под каждым рисунком указано время с момента заполнения сосуда горячей водой и температура воды на указанный момент времени.

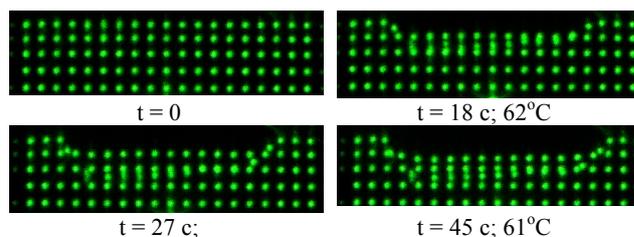


Рис.5а. Экспериментальные рефрактограммы точечного СЛИ.

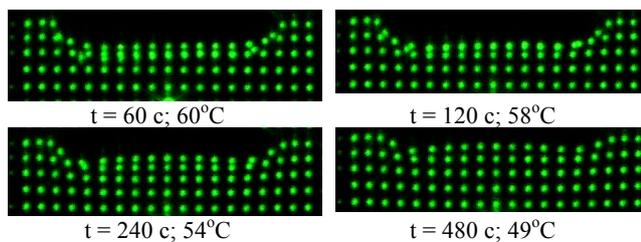


Рис. 5.6. Экспериментальные рефрактограммы точечного СЛИ, показанные крупным планом.

Приведен участок сечения СЛИ, деформированный из-за наличия температурного градиента в верхней части образца.

В рассматриваемой задаче температурное поле можно считать двумерным, так как градиент температуры в направлении перпендикулярном поверхности образца превышает градиенты температуры в других направлениях. Это означает, что в данных экспериментах имеет место слоистая неоднородность распределения температуры в исследуемом образце.

Для восстановления распределения температуры в пограничном слое оргстекла используется известное [7] соотношение, связывающее температуру T и показатель преломления n в оргстекле:

$$T(x, y) = T_0 + (n_0 - n(x, y)) / \alpha, \quad (1)$$

где коэффициент $\alpha=0,00015$, n_0 - показатель преломления оргстекла при температуре T_0 . Зависимость показателя преломления $n(x, y)$ от вертикальной координаты x и горизонтальной координаты y может быть найдена в результате решения обратной задачи рефракции на основе метода минимизации целевой функции. В качестве целевой функции принимается среднеквадратичное отклонение элементов расчетной рефрактограммы от элементов оцифрованной экспериментальной рефрактограммы. Расчет рефрактограммы проводится на основе параметрической модели поля показателя преломления вида:

$$n(x, y) = n_0 + \frac{(n_1(1+ky) - n_0)}{2} (1 + th(\frac{x-x_0}{h+ay})) (\sigma(y-b_1) - \sigma(y-b_2)) \quad (2)$$

где x, y - вертикальная и горизонтальная координаты, соответственно, параметр n_1 характеризует величину градиента в нагретом слое, h - характерную толщину слоя, x_0 задает положение слоя на оси x , параметры k и a характеризуют неоднородность слоя вдоль y , b_1 и b_2 задают границы слоя, наличие которых обусловлено конечными размерами сосуда и его положением, σ -сигма функция. Искомыми являются параметры модели, соответствующие минимуму целевой функции. Для описания измеренного распределения показателя преломления используется формула (2) с найденными параметрами. Однозначность решения обратной задачи для монотонных неоднородностей вида (2) обоснована в [1]. На рис. 6 показаны расчетные рефрактограммы, наилучшим образом, по выбранному критерию, соответствующие экспериментальным. Некоторое их отличие от экспериментальных объясняется тем, что при построении расчетных рефрактограмм не учитывались эффекты, обусловленные закруглением краев сосуда.

Рис. 7 демонстрирует восстановленное распределение температуры в слое в зависимости от времени. Наблюдается некоторая асимметрия распределения температуры вдоль горизонтальной оси, обусловленная, вероятно неравномерным заполнением сосуда горячей водой в начальный момент времени. С течением времени эта температура дна сосуда выравнивается, и асимметрия

исчезает. Градиент температуры по вертикальной оси много больше, чем по горизонтальной.

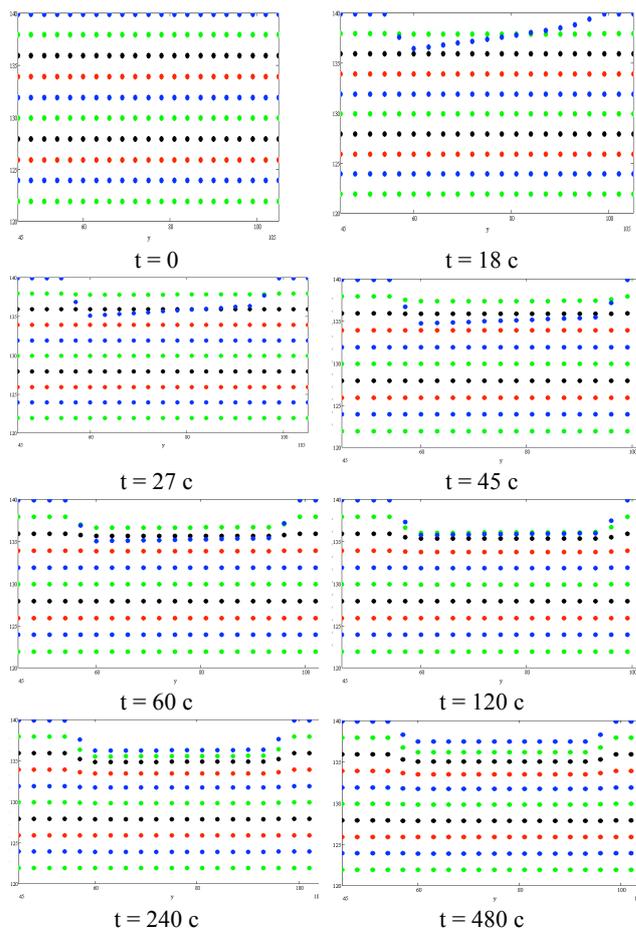


Рис. 6. Расчетные рефрактограммы, соответствующие экспериментальным.

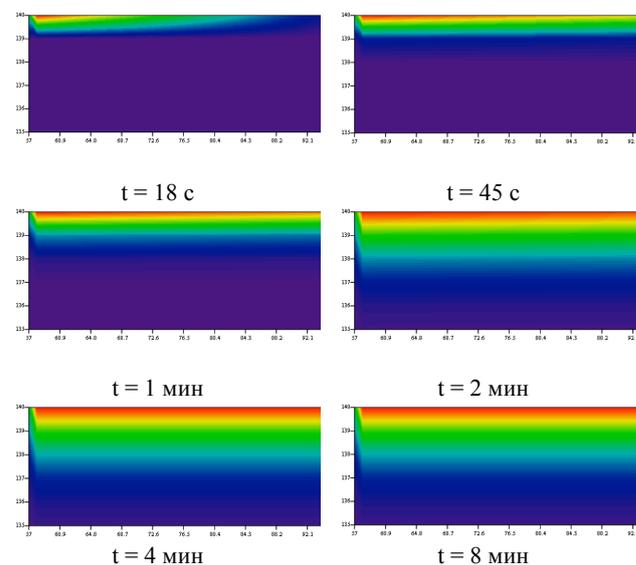


Рис. 7. Восстановленное распределение температуры в нагретом слое.

Зависимость от времени характерной толщины слоя и максимальной температуры в выбранном сечении показаны на рис. 8. Из рис. 8а следует, что с течением времени происходит рост толщины слоя, который постепенно

замедляется и фактически прекращается, а затем, как следует из рис. 8б, начинается остывание слоя.

Зависимость от времени характерной толщины слоя и максимальной температуры в выбранном сечении показаны на рис. 8. Из рис. 8а следует, что с течением времени происходит рост толщины слоя, который постепенно замедляется и фактически прекращается, а затем, как следует из рис. 8б, начинается остывание слоя.

Зависимость от времени характерной толщины слоя и максимальной температуры в выбранном сечении показаны на рис. 8. Из рис. 8а следует, что с течением времени происходит рост толщины слоя, который постепенно замедляется и фактически прекращается, а затем, как следует из рис. 8б, начинается остывание слоя.

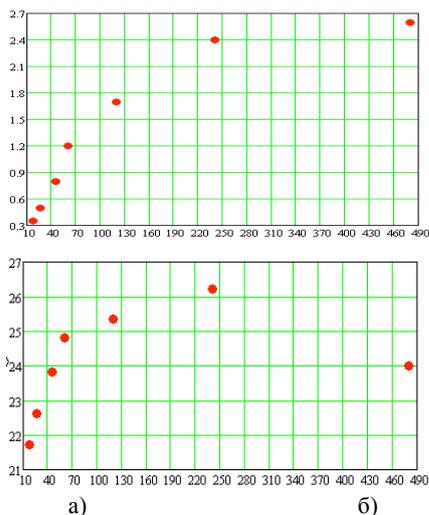


Рис. 8. Изменение характерной толщины (а) нагретого слоя оргстекла и температуры (б) в сечении $x=140$ мм в зависимости от времени наблюдения.

На рис. 9 показаны примеры восстановленных распределений температуры в одном из вертикальных сечений ($y=60$ мм) нагретого слоя оргстекла для различных моментов времени.

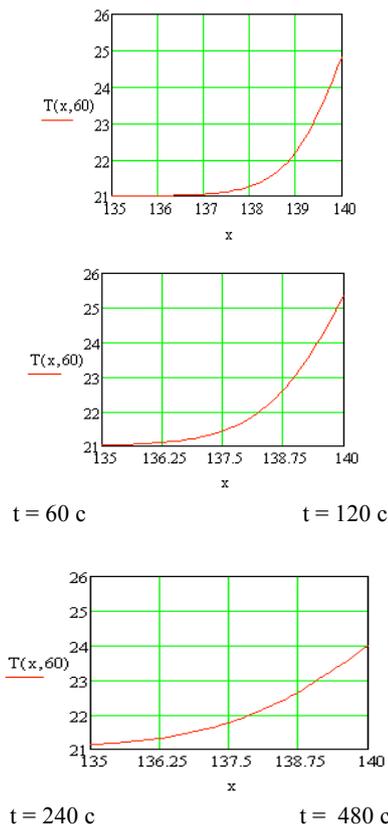
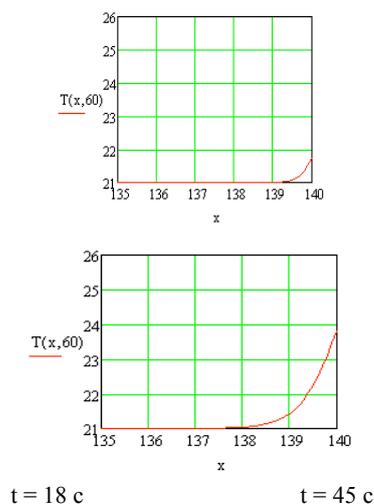


Рис. 9. Восстановленное распределение температуры в одном из вертикальных сечений ($y=60$ мм) нагретого слоя оргстекла.

3. Заключение

Разработана экспериментальная методика визуализации и количественной диагностики оптических неоднородностей в твердых телах на основе рефракции структурированного лазерного излучения. Методика реализована для случая неоднородного нагрева и остывания оргстекла. Визуализированы двумерные распределения температуры в контактирующем с горячим телом слое оргстекла и исследованы временные зависимости характеристик нагреваемого слоя.

4. Благодарности

Работа выполнена при финансовой поддержке Российского Фонда Фундаментальных Исследований (проект № 14-08-00948а).

5. Литература

- [1] Evtikhieva O.A., Rinkevichyus B.S., Raskovskaya I.L., Laser Refractography. Springer, New York, 2010.
- [2] Grechikhin V.A., Raskovskaya I.L., Rinkevichyus B.S. Influence of the acousto-optic effect on measurement error of particle vibration velocity in using the laser doppler anemometer. //Автометрия. 2000. № 5. С. 92-101.
- [3] Raskovskaya I.L., Rinkevichyus B.S., Tolkachev A.V. Structured beams in laser refractography applications // In The book Laser Beams: Theory, Properties and Applications 2011. Pp. 399-413.
- [4] Rinkevichyus B. S. Optical refractography: principles and applications. // In: The 2013 International Doctoral Forum

- of Electronic Science and Technology. Beijing, China, 9-13 July, 2013, 2013. Pp. 25 – 32.
- [5] Расковская И.Л. Волновая модель рефракции лазерных пучков с дискретным изменением интенсивности по поперечному сечению и их применение для диагностики протяженных нестационарных фазовых объектов. // Квантовая электроника. 2015. 45, № 8. С.765-770.
- [6] Расковская И.Л., Ринкевичюс Б.С., Толкачев А.В. Лазерный рефракционный метод визуализации и количественной диагностики диффузионного слоя жидкости при наличии внутренних волн // Научная визуализация. 2015. Т. 7. № 4. С. 76-86.
- [7] Расковская И.Л., Ринкевичюс Б.С., Толкачев А.В. Лазерная рефракционная термометрия прозрачных твердых тел при неоднородном нагреве // Измерительная техника, 2016, №10, С.42-44.
- [8] Расковская И.Л., Ринкевичюс Б.С., Толкачев А.В., Ширинская Е.С. Рефракция цилиндрического лазерного пучка в температурном пограничном слое. // Оптика и спектроскопия. 2009. Т. 106. № 6. С. 992-997.
- [9] Ринкевичюс Б.С., Ведяшкина А.В., Расковская И.Л., Толкачев А.В., Пудовиков Д.Е. 3d-визуализация и количественная диагностика процессов тепло-и массообмена с помощью метода лазерной рефрактографии. В сборнике: ГРАФИКОН'2016, Труды 26-й Международной научной конференции. 2016. С. 156-160.
- [10] Ринкевичюс Б.С., Есин М.В., Расковская И.Л., Толкачев А.В. Способ визуализации и измерения параметров физических процессов в жидкой среде. Патент на полезную модель RUS 105433 30.11.2010.
- [11] Павлов И.Н., Ринкевичюс Б.С., Толкачев А.В. Установка для визуализации испарения капли жидкости методом нарушенного полного внутреннего отражения лазерного пучка // Приборы и техника эксперимента, 2013, № 2. С. 130–135.
- [12] Павлов И.Н., Ринкевичюс Б.С., Толкачев А.В. Визуализация процессов кристаллизации в пристеночном слое капли воды // Метрология, 2013. №3. С. 11-17.
- [13] И.Н. Павлов, И.Л. Расковская, А.В. Толкачев, Структура микрорельефа поверхности испаряющейся с шероховатой подложки капли как возможная причина гистерезиса краевого угла, ЖЭТФ, 2017, Т. 151, №4, 670–681.
- [14] Павлов И.Н., Расковская И.Л., Ринкевичюс Б.С. Восстановление профиля поверхности капли жидкости на основе послыонного лазерного зондирования // ПЖТФ, 2017, Т. 43, № 13, С.19-25.

Об авторах

Ринкевичюс Бронюс Симович, д.ф.-м.н., профессор кафедры физики им. В.А. Фабриканта НИУ «МЭИ». E-mail: rinkevbs@mail.ru.

Павлов Илья Николаевич, к.т.н., доцент кафедры физики им. В.А. Фабриканта НИУ «МЭИ». E-mail: inpavlov@bk.ru.

Расковская Ирина Львовна, к.ф.-м.н., доцент кафедры физики им. В.А. Фабриканта НИУ «МЭИ». E-mail: raskovskail@mail.ru.

Толкачев Александр Владимирович, к.т.н., доцент кафедры физики им. В.А. Фабриканта НИУ «МЭИ». E-mail: avtolk2008@yandex.ru.

Интеграция инструментария научной визуализации SciVi с информационной системой Семограф

К.В. Рябинин¹, Д.А. Баранов², К.И. Белоусов¹

belousovki@gmail.com|kostya.ryabinin@gmail.com|baranov@semograph.com

¹Пермский государственный национальный исследовательский университет, Пермь, Россия;

²Оренбургский государственный университет, Оренбург, Россия

Данная работа посвящена вопросу выделения из адаптивной мультиплатформенной клиент-серверной системы научной визуализации SciVi настраиваемого модуля визуализации, пригодного для включения в состав сторонних по отношению к этой системе программных продуктов, в частности, в состав информационной системы графосемантического моделирования Семограф. Выделенный модуль, являясь, по сути, самостоятельным программным средством, тем не менее остаётся частью системы SciVi, наследуя от неё управляемость онтологической базой знаний о доступных графических объектах, типах сцен и семантических фильтрах. Как следствие, он имеет возможность расширения функциональности и гибкой настройки на различные задачи визуализации. В этот модуль добавлены средства для отображения графов, успешно протестированные на реальных данных, полученных из информационной системы Семограф. Детально проработана концепция встраивания созданного модуля в состав системы Семограф. Следующим этапом работы в рамках описанного проекта является интеграция указанных программных продуктов, а также расширение системы SciVi новыми графическими возможностями, в частности, для поддержки отображения геоинформационных данных, включающих как пространственные, так и временные координаты.

Ключевые слова: научная визуализация, информационная система, графосемантическое моделирование, Web-приложение.

Integration of Scientific Visualization Toolset SciVi with Information System Semograph

K.V. Ryabinin¹, D.A. Baranov², K.I. Belousov¹

belousovki@gmail.com|kostya.ryabinin@gmail.com|baranov@semograph.com

¹Perm State University, Perm, Russia

²Orenburg State University, Orenburg, Russia

This paper describes issues of extracting data rendering module from adaptive multiplatform client-server scientific visualization system SciVi to integrate it into third-party software. The concept of integration with graphosemantic modelling information system Semograph is elaborated. While the extracted rendering module is actually stand-alone application, it still remains to be a part of SciVi and therefore is driven by ontology knowledge base, which contains knowledge about graphical objects, scenes' types and semantic filters available. This module can be easily extended and set up for arbitrary visualization tasks. Currently it was tuned to visualize graphs and was tested on the real data obtained from Semograph. The next step of this project's work is integration of this module into Semograph and extending its rendering capabilities with the tools ready to visualize geoinformation data including spatial and temporal coordinates.

Keywords: scientific visualization, information system, graphosemantic modelling, Web-application.

1. Введение

При разработке информационных систем (ИС), предназначенных для работы со сложноорганизованными данными в различных предметных областях, часто возникает задача поддержки в составе функциональности этих систем средств визуализации. Для максимально полного покрытия задач отображения данных зачастую требуется реализация модулей визуализации с нуля, либо серьёзная модификация существующих библиотек рендеринга. В обоих случаях трудоёмкость требуемых доработок оказывается достаточно высокой, особенно если структуры данных, подлежащих наглядному изображению, нетривиальны.

Возможным вариантом экономии трудовых затрат является использование адаптивных систем научной визуализации [3] – специализированного программного обеспечения (ПО), взаимодействующего с разрабатываемой ИС по какому-либо стандартизированному протоколу (ТСР/IP стек, общая база данных, текстовые файлы, в которые производится экспорт подлежащих визуализации данных и т. п.). Такой подход оказывается очень удобным в контексте узкоспециализированных решателей (англ.

Solvers) и хранилищ данных, то есть средств, играющих роль исключительно источников данных.

Однако в контексте построения сложных ИС, обладающих собственным многофункциональным интерфейсом обработки данных, такой подход оказывается недостаточно эргономичным: пользователю ИС придётся прерывать работу с ней всякий раз, когда ему необходима будет визуализация. В лучшем случае прерывание будет состоять в простом переходе к другому приложению на уровне смены активного окна, а в худшем – в совершении каких-то подготовительных действий, например, экспорта данных.

С точки зрения эргономики значительно лучше оказывается включить визуализатор в состав самой ИС так, чтобы пользователь имел возможность совершать все интересующие его действия как по обработке, так и по визуализации данных в рамках единого интерфейса.

Для обеспечения такой возможности авторами было принято решение доработать предложенную ранее концепцию построения адаптивных мультиплатформенных систем научной визуализации [3] с учётом поддержки встраивания этих систем как отдельных модулей в состав другого ПО.

В результате предыдущих исследований была создана адаптивная система научной визуализации SciVi [8], основанная на принципах онтологического инжиниринга [1, 9]. Её поведение гибко конфигурируется посредством пополнения и/или изменения онтологической базы знаний о языках программирования совместимых с ней решателей, поддерживаемых семантических фильтрах входных данных, визуальных объектах и типах графических сцен. Благодаря этому данная система может быть быстро настроена на решение задач визуализации из любой предметной области и на взаимодействие с широким спектром различных источников данных.

Параллельно с разработкой системы SciVi развивалась ИС Семограф [5]. ИС Семограф предназначена для решения широкого спектра научных, образовательных и прикладных задач, связанных с обработкой и анализом языковых и текстовых данных. Типовыми задачами, решаемыми с помощью инструментария системы, являются: создание классификаторов и тезаурусов предметных областей, проведение психолингвистических, социолингвистических и пр. экспериментов (а также анализ полученных данных), создание и разметка выборок и корпусов текстов, построение информационных моделей предметных областей, и другие задачи, возникающие в ходе анализа текстового контента.

Информационная система реализована как Web-приложение и в настоящее время состоит из модулей, обеспечивающих следующую функциональность:

1. Поиск (с помощью поискового робота и парсера, созданного на основе фреймворк Scrapy на языке Python).
2. Импорт данных (с помощью поискового сервера на основе Apache Solr; кроме того поддерживается импорт файлов формата CSV).
3. Экспорт результатов во внешние приложения, в частности, R (статистическую среду анализа данных), Gephi (средство построения и анализа графов), а также в табличные форматы.
4. Администрирование и организационная деятельность (добавление приглашённых пользователей системы к проектам, управление правами доступа, создание открытых проектов, коммуникация).
5. Исследование (набор инструментов проведения экспериментов, анализа языкового контента, разметки корпусов, классификационной деятельности и др.).

ИС Семограф имеет API, который может использоваться для работы с внешними приложениями или разработки приложений на основе этой ИС с расширенной в каком-либо аспекте функциональностью.

Целью данной работы является интеграция систем SciVi и Семограф для создания отдельного модуля научной визуализации в ИС Семограф. Внедрение инструментария, позволяющего пользователям работать с языковыми и текстовыми данными, обогащёнными наглядной визуализацией, позволит реализовать возможность осуществления полного цикла исследования: от постановки проблемы до генерации отчёта по научному, образовательному или прикладному проекту.

2. Встраиваемая версия системы SciVi

Система научной визуализации SciVi состоит из серверной и клиентской частей. Сервер служит в первую очередь для связи с источником данных и их тяжеловесной предобработки (если таковая требуется при решении конкретной задачи визуализации). Клиент отвечает за связь с пользователем, предоставляя в первую очередь графический интерфейс системы визуализации. На данный момент активно развиваются толстый и тонкий клиенты.

Рендеринг данных адаптивно распределяется между клиентом и сервером с учётом производительности клиентской машины, загрузки сервера и скорости сетевого соединения.

Для включения системы SciVi в виде отдельного компонента в состав стороннего по отношению к ней ПО требуется возможность объединения в рамках этого компонента необходимой функциональности сервера и клиента. Это позволит избавиться от внешних зависимостей и минимизировать накладные расходы на пересылку данных внутри итоговой ИС.

Сервер и клиент SciVi имеют модульную архитектуру. Их функциональные блоки полностью изолированы друг от друга и взаимодействуют между собой посредством вызова интерфейсных методов, использования функций обратного вызова (англ. *Callback*) и передачи декларативных описаний (онтологических профилей [4]) обрабатываемых сущностей. Благодаря этому сборка на основе системы SciVi встраиваемых компонентов визуализации оказывается возможной и не требует глубинных изменений реализованной и отлаженной ранее базовой функциональности, такой, например, как механизмы взаимодействия с источником данных, рендеринг графических сцен и т.п.

3. Концепция интеграции с системой Семограф

ИС Семограф реализована с использованием современных практик разработки Web-приложений. В основе лежит принцип SPA (англ. *Single Page Application* – одностраничное приложение), что означает отсутствие лишних переходов между страницами, характерных для большинства обычных сайтов. Вместо этого приложение подгружает необходимые данные по мере надобности с помощью AJAX (англ. *Asynchronous Javascript and XML*). Благодаря этому достигается высокая отзывчивость пользовательского интерфейса при минимальном использовании сетевого трафика.

В качестве ключевых решений для организации клиентской части используются библиотеки React и Redux, для серверной – фреймворк Flask и СУБД PostgreSQL. Клиентская часть использует стандарт ECMAScript 6 и при сборке преобразуется в ECMAScript 5 с помощью пакета Babel. Для сборки клиентской части используется Webpack. Такой подход повышает скорость разработки, но накладывает ряд ограничений на используемые библиотеки, например, они должны поддерживать экспорт методов в формате CommonJS или совместимом с ним.

Для интеграции системы SciVi с ИС Семограф предполагается разработка компонента (в терминологии React), адаптирующего программный интерфейс тонкого клиента SciVi к стандартам приложений, разработанных с использованием React и Redux. При таком подходе данные будут следовать стандартному циклу ИС Семограф – загружаться по требованию с сервера через API системы Семограф, помещаться в состояние Redux, при необходимости обрабатываться и передаваться в новый компонент. Большим преимуществом в данном случае является отсутствие необходимости модифицировать сервер или его API, а также возможность использовать существующую систему контроля прав доступа.

Жизненный цикл компонента-адаптера должен включать конструирование, передачу данных и управление модулем визуализации на основе тонкого клиента SciVi. Также данный компонент должен обеспечить передачу управляющих команд от модуля визуализации в основное приложение посредством механизма функций обратного вызова. Таким образом, компонент должен реализовать как

функции представления, так и контроллера (согласно паттерну проектирования Model-View-Controller).

Полученный компонент может быть обособлен от ИС Семограф, унифицирован и собран в виде отдельного модуля, что позволит применять его в других приложениях, базирующихся на React и Redux.

4. Пополнение набора средств визуализации системы SciVi

Многие аналитические задачи, решаемые в ИС Семограф, предполагают работу с данными о связности различных сущностей. Такие данные нагляднее всего представимы в виде графов [6, 7].

Для обработки и отображения графов в систему SciVi были добавлены новые семантические фильтры (позволяющие по различным правилам вычислять положения вершин и характеристики связей графа) и визуальные объекты (позволяющие отображать именованные вершины и связи между ними). Добавление осуществлено путём пополнения соответствующих онтологий, входящих в базу знаний системы, а также путём подключения дополнительной библиотеки визуализации.

Каждая вершина визуализируемого графа характеризуется названием, идентификатором кластера, в который входит связанная с вершиной сущность, и весом. Кластеризация сущностей, отображаемых в виде графа, происходит на стороне ИС Семограф; на данный момент эта функциональность находится в стадии разработки.

Названия вершин отображаются в виде текстовых меток, повернутых под определённым углом, который для каждой вершины определяется вычисляющим её положение фильтром.

Принадлежность вершины к конкретному кластеру отображается цветом: каждому кластеру назначается свой определённый цвет из заданной по умолчанию таблицы; далее у пользователя есть возможность изменять эти цвета. Вершины располагаются радиально и группируются по принципу принадлежности к одному кластеру.

Вес вершин отображается в виде гистограммы, каждый столбец которой рисуется под текстовой меткой соответствующей ему вершины.

Каждая связь в графе характеризуется весом и отображается в виде квадратичной кривой Безье, толщина которой пропорциональна её весу. Цвет связи определяется цветами соединяемых ей вершин (если вершины имеют разные цвета, связь окрашивается градиентом).

В дальнейшем набор атрибутов вершин и связей может быть расширен.

Визуализатор графов поддерживает выделение вершины (с автоматическим подсвечиванием более ярким цветом всех инцидентных ей связей и всех смежных с ней вершин), а также произвольное перемещение вершин курсором мыши (либо пальцем в случае сенсорного экрана мобильного устройства). Таким образом пользователь может осуществлять перегруппировку вершин графа и исследовать их связность.

Настройка конкретного способа визуализации осуществляется пользователем путём редактирования диаграммы потока данных средствами высокоуровневого графического Web-интерфейса [8]. Эта диаграмма определяет последовательность шагов предобработки данных (семантической фильтрации), а также связь данных с теми или иными атрибутами поддерживаемых системой визуализации графических объектов. Набор доступных для построения диаграммы вершин, олицетворяющих семантические фильтры, графические объекты и типы сцен, описан в соответствующих онтологиях системы.

На рис. 1 представлен пример диаграммы потока данных, задающей настройку визуализации графа с радиальным расположением вершин.

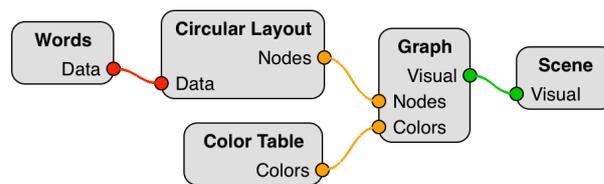


Рис. 1. Диаграмма потока данных, определяющая настройки визуализации графа понятий в системе научной визуализации SciVi.

Результат визуализации графа, состоящего из 166 вершин и 4917 связей, показан на рис. 2. На текущий момент это единственный реализованный способ визуализации графов, однако в дальнейшем, при необходимости, функциональность SciVi по отображению связанных и иерархических сущностей может быть расширена.

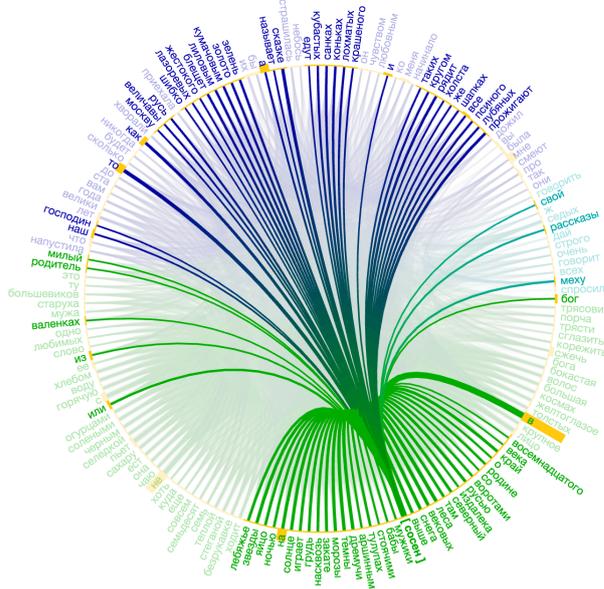


Рис. 2. Визуализация графа понятий в системе научной визуализации SciVi, построенного по данным, извлечённым из ИС Семограф.

Входные данные представляют собой результаты психолингвистического эксперимента по тематическому картированию текста. Информантам предъявляется текст и ставится задача определить его тему, выделить микротемы и к каждой микротеме отнести слова, представляющие её в тексте. Эксперимент проводится в ИС Семограф; одним из результатов исследования становится генерация семантической карты текста, представляющей собой матрицу $n \times n$, где n – число слов в тексте. В ячейках отображаются значения совместной встречаемости двух слов во всех выделенных микротемах. Особенностью матрицы является фиксированный порядок столбцов и строк, отражающий порядок слов в анализируемом тексте.

В результате эксперимента с визуализацией было выяснено, что приемлемой производительности (отклика быстрее, чем за 125 мс [2]) на желаемых объёмах данных удаётся достичь только используя аппаратное ускорение рендеринга. Так, например, использование для визуализации графа стандарта SVG (посредством библиотеки D3.js) в браузерах Mozilla Firefox (версия 51) и

Google Chrome (версия 56) на компьютере с процессором Intel Core i7 (4 ядра тактовой частотой 2 ГГц, без поддержки гипертрединга), объёмом оперативной памяти 8 Гб и видеокарткой Intel Iris Pro с объёмом видеопамати 1536 Мб перерисовка графа при перемещении курсора над текстовыми метками длится в среднем 1000 мс.

В связи с этим для рендеринга графа было решено использовать библиотеку PixiJS, предоставляющую удобную в использовании обёртку над низкоуровневым графическим API WebGL. Данная библиотека позволила добиться средней скорости отклика в 20 мс, что вполне удовлетворяет требованиям, предъявляемым к производительности.

5. Заключение

В результате проделанной работы сформулирована концепция встраивания системы научной визуализации SciVi в сторонние по отношению к ней приложения, а также произведены подготовительные действия по интеграции этой системы с ИС Семограф. Детально проработана последовательность шагов, необходимая для такой интеграции. В систему SciVi добавлены новые средства визуализации, служащие для наглядного представления данных, обрабатываемых в ИС Семограф, в частности средства для отображения графов. Из тонкого клиента системы SciVi выделен автономный модуль, готовый к встраиванию в сторонние Web-приложения. В целях обеспечения высокой эффективности рендеринга в нём использованы механизмы аппаратного ускорения графики, основанные на API WebGL. Работоспособность данного модуля успешно протестирована на реальных данных, извлечённых из ИС Семограф.

Полученный модуль визуализации может быть использован как автономное программное средство, но при этом остаётся составной частью системы SciVi. Это позволяет развивать его функциональность одновременно с развитием системы SciVi, тестировать новые средства визуализации, используя адаптационные механизмы этой системы, пополнять репозиторий доступных визуальных объектов, типов графических сцен и семантических фильтров путём расширения онтологий, составляющих базу знаний SciVi.

В дальнейшем планируется встроить созданный модуль визуализации в ИС Семограф, а также по мере необходимости пополнять его новыми графическими возможностями. В частности, приоритетным является реализация средств отображения геоинформационных данных с поддержкой временных срезов.

6. Благодарности

Работа выполнена в рамках государственного задания Минобрнауки России (проект 34.1505.2017/4.6).

7. Литература

- [1] Chuprina S., Nasraoui O. Using Ontology-based Adaptable Scientific Visualization and Cognitive Graphics Tools to Transform Traditional Information Systems into Intelligent Systems // *Scientific Visualization*. – М.: National Research Nuclear University "MEPhI", 2016. – Q. 1, V. 8, No1. – С. 23–44.
- [2] Keval H., Sasse M.A. To catch a thief – you need at least 8 frames per second: the impact of frame rates on user performance in a CCTV detection task // *Proceedings of the 16th ACM international conference on Multimedia*. – ACM, 2008. – P. 941–944.

- [3] Ryabinin K., Chuprina S. Development of Ontology-Based Multiplatform Adaptive Scientific Visualization System // *Journal of Computational Science*. – Elsevier, 2015. – Vol. 10. – P. 370–381.
- [4] Ryabinin K., Chuprina S. High-Level Toolset For Comprehensive Visual Data Analysis and Model Validation // *Procedia Computer Science*. – Elsevier, 2017. – Vol. 108. – P. 2090–2099.
- [5] Баранов Д.А., Белоусов К.И., Влацкая И.В., Зелянская Н.Л. Система графосемантического моделирования [программа для ЭВМ] – М.: Свидетельство о государственной регистрации в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам. Зарегистрировано в Реестре программ для ЭВМ № 20111617192 от 15.09.2011.
- [6] Касьянов В.Н., Золотухин Т.А. Visual Graph — система для визуализации сложно структурированной информации большого объема на основе графовых моделей // *Научная визуализация*. – М.: Национальный исследовательский ядерный университет МИФИ, 2015. – К. 4, Т. 7, №4. – С. 44–59.
- [7] Касьянов В., Касьянова Е. Визуализация информации на основе графовых моделей // *Научная визуализация*. – М.: Национальный исследовательский ядерный университет МИФИ, 2014. – К. 1, Т. 6, №1. – С. 31–50.
- [8] Рябинин К.В., Чуприна С.И., Бортников А.Ю. Автоматизация настройки систем научной визуализации на специфичу разнообразных источников данных // *Научная визуализация*. – М.: Национальный исследовательский ядерный университет МИФИ, 2016. – К. 4, Т. 8, №4. – С. 1–14.
- [9] Чуприна С.И., Зиненко Д.В. ОНТОЛИС: адаптируемый визуальный редактор онтологий // *Вестник Перм. нац. исследоват. поли-техн. ун-та*. – 2013. – №3(22). – С. 106–110.

Об авторах

Константин Валентинович Рябинин, к.ф.-м.н., доцент кафедры математического обеспечения вычислительных систем Пермского государственного национального исследовательского университета. Его электронный адрес kostya.ryabinin@gmail.com.

Дмитрий Александрович Баранов, преподаватель кафедры компьютерной безопасности и математического обеспечения информационных систем Оренбургского государственного университета. Его электронный адрес baranov@semograph.com.

Константин Игоревич Белоусов, докт. филол. н., профессор кафедры теоретического и прикладного языкознания Пермского государственного национального исследовательского университета. Его электронный адрес belousovki@gmail.com.

Адаптация систем научной визуализации к задачам мониторинга данных от легковесных аппаратных робототехнических устройств

К.В. Рябинин¹, С.И. Чуприна¹

kostya.ryabinin@gmail.com|chuprinas@inbox.ru

¹Пермский государственный национальный исследовательский университет, Пермь, Россия

Данная работа посвящена вопросам адаптации систем научной визуализации к решению задач мониторинга данных, генерируемых в реальном времени легковесными роботизированными устройствами, на примере мультиплатформенной клиент-серверной системы SciVi. Система научной визуализации строится на принципах модельно-ориентированного подхода. Поведение системы управляется онтологической базой знаний, что позволяет без модификации исходного кода ядра этой системы настроить её на решение задач визуализации данных из любой предметной области. За получение данных из стороннего источника и первичную их обработку отвечает серверная часть системы; результат визуализации демонстрируется пользователю на стороне клиента. С целью обеспечения эффективного взаимодействия с роботизированными устройствами в клиентской части SciVi был поддержан режим работы, позволяющий получать данные напрямую из источника без необходимости соединения с сервером. Описываются результаты тестирования системы в этом режиме на примере мониторинга углов курсовертикали.

Ключевые слова: научная визуализация, онтологический инжиниринг, системы реального времени, робототехника.

Tackle Lightweight Hardware Robotic Devices Data Monitoring Problems by Means of Scientific Visualization Systems

K. V. Ryabinin¹, S. I. Chuprina¹

kostya.ryabinin@gmail.com|chuprinas@inbox.ru

¹Perm State University, Perm, Russia

This paper is devoted to improving of the multiplatform client-server scientific visualization system SciVi adaptation capabilities to tackle lightweight hardware robotic devices data monitoring problems. SciVi is an ontology based model-driven system, which allows tuning to arbitrary application domain without core source code modification. The SciVi server is responsible for data retrieving and preprocessing, the client demonstrates the final rendering result to the user. The client has been improved to support direct communication with data source without server connection to ensure efficient interaction with external robotics device. This operating mode was tested on the attitude and heading reference system's orientation monitoring.

Keywords: scientific visualization, ontology engineering, real-time systems, robotics.

1. Введение

В настоящее время благодаря удешевлению элементной базы (управляющих и исполнительных элементов, а также источников электропитания), литьевых и формовочных материалов (пластиков и силиконов), а также распространению средств быстрого прототипирования (3D-печати, лазерной резки и т. п.) робототехника стала объектом пристального интереса не только крупных научно-исследовательских центров и производств, но и широких масс населения. Например, широкое распространение получили роботы-пылесосы, интерактивные игрушки и т. п. Кроме того, выделилось целое направление хобби-робототехники, предполагающее сборку роботизированных устройств в домашних условиях, а основы робототехники как науки и ремесла были включены в школьную программу в виде факультативных занятий [8].

На базе робототехники существуют различные концепции развития вычислительной техники и инфраструктуры её взаимодействия, например Интернет вещей (англ. *Internet of Things, IoT*) [3], умный дом (англ. *Smart Home*) [7], электронный текстиль (англ. *E-Textiles*) [4] и др.

В сфере непромышленной робототехники существует множество доступных по цене и лёгких в освоении аппаратных решений. Наиболее популярные из них стали на сегодняшний день полноценными легковесными платформами для разработки роботизированных устройств. К таким платформам относятся, например, Arduino [6],

предназначенная для решения широкого спектра задач автоматизации, и используемые в основном для решения задач в области Интернета вещей сборки на основе микроконтроллеров линейки ESP [2].

Одновременно с ростом популярности аппаратных средств для построения роботов и автоматизированных систем на их основе, возрос спрос и на программное обеспечение, позволяющее моделировать, программировать, мониторить и контролировать соответствующие устройства. Когда речь идёт о сборе, обработке и интерпретации данных, получаемых от роботизированной системы (например, комплексных параметров состояния жилых и нежилых помещений, снимаемых в реальном времени датчиками умного дома), в роли подобного программного обеспечения могут выступать системы научной визуализации.

Ранее авторами была предложена концепция создания адаптивных мультиплатформенных систем научной визуализации на основе методов и средств онтологического инжиниринга [9]. Эта концепция была воплощена на практике в виде программной системы SciVi.

Целью данной работы является дальнейшее совершенствование адаптационных возможностей систем научной визуализации для решения задач мониторинга легковесных непромышленных роботизированных устройств на примере системы SciVi. Это позволит применять систему SciVi для прототипирования и малой автоматизации целого ряда процессов как в научно-технической, так и в бытовой сфере.

2. Проблема взаимодействия с системами реального времени

Благодаря модельно-ориентированному подходу, лежащему в основе архитектуры системы SciVi, эта система при помощи средств высокоуровневого графического интерфейса может быть адаптирована к стороннему источнику данных без модификации исходного кода её ядра [9]. Взаимодействие с источником данных осуществляется через серверную часть SciVi, тогда как демонстрация результатов визуализации происходит средствами клиентской части. В зависимости от сложности данных и вычислительной мощности клиента, визуализация может происходить как на стороне сервера, так и на стороне клиента, либо быть распределена между этими вычислительными узлами.

Такой подход обеспечивает достаточную универсальность при визуализации данных из произвольных источников. Однако в некоторых частных случаях этот подход может оказаться неоптимальным, например, при взаимодействии с системами реального времени, к которым относится большинство робототехнических устройств.

Согласно стандарту POSIX 1003.1, система реального времени – это система, способная обеспечить требуемый уровень сервиса в определённый промежуток времени, т. е. реагировать в предсказуемое время на непредсказуемое появление внешних событий. При визуализации данных, получаемых от таких систем, время их сбора и рендеринга не должно превышать времени их генерации. Только в этом случае пользователь сможет производить адекватный мониторинг и своевременно реагировать на произошедшие события. Соответственно, наиболее оптимальным оказывается сценарий прямого взаимодействия визуализатора с источником данных, по возможности минуя посредников.

В контексте взаимодействия с легковесными робототехническими платформами речь, как правило, идёт об относительно небольших объёмах данных, подлежащих визуализации. Это обусловлено малой вычислительной мощностью аппаратной части. Так, например, тактовая частота используемых в этих платформах микроконтроллеров обычно не превышает 100 МГц, а объём оперативной памяти чаще всего оказывается менее 1 Мб.

В связи с этим даже в случае, когда в роли клиента системы SciVi выступает маломощное мобильное устройство, система окажется в состоянии обработать и визуализировать весь требуемый объём данных. Следовательно, более оптимальным в этом случае оказывается сбор данных на стороне клиента, минуя сервер.

Для пользователя исключение сервера SciVi из цепочки обработки данных при работе с робототехническими устройствами упрощает работу с системой визуализации, так как делает инфраструктуру вычислительной сети более тривиальной и интуитивно-понятной. Кроме того, этот режим позволяет системе SciVi, развёрнутой, например, на смартфоне, планшетном компьютере или ноутбуке, взаимодействовать с робототехническими устройствами, работающими в режиме точки доступа WiFi.

Для поддержки эффективного взаимодействия системы SciVi с легковесными робототехническими устройствами были выполнены следующие расширения:

1. Настройка на новые виды источников данных.
2. Реализация компонента получения данных напрямую с робототехнического устройства.

3. Настройка системы SciVi на взаимодействие с робототехническим устройством

Настройка системы SciVi на сторонний источник данных осуществляется посредством Web-интерфейса её серверной части. В том случае, если источник данных представлен программно-аппаратным решателем и пользователь предоставляет доступ к исходному коду программной части решателя, процесс настройки оказывается во многом автоматизирован. Система автоматически генерирует синтаксический анализатор для поданного на вход кода решателя на основе онтологии синтаксических конструкций ввода-вывода языков программирования [9]. В результате анализа из этого кода извлекаются описания входных и выходных данных, и пользователю предоставляется высокоуровневый графический интерфейс для выбора интересующих его параметров. В результате такой настройки формируется онтологический профиль источника данных, используемый затем для автоматического построения входной вершины в диаграмме потока данных, используемой в качестве спецификации внутреннего конвейера визуализации [9].

Если исходный код недоступен, пользователь может самостоятельно описать входные и выходные данные посредством высокоуровневого графического Web-интерфейса.

В случае настройки на работу с роботизированным устройством исходным кодом решателя выступает код прошивки соответствующего устройства. Как правило, этот код бывает доступен пользователю, особенно, если речь идёт о робототехнике, используемой для целей образовательного процесса, или о хобби-робототехнике: пользователь либо сам пишет этот код, либо скачивает прошивку для создаваемого им устройства из сети Интернет. Большинство популярных готовых прошивок распространяются в виде исходного кода (как правило, по OpenSource-лицензиям).

Взаимодействие с легковесными встраиваемыми системами осуществляется через проводные и беспроводные интерфейсы. К наиболее популярным проводным интерфейсам относятся I²C, SPI и UART. Беспроводная связь организуется посредством WiFi и Bluetooth. Для каждого способа коммуникации есть типовые операции ввода-вывода, реализованные для наиболее популярных платформ в виде библиотек, ставших стандартом де-факто. К таким библиотекам относятся, например, Serial и WiFi, предназначенные для коммуникации по стандартам RS-232 (UART) и IEEE 802.11 (WiFi) соответственно. Эти библиотеки изначально были реализованы на языке C++ для платформы Arduino, а затем портированы и на другие популярные платформы (например платформы на базе микроконтроллеров линейки ESP).

Основой организации поддержки взаимодействия SciVi с робототехническими платформами стало расширение онтологии синтаксических конструкций ввода-вывода описанием соответствующих вызовов популярных коммуникационных библиотек. Для реализации тестового примера и проверки жизнеспособности подхода были поддержаны программные интерфейсы библиотек Serial и WiFi на языке C++. Для библиотеки WiFi поддержаны функции приёма и передачи данных по протоколам TCP и UDP, при этом используемый протокол сохраняется в онтологическом профиле источника данных в виде метайнформации.

Модификации, связанные с обеспечением возможности настройки на соответствующий источник данных (робототехническое устройство), в соответствии с

модельно-ориентированным подходом, свелись к пополнению онтологии синтаксических конструкций ввода-вывода, а также добавлению в набор настроек единственной новой опции – флага использования режима прямого доступа к источнику данных (минуя сервер). Значение этой опции сохраняется в онтологический профиль источника данных в виде метаданных.

Этап настройки по-прежнему осуществляется на стороне сервера SciVi, так как перенос его на сторону клиента для мультиплатформенных систем не оправдан с технической точки зрения. Например, в случае, если клиент работает под управлением ОС iOS, для загрузки в систему файлов исходного кода решателя (с целью осуществления настройки на стороне клиента) потребовалось бы использование сторонних сервисов хранения (таких как Dropbox или Яндекс.Диск).

4. Режим работы системы SciVi без сервера

Для обеспечения поддержки прямого взаимодействия клиента SciVi с роботизированным устройством в состав клиента был добавлен расширяемый модуль связи, названный SciViATRobot, работающий с наиболее популярными средствами коммуникации – UART, WiFi и Bluetooth. Архитектура модуля схематично представлена на рис. 1.

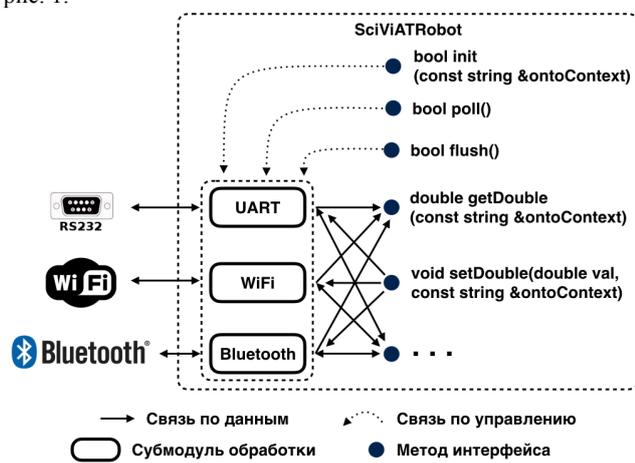


Рис. 1. Схема модуля SciViATRobot.

На программном уровне модуль имеет единообразный интерфейс для всех коммуникационных средств (в связи с чем при необходимости его функциональность может быть расширена). Этот интерфейс включает в себя метод инициализации соединения (init), а также методы запроса и отправки данных всех поддерживаемых базовых типов – целых и вещественных чисел, строк, векторов и т. п. (например, методы getDouble и setDouble для оперирования вещественными числами). Позиции запрашиваемых и отправляемых значений во входном и выходном сообщениях, а также используемый протокол передачи данных (например, для WiFi – TCP или UDP) определяются на основе онтологического профиля источника данных, созданного сервером системы SciVi. Ввод в вывод буферизированы, в связи с чем интерфейсом дополнительно предусмотрены методы опроса внешнего устройства (poll) и отправки сообщений (flush).

В том случае, если при настройке на источник данных пользователем был выбран режим прямого доступа, все входные данные системы визуализации с заданной периодичностью запрашиваются путём обращения к интерфейсу модуля SciViATRobot. Периодичность зависит от скорости передачи данных (декларированной конкретным стандартом связи) и скорости перерисовки графической сцены.

Следует отметить, что реализация и возможности этого модуля зависят от программно-аппаратной платформы клиента в силу различий в наборах коммуникационных средств (например, на современных мобильных устройствах отсутствует аппаратный UART).

5. Пример мониторинга системы реального времени

В качестве тестового примера системы реального времени на основе робототехнической платформы, авторами была собрана курсоверткаль (англ. *Attitude and Heading Reference System, AHRS*) [10] – прибор для определения углов Тейта-Брайана: крена (англ. *Roll*), тангажа (англ. *Pitch*) и рыскания (англ. *Yaw*). Основной элементной базой этого устройства являются следующие компоненты:

1. Трёхосевой гироскоп марки MPU3200, служащий для определения углов крена и тангажа.
2. Трёхосевой акселерометр марки ADXL345, служащий для компенсации погрешности измерений гироскопа.
3. Трёхосевой магнитометр марки HMC5883L, служащий для определения угла рыскания (на основе известных углов крена и тангажа).
4. WiFi-модуль ESP-07 на базе микроконтроллера ESP8266, служащий для управления устройством и передачи данных в эфир.

Устройство является полностью автономным (источником питания служат элементы типа AAA) и настроено на коммуникацию с внешними системами посредством сети WiFi, работая в режиме точки доступа.

Прошивка устройства была написана авторами с использованием библиотек [1], реализованных сообществом ESP8266-разработчиков. Микроконтроллер ESP8266 – самый популярный представитель линейки ESP, имеющий несколько вариантов компоновки, в частности, модель ESP-07 (производимая компанией AI-Thinker), отличающаяся очень малым размером, но, при этом, наличием встроенной керамической WiFi-антенны, обеспечивающей стабильную связь на расстоянии до 300 метров. Для него силами международного сообщества разработчиков сформирована обширная программная инфраструктура и удобный инструментальный для написания прошивок.

Технические характеристики данного микроконтроллера позволяют решать с его помощью достаточно широкий спектр задач, включая достаточно сложную обработку данных, получаемых с периферических сенсоров. Его тактовая частота составляет 80 МГц, объём оперативной памяти – 64 Кб для кода и 96 Кб для данных (в соответствии с Гарвардской архитектурой, на принципах которой построен этот микроконтроллер, исполняемый код и данные находятся на различных запоминающих устройствах). На аппаратном уровне поддерживаются интерфейсы SPI, I²C, I²S и UART, а также аналого-цифровой преобразователь разрядностью 10 бит. Для подключения периферических сенсоров и других устройств имеется 10 свободных пинов ввода-вывода.

Созданная для ESP-07 прошивка включает в себя функции опроса гироскопа, акселерометра и магнетометра (связанных с микроконтроллером по интерфейсу I²C), определение по полученным данным углов крена, тангажа и рыскания, и сглаживание осцилляций этих углов (неизбежно возникающих в силу случайных шумов на сенсорах) при помощи фильтра Калмана [5]. Кроме того, предусмотрена функция выставки курсоверткаль – задания текущей ориентации устройства в качестве отсчётной (ориентации, при которой выдаваемые устройством углы одновременно равны нулю). Выставка

осуществляется нажатием тактовой кнопки. Радианные меры углов крена, тангажа и рыскания (с учётом выставки) с частотой 50 Гц передаются по протоколу UDP в сеть (в виде троек чисел с плавающей точкой двойной точности), для чего используется библиотека WiFi, портированная сообществом на ESP8266.

На внутреннем уровне текущая ориентация представляется в виде кватерниона q_c , который строится по обработанным фильтром Калмана значениям крена, тангажа и рыскания, получаемым с соответствующих сенсоров. Такой подход упрощает выставку: в момент нажатия соответствующей тактовой кнопки текущий кватернион q_c присваивается кватерниону q_r , задающему отсчётную ориентацию: $q_r = q_c$. Ориентация с учётом выставки (ориентация, углы которой отправляются в сеть) задаётся кватернионом $q_s = q_c \cdot q_r^{-1}$.

Настройка системы SciVi на мониторинг данных, генерируемых собранным устройством, была осуществлена путём подачи на вход системы исходного кода прошивки этого устройства (а также указания режима прямого доступа к источнику данных). В результате автоматически сформировалось описание источника данных как триплета вещественных чисел, задающих крен, тангаж и рыскание. На этапе настройки графической сцены углы ориентации были связаны с 3D-моделью, которая в процессе визуализации принимает получаемую от курсовертикали ориентацию. Диаграмма потока данных, используемая на клиенте SciVi, а также соответствующий результат визуализации представлены на рис. 2.

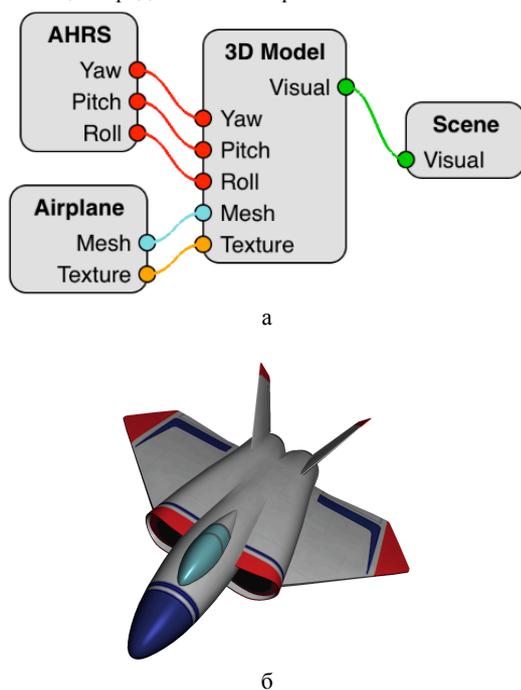


Рис. 2. Диаграмма потока данных (а) и результат визуализации ориентированной 3D-модели по данным с курсовертикали (б) на клиенте SciVi.

Вершина «AHRS» на диаграмме потока данных обозначает точку связи с курсовертикалью. Вершина «3D Model» отвечает за формирование визуального образа трёхмерной модели, имеющей ориентацию в пространстве (выраженную углами Тейта-Брайана), каркас и текстуру (для упрощения диаграммы другие параметры модели, такие, например, как положение в пространстве или свойства материала, были выставлены в значения по умолчанию и удалены из множества входных параметров на уровне онтологии визуальных объектов). Вершина «Airplane» определяет каркас и текстуру модели самолёта.

Вершина «Scene» служит для формирования графической сцены из поданных ей на вход визуальных объектов.

6. Заключение

В данной работе описаны результаты успешной адаптации системы SciVi к решению задач мониторинга данных, получаемых от легковесных робототехнических устройств, эффективного с точки зрения скорости и удобства настройки. Для этого в состав клиентской части системы был включён программный модуль прямой связи с роботизированным устройством, не нуждающийся во время своей работы в подключении к серверу SciVi.

Первичная адаптация к роботизированному устройству по-прежнему производится на стороне сервера, так как включает анализ исходного кода прошивки устройства, однако после этого этапа подключение к серверу более не требуется. Так, например, если сервер доступен в виде сервиса (англ. *Software as a Service, SaaS*) в сети Интернет, пользователь не должен будет заботиться о его развёртывании и обслуживании. Для формирования онтологического профиля соответствующего роботизированного устройства пользователю достаточно лишь кратковременного подключения к серверу для выполнения настройки. Затем можно отсоединиться от сети Интернет и, подключившись к локальной сети (возможно, созданной самим роботизированным устройством), осуществлять мониторинг аппаратной системы в реальном времени.

Взаимодействие системы SciVi с роботизированными устройствами было протестировано на примере мониторинга углов курсовертикали. В будущем планируется также отладить обратную связь SciVi с внешними устройствами, что позволит использовать её в режиме монитора, совмещённого с пультом управления робототехнической системой.

Следует также отметить, что в режиме работы без сервера система SciVi может эффективно взаимодействовать и с различными сторонними Web-сервисами, развёрнутыми в сети Интернет, например, сервисами погоды, курсов валют и т. п. Таким образом, разработанный режим в значительной степени увеличивает сферы её применимости.

7. Литература

- [1] Arduino IDE для ESP8266 // ESP8266 - Сообщество разработчиков. – 2014 [Электронный ресурс]. URL: <https://esp8266.ru/arduino-ide-esp8266/> (дата обращения 19.05.2017).
- [2] Brown E. Linux and Open Source Hardware for IoT // Linux.com. – 2016 [Электронный ресурс]. URL: <https://www.linux.com/news/linux-and-open-source-hardware-iot> (дата обращения 19.05.2017).
- [3] Brown E. Who Needs the Internet of Things? // Linux.com. – 2016 [Электронный ресурс]. URL: <https://www.linux.com/news/who-needs-internet-things> (дата обращения 19.05.2017).
- [4] Jamadar S. Applications of Smart and Interactive Textiles // Textile Learner. – 2013 [Электронный ресурс]. URL: <http://textilelearner.blogspot.ru/2013/04/applications-of-smart-and-interactive.html> (дата обращения 19.05.2017).
- [5] Kalman R.E. A New Approach to Linear Filtering and Prediction Problems // Journal of Basic Engineering. – 1960. – No. 82. – 11 p.
- [6] Kushner D. The Making of Arduino // IEEE Spectrum. – 2011 [Электронный ресурс]. URL:

<http://spectrum.ieee.org/geek-life/hands-on/the-making-of-arduino> (дата обращения 19.05.2017).

- [7] Yi Man Li R., Ching Yu Li H., Kei Mak C., Beiqi Tang T. Sustainable Smart Home and Home Automation: Big Data Analytics Approach // *International Journal of Smart Home*. – 2016. – Vol. 10, No. 8. – PP. 177–198.
- [8] Жупиков Ю.Ю. Робототехника как фактор формирования исследовательской компетентности учащихся в рамках внеурочной деятельности // *Сборник тезисов VI Всероссийской конференции «Современное технологическое обучение: от компьютера к роботу»*. – СПб., 2016. – С. 18–19.
- [9] Рябинин К.В., Чуприна С.И., Бортников А.Ю. Автоматизация настройки систем научной визуализации на специфику разнообразных источников данных // *Научная визуализация*. – М.: Национальный исследовательский ядерный университет МИФИ, 2016. – К. 4, Т. 8, №4. – С. 1–14.
- [10] Соловьёв В.И., Шабалов П.Г. Инерциальные навигационные системы // *Учеб. пособие; Самар. гос. нац. исслед. аэрокосм. ун-т. им. С.П. Королёва*. – Самара: СГАУ, 2011. – 74 с.

Об авторах

Константин Валентинович Рябинин, к.ф.-м.н., доцент кафедры математического обеспечения вычислительных систем Пермского государственного национального исследовательского университета. Его электронный адрес kostya.ryabinin@gmail.com.

Светлана Игоревна Чуприна, к.ф.-м.н., заведующая кафедрой математического обеспечения вычислительных систем Пермского государственного национального исследовательского университета. Её электронный адрес chuprinas@inbox.ru.

Особенности реализации естественных интерфейсов для прикладных задач

И.С. Стародубцев^{1,2}, Г.А. Зуева²

StarodubtsevIS@imm.uran.ru|g_z11@mail.ru

¹Институт математики и механики им. Н.Н.Красовского (ИММ УрО РАН), Екатеринбург, Россия;

²Уральский федеральный университет (УрФУ), Екатеринбург, Россия

Работа посвящена разработке жестовых интерфейсов для специализированных систем научной визуализации с использованием виртуальной реальности а так же проектированию естественных интерфейсов управления. В работе описывается реализация системы взаимодействия с виртуальными объектами, основанные на жестах, а так же общие проблемы взаимодействия с виртуальными объектами, особые требования, предъявляемые к интерфейсам для сред виртуальной и расширенной реальности и обосновываются предложенные методы.

Ключевые слова: естественные интерфейсы, жестовые интерфейсы, жесты, человеко-машинное взаимодействие.

Features for implementing natural interfaces for applications

I.S. Starodubtsev^{1,2}, G.A. Zueva²

StarodubtsevIS@imm.uran.ru|g_z11@mail.ru

¹Krasovskii Institute of Mathematics and Mechanics (IMM UB RAS), Yekaterinburg, Russia;

²Ural Federal University (UrFU), Yekaterinburg, Russia

The paper describes the implementation of the system of interaction with virtual objects based on gestures. The paper describes the common problems of interaction with virtual objects, specific requirements for the interfaces for virtual and augmented reality.

Keywords: natural interface, gesture-based interface, gesture, NUI, human-computer interaction.

1. Введение

Со времени создания компьютеры прошли большой путь в плане обеспечения *Человеко-Машинного Взаимодействия* (HCI). Были использованы и механические системы взаимодействия и аналоговые. На сегодняшний день самыми распространёнными являются *Графические Пользовательские Интерфейсы* (GUI), в которых пользователь использует классические средства ввода – мышку, клавиатуру или тачпад/тачскрин и графическое отображение результатов взаимодействия с компьютером. И, хотя GUI многое дал нам, он имеет и некоторые ограничения. Пришло время для следующего шага – *Естественное Человеко-Компьютерное Взаимодействие* (NUI), которое подразумевает использование средств, данных нам природой, своего собственного тела.

В современном мире компьютеризация широко проникает во все аспекты жизни. Сегодня компьютеры это не только то, к чему можно присоединить мышшь и клавиатуру. Это телефоны, телевизоры, игровые консоли и даже холодильники и микроволновки. И в некоторых случаях, выразительные средства мыши и клавиатуры оказываются избыточны, особенно если необходимо совершать небольшой набор несложных действий.

С другой стороны, встречается и обратная ситуация, когда подключение мыши и клавиатуры невозможно по тем или иным причинам. Например, управление медицинским оборудованием во время проведения операции. Использование традиционных средств

ввода приведёт к выходу из режима стерильности. Или же взаимодействие с объектами в средах виртуальной и расширенной реальности. В этом случае применение классических средств ввода ограничено их недостаточной выразительной способностью при работе с трехмерными объектами.

В этих случаях использование естественных интерфейсов может предложить решение, в котором пользователь управляет приложением непосредственно руками, используя определённый набор движений и жестов.

Разработку интерфейсов можно разделить на два направления: разработка массовых интерфейсов общего назначения и специализированных интерфейсов. От первых требуется максимальная универсальность, возможно и в ущерб удобству в конкретных ситуациях. В качестве примера можно привести интерфейсы командной строки и некоторые виды графических интерфейсов, использующих классические средства ввода, которые с разной степенью удобства подходят для решения самого разнообразного круга задач.

Вторые же разрабатываются исходя из конкретных задач и условий использования и ставят во главу угла именно удобство для решения конкретной задачи, использования в конкретных условиях или даже конкретным пользователем. В ущерб универсальности.

Далее будут рассмотрены результаты в области исследования и разработки специализированных интерфейсов, использующих NUI как основу. В частности,

будут рассмотрены требования к интерфейсу при использовании его для конкретных задач.

2. Практические задачи

При выполнении данных исследований перед нами ставились такие реальные задачи:

2.1. Навигация в абстрактном пространстве

Используемые при численном моделировании сетки иногда имеют дефекты, такие как вырожденные ячейки, самопересечения, нарушения структурной топологии, которые сложно искать и исправлять в автоматическом режиме. Несмотря на то, что работы в этой области ведутся очень активно и представлено большое количество результатов по автоматизации этого процесса, многие ситуации требуют ручной обработки специалистом.

В связи с этим родилась идея визуализации сеток с применением технологий виртуальной реальности. Погрузившись «внутри» сетки и имея возможность проанализировать дефект с разных ракурсов, в том числе и изнутри, если необходимо, специалист получит более качественное представление о проблеме, в сравнении с использованием 2D представления сетки.

Но в ходе исследований было выявлено, что использование клавиатуры и мыши во время пребывания в виртуальной реальности приводит к постоянному переключению внимания между виртуальным и реальным мирами, что крайне неудобно, и, более того, приводит к нарушению эффекта присутствия, что сильно снижает производительность труда.

Поэтому было принято решение об исследовании и реализации системы навигации (в сеточном пространстве) и взаимодействия с виртуальными объектами с использованием жестового интерфейса, как частного случая NUI.

Требование к такому интерфейсу:

1. Нативное позиционирование и взаимодействие с трёхмерными объектами в трёхмерном пространстве.
2. Отсутствие необходимости выходить из виртуальной реальности для осуществления манипуляции.

2.2. Управление медицинским оборудованием

Другая задача пришла из области медицины, кардиохирургии. Во время выполнения операции специалисту необходимо контролировать работу и следить за показаниями множества систем. При этом, находясь в режиме стерильности, он не может взаимодействовать с устройствами ввода напрямую и ему необходим посредник, ассистент. Это приводит к необходимости личностного коммуникативного общения, что в условиях повышенного стресса во время операции приводит к увеличению нагрузки на специалиста и, как следствие, вероятности ошибки.

В связи с этим начата разработка жестового интерфейса для хирурга, позволяющего бесконтактно управ-

лять приложениями и переключаться между ними, не выходя при этом из режима стерильности.

В дальнейшем задача была расширена до создания системы команд и языка для управления реальным медицинским оборудованием, в т.ч. ангиографом и электродом при малоинвазивной операции на сердце.

Выделенные требования:

1. Бесконтактный интерфейс для сохранения режима стерильности.
2. Трёхмерное позиционирование для управления оборудованием в пространстве.
3. Повышенная надёжность и отказоустойчивость.
4. Гибкость языка жестов для персонализации и индивидуальной настройки для конкретного хирурга.

2.3. Требования к интерфейсу

Таким образом, обобщая вышесказанное, мы приходим к следующим требованиям: необходима разработка интерфейса, использующего жесты для ввода информации, как частный случай NUI. При этом интерфейс должен быть:

1. Надёжным. Должен отвечать повышенным требованиям надёжности и отказоустойчивости, предъявляемого к медицинскому оборудованию.
2. Бесконтактным.
3. Нетребовательным. Интерфейс не должен предъявлять особые требования к пользователю (например, использовать какие-нибудь предметы, манипуляторы или перчатки) или к окружающей обстановке (например, определённый уровень освещения или акустического шума).

Кроме самого самого интерфейса так же необходима разработка специального языка взаимодействия для него. Язык так же должен удовлетворять ряду критериев, основанных на идее деятельностного подхода. В частности, слова языка (в данном случае – жесты) должны быть:

1. Простыми в исполнении.
2. Не противоречить основной деятельности пользователя.
3. Легко повторимыми с одной стороны, и надёжно распознаваемыми с другой.
4. Иметь большие выразительные возможности при небольшом количестве самих жестов.

3. Результаты и обсуждение

Для реализации такого интерфейса были разработаны методы захвата движения на основе карт глубины сцены [1, 2], которые легли в основу для создания программно-аппаратного комплекса «VirtualHand». Этот комплекс представляет реализацию естественно-человеко-компьютерного интерфейса, основанного на жестах. В отличие от близких по подходу разработок MS Kinect SDK [3], OpenNI [4] и Intel RealSense, комплекс «VirtualHand» имеет ряд особенностей:

1. Он не привязан к одному типу оборудования. Благодаря модульной структуре, он может использовать практически любое оборудование, которое

может генерировать карту глубин. Были проведены эксперименты с камерами Kinect, стереокамерами, скоростным сканирующим лидаром и камерой LeapMotion.

- Он не использует предопределенный неизменяемый набор жестов. Вместо этого он включает фреймворк для разработки языка жестов, позволяющий гибко реализовать специализированный интерфейс с учетом особенностей использования.

В работах [5, 6] приводятся примеры использования в реальных задачах. В [7] приводится пример разработки языка жестов для задачи навигации и взаимодействия с большими расчетными сетками в виртуальной реальности.

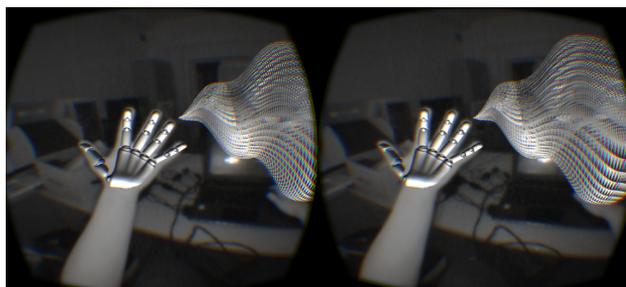


Рис. 1. Пример использования VirtualHand для взаимодействия с виртуальными объектами (рассчитанная сетка) при помощи жестов в расширенной реальности.

Так же в заключении можно добавить, что данное направление исследований является востребованным не только в научной визуализации и медицинских приложениях.

Оно так же востребовано в таких областях как:

1. Мобильная и промышленная робототехника. Использование трехмерных жестов делает управление колесными и летающими дронами более естественным.
2. Обучение. Манипуляция с виртуальными объектами в режиме совместного доступа может быть использована в процессе обучения или для демонстрации.
3. Искусство. Использование пространственных жестов может моделировать наиболее точно моделировать процесс создания скульптур и лепки.

4. Благодарности

Работа опубликована при финансовой поддержке РФФИ, грант 15-07-20347.

5. Литература

- [1] Стародубцев И.С. Инициализируемый захват движения и слежение на основе карт глубин для жестового интерфейса // Научно-технический вестник Поволжья – 2013. – № 3. – С. 264–268.
- [2] Стародубцев И.С. Захват движения на основе иерархического гибкого скелета // Научно-

технический вестник Поволжья – 2014. – № 1. – С. 159–162.

- [3] Zhang Zhengyou Microsoft Kinect Sensor and Its Effect // IEEE MultiMedia – 2012. – 19. – P. 4–12.
- [4] Lu Xia, Chia-Chih Chen, Aggarwal J.K. Human detection using depth information by Kinect // CVPR 2011 WORKSHOPS – 2011. – P. 15–22. DOI: 10.1109/CVPRW.2011.5981811
- [5] Авербух В. Л., Авербух Н. В., Стародубцев И. С., Тоболин, Д. Ю. Использование жестовых интерфейсов при взаимодействии с объектами // Научная Перспектива – 2014. – 56(10). – С. 57–66.
- [6] Starodubtsev I., Averbukh V., Tobolin D., Averbukh N. Professional Natural Interfaces for Medicine Applications // ed. Stephanidis C. / – HCI International 2014 - Posters' Extended Abstracts, Communications in Computer and Information Science, Springer International Publishing – 2014. – vol. 435 – P. 435–439.
- [7] Пестова М.С., Стародубцев И.С. Разработка специализированных жестовых интерфейсов для системы научной визуализации // Труды международной конференции GraphiCon'2016 – 2016. – С. 369–372.



Компьютерное зрение

Computer Vision

GraphiCon 2017

24–28 сентября 2017
Пермь, Россия

Gait Recognition Based on Relative Distance and Motion Features

S. Arseev¹, A. Konushin^{1,2}

¹Lomonosov Moscow State University, Moscow, Russia;

²NRU Higher School of Economics, Moscow, Russia

In this paper we present a human gait recognition system for human identification in video sequences. It uses the structural approach, i.e. pose estimation to extract features from the sequence. Three kinds of features are used: anthropometric features, based on the length of the skeleton segments; relative distance features, based on relative distances between the skeleton joints; and motion features, based on the movement of a joint between two frames. Two versions of the algorithm are presented: the first one uses the depth data alongside with the images while the other one uses only the video sequence. We performed training and evaluation on two datasets, experimental results are presented.

Keywords: biometrics, gait recognition, gait features, anthropometric features, relative distance features, motion features.

1. INTRODUCTION

Gait recognition task is a subset of the biometric human recognition task. Human recognition in video can be applied in many different areas, such as security and automatic dataset markup for further machine learning tasks. However, all existing methods have limitations, preventing them from being used in certain cases. For example, the most popular method of human recognition is the face recognition, but it cannot be performed when the face is covered or if the person is facing away from the camera. Gait recognition, i.e. recognition based on features acquired from analyzing the walking subject's movement, can be performed even on bad quality video, and the only requirement is the visibility of a silhouette.

Gait recognition methods use two main distinct approaches to the task: the silhouette approach and the structural approach. The silhouette is a binary mask which contains pixels that belong to the person. Methods that use the silhouette as a main information source typically have low computational complexity (and, therefore, are fast), but they are very sensitive to the silhouette change, so if the silhouette is changed by a bag or even a long coat, it can seriously affect the recognition result.

The structural approach is based on pose estimation and construction of a pose model, typically a skeleton, i.e. graph with vertices corresponding to the joints of a human body. These methods have higher complexity due to the model construction, but are much less sensitive to the silhouette change due to the fact that if the model is produced correctly, it is not affected by that.

2. PROPOSED METHOD

The proposed method is based on the method proposed by Ke Yang et al [3]. Each sequence is pre-processed to produce a pose estimation model for each frame (a skeleton) which is used by the identification algorithm itself. The skeleton produced by Microsoft Kinect sensor is marked as shown on Figure 1.

Three kinds of features are extracted from each sequence of skeletons: anthropometric features (AF), relative distance features (RDF) and motion features (MF). Then, these features are classified using an ensemble of K-nearest neighbour (KNN) classifiers.

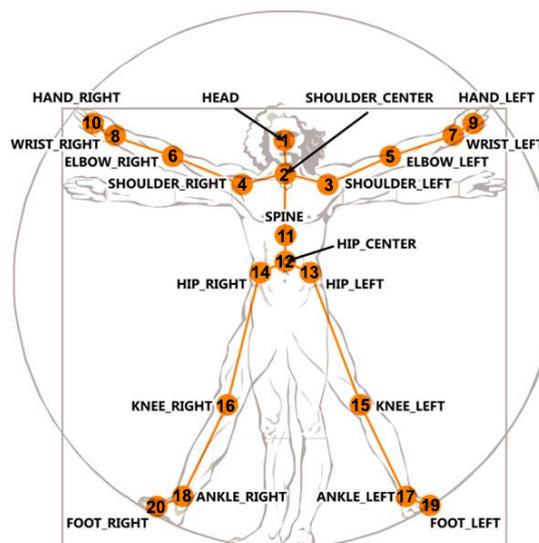


Fig. 1. Kinect skeleton.

2.1 Feature extraction

2.1.1 Anthropometric features

Anthropometric features represent skeleton segments lengths and the total height of the person. $Len(a, b)$ is the euclidean distance between joints a and b .

For the Kinect skeleton.

$$Height = Len(1, 2) + Len(2, 11) + Len(11, 12) + \frac{Len(14, 16) + Len(16, 18) + Len(13, 15) + Len(15, 17)}{2}$$

19 Len values and the $Height$ value produce the anthropometric features (AF) vector. To improve measurement precision, mean and standard deviation values are calculated for each AF component over the video sequence. Then all values which differ from the mean by more than two standard deviations are dropped out, and the mean values of remaining elements produce the final AF vector.

2.1.2 Relative distance features

Each joint a of the skeleton is described by its coordinates: x_a , y_a and z_a for skeletons with the depth map. Relative distance features represent the distances between skeleton joints. They are calculated for each axis separately, and the RDF vector consists of all distances represented as $|x_a - x_b|$, where a and b are the numbers of the paired joints, such as 3 and 4 or 15 and 16, or a is a joint in the pair and b is the corresponding middle joint (2 for a 3-10 and 12 for 13-20). It also includes features represented as $|x_c - \frac{x_a + x_b}{2}|$, where a and b are the numbers of the paired joints and c is either 1 (head joint) or 11 (spine joint). The mean and standard deviation

value of each distance over the sequence produce the final RDF vector.

For three-dimensional Kinect skeletons, the RDF vector consists of 240 elements: there are 16 pairs of joints and each pair produces 5 features for each of the 3 coordinate axis.

2.1.3 Motion features

Motion features are based on the joint shift between frames representing the joint movement speed. The shift is calculated in relative coordinates to ignore the overall movement of the person.

These features are calculated on all axis separately: x , y and z . In the two-dimensional version of the algorithm there are only two axis since there is no depth information.

$$Mx_{a,t} = |x_{a,t} - x_{b,t} - (x_{a,t-1} - x_{b,t-1})|$$

Here, a is a joint on a limb (joints 3-10 and 13-20) and b is the corresponding middle joint (as in 2.1.2).

The mean (by t) and standard deviation values for each joint and each axis produce the MF vector. For Kinect skeletons, it consists of 96 elements: each of the 16 pairs produces 2 features for each of the 3 axis.

The difference between these three kinds of features extracted from a skeleton sequence is shown in Figure 2.



Fig. 2. Three types of features.

2.2 Sequence classification

The resulting feature vector, consisting of AF, RDF and MF vectors, is classified using an ensemble of 100 K-nearest neighbour classifiers with the city block metric. Due to the large size of the feature vector (356 features), each classifier uses a subset of 10 features.

Classifiers were selected to the ensemble via an evaluation procedure. First, the KNN classifier was trained using a random subset of 10 features. Then, for each element of the testing set, five nearest neighbours from the training set were selected and the classifier was assigned a score if any of them had the same class as the test sample: 5 points if the nearest neighbour was of the same class, 4 points for the second nearest neighbour and so on with 1 point for the fifth nearest neighbour being of the same class as the test sample.

Then the classifier was compared to the classifiers that were already in the ensemble and penalized for each feature that was used by another classifier in the ensemble to avoid using similar classifiers and neglecting features. After evaluating a batch of the random classifiers, the ensemble was reformed

using 100 classifiers with best scores, including both the old ones that were already in the ensemble and the new ones. The final ensemble was built after evaluation of 2000 random classifiers.

Figure 3 shows the usage of the features by the ensemble for the three-dimensional version of the algorithm. The X axis represents the index of the feature in the feature vector with first 240 being RDF, next 20 being AF and the rest being MF. The Y axis represents the number of classifiers that used this feature. As seen from the diagram, almost all features are used in the ensemble, and some of them show clear prominence over the others. While the AF vector is the shortest of the three, its features are not dropped out and all contribute to the result.

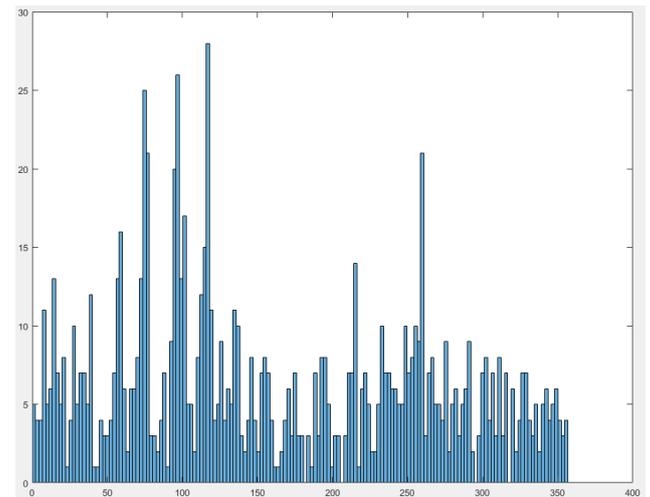


Fig. 3. Feature usage.

The classification results were produced using the weighted vote from the ensemble. Each classifier produced 5 most probable labels, and the class label was assigned a score of 5 for the first place, 4 for the second, and so on until the score of 1 if it was the fifth most probable. Then the total score of each class label was calculated, producing the ranked classification result.

3. EXPERIMENTAL EVALUATION

The algorithm has been evaluated on two datasets: the dataset used by Andersson and Araujo [1] and the TUM GAID dataset [2]. The first one was used to evaluate the three-dimensional version of the dataset and the second one was used to evaluate the two-dimensional version. The results are presented in the table.

	Rank 1	Rank 5
Kinect	93.9	100
GAID-N	95.8	98.7
GAID-B	76.5	93.2
GAID-S	87.8	94.5

Table 1. Experimental results

The results are given as the percentage of correctly classified samples. Rank-5 metric counts the sample as correctly classified if any of the 5 most probable classes was the correct one.

The “Kinect” line represents the dataset used in [1] and [3], and the following lines represent different subsets of the training set in GAID dataset: GAID-N are normal walking sequences,

GAID-B are sequences with backpack and GAID-S are sequences where the person is wearing coating shoes. As seen from the table, the algorithm shows good recognition rate both with the depth map and without it. That means that the recognition quality depends mostly on the skeleton extraction quality while the additional depth information is not strictly necessary for this method.

4. CONCLUSION

A new gait recognition algorithm based on the work [3] was presented in this paper. It uses three kinds of features for classification: anthropometric features, relative distance features and motion features. These features are then classified using an ensemble of KNN classifiers, each of which uses a subspace of these features. The algorithm has been evaluated on two datasets and shows good recognition accuracy both with and without the additional information from the depth map.

5. REFERENCES

- [1] Andersson V. O., Araujo R. M. Person Identification Using Anthropometric and Gait Data from Kinect Sensor // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015. 1. P. 425-431
- [2] Hoffman M., Geiger J., Bachmann S., Schuller B., Rigoll G. The TUM Gait from Audio, Image and Depth (GAID) Database: Multimodal Recognition of Subjects and Traits // Journal of Visual Communication and Image Representation, Special Issue on Visual Understanding and Applications with RGB-D Cameras. 2014. 25. N1. P. 195-206
- [3] Yang K., Dou Y., Lv S., Zhang F., Lv Q. Relative Distance Features for Gait Recognition with Kinect // Journal of Visual Communication and Image Representation. 2016. 39. P. 209-217

Abandoned Objects Detection in Video Sequences

A. Kukleva¹, V. Konushin², A. Konushin^{1 3}

anna.kukleva@graphics.cs.msu.ru|vadim@tevian.ru|ktosh@graphics.cs.msu.ru

¹Lomonosov Moscow State University, Moscow, Russia;

²Video Analysis Technologies, LLC, Moscow, Russia;

³NRU Higher School of Economics, Moscow, Russia

In this paper we handle a problem of abandoned objects detection in video sequences by using a deep convolutional neural network (CNN). At first, a background subtraction based detection [6] is used, which shows high recall and computational speed. Then CNN, which is trained on generated synthetic data, is used to filter out false positives and classify abandoned objects types. The proposed algorithm was tested on a privately collected video collection and showed good performance on long videos.

Keywords: abandoned objects, object detection, neural network classifier.

1. Introduction

Currently, the automatic video sequence analysis is a developing field due to the increase in safety requirements in public places. A city-scale video surveillance system includes the collection, processing and storage of videodata. In these systems hundreds of thousands of cameras are mounted in all public places, including staircase landings, courtyards, places of mass congestion of citizens, educational institutions, private enterprises, subways. The number of cameras is increasing every year, and it becomes impossible for operators to process incoming information in real time. Automatic video analysis tools can help personnel to focus only on dangerous or strange occurrences.

One of the main goals of automatic video surveillance is the detection of bags and luggage that were left without attendance in public places, such as airports. Such object are called "abandoned". It can also be noted that algorithms that solve this problem can be applied in other fields of video analysis as well: for example Fig. 1, monitoring the safety of valuable objects and when a parking lot is full.



Fig. 1. Examples of abandoned objects
Left – parked car, right – abandoned bag.

Currently existing algorithms, which are based on object tracking or background modeling, cannot ensure a sufficiently high accuracy under real conditions to provide a good reliability level. In recent years in many computer vision problems the rapid progress is demonstrated by methods, which are based on convolutional neural networks. But training a convolutional neural network requires a large amount of training data. It is important problem, since for some tasks it is very difficult to obtain a sufficient amount of training data. The detection of abandoned objects is an example of such task. If we provide enough training data

for implementation of neural network, we can expect a significant increase in abandoned objects detection accuracy.

2. Related Works

Most of successful abandoned object detection algorithms developed in the previous years rely on background subtraction. Two different approaches can be distinguished.

The first family of works relies on using combination of existing background subtraction algorithms and other independent methods such as object tracking, cumulative mask formation for foreground objects, people detection [9–14].

The second family of works modifies background subtraction algorithms for solving this problem. For instance modifications can relate to alteration of one or several distribution of Gaussian mixture model or integration of a finite state machine into background model [1–5].

There are plenty of algorithms for classification of the founded objects: decision trees, linear classifier, SVM, a bag of words, etc. In recent years deep convolution neural networks demonstrate the best results in the classification problems.

The similar approach [12] to one described in this paper was published concurrently with this article in summer 2017. There are the following differences. First in [12] Gaussian Mixture Model is used for background modeling and static object detection while we employ the algorithm based on accumulation and analysis pixel-by-pixel masks. Second, we use different architecture of convolutional neural network for the classifier of abandoned objects and false detections.

3. Proposed algorithm

Our proposed algorithm consists of two main steps:

1. Detection of a potential abandoned object (hypotheses generation) with background subtraction algorithm.
2. Classification of the found hypotheses with neural network.

4. Hypotheses search algorithm

After analyzing the existing approaches for detecting static objects and those removed from a scene, we selected as a baseline the algorithm [6], which demonstrated a high

recall. The algorithm takes sparse video frames as input, about one frame per second, which allows it to work in real-time.

The output of the baseline algorithm is a set of static foreground regions (non-moving for several previous frames and different from background).

Several masks are generated to handle each frame:

- Frame difference accumulation mask – pixel-by-pixel difference between two neighbor frames.
- Background accumulation mask – pixel-by-pixel comparison of the gradient of a current frame and a background model.
- The intersection of the previous two masks is a stationary pixel mask, and a pixel is implanted into the background model upon reaching a certain threshold.

To obtain hypotheses on the stationary pixel mask, connected components are marked by bounding boxes. Fig. 2 shows baseline algorithm as a block diagram.

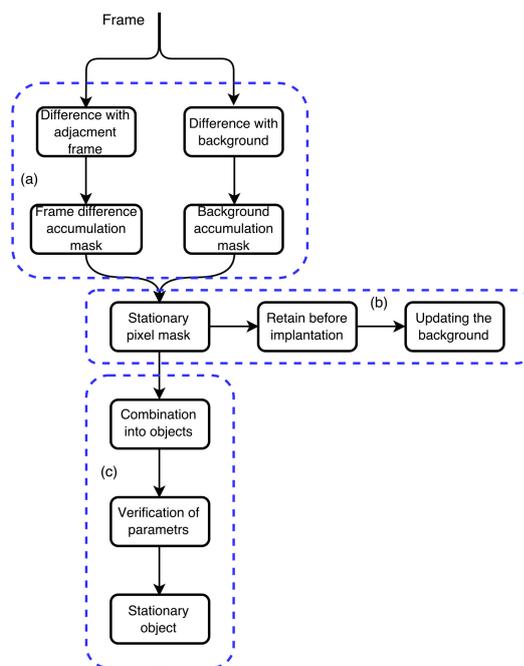


Fig. 2. Block diagram of the baseline algorithm.

An image gradient is used to reduce the impact of illumination intensity instead of the original image. The pixel difference is estimated considering the noise parameters of a particular video. These parameters are determined before the main algorithm starts working.

We propose to automate noise level estimation for each video.

We need to acquire the noise level in a video sequence from input images. We can assume that the noise is white – a stationary noise, the spectral components of which are

evenly distributed over the entire range of frequencies involved, and we can minimize the problem up to the determination of the root-mean-square noise deviation.

We have selected the noise level estimation method on the basis of image blocks from [8]. It is easy to implement, requires a small amount of memory, and showed fairly high performance in the tests.

A frame sequence is selected from each video and then equally spaced geometrically rectangles with non-moving scenes were cut out. Since moving objects distort the evaluation, they should be absent in these areas. This is achieved by applying a background subtraction algorithm and generating an accumulating background mask.

$$AccMask_t(x) = \begin{cases} 0, & \text{foreground pixel;} \\ AccMask_{t-1}(x) + 1, & \text{otherwise.} \end{cases}$$

The Gaussian mixture model was chosen as background subtraction algorithm. To avoid random noise, a morphological erosion operator is applied to each background mask.

Every 30th iteration, a sliding window algorithm searches for the largest area of an accumulating background mask, which remains static at least 50 consecutive frames.

An image gradient is used to extract noise parameters. The mean value is calculated for the entire selected sequence of areas with 3x3 filter and the mean value of the root-mean-square noise deviation for N frames is calculated using the following formula:

$$f(x) = \sqrt{\frac{1}{N} \sum_{k=1}^N \sigma_k^2},$$

where σ – Root-mean-square deviation of frame number k from mean.

5. Classification

5.1. Data preparation

To validate our approach in challenging realistic environments, we collected data from several cameras monitoring busy university hall, parking and subway with a lot of loitering people or cars within about 40 hours of video. All objects were manually classified in this data set. It contains 234 ground truth objects which is not enough for training deep neural network. As a result, we created synthetic data (Fig. 3). We have made the collection of background images, which reflects all kinds of conditions: cabinet, park, street, shop and others, and the collection of bags composed of bags with different colors and shapes. The average bag size corresponding to the real background conditions was determined manually for each background image.

For easy embedding of bags into background, it is necessary that its background is white, without noises and shadows. The image with a bag is binarized, the contour of an object is located, the alpha channel of a background around the object is equated to 0, the object is combined

with a randomly selected part of a background image. The size of the generated image depends on the manual mapping of the mean suitable size for the background and on the stretch factors of the chosen bag along the vertical and horizontal directions, the length of each side can range from 20 to 500 pixels.

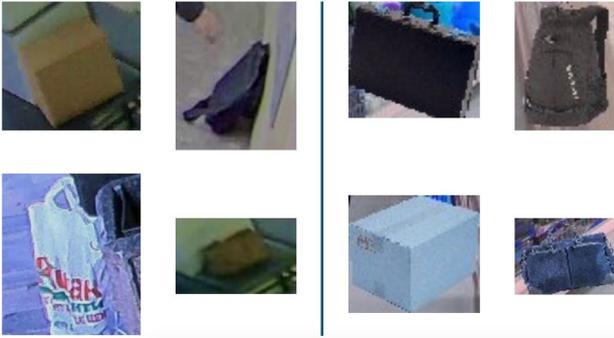


Fig. 3. Left – real data
Right – synthetic data.

Thereby more than 15 thousand synthetic bags were produced.

5.2. Convolutional neural network

For hypotheses classification we used a CaffeNet, which is a replication of AlexNet [7], in which the order of normalization layers and pooling is switched. It is pre-trained on ImageNet, a large image database, convolution layers capture well generalized features at lower levels and specific features for a specific task at higher levels of the architecture.

Two approaches are employed for training common convolutional neural network in this paper: fine-tuning the CaffeNet network pre-trained on Imagenet and training from scratch the same architecture. Mini-batch consists of 20 images, initial learning rate is 0.01 multiplying by 0.1 every 2500 iterations. Activation function is ReLU.

6. Experimental Results

Firstly, we compared the base method [6] with its modification consisting of the base algorithm and the block noise level estimation [8]. Results are in the table 1. P stands for precision and R stands for recall.

PETS2006	P	R
Base algorithm	1.00	1.00
Proposed algorithm	1.00	1.00

Table 1. Dataset PETS2006. Results.

As you can see from the table 1 both of the algorithms ideally detect all abandoned objects from PETS2006 [2].

For algorithm evaluation such number of videos and overall duration isn't enough. A proper dataset was collected and composed of videos with total duration of 40 hours. There are various background scenes such as university hall, classrooms, underground, parking places and

etc. Further experimental estimation was carried out on this dataset.

Assembled dataset	P	R
Base algorithm	0.13	0.89
Proposed algorithm	0.09	0.96

Table 2. Assembled dataset. Results.

As seen in the table above (Table 2), precision was significantly decreased but this algorithm should only show high recall as it is followed by CNN classification.

Classification accuracy was measured on different subsets of the data. 2 classes¹ – bags and false alarms, 3 classes² – bags, cars and false alarms. *CaffeNet*₀ – trained from scratch, *CaffeNet*_{ft} – fine-tuned.

Network	2 cl. ¹		3 cl. ²	
	P	R	P	R
<i>CaffeNet</i> ₀	0.91	0.95	0.72	0.86
<i>CaffeNet</i> _{ft}	0.93	0.95	0.77	0.87

Table 3. Classification result.

Fine-tuning shows better results (Table 3) than training from scratch. Apparently there is no such diversity of synthetic data as to ImageNet dataset.

Proposed method is a superposition of hypotheses search and images classification algorithms. The total recall of abandoned objects detection for long videos can be found in the table 4.

Network	2 cl. ¹		3 cl. ²	
	P	R	P	R
<i>CaffeNet</i> ₀	0.91	0.91	0.72	0.81
<i>CaffeNet</i> _{ft}	0.93	0.91	0.77	0.83

Table 4. Result of the whole algorithms.

7. Conclusion

In this work we proposed a method for abandoned object detection, which combines the hypotheses search algorithm and the neural network classifier. We employed algorithm [6] for detection of potential abandoned objects and implemented automated noise estimation in the baseline algorithm. We used convolutional neural network as a classifier. For training classifier we generated synthetic data. This approach demonstrates good results on long real videos where false alarms occur much more often than true positive abandoned objects. Fine-tuning has a slight advantage over training from scratch.

As future work, we will explore the use of different approach for background subtraction [10]. It can help to generate more qualitative hypotheses and decrease amount of ghosts. This algorithm can be integrated with people tracking algorithm [11] to define object owner.

8. References

- [1] Bayona A., SanMiguel J.C., Martinez J.M. Stationary foreground detection using background subtraction and temporal difference in video surveillance. IEEE International Conference on Image Processing, 2010.
- [2] Dataset Pets2006. <http://www.cvg.reading.ac.uk/PETS2006/data.html>
- [3] Evangelio R.H., Sikora T. Complementary background models for the detection of static and moving objects in crowded environments. 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2011. – pp 71–76.
- [4] Fan Q., Gabbur P., Pankanti S. Relative Attributes For Large-scale Abandoned Object Detection. ICCV2013, 2013, – pp 2736–2743.
- [5] Ferrando S., Gera G., Regazzoni C. Classification of unattended and stolen objects in video-surveillance system. 2006 IEEE International Conference on Video and Signal Based Surveillance, 2006, – p 21.
- [6] Kireev D., Konushin V. Stationary object detection. Conference "Lomonosov – 2016", 2016. (in Russian).
- [7] Krizhevsky A., Sutskever I., Hinton G. ImageNet Classification with Deep Convolutional Networks. In NIPS, 2012.
- [8] Lapshenkov E. Non-standard noise level estimation of digital image based on harmonic analysis. Computer optic, 2012, vol 36, №3. (in Russian).
- [9] Miguel J.C.S., Martinez J.M. Robust unattended and stolen object detection by fusing simple algorithms. IEEE International Conference on Advanced Video and Signal Based Surveillance, 2008.
- [10] Morozov F., Konushin A.S. Background subtraction using a convolutional neural network. Proceedings of the 26th International Conference on Computer Graphics and Vision GraphiCon'2016, pp. 445–447.
- [11] Shalnov E., Konushin V., Konushin A. An improvement on an MCMC-based video tracking algorithm. Pattern Recognition and Image Analysis, Allen Press Inc. (United States), vol 25, pp. 532–540.
- [12] Sidiyakin S.V., Vishnyakov B.V. Real-time detection of abandoned bags using CNN. Proc. of SPIE Vol. 10334 103340J-2.
- [13] Tian Y.L., Feris R., Liu H. Robust detection of abandoned and removed objects in complex surveillance videos. IEEE Transactions on Systems, Man, and Cybernetics, 2010. – 41(5) – pp 565–576.
- [14] Wen J., Gong H., Zhang X., Hu W. Generative model for abandoned object detection. IEEE International Conference on Image Processing, 2009.

Background subtraction with convolutional neural network and oversegmentation

F. Morozov¹, A. Konushin^{1,2}

fedor.morozov@graphics.cs.msu.ru|anton.konushin@graphics.cs.msu.ru

¹ Lomonosov Moscow State University, Moscow, Russia;

² NRU Higher School of Economics, Moscow, Russia

Background subtraction in video is an important computer vision problem as it serves as a first step in many video analysis algorithms, yet it cannot be considered to be fully solved due to complexity and variety of input data from surveillance cameras. In this work we propose a background subtraction method based on a random sample background model [14] using a convolutional neural network to compare image patches with the background model. To improve the quality of stopped objects detection we utilize oversegmentation of input frames, obtained superpixels are also used at postprocessing step. Experimental evaluation on ChangeDetection.net dataset shows superiority of the proposed algorithm compared to analogues.

Keywords: background subtraction, convolutional neural networks, oversegmentation.

1. Introduction

This work considers the problem of background subtraction in video, it is formulated as follows: each pixel of video sequence frame has to be classified as belonging either to background or to a moving object. In video analysis moving objects represent the greatest interest so they have to be separated from the mostly static background.

Background subtraction is often used as one of the first steps in video analysis algorithms based on object tracking and recognition. It can be used for detection of objects left behind and objects removed from the scene, people counting and crowd detection, tracking objects in security systems, control of parking spaces or traffic analysis such as congestion detection. In figure 1 an example of object left behind detection is presented with corresponding background subtraction mask.



Fig. 1. Object left behind detection with background subtraction mask.

Many of the mentioned problems use video from surveillance cameras as an input. The use of video analysis algorithms allows to automatically detect potentially dangerous situations and respond to them in time. It also substantially reduces the amount of works that has to be done by surveillance systems operators and improve the recall of detected incidents. That is especially important considering that the amount of video data to be processed is rapidly increasing.

Background subtraction algorithms have to be robust to sudden or gradual lighting changes, detect camouflaging objects and objects moving in shadows, while glares, shadows and reflections shouldn't cause false positive detections. Scenes that feature periodic background motion like waving leaves or water ripple or intermittent object motion are of particular complexity as are scenes with small camera displacements caused by wind, vibration or other reasons. Videos captured in difficult weather conditions like rain and snow or videos obtained at night with artificial illumination are also to be considered.

Result of state of the art algorithms are still inferior to manual labeling by experts due to the challenges described above so the problem of background subtraction cannot be considered to be fully solved.

Methods based on artificial neural networks are broadly used for various computer vision problems and show state of the art results in object detection, image classification, semantic segmentation, face recognition etc. Use of deep learning for background subtraction has not been widely researched yet while it may significantly improve existing algorithms or provide some novel approaches.

2. Related work

Most background subtraction algorithms rely on building a background model at each pixel. Then every pixel of a frame being processed is compared to the model and classified as background if it fits the model and as a moving object otherwise. Various algorithms differ in model construction, matching and updating approaches. They can be separated into the following groups based on background modeling:

1. Methods based on mixture of Gaussians [9]
2. Methods based on a set of samples [14]
3. Methods based on codewords [7].

Gaussian mixture models estimate the mean and variance of a number of Gaussians which also have weights indicating their persistence. If observed value matches a Gaussian, it's weight increases and parameters are updated using running average. If the sum of matched Gaussians' weights is above a given threshold, the pixel is classified as background.

Algorithm [15] uses Gaussian mixture modeling augmented by flux tensor motion detection which significantly improves the quality of stopped objects detection. Algorithm [2] performs matching in rectangular area around the pixel and random update strategy. In [5] it is proposed to use local binary patterns to leverage neighborhood information, algorithm [3] uses minimal spanning trees to average results of similar pixels.

Algorithms based on a set of values store a number of previously observed values for each point. One of the most popular models proposed in [14] uses random update policy and consensus based classification. If the observed value is similar to a given number of model values it is classified as background and may replace one of the stored elements. This way the probability that a sample will be presented in the model decreases monotonically with time.

Algorithm [1] proposes a number of heuristics improving the base model: connected components analysis, model update limitation, blinking pixel detection and adaptive thresholds. In [6] an adaptive model is proposed that tunes parameters of the algorithm at the runtime.

Algorithms [10] and [11] use a local feature descriptor – local binary similarity patterns. In [10] it is demonstrated that the use of area descriptor can drastically improve the performance of sample-based algorithm while [11] combines this idea with adaptive parameter tuning.

Models based on code words consist of complex elements – background words with some aggregate statistics like frequency, mean values, maximum negative run length etc. These statistics are used to estimate persistence of each background word similar to Gaussian mixture methods. Algorithm [16] uses a small number of code words obtained by clustering observed values using running average and efficacy counters. Algorithm [12] uses mixed code words with both color and texture information in the form of LBSP descriptors.

Several methods based on convolutional neural networks were proposed for segmentation and matching problems. In [17] different neural network architectures are compared for the task of local feature matching. In [18] a Siamese neural network is used to compare image patches for stereo matching. In [8] a similar approach was proposed for background subtraction with a trivial model, in this article we expand this idea with a more complex algorithm.

3. Proposed method

Proposed algorithm is based on background model proposed in [14]. This model is robust in presence of many typical challenges and has relatively small number of parameters that require tuning. It has been thoroughly studied in literature with many modifications and implementation details available. The background model does not use any form of clustering and allows almost voluntary feature representations.

The background model consists of a fixed number of samples at each pixel location. In [14] it is initialized using only the first frames which leads to errors if there are moving objects present at the moment of initialization. We propose to initialize the model using a number of first frames of video sequence and choose only persistent background values.

On each iteration a frame being processed is compared to the elements of the background model at corresponding point. Is its similar to a given number of elements according to the chosen metric, a background label is selected. Otherwise the pixel is classified as part of a moving object.

Since values of individual pixels are sensitive to noise and illumination changes we leverage information stored in a neighborhood around the pixels. To compare these image fragments a Siamese neural network [8] described in section 4 is used, that calculates descriptors for each patch and compares them using dot product.

Pixels classified as background are used to update the background model. Since the original values of patch pixels are not needed for matching only calculated descriptors can be stored for each pixel. Following [14] we use a random update policy: with a given probability a randomly selected background model element is discarded and replaced with the observed descriptor value.

For better tolerance to dynamic background and illumination changes a background pixel may replace a random element in one of neighboring background models with the same probability. This propagation is limited using superpixel boundaries as described in section 6. For postprocessing we use a median filter with kernel size equal to 5, the results are further corrected by filling homogenous superpixels.

4. Neural network descriptor

To compare image patches we use a Siamese neural network proposed in our previous work [8]. Its architecture is illustrated in figure 2: the network consists of two branches with shared weights, each branch has three convolutional layers with rectified linear activations and calculates a descriptor of length 16 for the input patch. To compare the patches a dot product is used, its result indicates a similarity score for the two patches.

The main advantage of the proposed architecture is that it allows to perform most of the computations (network branches) once for each fragment and each time we match two fragments a relatively simple dot product operation is only required. This is especially important for background subtraction, because multiple matching with background model elements is required for each pixel and descriptor values can be stored instead of the original patches.

To train the network mini-batch gradient descent is used with batch size of 128 and 0.001 learning rate. Pairs of negative and positive examples are used to train the network optimizing margin loss with margin value of 0.2.

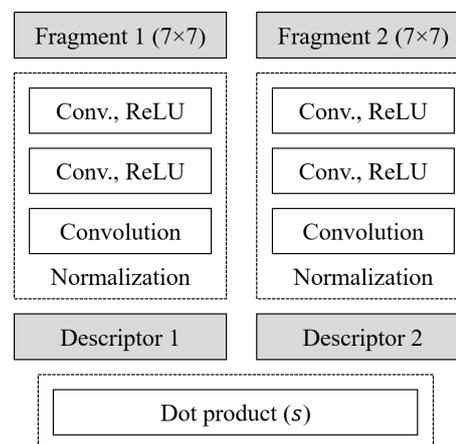


Fig. 2. Proposed neural network architecture.

5. Training the network

To train the network one needs pairs of positive and negative examples. We use a dataset of video sequences with ground truth labeling to generate pairs of patches. Pairs of background patches in the same point at different moments in time form the positive examples – assumingly similar patches, while pairs of background and a moving object patches in the same point form the negative ones, that the algorithm has to differentiate.

The number of background pixels is much greater than the number of moving objects' pixels so we construct as many negative examples as possible and generate the same amount of positive ones. So for every patch centered around a moving object pixel we need to pick two background patches around the same point. These fragments also shouldn't be too far in time since the background may change significantly over time, and may be centered around neighboring pixels to improve robustness in cases of dynamic background and camera displacement.

To generate the patches, it is proposed to use a sample-based background model. The use of background models allows to select background patches on the fly while random update makes the data more diverse and removes older samples, neighbor updates improve displacement tolerance.

6. Oversegmentation

Updating models in neighboring pixels allows the base background model to better cope with noise and illumination changes. On the other hand, this approach may lead to degradation in the presence of intermittent object motion: stopped objects are rapidly incorporated into the background models, detection errors appear on object boundaries.

One of the ways to deal with this problem is to limit neighbor updates on object boundaries as proposed in [1]. We propose to use oversegmentation to achieve this and use superpixel boundaries to limit the model elements propagation. For oversegmentation an algorithm [13] is used as it allows

real-time video processing with state of the art segmentation results.

The resulting superpixels are also used for postprocessing after median filtering: if a large enough share (0.96) of pixels grouped in a superpixel have similar classification results, the whole superpixel is classified with the prevailing value. An example of superpixel segmentation is presented in figure 3.

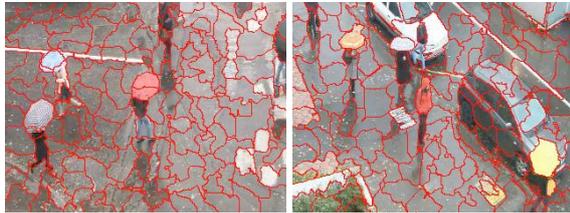


Fig. 3. Examples of oversegmentation.

7. Evaluation

To evaluate the proposed algorithm, we used ChangeDetection.net dataset [4]. This dataset consists of 31 video sequences with manual pixel perfect ground truth labeling split into five categories corresponding to typical background subtraction challenges (shadow, dynamic background, intermittent object motion, camera jitter, thermal) and a baseline category.

In table 1 are presented F-measure values calculated for detection results by the proposed method, base algorithm [14] and two its modifications: [10] that uses local binary similarity patterns and [6] that proposes adaptive parameter tuning. Proposed method outperforms analogues in all the dataset categories and in average.

We used OpenCV library for algorithm implementation and caffe framework for neural network training. Performance of the prototype version is about 6 frames per second for 320×240 resolution.

	ViBE [14]	LOBSTER [10]	PBAS [6]	Proposed
Baseline	0,87	0,92	0,92	0,94
Shadow	0,80	0,87	0,68	0,90
Dynamic background	0,57	0,57	0,72	0,78
Intermittent object motion	0,51	0,58	0,57	0,63
Camera jitter	0,60	0,74	0,81	0,82
Thermal	0,66	0,82	0,76	0,84
Average	0,67	0,75	0,75	0,82

Table 1. F-measure values on ChangeDetection.net dataset.

8. Conclusion

In this work a new background subtraction method is proposed based on sample background model with random update [14]. Neural network descriptor is used to compare patches on the frame being processed with the background model representations. Oversegmentation is used at model update and postprocessing steps to improve quality at object boundaries and fill in erroneous regions.

The proposed modifications of the base algorithm provide superiority compared to the base algorithm and its modifications in terms of f-measure as demonstrated on ChangeDetection.net. Examples of background subtraction with the proposed algorithm are presented in figure 4.

In the future we are planning to develop a more optimized implementation of the algorithm and further improve its precision and recall with data augmentation and adaptive parameter tuning.



Fig. 4. Examples of background subtraction using proposed algorithm.

9. Acknowledgements

This work was supported by RFBR grant no. 14-01-00849 and by the Skolkovo Institute of Science and Technology, the contract 081-R, Annex A2.

10. References

- [1] Barnich O., Van Droogenbroeck M. Background subtraction: Experiments and improvements for ViBe.
- [2] Chen Y., Wang J., Lu H. Learning sharable models for robust background subtraction.
- [3] Chen M., Yang Q., Li Q., Wang G., Yang, M. Spatiotemporal background subtraction using minimum spanning tree and optical flow.
- [4] Goyette N. Changedetection. net: A new change detection benchmark dataset.
- [5] Heikkilä M., Pietikäinen M., Heikkilä J. A texture-based method for detecting moving objects.
- [6] Hofmann M., Tiefenbacher P., Rigoll G. Background segmentation with feedback: The pixel-based adaptive segmenter.
- [7] Kim K. Real-time foreground-background segmentation using codebook model.
- [8] Morozov F., Konushin A. Background subtraction using a convolutional neural network
- [9] Stauffer C., Grimson W. Adaptive background mixture models for real-time tracking.
- [10] St-Charles P., Bilodeau G. Improving background subtraction using local binary similarity patterns.
- [11] St-Charles P. L., Bilodeau G. A., Bergevin R. Flexible background subtraction with self-balanced local sensitivity.
- [12] St-Charles P. L., Bilodeau G. A., Bergevin R. A self-adjusting approach to change detection based on background word consensus.
- [13] Van den Bergh M. et al. SEEDS: Superpixels extracted via energy-driven sampling.
- [14] Van Droogenbroeck M., Paquot O. ViBe: a powerful random technique to estimate the background in video sequences.
- [15] Wang R. Static and moving object detection using flux tensor with split gaussian models.
- [16] Wang B., Dudek P. A fast self-tuning background subtraction algorithm.
- [17] Zagoruyko S., Komodakis N. Learning to compare image patches via convolutional neural networks.
- [18] Zbontar J., LeCun Y. Computing the stereo matching cost with a convolutional neural network.

About the authors

Fedor Morozov is a Master student at Lomonosov Moscow State University, Department of Computational Mathematics and Cybernetics. His contact email is fedor.morozov@graphics.cs.msu.ru.

Anton Konushin is an associate professor at Lomonosov Moscow State University and at National Research University Higher School of Economics. His contact email is anton.konushin@graphics.cs.msu.ru.

Палеографическое датирование берестяных грамот

К. А. Сидоров
sidorovk@cardiff.ac.uk
Cardiff University, Cardiff, UK

В статье предлагается новый метод для автоматического датирования берестяных грамот по виду графем (палеографически). Предлагаемый метод показывает среднюю абсолютную ошибку всего в 18,9 лет, то есть не хуже точности датирования экспертами.

Ключевые слова: анализ почерка, анализ документов, палеография, вычислительная археология, морфометрика, распознавание образов, берестяные грамоты.

Paleographic Dating of Birch Bark Manuscripts

K. A. Sidorov
Cardiff University, Cardiff, UK

We propose a novel method for automatically estimating the age of birch bark manuscripts based solely on the appearance of graphemes (paleographic dating). Our method achieves mean absolute accuracy of 18.9 years which is comparable to or surpasses the performance of human experts and of other computational paleography studies.

Keywords: handwriting analysis, document analysis, paleography, computational archaeology, morphometrics, pattern recognition, birch bark manuscripts.

10 We address the problem of estimating the age of birch bark manuscripts [22] from their appearance (paleographically). These birch bark manuscripts (BBMs) are one of the most valuable corpora of Old Russian texts, and a crucially important source of information about mediæval history and evolution of the language. Accurate dating of BBMs is necessary in order to place them in a correct historical context, before historians and linguists can take advantage of their valuable contents.

1. Background

Paleographic dating is possible due to the fact that the appearance of graphemes does not remain constant over time. For BBMs, a considerable effort has been undertaken by Zaliznyak *et al.* [22] (pp. 134–429) to codify paleographic expertise as *objectively* as possible. They have constructed paleographic tables [22] for each grapheme, relating the occurrences of graphemes' appearance features to dates at which these are known to occur. Recently, using computer vision and machine learning techniques, He *et al.* [7–11] have addressed the problem of direct paleographic dating of mediæval charters, with accuracy in the order of decades. In [7], texture-level features, previously used for writer identification (Hinge, Fraglets), are used to form descriptors for the entire documents, and dating is done with two-level (global and local) support vector regression. The approach in [11] is based on a histogram of orientations of strokes as a descriptor, and a procedure based on 3D self-organising map to discover correlations between features and dates. In [21] a bag-of-words approach on shape context vectors extracted from edge maps in manuscripts is used for dating Swedish mediæval charters.

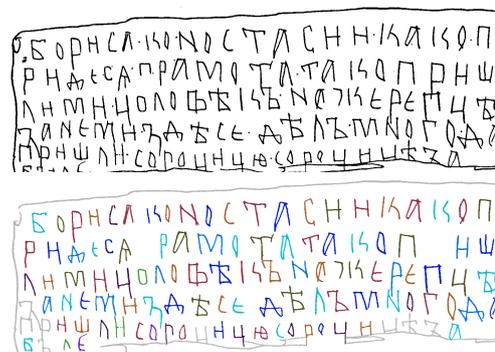


Fig. 1. Top: outline drawing of BBM №43. Bottom: the result of segmentation into graphemes.

2. The Proposed Approach

We have extracted (see Fig. 1) and annotated 48,870 graphemes from the outline drawings in the BBM corpus (data were predominately taken from the digital archive [1] and from digitised reproductions in the relevant volumes of series [22]). A total of 814 manuscripts, from which at least one grapheme could be reliably extracted, were processed.

Following [22], we reduce the problem of dating a manuscript to the problem of dating individual graphemes in it, and then aggregating the results. We attack the problem on two fronts: first, we use a model of grapheme deformation, based on groupwise nonrigid registration of grapheme images; second, we employ a model based on convolutional neural networks to capture nuanced details in appearance variation.

Deformable Model of Graphemes. As the first angle of attack, we parameterise the large-scale variation in grapheme shapes with a low-dimensional representation. We employ the classic technique of decomposing the appearance into deformable “shape” and the corresponding “texture” and modelling them

jointly. This technique has been variously known as Morphable Model or Appearance Model (AM) [5]. This approach allows us to describe as much as possible of the variation in graphemes' appearance as a smooth deformation of their shape, yet describe the residual variation (subtle nuances that cannot be explained by smooth elastic deformation, *e.g.* changes in topology) as variation in "texture". The main challenge involved in building such appearance models is finding correspondences between analogous parts of all deformable exemplars: the problem of groupwise non-rigid registration, for which several approaches have been developed [4, 15]. The standard idea is to iteratively align all exemplars to a common reference model, which itself is iteratively evolved by averaging aligned exemplars. However, we found that intensity-based image registration techniques [3, 4, 15] do not work well with binary grapheme images, as the objective function is not smooth and/or convex enough.

Algorithm 1 Register grapheme images to build AM

Require: Images I^i ($i \in \{1 \dots n\}$); tolerance ε_{tol} ; max epochs t_{max} ; buffer dimensions w, h ; point cloud size p .

- 1 **function** RegisterGraphemes($I, \varepsilon_{\text{tol}}, t_{\text{max}}, w, h, p$)
- 2 **initialisation:**
- 3 $a_{\text{orig}} \leftarrow 0$ ▷ total area
- 4 **for** $i \leftarrow 1$ to n
- 5 triangulate images:
- 6 $\{V_{2 \times v_i}^i, T_{3 \times t_i}^i\} \leftarrow \text{Triangulate}(I_i)$
- 7 $D^i \leftarrow \mathbf{0}_{2 \times v_i}$ ▷ initialise deformations
- 8 $a_{\text{orig}} \leftarrow a_{\text{total}} + \text{Area}(V^i, T^i)$
- 9 **iterative alignment:**
- 10 **for** $t \leftarrow 1$ to t_{max}
- 11 $R \leftarrow \frac{1}{n} \sum_{i=1}^n \text{Rasterise}(V^i + D^i, T^i)$ ▷ reference
- 12 **if** $\text{Error}(R, V, T, D) < \varepsilon_{\text{tol}}$ **then**
- 13 **break** ▷ tolerance reached
- 14 $a \leftarrow 0$
- 15 **for** $i \leftarrow 1$ to n ▷ align each grapheme to R
- 16 remove i -th image from reference:
- 17 $\hat{R} \leftarrow (nR - \text{Rasterise}(V^i + D^i, T^i)) / (n - 1)$
- 18 $P_{2 \times p} \leftarrow \text{Sample}(\hat{R} / \sum \hat{R}, p)$ ▷ sample points
- 19 $D^i \leftarrow \text{CPD}(P, V^i + D^i)$ ▷ align with [16]
- 20 $a \leftarrow a + \text{Area}(V^i + D^i, T^i)$
- 21 ensure area is preserved:
- 22 $c \leftarrow (\sum_{i=1}^n \sum_{j=1}^{v_i} (V^i(:, j) + D^i(:, j))) / \sum_{i=1}^n v_i$
- 23 **for** $i \leftarrow 1$ to n
- 24 $D^i \leftarrow c - V^i + (V^i + D^i - c) \cdot \sqrt{a_{\text{orig}}/a}$
- 25 **compute shape and texture for appearance model:**
- 26 $X_{2 \times s} \leftarrow \forall(x, y) \mid R(y, x) > 0$
- 27 **for** $i \leftarrow 1$ to n ▷ align R to each grapheme
- 28 $P_{2 \times p} \leftarrow \text{Sample}(R / \sum R, p)$ ▷ sample points
- 29 $D_{\text{ref}}^i \leftarrow \text{CPD}(V^i, P)$ ▷ align with [16]
- 30 $S_{2 \times s}^i \leftarrow \text{Extrapolate}(D_{\text{ref}}^i \text{ on } X)$ ▷ shape
- 31 $T_{1 \times s}^i \leftarrow I^i(S^i)$ ▷ texture
- 32 **return** S, T
- 33 **function** Error(R, V, T, D)
- 34 $\varepsilon \leftarrow 0$
- 35 **for** $i \leftarrow 1$ to n
- 36 $\hat{I} \leftarrow \text{Rasterise}(V^i + D^i, T^i, w, h)$
- 37 $\varepsilon \leftarrow \varepsilon + \sum_{x=1}^w \sum_{y=1}^h |R(y, x) - \hat{I}(y, x)|$
- 38 **return** ε/n

We, therefore, employ a hybrid approach: we evolve the reference model as an image, yet use point cloud registration to align graphemes to it. The procedure is summarised in Algorithm 1. The input grapheme im-

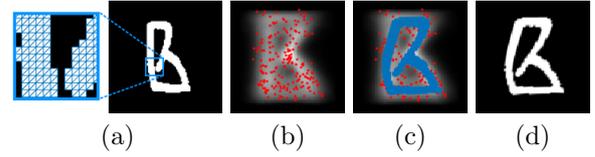


Fig. 2. Registration of graphemes to the reference.

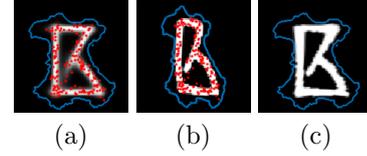


Fig. 3. Registration of the reference to graphemes and sampling of the texture.

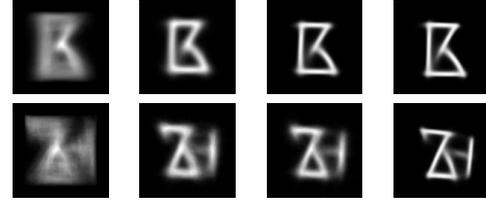


Fig. 4. The average shape and texture of the ensemble as the registration progresses.

ages are first densely triangulated into meshes (line 6, see Fig. 2 (a)) and the total area of all meshes in the ensemble is remembered (lines 3, 8). The algorithm proceeds in epochs (lines 10–24). In each epoch, the reference R is first computed (line 11) by averaging the rasterised warped meshes. The algorithm proceeds to align each grapheme to the current reference (lines 15–24). (To avoid problems with local minima, the grapheme being aligned is excluded (line 17) from the reference.) To do this alignment, the reference R is treated as a probability mass function of some distribution, and p points $P_{2 \times p}$ are randomly drawn from it (line 18) (importantly, different points are drawn at each iteration). Figure 2 (b) shows the reference R at some iteration, with sampled points P overlaid (in red). The vertices V^i of the grapheme's mesh are aligned to these points P (line 19), yielding vertex displacements D^i . To align the two point clouds V^i and P we use the Coherent Point Drift (CPD) algorithm described in [16]. We use the following parameters for CPD: Gaussian width $\beta = 1.4$, regularisation weight $\lambda = 8$, outliers ratio 0.7, and we run it for maximum of 100 iterations (for more detail on these parameters see [16]). Figure 2 shows a grapheme $\{V^i, T^i\}$ (a) being aligned to reference points P (b) yielding a warped mesh $\{V^i + D^i, T^i\}$ (c) rasterised as (d). To avoid a run-away effect to the trivial solution, whereby all meshes may become compressed into a point, after each epoch we re-scale the meshes to preserve the total original area (lines 21–24). Figure 4 shows the reference R as the registration proceeds. Note how as the alignment becomes better, the average grapheme appearance R becomes progressively crisper, indicating good alignment. The registration terminates either when the maximum number of epochs t_{max} has

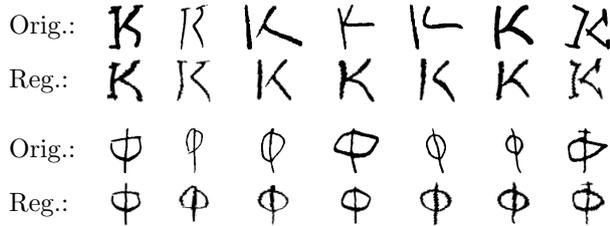


Fig. 5. Examples of graphemes before and after registration (shape normalised).

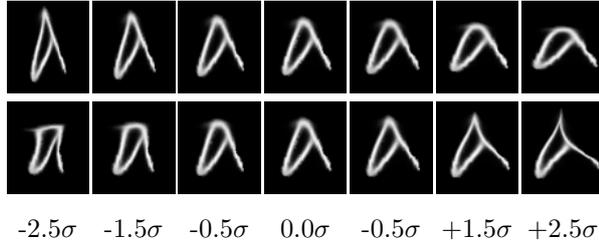


Fig. 6. Most significant of appearance variation for some of the graphemes.

exceeded or when the registration error is below acceptable ε_{thr} (lines 12–13, 33–38).

Finally, we compute the shape and texture describing each grapheme, by finding an optimal alignment of the reference to the graphemes. To do so, we find (line 26) the union X of all aligned graphemes (its boundary is shown in blue in Fig. 3). For each grapheme, as above, points are sampled from the reference, but this time the reference points are aligned to the grapheme mesh (line 29). The thus computed deformation is extrapolated to the entire domain X using radial basis functions (line 30) to find the shape deformation S^i that best explains a grapheme in terms of the warped reference. Given the extrapolated deformations, the texture T^i is found by appropriately sampling the original grapheme image (line 31). Figure 3 shows the reference shape (a) warped to align with a grapheme (b), and the texture sampled from the grapheme into reference shape (c). Figure 5 shows the results of the groupwise registration of graphemes.

Having obtained samples of the deformed shapes S^i and the corresponding textures T^i , we seek to obtain their low-dimensional representation. In this work we employ basic linear dimensionality reduction with Principal Component Analysis (PCA), as done in [5], yielding a twice-linear model

$$E\mathbf{a} = \begin{pmatrix} E_{cs} \\ E_{ct} \end{pmatrix} \mathbf{a} = \begin{pmatrix} \mathbf{c}_s \\ W_t \mathbf{c}_t \end{pmatrix} = \begin{pmatrix} E_s^T (\mathbf{s} - \boldsymbol{\mu}_s) \\ W_t E_t^T (\mathbf{t} - \boldsymbol{\mu}_t) \end{pmatrix},$$

with the shape model $\mathbf{s} = E_s \mathbf{c}_s + \boldsymbol{\mu}_s = E_s E_{cs} \mathbf{a} + \boldsymbol{\mu}_s$ and the texture model $\mathbf{t} = E_t \mathbf{c}_t + \boldsymbol{\mu}_t = E_t W_t E_{ct} \mathbf{a} + \boldsymbol{\mu}_t$, where \mathbf{a} , \mathbf{s} , and \mathbf{t} are the appearance, shape, and texture parameters (feature vectors) respectively; E , E_s , E_t are the corresponding eigenvectors, and W_t are normalisation weights [5]. We preserve 128 dimensions in appearance feature vectors.

№	Type	Output	№	Type	Output
0	Input	$64 \times 64 \times 1$	4	Conv $3 \times 3 @ 128$	$4 \times 4 \times 128$
1	Conv $3 \times 3 @ 16$	$32 \times 32 \times 16$	5	Conv $3 \times 3 @ 256$	$2 \times 2 \times 256$
2	Conv $3 \times 3 @ 32$	$16 \times 16 \times 32$	6	Conv $1 \times 1 @ 512$	$1 \times 1 \times 512$
3	Conv $3 \times 3 @ 64$	$8 \times 8 \times 64$	7	Conv $1 \times 1 @ 43$	$1 \times 1 \times 43$
			8	Softmax Loss	

Table 1. Structure of our CNNs.

Analysis with CNNs. Deep convolutional neural networks (CNNs) have recently demonstrated excellent performance on diverse tasks ranging from image classification, to speech recognition, to natural language processing [6, 13]. The main challenge in our case is that the number of training samples available per each of the 43 alphabetic classes divided between 19 temporal classes is orders of magnitude too scarce for naïvely training CNNs. We employ two strategies to tackle sample deficit. First, we train a CNN on the task of classifying graphemes into alphabetic classes (OCR), without subdivision into temporal classes. (This OCR CNN was also used in annotation of the dataset.) We discovered that it is possible to achieve this with a very small CNN that does not over-fit despite sample deficit. Having trained the OCR CNN, we reuse the low- and mid-level representations learned by it for the task of dating. This technique is known as transfer learning [2, 18]. Second, instead of pursuing a classifier with 19 temporal classes, we reduce the problem to an ensemble of binary classifiers, as will be discussed below. In this regime there are substantially more samples available per class.

The structure of the network is shown in Table 1. It consists of 9 convolutional blocks. Each block consists of a convolutional layer (see Table 1 for filter dimensions and numbers; we use zero-padding throughout), followed by a dropout layer [17] with 0.2 rate, followed by a ReLU non-linearity, followed by 2×2 max-pooling (except in the last two blocks). The total number of parameters is 1.6×10^6 .

The network takes as an input 64×64 images. All grapheme images were resized to 48×48 , preserving the aspect ratio, and leaving 8 pixel border on each side (necessary for augmentation, see below). The size of the images was chosen by hyper-optimisation. We observed that much smaller images result in poor discrimination, while much larger sizes do not yield any appreciable increase in performance. Training was done for 1500 epochs using the standard stochastic gradient descent optimiser, with an initial learning rate of 0.001, descending to 0.0001 after 750 epochs. The batch size of 100 was chosen. We did not observe any increase in performance on the development set with longer training. To partially alleviate the imbalance between classes, batches are formed in a stratifying manner, sampling the more rare classes proportionately more frequently.

In order to decrease the sensitivity of the CNN to variations that may occur in the data yet do not have paleographic significance, the images in each batch were subject to augmentation by random affine trans-

forms, and random alteration of the stroke widths by morphological erosion and dilation, which we perform on-the-fly. This augmentation serves as regularisation and helps alleviate over-fitting. For all experiments with CNNs we used the MatConvNet toolbox [20]. The error rate of the CNN on the OCR task was 2.04% on the testing set (1/8 of all letters) and 0.98% on the training set: over-fitting is very insignificant. Investigation of the misclassified samples shows that most of them are either highly distorted or mis-written letters.

We take the 512-dimensional output of the penultimate fully connected layer (block 7) as the high-level feature representation of the graphemes. As will be discussed below, multiple networks we re-trained (fine-tuned) as binary classifiers, to further improve their dating performance. The tuning proceeded as above (with block 8 replaced for the case of two labels), for 100 epochs and with a very small learning rate of 1×10^{-5} .

Ensembles of Classifiers. As in [22], we discretise the entire time line into $N_t = 19$ bins (temporal classes): between the years 1100–1400 each bin is 20 years long, in the 15th century the bins are 1400–1410, 1410–1420, 1420–1450, and in the 11th century, there is only one bin 1025–1100.

We proceed by reducing the problem of multiclass classification to training an ensemble of binary classifiers and combining their results. Two most common approaches are: one-vs-all (N classifiers are trained, to differentiate each class from all other classes) and one-vs-one ($N(N-1)/2$ classifiers to differentiate between all possible pairs of classes) [12]. The relevant more general theory here is the method of classifier ensembles based on error-correcting codes (ECOC) [12], which considers arbitrary dichotomies of a multi-class set into two-class sets. Using the ECOC terminology [12], we represent a dichotomy of an N_t -class set as a binary vector $\mathbf{d} \in \{0, 1\}^{N_t}$, whose elements represent to which of the two new classes the original classes correspond. Similarly, ensemble of N_c classifiers, each with its own dichotomy, can be represented by a binary *coding matrix* $D_{N_t \times N_c}$. Standard ECOC coding matrices include [12] the above one-vs-all and one-vs-one schemes (as a generalisation), random codes, and even exhaustive codes (prohibitive in our case).

Zaliznyak [22] notes that, instead of gradually evolving over time, writing style features first *abruptly* appear (get invented), then remain in use for a number of decades (co-existing and competing with other styles for their relative frequency), then finally die down [22]. We take advantage of this fact and train an ensemble where each classifier distinguishes between a particular *continuous* segment in time and all other dates, with one classifier for each possible segment: $[1 \dots 1]$, $[1 \dots 2]$, \dots , $[1 \dots (N_t - 1)]$, $[2 \dots 2]$, $[2 \dots 3]$, \dots , $[2 \dots (N_t - 1)]$, \dots , $[(N_t - 2) \dots (N_t - 1)]$, $[(N_t - 1) \dots (N_t - 1)]$. It is easy to see that there are $N_t(N_t + 1)/2 - N_t/2 = N_t(N_t - 1)/2 = 171$ such classifiers. With this arrangement, we expect that for some of the classifiers in the ensemble,

their corresponding segments will coincide *exactly* (modulo discretisation) with the lifetime of some paleographically significant features, hence making (at least some of) the classifiers much more sensitive than in any other regime.

We have trained, for each grapheme in the alphabet, a full set of 171 classifiers using a basic linear SVM using appearance model features, and a further ensemble using CNN features. For each classifier, we estimated its performance by computing the confusion matrices $C_{2 \times 2}$ on a small validation subset of the training set. ($C(i, j)$ shows the number of samples known to be from class i classified as belonging to class j .) We have also computed the unweighted average recall for each classifier:

$$\text{AR}(C_{K \times K}) = \frac{1}{K} \sum_{i=1}^K \frac{C(i, 1)}{\sum_{j=1}^K C(i, j)}. \quad (\text{here, } K = 2) \quad (1)$$

We then proceeded to fine-tune an ensemble of CNNs as binary classifiers, for those dichotomies for which corresponding SVM classifiers yielded a “promising” performance ($\text{AR} \geq 60\%$). We then re-trained the ensemble of SVM classifiers from the thus updated CNN features. The rationale for tuning is as follows: the generic OCR network serves as a good initialisation (and its features, as we noted, are already useful for dating), and tuning the ensemble of networks, each to its assigned time segment, further improves the sensitivity by improving the learned features (at all levels) that may be specific to a particular time segment.

Aggregating Evidence. We adopt Bayesian inference as the most natural framework in which to combine the evidence from multiple graphemes, and for each grapheme — the predictions of multiple classifiers. Given an unseen manuscript to be dated, we extract all the graphemes and evaluate each of them with the corresponding classifier ensemble. To fuse the evidence from all classifiers, we maintain a discrete probability mass function $P_{1 \times N_t}$, $P(a) = P(a|L_1, L_2, \dots)$, which depends on observing graphemes L_1, L_2, \dots , over the discrete set of $N_t = 19$ temporal classes $a \in \{1 \dots N_t\}$.

Out of necessity, we make the assumption that observing each next grapheme L_i in the manuscript is an event independent from observing other graphemes in the same manuscript.

Algorithm 2 Bayesian aggregation step

Require: Prior distribution $P_{N_t \times 1}$; classifier’s dichotomy $\mathbf{d}_{N_t \times 1} \in \{0, 1\}^{N_t}$, vote $e \in \{0, 1\}$, normalised confusion matrix $C_{2 \times 2}$, smoothing parameter α , threshold AR_{thr} .

- 1 **function** UpdateBelief(P, \mathbf{d}, v, C)
- 2 **if** $\sum_{i=1}^2 C(1, i) = 0$ **or** $\sum_{i=1}^2 C(2, i) = 0$ **then**
- 3 **return** P ▷ uninformative classifier, abort
- 4 **if** $\frac{1}{2}(C(1, 1) + C(2, 2)) < \text{AR}_{\text{thr}}$ **then**
- 5 **return** P ▷ below AR threshold, abort
- 6 $P_{\text{new}}(i) \leftarrow \mathbf{0}_{N_t \times 1}$
- 7 **for** $i \leftarrow 1$ to N_t ▷ update belief
- 8 $P_{\text{new}}(i) \leftarrow P(i) \cdot \hat{C}(\mathbf{d}(i) + 1, e + 1)$
- 9 $P_{\text{new}} \leftarrow P_{\text{new}} / \sum_i P_{\text{new}}(i)$ ▷ normalise
- 10 **return** $(1 - \alpha)P_{\text{new}} + \alpha P$ ▷ apply smoothing

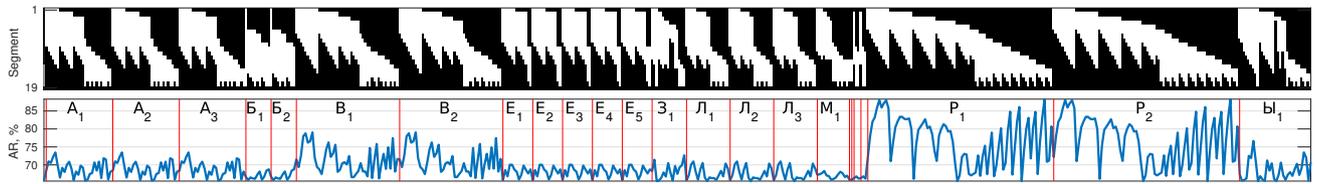


Fig. 7. Example of dating with a classifier ensemble (BBM 436). *Top*: map of classifier segments. *Bottom*: unweighted average recall of the classifiers. Red lines delineate individual letters (25 letters survived thresholding in this example).

Using the independence assumption, we thus perform Bayesian update of P for each evidence $e \in \{0, 1\}$ from each classifier from each grapheme as follows:

$$P_t(a|e_t) \leftarrow \frac{P(e_t|a)P_{t-1}(a)}{P(e_t)} \quad (\forall a \in \{1 \dots N_t\}). \quad (2)$$

We can estimate $P(e_i|a)$ and $P(e_i)$ from the classifiers' confusion matrices. Let C be the row-normalised confusion matrix for some classifier with coding vector $\mathbf{d} \in \{0, 1\}^{N_t}$, and assume the classifier outputted evidence $e \in \{0, 1\}$, then Eq. (2) becomes:

$$P_t(a|e_t) \leftarrow \frac{C(\mathbf{d}(a) + 1, e_t + 1)P_{t-1}(a)}{\sum_{i=1}^{N_t} C(\mathbf{d}(i) + 1, e_t + 1)P_{t-1}(i)}. \quad (3)$$

The Bayesian update step is summarised in Algorithm 2. As an additional measure to improve the quality of evidence, we reject all classifiers for which the average recall is below a threshold AR_{thr} (lines 4–5), as well as discarding completely uninformative classifiers (lines 2–3). To partially compensate for the violation of the independence assumption, we introduce a smoothing parameter α which reduces the amount of information each next classifier contributes (line 10). (We optimised the smoothing parameter α and the rejection threshold AR_{thr} on a small (validation) subset of the training set.) Algorithm 2 is called to update P for each grapheme in a manuscript, for each classifier in the grapheme's ensemble, yielding the estimated date distribution for the manuscript.

Figure 7 illustrates the ensemble dating (Bayesian aggregation) on an example of a particular manuscript (BBM №436). It can be seen, in this example, that graphemes “ρ”, “Б”, and “А” were among the most significant for dating.

3. Results and Evaluation

We applied our method to the task of dating the BBM corpus. We evaluate the performance of our method using eight-fold cross-validation (7/8th of the data in the training set, 1/8th in testing). Special care has been taken to ensure that documents (or parts of a document) written by the same hand are always in the same set. To qualitatively measure the dating performance of our system, we compute the mean absolute error (MAE), and the cumulative score (CS_α) which measures the fraction of test documents for which the date estimation error is no greater than α , as done in [7–11]. We additionally compute the Kullback–Leibler divergence (KDL) and the earth mover's dis-

tance (EMD) between the ground truth and the predicted distributions.

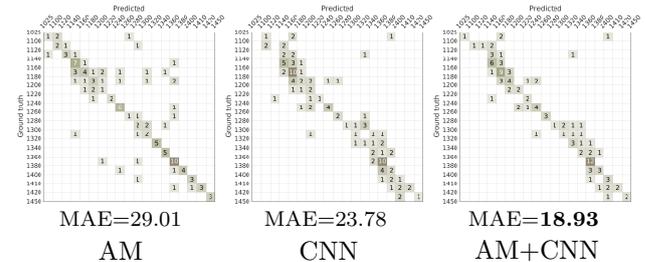


Fig. 8. Confusion matrices for the different combination of features.

We have evaluated our approach with both types of features, AM and CNN, and their combinations: at the feature level (AMUCNN) and at the decision level (concatenating classifier ensembles), AM+CNN. We have done this for both simple majority voting (as a baseline) and Bayesian aggregation, with flat and empirical priors. As a baseline, we also performed aggregation by simple majority voting. The results are summarised in Table 2, the corresponding cumulative scores in Fig. 9 (left), and the resulting confusion matrices in Fig. 8. The best MAE achieved was **18.93** years with AM+CNN features and empirical prior. (Overall, using empirical prior gave minor improvement over flat prior, except for AMUCNN features.) Bayesian aggregation overall yielded much better results than majority voting. Individually, CNN features performed better than AM. Combination at the decision level produced better result than at the feature level. We compare of our method (Table 3) with other relevant computer paleography studies [7–9, 11, 14], however exact comparison is difficult due to the different nature and amount of material

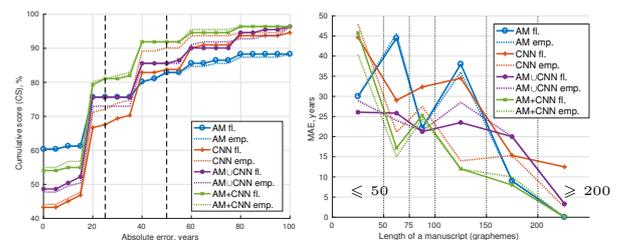


Fig. 9. *Left*: Cumulative scores (CS) for the different features. *Right*: Dating error as a function of the manuscript length.

Feat.	AM				CNN				AMUCNN				AM+CNN			
	voting		Bayes		voting		Bayes		voting		Bayes		voting		Bayes	
	fl.	emp.	fl.	emp.	fl.	emp.	fl.	emp.	fl.	emp.	fl.	emp.	fl.	emp.	fl.	emp.
MAE	96.71	77.39	28.83	29.01	41.19	47.84	24.10	23.78	46.80	51.80	22.41	22.68	50.23	42.43	20.34	18.93
EMD	121.15	116.47	31.08	30.83	110.53	104.13	25.15	24.25	118.54	113.54	23.99	23.22	108.16	95.26	20.30	18.91
KLD	9.11	9.07	5.36	5.33	9.01	8.92	7.88	7.75	9.07	9.05	7.14	7.04	8.99	9.07	6.47	6.28
CS ₂₅	37.84	36.94	75.68	75.68	60.36	47.75	67.57	72.07	56.76	45.05	75.68	72.97	54.95	47.75	81.08	81.08
CS ₅₀	49.55	54.05	82.88	82.88	74.77	65.77	83.78	90.09	70.27	63.96	85.59	85.59	69.37	68.47	91.89	91.89

Table 2. Summary of the results. The dating performance is shown for the appearance (AM) and neural network (CNN) features, and their combination at feature level (AMUCNN) and decision level (AM+CNN). The results are shown for both voting and Bayesian aggregation, with flat (fl.) and empirical (emp.) priors.

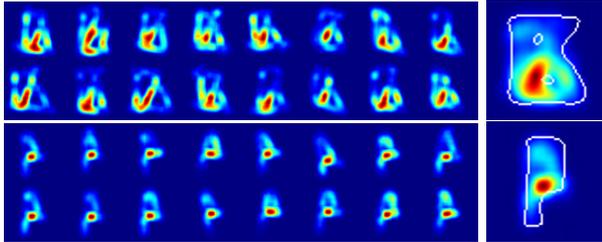


Fig. 10. Importance map of grapheme parts as measured by the dating CNN response. *Left:* individual importance maps for 16 graphemes that cause the strongest response. *Right:* average map with grapheme outline superimposed. *Odd rows:* response to ≤ 1300 , *even rows:* > 1300 .

Study	MAE	CS ₂₅	CS ₅₀
[7]	35.4	63.5%	($\approx 85\%$) [*]
[9]	20.9	77.5	88.5%
[14]	20.5	—	—
[8]	14.1/41.0 [‡]	74.3%/53.3%	($\approx 80\%/70\%$)
[11]	15.9	85.4%	($\approx 90\%$)
Our	18.9	81.1%	91.9%

Table 3. Comparison with other studies.

^(*)Numbers in brackets are not given in the paper, but estimated from the plots.

(our BBMs are orders of magnitude shorter) and the differences in time scales. Our method performs better than [7, 9, 14] both in terms of MAE and CS. The MAE in [8] (14.1) and [11] (15.9) are lower than our best result (18.9), but our CS scores are substantially better than [8], indicating higher reliability. Further, results in [8] drastically drop to MAE=41.0 ([‡]) when they apply dating across cities. Only the study [11] surpasses our result in MAE and CS₂₅, but we note that their range of dates is narrower and random guess in [8, 11] yields MAE=85.3 (in our case random guess MAE=124.2 years), therefore it is more meaningful to compare the results as percentages of the random MAE: scaled to our range, the results of [8, 11] would yield 20.5 and 23.2 years respectively — thus the relative accuracy with our method is still better.

Given that the success in estimating age depends on the amount of available evidence (graphemes), we plot the MAE as a function of manuscript length in Figure 9 (right). It is also prudent to visualise the separ-

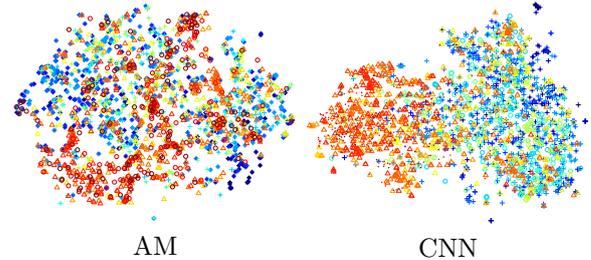


Fig. 11. Embeddings of the CNN and AM features in \mathbb{R}^2 using t-SNE [19] (grapheme “B”). Colour indicates age.

ability of classes by plotting the embedding of feature vectors into a low-dimensional space. We accomplish this with t-SNE algorithm [19]. Figure 11 shows the embeddings for AM and CNN features (grapheme “B”). At least in this example, CNN features appear more clearly separable. Importantly, the different topologies of the embeddings suggest that the AM and CNN features capture different aspects of graphemes’ appearance and are complementary. Further, having fine-tuned the CNNs on binary dating task, we can explore what features in the graphemes contribute to the network’s response, thus potentially revealing human-interpretable features. We have done so by systematically occluding parts of a grapheme image with a 8×8 black box, and measuring the change in activations in the final classification layer [23]. We show the results in Fig. 10 (for the 16 samples that produce the strongest responses (*left*), and the average (*right*)). Remarkably, the most significant areas approximately correspond to the location of some of the features in paleographic tables [22], *e.g.* a triangle-shaped loop in “p” is characteristic of the later period, and for “B” the connection of the diagonal stroke to the vertical stem.

Conclusion We have investigated paleographic dating using an elastic model of graphemes’ deformation, features learned by CNNs, and Bayesian aggregation of results from an ensemble of specially tuned classifiers. The experimental results clearly demonstrate the efficiency of our approach (MAE=18.93 years). Our method may corroborate and refine the existing paleographic analysis by human experts.

4. Bibliography

- [1] “Birchbark literacy from Medieval Rus: Contents and contexts (digital archive),” <http://gramoty.ru>.
- [2] Y. Aytar and A. Zisserman, “Tabula rasa: Model transfer for object category detection,” in *ICCV*, Nov 2011, pp. 2252–2259.
- [3] T. F. Cootes *et al.*, “Groupwise construction of appearance models using piece-wise affine deformations,” in *BMVC*, 2005, pp. 879–888.
- [4] T. F. Cootes *et al.*, “Computing accurate correspondences across groups of images,” *IEEE PAMI*, vol. 32, no. 11, pp. 1994–2005, Nov 2010.
- [5] T. F. Cootes *et al.*, “Active appearance models,” *IEEE PAMI*, vol. 23, no. 6, pp. 681–685, 2001.
- [6] J. Gu *et al.*, “Recent advances in convolutional neural networks,” *CoRR*, vol. abs/1512.07108, 2015.
- [7] S. He *et al.*, “Towards style-based dating of historical documents,” in *ICFHR*, Sept 2014, pp. 265–270.
- [8] S. He *et al.*, “Discovering visual element evolutions for historical document dating,” in *ICFHR*, Oct 2016, pp. 7–12.
- [9] S. He and L. Schomaker, “A polar stroke descriptor for classification of historical documents,” in *ICDAR*, Aug 2015, pp. 6–10.
- [10] S. He *et al.*, “Historical document dating using unsupervised attribute learning,” in *DAS*. IEEE Computer Society, 2016, pp. 36–41.
- [11] S. He *et al.*, “A multiple-label guided clustering algorithm for historical document dating and localization,” *IEEE Trans. Image Processing*, vol. 25, no. 11, pp. 5252–5265, 2016.
- [12] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [13] Y. LeCun *et al.*, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [14] Y. Li *et al.*, “Publication date estimation for printed historical documents using convolutional neural networks,” in *Proc. 3rd Int. Workshop on Historical Document Imaging and Processing*, ser. HIP ’15, 2015, pp. 99–106.
- [15] K. Sidorov *et al.*, “An efficient stochastic approach to groupwise non-rigid image registration,” in *CVPR*, 2009, pp. 2208–2213.
- [16] X. Song and A. Myronenko, “Point set registration: Coherent point drift,” *IEEE PAMI*, vol. 32, pp. 2262–2275, 2010.
- [17] N. Srivastava *et al.*, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [18] T. Tommasi *et al.*, “Safety in numbers: Learning categories from few examples with multi model knowledge transfer,” in *CVPR*, 2010, pp. 3081–3088.
- [19] L. van der Maaten and G. E. Hinton, “Visualizing high-dimensional data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [20] A. Vedaldi and K. Lenc, “MatConvNet: Convolutional neural networks for MATLAB,” in *ACM Int. Conf. on Multimedia*, 2015.
- [21] F. Wahlberg *et al.*, “Large scale style based dating of medieval manuscripts,” in *IWHDDIP*, 2015, pp. 107–114.
- [22] V. L. Yanin and A. A. Zaliznyal, *Novorodskiye gramoty na bereste (iz raskopok 1990–1996 godov)*. Moscow: Nauka, 2000.
- [23] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *ECCV*, 2014, pp. 818–833.

Towards cylindrical primitive extraction from industrial laser-scan data by local approximation with refinement

S. Slyadnev, V. Turlapov
sergey.slyadnev@gmail.com, vadim.turlapov@itmm.unn.ru
Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

We present a novel method for semi-automatic reconstruction of cylindrical primitives from three-dimensional dense point clouds corresponding to the results of laser-scan acquisition. Our method is a part of the general reverse engineering procedure aimed at reconstructing a complete 3D model of an industrial site, such as a petrochemical factory. We use differential properties of a point cloud to restore cross-sections of a cylindrical primitive representing a straight pipe segment. Our method requires user interaction for the selection of cross-sectional circles of the candidate cylinders. In order to make reconstruction more accurate, we refine the extracted radius and axis by taking into account the extended neighborhood of the sample point.

Keywords: *As-built reconstruction, reverse engineering, cylinder fitting, industrial reconstruction.*

1. Introduction

Digital models of industrial facilities, such as refineries, are important in many scenarios, including training, maintenance, renovation of equipment, simulation of different operational situations, data exchange, etc. Existing industrial facilities can be accurately measured with the use of terrestrial laser scanners. Laser scanning systems produce large-scale point clouds with a point density large enough to cover even small details of a measured facility. Typical point clouds representing industrial installations may contain hundreds of millions of points. Automated and accurate transformation of such point clouds to compact 3D representations remains an open problem which is known as *reverse engineering*. A class of reverse engineering problems dealing with large amounts of discrete data representing existing industrial facilities is called *as-built reconstruction*. Oil platforms, factories, buildings are just some of the examples of the facilities that can be reengineered in this field.

There is a growing demand for as-built reconstruction of industrial sites using laser-scan data. As-built reconstruction is aimed to produce a compact three-dimensional model of an operational facility which can be subsequently used in different design and management software (e.g. AVEVA PDMS). Our paper is focused on the reconstruction of cylindrical pipelines. As mentioned by [Rabbani 2006] and [Qiu et al. 2014], a cylinder is one of the most frequently used primitives in industrial engineering. This is especially true for processing industries such as petrochemical plants, refineries, nuclear power plants, etc. As a result, automatic and robust methods for detecting and fitting cylindrical primitives are essential for the reconstruction of such sites.

We present a reconstruction procedure which employs differential characteristics of a point cloud as proposed by [Son and Kim 2016]. We demonstrate that the use of differential heuristics alone may introduce a variety of subtle phenomena which complicate the reconstruction process. These problems can be solved by refining the extracted differential characteristics taking into account global cues. Our work is focused on the reconstruction of straight pipe segments. We do not examine pipe connection features such as elbows or tee pipes. Reconstruction of pipe connection features and other primitive types is left for future research.

2. Related work

[Son et al. 2015] conducted a survey on as-built reconstruction methods for industrial facilities, such as

petrochemical plants. Their study covers both automatic and interactive reconstruction techniques including approaches adopted in commercial systems. In particular, the authors mention that the *majority of existing commercial tools are interactive in nature*. Although, according to [Benko et al. 2001], in respect of common engineering objects (i.e. those bounded by simple analytic surfaces) significant research efforts are directed at automatic reconstruction with little or no user interaction.

Reverse engineering typically starts with segmentation of an input point cloud into meaningful regions. In general, the segmentation problem is ill-posed, as there is no any objective criterion to measure the segmentation quality. A particular segmentation technique is chosen in function of the target application domain. [Le and Duan 2017] proposed a segmentation framework for the reconstruction of mechanical CAD parts. The authors state that their approach is more robust than the widely adopted random sample consensus (RANSAC) method introduced by [Schnabel et al. 2007]. They identify the model's major directions and perform dimensional reduction before the extraction of various types of geometric primitives. The method is based on the observation that many mechanical models have only a moderate number of primary orientations around which the CAD features reside. These orientations are extracted from Gaussian maps corresponding to plane normals. As-built reconstruction methods actively use Gaussian maps for the recognition of principal axes of an object.

For the automatic decomposition of large point clouds into connected regions, [Rabbani et al. 2006] used segmentation by smoothness constraint which is based on the pre-computed normal field. Segmentation comprises two steps: normal estimation and region growing. This simple technique has found broad application in industrial reconstruction. Region growing is one of the initial stages of our method.

[Benko et al. 2001] utilized the so-called direct segmentation which is an efficient, non-iterative approach. Their segmentation technique is based on the selection of "stable" points within a candidate region. If local characteristics of points within a selected area are unstable, the algorithm deduces that the neighborhood lies between two or more regions.

A popular technique for 3D reconstruction from a point cloud is mesh generation (see [Hoppe et al. 1992]). However, mesh alone is rarely enough, as it lacks information on semantics (CAD features) and does not offer such compact representation as planar, cylindrical, conical, or toroidal primitives. There are hybrid approaches which employ mesh generation at the pre-processing stage for accurate

reconstruction. [Masuda and Tanaka Ichiro 2010] presented a system which is based on the conversion of input point clouds into mesh models. After mesh generation the user may perform interactive picking of a seed region which is automatically expanded by the system to one of the conventional primitive types.

[Qiu et al. 2014] presented a method for the reconstruction of pipelines and their joints from industrial point clouds. According to the authors, RANSAC method is not reliable enough to drive the automated as-built reconstruction process. The same applies to the GlobFit method described by [Li et al. 2011], as, according to [Qiu et al. 2014], the GlobFit method may suffer from unreliable initial primitive detection.

[Liu et al. 2013] classify all pipelines from laser scan data as those perpendicular to the ground and those parallel to the ground. They reduce 3D reconstruction problem into a simpler 2D circle detection problem. However, the employed prior knowledge of the ground plane and the assumption about the pipes orientation with respect to this plane may be false.

[Tran et al. 2015] extract cylindrical primitives automatically. They use a dedicated validation stage to accept or reject the fitting results. The authors notice that acquisition noise and point cloud structure may lead to convergence to a wrong solution. Therefore, multiple fitting attempts are necessary to achieve the expected results.

[Son and Kim 2016] presented a fully automatic pipeline reconstruction procedure. Their method employs local approximation with a B-spline surface to derive differential properties of a point cloud. It looks promising; however, the described method may suffer from fluctuations in extracted differential properties due to imperfections of a locally approximating surface. We have reused the ideas presented by [Son and Kim 2016] and enriched them with additional refinement stages aimed to offset the negative effects of local fluctuations.

3. Algorithm

Regardless of whether a software system is interactive or not, there is a common pool of algorithms serving the reconstruction procedure. The “gentleman’s set” includes noise smoothing, point cloud filtering, region growing, primitives fitting, feature recognition, and many other utilities. Our intention was to develop a set of auxiliary methods which would help solve a broad range of as-built reconstruction problems. The output of reconstruction should be transferable to any popular CAD system for accurate manual operation. In order to achieve this goal, we have created an interactive system for as-built reconstruction. The system works as a test bench for different reverse engineering techniques, including those which allow for user interaction. In this section, we present the basic implemented approach and highlight several subtle problems that may arise.

3.1 Outline of the algorithm

Our algorithm is outlined as follows:

1. Prepare point cloud:
 - a. Filter out near-coincident points in 3D.
 - b. Build k-d tree for the initial point cloud.
 - c. Build normals for the initial point cloud.
2. Perform coarse segmentation by region growing.
3. Prepare cross sections:

- a. Let the user pick interactively seed points within the area where local characteristics of points are stable.
 - b. Approximate each picked point’s neighborhood with a B-spline surface.
 - c. Build normal plane using differential properties of local approximation surface.
 - d. Find points which lie within a certain threshold distance to the normal plane. We call such points a *neighbor band*.
 - e. Refine normal plane with respect to the Gauss map of the normal vectors corresponding to the neighbor band points.
 - f. Calculate the normal section curve by intersecting the approximation surface with a refined normal plane.
 - g. Build an osculating circle for the normal section curve.
 - h. Project neighbor band points to the normal plane.
 - i. Refine the osculating circle to fit the projected points.
 - j. Snap the resulting radius value to one of the predefined radius values of Piping and Instrumentation Diagrams (P&ID) available for the industrial facility.
4. Build a cylindrical surface between the reconstructed circular sections.
 5. Trim the cylindrical surface by projecting the points of a region and filtering out the outliers.

For fast nearest neighbor matching we have used the open-source FLANN library presented by [Muja and Lowe 2014]. Normal vectors are calculated by local plane approximation with the use of covariance matrices at each point as described by [Hoppe et al. 1992]. Normal orientation calculated in this way is ambiguous (see Figure 1). However, region growing segmentation and other steps of our method are not sensitive to such ambiguity.

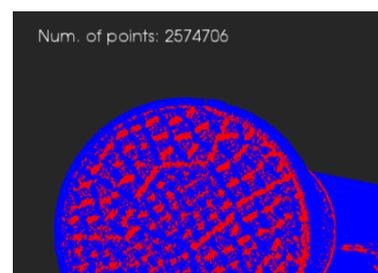


Fig. 1. Ambiguity in the orientation of normal vectors (blue). The original point cloud is shown in red color.

Region growing has been used for decomposition of the initial point cloud into the segments which can be processed individually. Such coarse segmentation is an essential step to enable the work with large industrial point clouds as it allows for piece-by-piece operation, thus relaxing the hardware’s requirements. The segmented point cloud is then interactively modified by the user. The user picks a set of seed points to make an assumption on cylindrical primitive’s location. The system analyzes local differential characteristics of the selected region and reconstructs a cross-sectional circle which passes through a refinement stage as described in paragraph 3.4. The

system is then asked to build an infinite cylindrical surface passing through the set of guess circles. Finally, the infinite surface is trimmed by a subset of tentative points of a region. The trimming is accomplished by checking if the distances from the points to the cylindrical surface are within a threshold value. Figure 2 shows some results obtained with the use of our method on synthetic models.

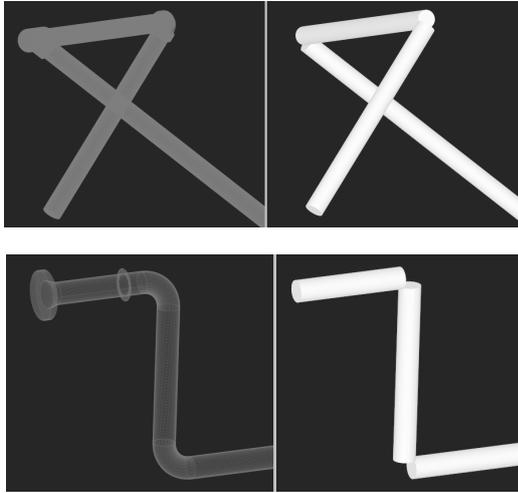


Fig. 2. Some examples of the reconstruction results obtained with the use of our method.

3.2 Data management

The reconstruction process involves several data abstractions which are conveniently represented in an object-oriented programming language such as C++. In addition to the data structures representing input point clouds and resulting primitives, it is convenient to have dedicated data structures for segmentation results, normal vectors, Gauss maps, curves and surfaces employed in reconstruction, etc. In our work, we used the open-source framework [Slyadnev 2017] for data management and visualization. Figure 3 shows the hierarchy of data objects employed in our reconstruction procedure.

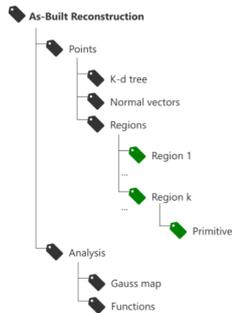


Fig. 3. Hierarchy of the employed data structures. Skeleton objects which persist throughout the reconstruction process are marked with black labels. Objects which are allocated during the reconstruction procedure are marked with green labels.

All inputs and outputs are placed under the *Points* object which represents the initial point cloud. The child objects store the k-d tree for fast neighbors access and the normal vectors calculated at each point. The results of the region growing segmentation are collected under the *Regions* group. A region is nothing but a set of indices pointing to the input cloud elements. Once the reconstruction is done, the resulting primitives can be saved to STEP (ISO 10303) file allowing for interoperability with any CAD system.

3.3 Extract differential properties

Some popular computer vision techniques for primitive fitting are RANSAC and Hough transform. These methods generally avoid using differential characteristics of a point cloud such as principal curvatures. In our method, we calculate the differential properties of an interactively picked area. Having these properties allows for *recognition prior to fitting*. Recognition lets the system “guess” whether the picked region constitutes a cylinder or some other type of primitive. Significant oscillations in the local characteristics let the system guide the user to choose another region which is more “stable”.

Differential properties of a point cloud segment can be estimated by local approximation. Such properties as principal curvatures are very sensitive to fairness of the used approximation surface. In this work, we have adopted *thin plate spline (TPS)* technique based on OpenCascade geometric kernel (see [Slyadnev 2014] for the overview). The used approach is similar to the one described by [Hegland et al. 1997].

3.4 Refinement of cylindrical section

According to [Kim and Son 2016], differential characteristics of a locally approximating surface patch can be used to extract the main properties of a reconstructed cylindrical surface, i.e. its radius and axis. We have found however that using differential properties alone is a troublesome technique. The reconstruction approach which is exclusively based on differential properties of a point cloud is impractically sensitive to the quality of a locally approximating surface. Intuitively, such an approach may not provide reliable results as it uses inherently local (thus unreliable) geometry characteristics to derive global shape properties. Depending on a surface approximation technique, it may happen that the obtained radius and the axis direction are significantly distorted even for a regular point cloud without any noise. Figure 4 illustrates the osculating circle which does not fit in the intuitively expected shape of the cylinder.

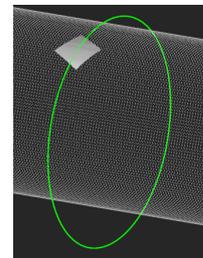


Fig. 4. An example of an inaccurate osculating circle whose quality suffers from oscillations in a local approximation surface. The radius of the osculating circle is greater than expected.

Figure 5 illustrates unexpected inclinations of normal planes. The inaccuracy of a normal plane means inaccuracy of the detected axis direction.

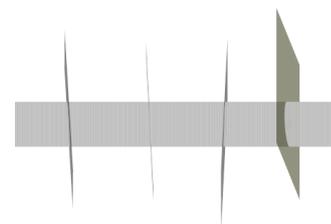


Fig. 5. Unexpected inclinations of normal planes due to small fluctuations of local approximation surfaces.

Both problems are caused by insufficient fairness of a locally approximating surface. It is known from geometric modeling practice that the formalization of such “fairness” is not straightforward. Summarizing the detected phenomena, we conclude that the differential properties of the approximating surface are “weak” heuristics which cannot be directly used in reasoning about the global pipe’s properties. Apparently, both the radius and the axis of the cylindrical primitive have to be refined taking into account a “more global” behavior of the point cloud in a sufficiently larger subregion.

3.4.1 Axis refinement

First, we sharpen the orientation of the normal plane. As a global cue for refinement, we use the neighbor points lying near the original section plane (Figure 6). The corresponding normal vectors constitute a Gauss map whose vectors should be predominantly orthogonal to the cylinder axis.

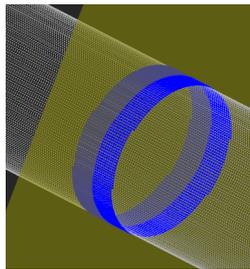


Fig. 6. Neighbor band points as a global cue for refinement.

For the unknown axis coordinates (x_1, x_2, x_3) , a residual value r for a single Gauss map element $\mathbf{u} = (u_1, u_2, u_3)$ can be calculated as a dot product $r = x_1u_1 + x_2u_2 + x_3u_3$. The perfect fit is achieved for $\sum_{i=1}^N r_i = 0$, where N is the number of elements in the Gauss map (the number of points in the neighbor band). The perfect fit criterion can be formulated for each component individually as $x_k \sum_{i=1}^N u_i = 0$ for $k = 1, 2, 3$. In matrix notation, this can be reformulated as a homogeneous linear system $A\mathbf{x} = 0$, where A is the diagonal matrix. Minimizing $|A\mathbf{x}|^2$ under the constraint $|\mathbf{x}| = 1$ using the Lagrange multipliers yields

$$\mathbf{x}^T A^T A \mathbf{x} + \lambda(\mathbf{x}^T \mathbf{x} - 1) = 0.$$

Derivation by \mathbf{x} yields

$$A^T A \mathbf{x} + \lambda \mathbf{x} = 0.$$

Therefore, we end up with the eigenvalue problem for $A^T A$ matrix. Figure 7 shows an example of axis refinement.

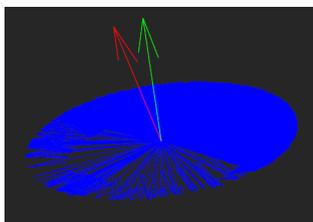


Fig. 7. Axis refinement using the Gauss map of neighbor band points. The original axis direction (red) is orthogonalized (green) in accordance with point cloud normals.

3.4.2 Cross-section refinement

The refined normal plane is used for computation of the intersection curve on the locally approximating surface. Inspired by [Son and Kim 2016], our intention was to use the osculating circle for this curve as a cross-section of the target

cylindrical primitive. However, such an approach is not sufficiently reliable due to possible oscillations of the approximating surface leading to a wild shape of the intersection curve (Figure 8).

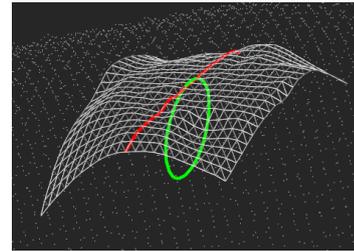


Fig. 8. Oscillating intersection curve (red) due to the insufficient fairness of the approximation surface (white wireframe). The osculating circle is shown in green color.

For circle refinement, we use simple gradient descent method with Armijo rule for adaptive step selection. The radius and the (u, v) center coordinates of the original osculating circle are passed to the optimizer as the initial guess. Other approaches, such as direct circle fitting proposed by [Coope 1993] also look promising. One example of osculating circle refinement is shown in Figure 9.

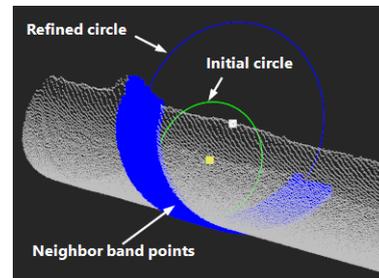


Fig 9. The result of refinement (blue circle) applied to the original osculating circle (green) used as the initial guess in local optimization.

4. Discussion and future research

Fully automatic as-built reconstruction is a challenging problem. We believe that the user’s intervention cannot be avoided in general case as no algorithm can reliably substitute an experienced human engineer. On the other hand, there is a huge amount of manual operations which consist in searching for stereotypical shapes in the input laser-scan data and selecting the best fitting primitives for the unambiguous regions. Such operations require many man-days of qualified human engineer’s work, so their automation is greatly beneficial. We have presented the algorithm for semi-automatic extraction of cylindrical primitives from dense laser-scan data. The algorithm is based on local analysis of a point cloud region with its user-driven completion into a segment of a straight cylindrical pipe.

We have found that differential characteristics are weak cues even if applied to synthetic point clouds without noise. Two such weak characteristics were used in our work: the osculating circle and the axis of the normal plane. Any such differential characteristic requires refinement by global cues. E.g. center and radius of osculating circle were refined by simple gradient descent method using the neighbor band points.

Figure 10 illustrates some reconstruction results obtained for regions of an industrial point cloud representing a real petrochemical factory.

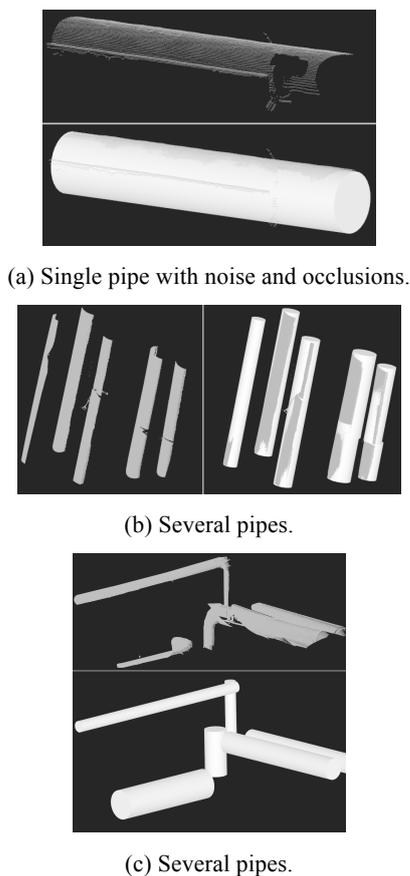


Fig 10. Some examples of reconstruction results. The input data set contains many occlusions and outliers.

In our experiments, we did not measure the deviation between the original points and the resulting cylinder since P&ID information was used to choose final radii. The precise surface deviation was of little interest in our setting, especially due to presence of huge geometric defects in the input clouds. Careful comparison of our method with other solutions is left for future research because we do not consider our method complete. Moreover, the methods which are reported as fully automatic and precise are not publicly available to make such comparison representative. However, we found interesting to compare our solution with a popular computer vision method RANSAC as reported by [Schnabel et al. 2007]. Our experiments show that RANSAC gives appropriate results for the unambiguously segmented point cloud regions. For industrial point clouds, such segmentation is hardly possible without intensive user interaction. Therefore, fully automatic RANSAC is not an option. The system which employs RANSAC should provide rich interactive means for point cloud segmentation.

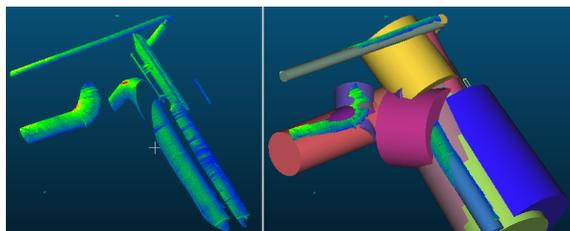


Fig 11. Recognition results using RANSAC method in system [CloudCompare 2017].

Figure 11 illustrates the results of RANSAC method launched for a piece of industrial point cloud (also shown in the Figure 10). The set of extracted cylinders contains many

redundant primitives which should be excluded from the result manually.

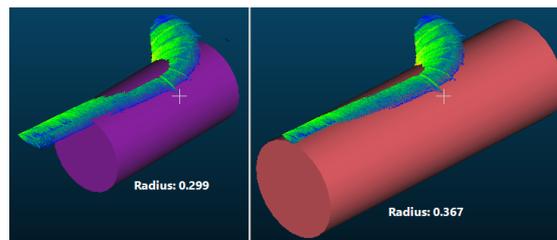


Fig 12. Recognition result using RANSAC method in system [CloudCompare 2017]. One portion of point cloud is recognized in two different ways.

Another observation is related to recognition of an individual primitive. Figure 12 shows that RANSAC method was able to detect two cylinders for a manually captured segment of a real industrial point cloud. The cylinders of different radii give multiple solution in a situation where only one solution is expected. The result of our method is shown in the Figure 13. Since several neighbor bands are used to detect and refine cylindrical cross-sections, our method extracts two radius values and averages them in order to return a single result.

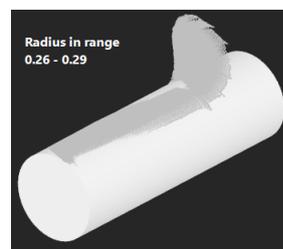


Fig 13. Recognition result using our method.

Another advantage of our method is that it enables *recognition prior to fitting*. Using differential characteristics of a local area, it is possible to decide algorithmically whether consequent fitting makes sense or not. Moreover, since differential characteristics of a point cloud are extracted by general-purpose TPS technique, our method can be extended to recognition and extraction of pipeline connection features.

5. References

- [1] Benko, P., Martin, R.R., and Varady, T. 2001. Algorithms for reverse engineering boundary representation models. *CAD Computer Aided Design* 33, 11, 839–851.
- [2] CloudCompare. 3D point cloud and mesh processing software. [online] Available at: <http://www.danielgm.net/cc> [Accessed 15 August 2017].
- [3] Coope, I.D. 1993. Circle fitting by linear and nonlinear least squares. *Journal of Optimization Theory and Applications* 76.
- [4] Hegland, M., Roberts, S., Altas, I., et al. 1997. Finite element thin plate splines for surface fitting. *Computational Techniques and Applications: CTAC97* 41, 1, 289–296.
- [5] Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W. 1992. Surface reconstruction from unorganized points. *ACM SIGGRAPH Computer Graphics* 26, 2, 71–78.
- [6] Le, T. and Duan, Y. 2017. A primitive-based 3D segmentation algorithm for mechanical CAD models. *Computer Aided Geometric Design*.
- [7] Li, Y., Wu, X., Chrysathou, Y., Sharf, A., Cohen-Or, D., and Mitra, N.J. 2011. GlobFit: Consistently Fitting

- Primitives by Discovering Global Relations. *ACM Trans. Graph.* 30, 4.
- [8] Liu, Y.J., Zhang, J. Bin, Hou, J.C., Ren, J.C., and Tang, W.Q. 2013. Cylinder detection in large-scale point cloud of pipeline plant. *IEEE Transactions on Visualization and Computer Graphics* 19, 10, 1700–1707.
- [9] Masuda, H. and Tanaka Ichiro. 2010. As-Built 3D Modeling of Large Facilities Based on Interactive Feature Editing. *Computer-Aided Design and Applications* 7, 1–10.
- [10] Muja, M. and Lowe, D.G. 2014. Scalable Nearest Neighbour Algorithms for High Dimensional Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 11, 2227–2240.
- [11] Qiu, R., Zhou, Q.Y., and Neumann, U. 2014. Pipe-run extraction and reconstruction from point clouds. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8691 LNCS, PART 3, 17–30.
- [12] Rabbani, T. 2006. Automatic Reconstruction of Industrial Installations Using Point Clouds and Images. *Publications on geodesy* 62, May, 7401–7410.
- [13] Rabbani, T., van den Heuvel, F. a, and Vosselman, G. 2006. Segmentation of point clouds using smoothness constraint. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences – Commission V Symposium “Image Engineering and Vision Metrology”* 36, 5, 248–253.
- [14] Schnabel, R., Wahl, R., and Klein, R. 2007. Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum* 26, 2, 214–226.
- [15] Slyadnev, S. 2014. Open CASCADE Technology Overview. [online] Available at: http://isicad.net/articles.php?article_num=17368 [Accessed 3 May 2017].
- [16] Slyadnev, S. 2017. Analysis Situs software for B-Rep inspection and prototyping CAD algorithms. [online] Available at: <http://www.analysis situs.org> [Accessed 3 May 2017].
- [17] Son, H., Bosché, F., and Kim, C. 2015. As-built data acquisition and its use in production monitoring and automated layout of civil infrastructure: A survey. *Advanced Engineering Informatics* 29, 2, 172–183.
- [18] Son, H. and Kim, C. 2016. Automatic segmentation and 3D modeling of pipelines into constituent parts from laser-scan data of the built environment. *Automation in Construction*.
- [19] Tran, T.T., Cao, V.T., and Laurendeau, D. 2015. Extraction of cylinders and estimation of their parameters from point clouds. *Computers and Graphics (Pergamon)* 46, 345–357.

About authors

Vadim Turlapov (vadim.turlapov@itmm.unn.ru): professor of Computer Science, Lobachevsky Nizhny Novgorod University.

Sergey Slyadnev (sergey.slyadnev@gmail.com): PhD student, Lobachevsky Nizhny Novgorod University.

Automatic calibration of surveillance video camera

I.A. Valuiskaia¹, E.V. Shalnov¹, A.S. Konushin^{1,2}

yana.valuyskaya@graphics.cs.msu.ru|eshalnov@graphics.cs.msu.ru|anton.konushin@graphics.cs.msu.ru

¹MSU, Faculty of Computational Mathematics and Cybernetics, Moscow, Russia

²HSE, Faculty of Computer Science, Moscow, Russia

This paper is devoted to the automatic calibration of surveillance video camera using objects in the scene. In this paper the problem of estimation of three extrinsic parameters (camera height, tilt angle, roll angle) is considered. The idea of baseline method, proposed in [14], is based on convolutional neural network. As input data, head bounding boxes and the camera focal length are used. In this paper, the modification of the baseline method was proposed, and also methods based on random forest and gradient boosting were studied in order to understand the necessity of using neural networks. An experimental evaluation of the proposed methods on synthetic and TownCenter datasets demonstrated their high efficiency. The best results was shown by the proposed method based on neural networks.

Keywords: automatic calibration, surveillance video camera, extrinsic camera parameters, neural networks, random forest, gradient boosting.

1. Introduction

The purpose of camera calibration is to determine the intrinsic and extrinsic calibration parameters of surveillance video camera. The classical formulation of this problem assumes the presence of calibration object with a known geometry in the scene. An example of such a calibration object can be a chessboard, Fig. 1.

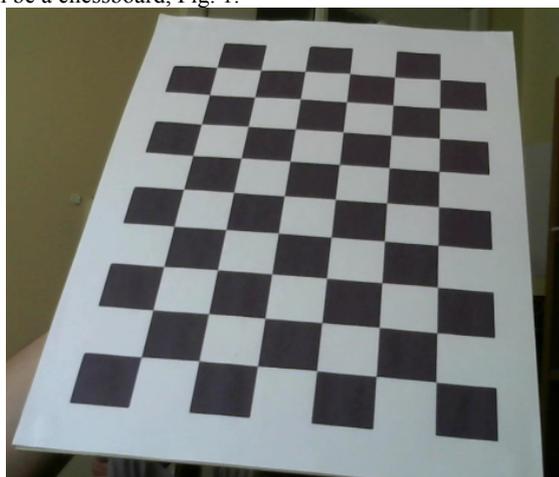


Figure 1. Camera calibration using special calibration object.

However, it is difficult to calibrate large amounts of cameras in video surveillance systems using the classical approach because it requires artificially putting a calibration object in the scene. Therefore, for video surveillance systems, automatic calibration, that is, analyzing objects in the scene (people, cars, buildings) is the most promising (Fig. 2).



Figure 2. Video surveillance footage.

Camera calibration is necessary to calculate the correspondences between 3D scene points and 2D points on the

image plane, which allows you to calculate the distances between objects, track changes in the size of objects, estimate the 3D position of a person on the ground [13], etc.

In addition, information about camera calibration parameters can be used to improve object detection and object tracking algorithms [11], for example, by filtering false detections, which are geometrically incorrect [12]. For example, person height depends on the camera position and orientation, so the information about camera calibration will allow finding regions of interest, where a person of a specific size can be located, which can increase the accuracy of object detection and reduce the computation time. Such filtering can be used alongside with other methods that use some other criteria, for example, estimation temporal consistency of detections [10].

In this paper, the problem of determining the following parameters is considered: camera height above the ground, tilt angle and roll angle (Fig. 3). The baseline method is based on convolutional neural network [14]. In this paper, the modification of the architecture of the baseline neural network will be proposed. In addition, two methods based on other machine learning algorithms will be studied in order to understand the necessity of using neural network to solve the problem of camera calibration. As input data, head bounding boxes and camera focal length are used.

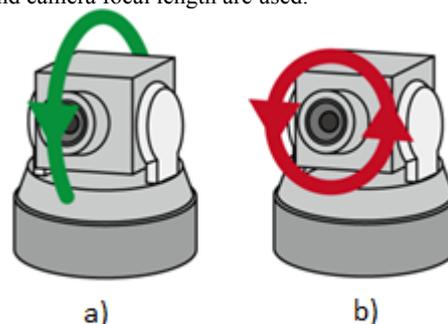


Figure 3. a) Tilt angle; b) Roll angle.

2. Related work

The existing approaches to the automatic calibration of surveillance video camera can be divided into two groups:

1. Analysis of the lines in the image.
2. Analysis of the objects' size.

The main idea of the first approach is to calculate and then analyze vanishing points and vanishing lines. Vanishing point is a point in the image plane where the projections of a set of parallel lines in space intersect. Each set of parallel lines in space defines its vanishing point. The set of such vanishing points is called a vanishing line, for example, horizon line.

This approach is used in articles [1, 2, 4, 6, 7, 9]. In articles [2, 7, 9] people are used as calibration objects. In the article [2] the method is based on calculating vertical vanishing point, which is specified by the position of feet and head of a person, as well as calculating vanishing point obtained by tracking person in several frames. Authors of the paper [9] consider the problem of calibrating video system, which consists of two cameras. Vanishing points and vanishing lines are computed by RANSAC/EM algorithm [8]. In the article [7], authors use some additional information on the distribution of human heights in the real world. Camera calibration parameters are found by maximizing the log likelihood function under the assumption that 90% of people in the image have height that differs from the average by less than 7.6%.

In [4], cars are used as calibration objects. However, there are some limitations on the scene conditions, which is necessary for calculating vanishing lines: essential part of car trajectory should be a straight line.

In papers [1, 6], buildings are used as calibration objects. In [6], authors proceed from the assumption that there are three orthogonal planes in the scene, so it is possible to calculate three vanishing points. In [1] authors consider the sequence of images of urban environment. Images should be overlapping so that for each pair of adjacent frames it is possible to find some feature point seen in both images.

The second approach is used in [3, 5, 14]. In [3] authors proposed the method for estimating camera calibration parameters for video surveillance system, which consists of several cameras with not overlapping viewing zones. The camera focal length and the tilt angle are obtained by analyzing the distribution of human heights depending on the feet location. However, additional information about person height in real world is necessary to determine the camera height above the ground.

In [14] convolutional neural network is used to estimate calibration parameters using head bounding boxes and camera focal length. For training synthetic data was used.

The main drawback of most methods is instability to changes of the scene or calibration object. The neural network method from [14] is chosen as baseline method, as the most promising and stable to changes of the scene. In this paper, its modification and some other methods based on random forest and gradient boosting will be proposed.

3. Proposed method

The input of the algorithm is a set of head bounding boxes $\{R_i\}_{i=1}^{N_j}$, and focal length f_j , where R_i is a head bounding box, which is given in image coordinates by three numbers: $R_i = \langle x_i, y_i, scale_i \rangle$, where (x_i, y_i) is the coordinates of top left corner of bounding box, $scale_i$ is the size of bounding box (assuming that the human head can be enclosed in square); N_j is the number of head bounding boxes for camera position in space K_j ; for each unique camera position there is a set of bounding boxes.

The output of the algorithm is a triple of parameters characterizing the position of camera in space K_j :

$$K_j = \langle \theta_j, \varphi_j, h_j \rangle,$$

where θ_j is a tilt angle in radians, $\theta_j \in \left[0, \frac{5\pi}{12}\right]$,

φ_j is a roll angle in radians, $\varphi_j \in \left[-\frac{\pi}{12}, \frac{\pi}{12}\right]$,

h_j is a camera height above the ground in meters, $h_j \in [0, 20]$.

3.1 Baseline method

Baseline method based on convolutional neural network was proposed in [14]. The problem of the lack of training data was solved by constructing a synthetic dataset, which will be described in section 4.1. The general scheme of this algorithm:

1. Preparation of input data.

At this step, bounding boxes $\{R_i\}_{i=1}^{N_j}$ are grouped by 64 observations for each unique camera position K_j and focal length f_j in matrices M_j of dimension $3 \times 8 \times 8$. Within each group, the bounding boxes are sorted by their sizes $scale_i, i = 1, \dots, 64$ in ascending order. Thus, convolutional layers can analyze both boxes that are close in sizes and far apart, that is, track the changes in the object size depending on its position relative to the camera.

2. Training neural network with the architecture shown in Fig. 4.

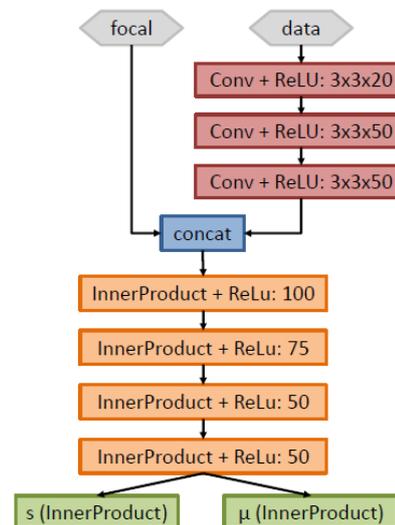


Figure 4. Neural network architecture of baseline method.

2.1. Two data input layers: focal length f_j and matrices M_j from step 1.

2.2. Convolutional layer with ReLU activation function and parameters: number of filters is 20; kernel size is 3×3 .

2.3. Convolutional layer with ReLU activation function and parameters: number of filters is 50; kernel size is 3×3 .

2.4. Convolutional layer with ReLU activation function and parameters: number of filters is 50; kernel size is 3×3 .

2.5. Concatenation layer, which concatenates the output of the last convolutional layer and input focal length layer.

2.6. Fully connected inner product layer with ReLU activation function and 100 neurons.

2.7. Fully connected inner product layer with ReLU activation function and 75 neurons.

2.8. Fully connected inner product layer with ReLU activation function and 50 neurons.

2.9. Fully connected inner product layer with ReLU activation function and 50 neurons.

2.10. Two output layers: mathematical expectation of parameters μ and logarithm of variance of normal distribution s .

Neural network was trained by minimizing the following loss function:

$$L(y|\mu, s) = -\log N(y|\mu, \text{diag}(e^s) + \varepsilon),$$

where $y = \langle t_{gt}, r_{gt}, h_{gt} \rangle$ are ground truth values of parameters (t is for tilt, r is for roll, h is for height);

$\mu = \langle t_{predicted}, r_{predicted}, h_{predicted} \rangle$ are predicted values of parameters;

$s = \langle s_t, s_r, s_h \rangle$ are logarithms of variance of normal distribution of parameters, $\text{diag}(e^s)$ – diagonal matrix of dimension 3×3 ; $N()$ – normal distribution density function;

$\varepsilon = 10^{-6}$ is used to prevent overfitting to a single train sample.

3.2 Modifications of baseline method

Since the input data in the baseline method of automatic calibration of surveillance video camera is a set of head bounding boxes and camera focal length and not the frames of the video sequence itself, the question about using convolutional layers in neural network architecture arises. Despite the sorting of bounding boxes by their sizes during the construction of matrices M_j , due to the wide variety of locations (and sizes) of human head in the image, it is not guaranteed that constructed matrices contain local features that could be effectively recognized by the convolutional filters. Thus, a modification with the following architecture of neural network was proposed (Fig. 5):

1. Convolutional layers were removed.
2. Fully connected inner product layer with ReLU activation function and 150 neurons was added.
3. Three output layers: one for each of the estimated parameters.
4. As a loss function, Euclidean distance is used.

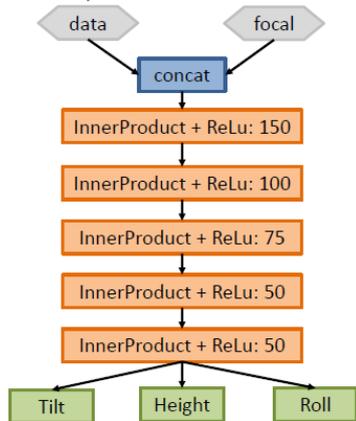


Figure 5. Neural network architecture of proposed method.

In addition, the neural network without focal length input layer was trained, because this information is not always available.

3.3 Other methods based on other machine learning algorithms

It was decided to develop and evaluate methods based on other machine learning algorithms in order to understand the necessity of using neural networks to solve the problem of automatic camera calibration.

In this paper, two methods were proposed: first method is based on random forest, second method is based on gradient decision tree boosting.

As input data, vector of 193 elements (64 observations of bounding boxes plus focal length) was used.

A method based on random forest: three models were trained, each has 60 tree with a depth of 18).

A method based on gradient decision tree boosting: three models were trained, each consists of 300 decision trees with a depth of 9.

4. Results

4.1 Datasets

To evaluate proposed methods two datasets were used:

1. Synthetic dataset.
2. TownCenter¹ dataset.

Synthetic dataset was generated using HumanShape² and neural network head detector FasterHog³:

¹http://www.robots.ox.ac.uk/~lav/Research/Projects/2009bбенf old_headpose/project.html

²<https://github.com/e-sha/pyhumanshape>

³<https://github.com/e-sha/fasterhog>

1. About 300000 samples were generated, each containing 64 observations of human head.
2. Samples were generated for camera calibration parameters if following ranges:
 $tilt \in [0, 75^\circ]$; $roll \in [-15^\circ, 15^\circ]$; $height \in [0, 20]$.
3. Outliers were artificially added to generated samples to emulate the incorrect detections of human head.
4. 80% of synthetic data were used for training purposes.

An example of the generated synthetic image is shown in Figure 6. A detail description of constructing synthetic dataset can be found in [14].



Figure 6. Image from the synthetic dataset.

Second dataset contains footage from real video surveillance camera. This TownCenter dataset contains 4500 frames, obtained from a 5-minutes video sequence. About 70000 head bounding boxes are marked on these frames. For the evaluation, 40000 samples were constructed (64 observations each). An example from a TownCenter dataset is shown in Fig. 7.



Figure 7. Image from TownCenter dataset.

4.2 Quality metrics

L1 metric was used to evaluate the quality of developed algorithms.

$$d(label, score) = \frac{1}{N} \sum_{j=1}^N |label_j - score_j|,$$

where $label_j$ is a true value of a parameter for the j sample, $score_j$ is a predicted value of a parameter for the j sample, N is a number of samples.

This metric is chosen because it allows easy interpretation of the results.

4.3 Experimental evaluation

The results of the experimental evaluation of synthetic dataset are given in Table 1.

All proposed algorithms (except the neural network based method without using focal length information) showed good quality on synthetic dataset in comparison with the baseline one. The best result for the tilt and roll angles was demonstrated by a modified neural network algorithm. The camera height

above the ground was best determined by method based on gradient decision tree boosting.

The results on TownCenter dataset are given in Tables 2 and 3. Table 2 shows the average error on all test samples. Table 3 shows the error for estimated camera calibration parameters, obtained by averaging over all samples.

On TownCenter dataset proposed neural network algorithm showed an advantage over the baseline one and all other proposed algorithms on all three estimated parameters. The fact that the results on TownCenter dataset are better than on synthetic dataset can be explained by the fact that TownCenter dataset does not contain incorrect head observations.

Experimental evaluation shows the applicability and effectiveness of methods based on machine learning algorithms in automatic calibration of surveillance video camera problem. However, the best result was shown by the modification of baseline neural network method. The proposed neural network method on synthetic dataset showed results that exceed results of the baseline algorithm by about 1.7 times, which confirms the hypothesis from section 3.2 that a wide variety of locations of human head in the image does not allow to use convolutional layers efficiently.

Table 1. Average error on synthetic dataset.

Estimated parameter	Baseline method	Proposed neural network method (with focal length)	Proposed neural network method (without focal length)	Proposed random forest based method	Proposed DT-boosting based method
Tilt angle, radians	0.084	0.04984	0.1096	0.0859	0.0555
Roll angle, radians	0.0925	0.0553	0.0575	0.0764	0.0586
Camera height, metres	0.8177	0.5226	1.1008	0.6140	0.4904

Table 2. Average error on TownCenter dataset.

Estimated parameter	Baseline method	Proposed neural network method (with focal length)	Proposed neural network method (without focal length)	Proposed random forest based method	Proposed DT-boosting based method
Tilt angle, radians	0.0196	0.0186	0.2591	0.0668	0.0298
Roll angle, radians	0.0398	0.0344	0.0487	0.0527	0.0367
Camera height, metres	0.6690	0.4692	0.6597	1.1187	0.8134

Table 3. Error for estimated camera calibration parameters on TownCenter dataset.

Estimated parameter	Baseline method	Proposed neural network method (with focal length)	Proposed neural network method (without focal length)	Proposed random forest based method	Proposed DT-boosting based method
Tilt angle, radians	0.0157	0.0037	0.2600	0.0667	0.0006
Roll angle, radians	0.0377	0.0014	0.0432	0.0511	0.0315
Camera height, metres	0.5562	0.1935	0.5191	1.1004	0.7091

6. References

- [1] Bobkov V.A., Ronshin Y.I., Kudryashov A.P. Line identification using uncalibrated urban environments (In Russian) //Information Technology and Computer Systems. – 2007. - №.1. – p.63-71.
- [2] Chen T. et al. Accurate self-calibration of two cameras by observations of a moving person on a ground plane //Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on. – IEEE, 2007. – p. 129-134.
- [3] den Hollander R. J. M. et al. Automatic inference of geometric camera parameters and intercamera topology in uncalibrated disjoint surveillance cameras //SPIE Security+ Defence. – International Society for Optics and Photonics, 2015. – p. 96520D-96520D-15.
- [4] Dubska M., Herout A., Sochor J. Automatic Camera Calibration for Traffic Understanding //BMVC. – 2014.
- [5] Hoiem D., Efros A. A., Hebert M. Putting objects in perspective //International Journal of Computer Vision. – 2008. – T. 80. – №. 1. – p. 3-15.
- [6] Li B. et al. Simultaneous vanishing point detection and camera calibration from single images //International

5. Conclusion

Compared with the existing methods, the machine learning based approach to solve automatic camera calibration problem has greater resistance to changes in the scene conditions and can be applied to different types of objects. The problem of the lack of training data can be solved using synthetic dataset.

In this article, the modification of the baseline [14] neural network method was proposed. As input data head bounding boxes as long as camera focal length were used. Some modifications of the architecture of the neural network were proposed. The experimental evaluation of the proposed method showed that it is more efficient than the baseline one.

In addition, other methods based on random forest and gradient boosting were proposed in order to evaluate the necessity of using neural network to solve the problem of automatic camera calibration. The experimental evaluation of these methods showed their applicability, however, the best results on both synthetic and TownCenter datasets were shown by the proposed neural network method.

- Symposium on Visual Computing. – Springer Berlin Heidelberg, 2010. – p. 151-160.
- [7] Liu J., Collins R. T., Liu Y. Surveillance camera autocalibration based on pedestrian height distributions //Proceedings of the British Machine Vision Conference. – 2011. – p. 144.
- [8] Pflugfelder R., Bischof H. Online auto-calibration in man-made worlds //Digital Image Computing: Techniques and Applications, 2005. DICTA'05. Proceedings 2005. – IEEE, 2005. – p. 75-75.
- [9] Pflugfelder R., Bischof H. People tracking across two distant self-calibrated cameras //Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on. – IEEE, 2007. – p. 393-398.
- [10] Sergeev A. E., Konushin A. S., Konushin V. S. Suppression of false positive detections in video streams of video surveillance systems (In Russian) // Computer Optics. — 2016. — T. 40, № 6. — p. 958–967
- [11] Shalnov E. V., Konushin V. S., Konushin A. S. An improvement on an MCMC-based video tracking algorithm //Pattern Recognition and Image Analysis. – 2015. – T. 25. – №. 3. – p. 532-540.
- [12] Shalnov E. V., Konushin A. S. Increasing accuracy of detectors using scene geometry (In Russian)// Software product and systems. — 2017. — T. 30, № 1. — p. 106–111.
- [13] Shalnov E. V., Gringauz A. D., Konushin A. S. Estimation of the people position in the world coordinate system for video surveillance // Programming and Computer Software. — 2016. — Vol. 42, no. 6. — p. 361–366
- [14] Shalnov E. V., Konushin A. S. Convolutional neural network for camera pose estimation from object detections // International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. – 2017. – T. 42.

About authors

Anton S. Konushin, PhD, associate professor at the Department of Computing Systems and Automation and head of the Graphics and Media Lab at the Faculty of Computational Mathematics and Cybernetics at Lomonosov Moscow State University, associate professor at the Big Data and Information Retrieval School at the Faculty of Computer Science at Higher School of Economics.

Email anton.konushin@graphics.cs.msu.ru.

Evgeny V. Shalnov, junior researcher at the Department of Computing Systems and Automation at the Faculty of Computational Mathematics and Cybernetics at Lomonosov Moscow State University.

Email eshalnov@graphics.cs.msu.ru.

Iana A. Valuiskaia, graduated from the Faculty of Computational Mathematics and Cybernetics at Lomonosov Moscow State University in 2017 (master degree).

Email yana.valuyskaya@graphics.cs.msu.ru.

Распознавание и классификация микроструктуры металлов и сплавов с использованием глубоких нейронных сетей

Г.В. Аристов¹, А.В. Клюев¹
kav-1@bk.ru|chacknoris128@ya.ru

¹Пермский национальный исследовательский политехнический университет, Россия

Исследуются возможности искусственных нейронных сетей третьего поколения в задачах распознавании микроструктуры материалов. Разработка новых методов распознавания микроструктуры металлов и сплавов является актуальной задачей для создания функциональных материалов. Новые методы открывают путь к построению высокоэффективных методов прогнозирования их физико-механических свойств. В работе предложена архитектура свёрточной нейронной сети (относится к сетям третьего поколения), способной к распознаванию элементов микроструктуры. Обучение и апробация сети выполнена на задаче распознавания наличия дендритных структур на снимках микрошлифов металлов. Достигнуты положительные результаты с точностью более 90%. Высокая достоверность полученных результатов говорит о перспективности применения нейронных сетей третьего поколения в задачах распознавания и классификации микроструктур металлов и сплавов.

Ключевые слова: распознавание, микроструктура материала, анализ микрошлифов, функциональные материалы.

Recognition and classification of microstructure of metals and alloys with the use of deep neural networks

G.V.Aristov¹, Klyuev A.V.¹
kav-1@bk.ru|chacknoris128@ya.ru

¹Perm State National Research Polytechnic University, Russia

The possibilities of artificial third generation neural networks in problems of microstructure material recognition are explored. The development of new methods for recognizing the microstructure of metals and alloys is an actual task for the creation of functional materials. New methods open the way to building highly effective methods for predicting their physico-mechanical properties. The architecture of a convolutional neural network (refers to third generation networks) capable of recognizing microstructure elements is proposed. Training and approbation of the network is performed on the task of recognizing the presence of dendritic structures on the images of microsections of metals. Positive results were achieved with an accuracy of more than 90%. High reliability of the obtained results testifies to the prospects of using third generation neural networks in problems of recognition and classification of microstructures of metals and alloys.

Keywords: recognition, microstructure of the material, analysis of microsections, functional materials.

1. Введение

Микроструктура металлов и сплавов оказывает существенное влияние на термомеханические свойства изделий и конструкций, изготовленных из них. Именно поэтому выяснение микроструктуры металлических изделий и конструкций при производстве и эксплуатации является одной из актуальных задач.

В настоящее время существует три основных метода проведения контроля качества материала: экспертный анализ, лабораторный контроль, неразрушающий контроль. У первых двух методов есть существенные недостатки, хотя они являются наиболее широко распространенными на практике. Кроме значительных затрат времени и средств на реализацию, лабораторные исследования могут нести частичное разрушение образцов изделий в ходе эксперимента. Недостаток способов оценки материала с помощью экспертного анализа заключается в возможных значительных ошибках, определенных субъективными мнениями экспертов.

Основными методами неразрушающего контроля качества являются: магнитный, электрический, вихретоковый, радиоволновой, тепловой, визуально-оптический, радиационный, акустический и контроль проникающими веществами.

Исследование особенностей методов неразрушающего и разрушающего контроля качества показывает, что результатами большинства этих методов являются снимки (изображения) объектов и их дефектных областей [11].

Поэтому современные методы и алгоритмы обработки цифровых изображений и технологии искусственных нейронных сетей могут быть успешно применены для автоматизации анализа снимков, получаемых посредством различных методов контроля качества.

2. Выбор метода распознавания

В качестве метода анализа цифровых снимков микрошлифов металла была выбрана свёрточная нейронная сеть. Свёрточные нейронные сети (СНС, Convolutional neural network (CNN)) - это специальная архитектура искусственных нейронных сетей (ИНС) третьего поколения, нацеленная на эффективное распознавание изображений. Такие сети принято называть глубокими в связи с наличием в них большого количества слоев. Например, в работе [5] упоминается исследование сверхглубокой сети, в которой более 1000 слоев. Большие успехи глубоких сетей в задачах классификации [3,5] и семантической сегментации [2,4] широко известны, но, к сожалению, авторам не известны работы, в которых CNN применяется для классификации микроструктуры в металлах и сплавах. Тем не менее, неглубокие нейронные сети второго поколения уже использовались ранее в таких задачах анализа микроструктуры как задачи определения объемного содержания фазы, анализа зерна, выявление дефектов сварных соединений [7,10,1].

Основная идея СНС заключается в том, что при обучении СНС на основе исследуемых изображений одного класса, она самостоятельно формирует наборы признаков

(так называемые карты признаков), которые в полной мере характеризуют этот самый класс, и в то же время, отделяют их от других изображений [6]. Всё это реализуется в сети, с помощью так называемых свёрточных слоёв, которые составляют карты признаков изображения, и слоёв подвыборки, которые уменьшают масштаб изображения. Чередование данных слоёв позволяет составить карты признаков, с помощью которых изображения классифицируются.

Данная технология была выбрана в связи с относительно низкой чувствительностью к искажениям входных сигналов по сравнению с другими методами распознавания изображений. Кроме того, авторы учли, что СНС показывает высокую степень точности в разнообразных задачах распознавания образов [9,13,12].

3. Архитектура сети и обучение

В рамках исследования была разработана и реализована система распознавания дендритных микроструктур на цифровых снимках микрошлифов. В системе распознавания была реализована модель свёрточной нейронной сети (рис. 1). Структура сети характеризовалась порядком расположения свёрточных слоёв и слоёв подвыборки.

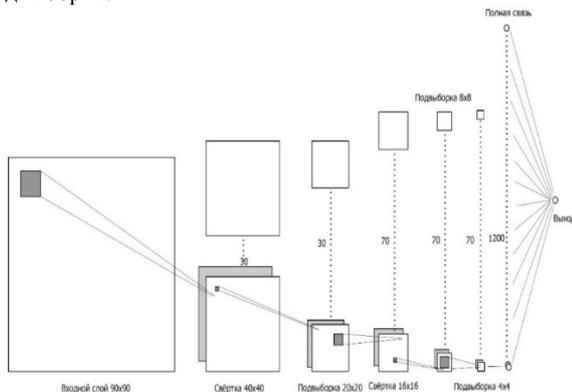


Рис. 1. Архитектура свёрточной нейронной сети.

Входной слой служит для подачи изображения в сеть. Размер входного слоя 90×90=8100 нейронов, что соответствует размеру исследуемого изображения в пикселях. Далее идут 2 слоя свертки и 3 слоя подвыборки. Седьмой слой – полносвязный слой. Данный слой служит для обеспечения классификации, после того как выделены все карты признаков. Каждый нейрон этого слоя полностью связан только с одним признаком предыдущего слоя. На выходе слоя 1200 простых сигмоидальных нейронов. Последний слой является выходным и состоит из одного нейрона, полностью связанного со всеми нейронами предыдущего слоя. В сумме сеть составила 92821 нейронов и 6332400 связей между ними.

Значение на выходе нейрона свёрточного слоя вычисляется с помощью выражения:

$$y_k^{(i,j)} = f \left(\sum_{s=1}^K \sum_{t=1}^K \sum_{k=1}^L b_k + w_k^{(s,t,k)} x^{((i-1+s),(j-1+t),k)} \right)$$

$$i = \overline{1, w_{out}}, j = \overline{1, h_{out}}, s, t = \overline{1, K},$$

где $y_k^{(i,j)}$ – выходной сигнал (нейрон), $w_k^{(s,t,k)}$ – веса, матрица синаптических коэффициентов рецептивных плоскостей всех карт текущего слоя, $x^{(s,t,k)}$ – входной сигнал, область входной плоскости, покрываемая синаптической маской, K – размер рецептивной плоскости (синаптического фильтра), b_k – порог, нейронное смещение k -ой карты признаков – отражает увеличение или уменьшение входного сигнала подаваемого на функцию

активации, k – номер карты признаков, L – общее количество карт признаков в слое.

Значение на выходе нейрона слоя подвыборки вычисляется с помощью выражения:

$$y_k^{(i,j)} = \max \{ x^{(i+sq_a, j+sq_b)} \}, \quad i = \overline{1, w_{out}}, j = \overline{1, h_{out}}, a, b = \overline{1, 2}$$

где $y_k^{(i,j)}$ – выходной сигнал (нейрон) k -ой карты признаков, $sq_a = (i, i - 1)$, $sq_b = (j, j - 1)$ – векторы с координатами элементов входной плоскости, $x^{(*)}$ – входной сигнал, область входной плоскости, покрываемая синаптической маской.

Механизм обучения СНС называется «обучение с учителем», что подразумевает следующее: на вход СНС вместе с обучающим множеством подается правильный ответ распознавания. Таким образом, СНС с каждым новым элементом обучающего множества, корректирует весовые коэффициенты, исходя из разницы значений между фактическим ответом и тем, который дал «учитель». Для измерения качества распознавания используется функция среднеквадратичной ошибки. Корректировка весовых коэффициентов происходит с помощью метода обратного распространения ошибки. Начальные значение весовых коэффициентов были выбраны случайными величинами с нормальным распределением.

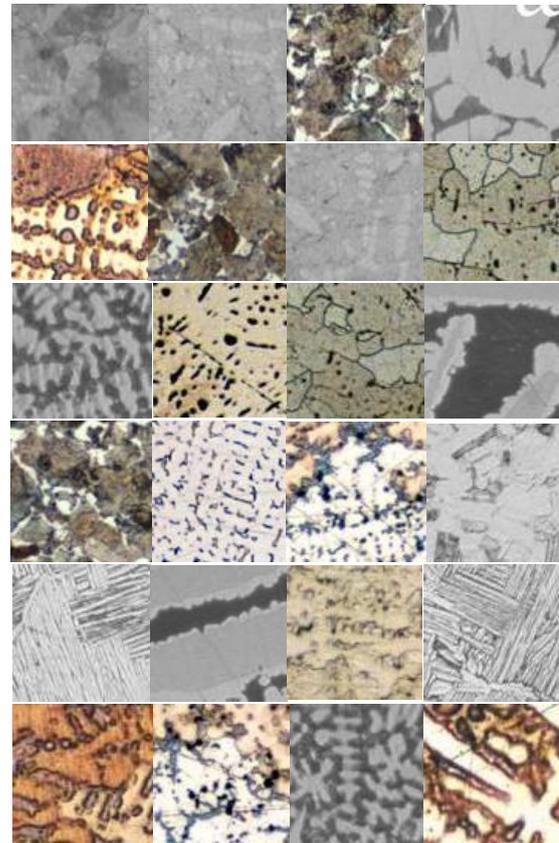


Рис. 2. Примеры снимков обучающей выборки.

В качестве инструмента использовался фреймворк Caffe. Перед обучением сети ее веса задавались случайным образом. Предобученная сеть не использовалась, ввиду особенности рассматриваемой задачи.

Обучение сети было выполнено с помощью обучающего множества объемом 3122 изображений. На рис.2 представлены характерные снимки из базы обучающего множества. Например, на снимке, расположенной в последнем ряду вторым справа, хорошо различима дендритная структура. Обучающее множество

было сформировано на основе микрофотографий проекта DoITPoMS (Dissemination of IT for the promotion of Materials Science) Кембриджского университета. Тестовый набор составил 500 изображений. Для того чтобы увеличить обучающие множество и точность распознавания, над оригинальными изображениями шлифов проводились различные операции обработки: зашумление, отражение, искажение.

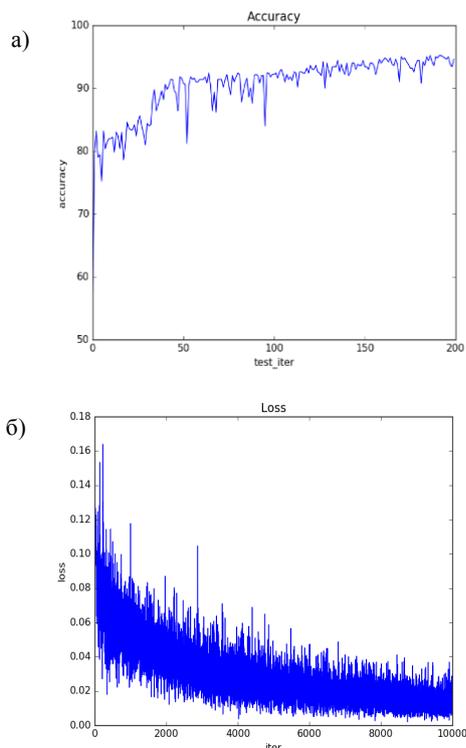


Рис. 3. Процесс обучение СНС: а) изменения точности, б) изменения функции ошибки.

Обучение сети длилось более 10000 итераций. В результате обучения сети из изображений были выделены карты признаков на каждом слое свёртке. Расчёт точности приводился каждые 50 итераций и равен проценту верно распознанных изображений из тестовой выборки. На рис.3,а представлен график изменения точности в процессе обучения. Значение 200 по оси абсцисс эквивалентно 10000-ам итераций. На момент окончания обучения точность распознавания сети достигла значения 94.6%. На рис.3,б представлена зависимость изменения функции ошибки от номера итерации.

Полученную точность классификации можно сравнить с точностью методов решения близкой задачи – задачи классификации текстур (распознавания поверхности материалов). В работе [8] исследуются несколько современных детекторов, включая многомасштабные, на известных наборах данных KTH-TIPS2 и FMD. Максимальная точность, которую удается достичь с помощью предложенного авторами детектора S3C (Spike-and-Slab Sparse Coding), составляет не более 72%.

4. Заключение

В рамках исследования была разработана структура свёрточной нейронной сети, произведено ее обучение для распознавания дендритной микроструктуры. Получены результаты распознавания на тестовой выборке снимков. Высокая достоверность полученных результатов говорит о перспективности применения СНС в задачах классификации микроструктур.

5. Поддержка

Исследование проводилось при финансовой поддержке Министерства образования и науки РФ в рамках проекта RFMEFI58617X0055.

6. Литература

- [1] Gang Wang, T. Warren Liao Automatic identification of different types of welding defects in radiographic images // NDT&E International. 2002. 35. pp 519–528.
- [2] Sithichok Chaichulee, Mauricio Villarroel, João Jorge, Carlos Arteta, Gabrielle Green, Kenny McCormick, Andrew Zisserman, Lionel Tarassenko, Multi-Task Convolutional Neural Network for Patient Detection and Skin Segmentation in Continuous Non-Contact Vital Sign Monitoring // Automatic Face & Gesture Recognition (FG 2017) 2017 12th IEEE International Conference on, 2017. pp. 266-272
- [3] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D, Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. P.1-9
- [4] Jonathan Long, Evan Shelhamer, Trevor Darrell Fully convolutional networks for semantic segmentation // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. pp. 3431-40
- [5] K. He, X. Zhang, S. Ren, J. Sun Deep Residual Learning for Image Recognition// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) — DC, USA: IEEE Computer Society Washington, 2016 — С. 2342—2354.
- [6] Le Cun Y., Bengio Y. Convolution networks for images, speech and time series // The handbook of brain theory and neural networks. 1998. V. 7. № 1. pp. 225 – 258.
- [7] Victor Hugo C de Albuquerque, Auzuir Ripardo de Alexandria, Paulo César Cortez, João Manuel RS Tavares, Evaluation of multilayer perceptron and self-organizing map neural network topologies applied on microstructure segmentation from metallographic images // NDT & E International, Vol.42(7), 2009, pp. 644-651.
- [8] Wenbin Li, Mario Fritz Learning Multi-Scale Representations for Material Classification // German Conference on Pattern Recognition, 2014, pp. 757-764
- [9] Друки А.А. Применение свёрточных нейронных сетей для выделения и распознавания автомобильных номерных знаков на изображениях со сложным фоном // Интеллектуальные системы. Известия Томского политехнического университета. 2014. Т. 324. № 5 С. 85 – 92.
- [10] Емельянов В.А. Моделирование нейронных сетей распознавания металлографических изображений для диагностики состояния сталей // Электротехнические и компьютерные системы. № 12 (88), 2013, С. 125-131.
- [11] Ермаков А.А., Орлов А.А., Садыков С.С., Стародубов Д.Н. Методы и алгоритмы обработки и анализа дефектоскопических и металлографических снимков. Владимир: Изд-во Владимир, гос. ун-та, 2008, 112 с.
- [12] Макаренко А. А., Калайда В.Е. Методика локализации изображения лица для систем видео контроля на основе нейронных сетей // Известия Томского политехнического университета. 2006. Т. 309. № 8. С. 113 – 118.
- [13] Солдатова О.П., Гаршин А.А. Применение свёрточных нейронных сетей для распознавания рукописных цифр // Компьютерная оптика том 34. № 2. 2010. С.252 – 259.

Об авторах

Клюев Андрей Владимирович, к.ф.-м.н., доцент кафедры «Вычислительная математика и механика» Пермского национального исследовательского политехнического университета. Его e-mail: kav-1@bk.ru.

Аристов Григорий Валерьевич, магистр, кафедра «Вычислительная математика и механика» Пермского национального исследовательского политехнического университета. Его e-mail: chacknotis128@ya.ru.

Выделение связного переднего плана на картах глубины с помощью медиального представления сцены

Д.С. Белобородов, Л.М. Местецкий
dmitribeloborodov@yandex.ru|mestlm@mail.ru
МГУ им. М.В. Ломоносова, Москва, Россия

В данной статье рассматривается задача выделения переднего плана на картах глубины. Предлагается подход к выделению переднего плана, основанный на медиальном представлении силуэтов объектов сцены. Поиск минимального разреза в графе медиального представления позволяет выделять связные области, отвечающие отдельным объектам переднего плана. Предложенный метод работает в режиме реального времени, не требует обучения и является устойчивым относительно движений камеры и объектов. Работа метода демонстрируется на картах глубины, полученных с помощью устройства Microsoft Kinect v2.

Ключевые слова: карты глубины, передний план, медиальное представление, сегментация, Kinect.

Connected foreground detection on depth maps using medial axes of the scene

D.S. Beloborodov, L.M. Mestetskiy
dmitribeloborodov@yandex.ru|mestlm@mail.ru
Lomonosov MSU, Moscow, Russia

This article considers the problem of foreground detection on depth maps. Proposed approach to foreground detection is based on medial representation of scene objects silhouettes. Minim cut calculation in medial axes graph allows to detect connected components responsible for separate foreground objects. Proposed method works in real time, it does not require learning procedures and is invariant to camera and objects movement. The method is demonstrated on depth maps obtained from Microsoft Kinect v2 device.

Keywords: depth maps, foreground, medial axes, segmentation, Kinect.

Выделение переднего плана и объектов интереса на изображениях и видео является одной из важных задач в областях компьютерного зрения, отслеживания и распознавания объектов, распознавания жестов и многих других. Эта задача заметно усложняется в условиях пестрого или сложно структурированного фона, при движении камеры, а также при наличии окклюзий среди объектов.

Карты глубины представляют собой массивы, хранящие информацию о расстоянии до камеры для точек сцены. Такая информация позволяет искать передний план в терминах взаимного расположения объектов сцены. На данный момент существуют устройства, позволяющие получать в реальном времени не только цветные изображения, но и соответствующие карты глубины сцен, например, Microsoft Kinect.

Известные подходы к выделению объектов интереса с использованием карт глубины включают:

Построение модели фона по нескольким подряд идущим картам глубины. Пиксели, которые плохо объясняются моделью фона, считаются передним планом. Такие подходы применяются, например, в статьях [4, 6]. Недостаток этого подхода заключается в том, что передним планом считаются только подвижные объекты, а метод чувствителен к движению камеры.

Подходы, основанные на вычислении потока сцены, например [9, 10], обладают теми же недостатками.

Другой подход — выделение в сцене плоскости пола либо полная сегментация сцены на планарные об-

ласти с помощью алгоритма RANSAC [7]. Этот подход предполагает наличие в сцене планарных областей, то есть, накладывает сильные ограничения на структуру сцены. Например, такой подход применяется в статье [3].

Также для сегментации сцены и выделения объектов интереса применяются различные классификаторы, например, глубинные нейронные сети используются в статье [8]. Такой подход является вычислительно сложным и требует большого объема размеченных данных.

В предыдущей работе [1] был предложен метод выделения переднего плана на картах глубины, основанный на анализе медиального представления силуэтов сцены. Он использует единственную карту глубины в качестве входных данных и не требует обучения, а также инвариантен к движению камеры и не накладывает ограничений на структуру объектов сцены. Данная статья является развитием предыдущей работы. Предложен новый способ определения принадлежности узлов к переднему плану, устойчивый к пропускам в картах глубины, которые появляются из-за особенностей работы сенсоров. Также используется поиск минимального разреза в графе медиального представления для выделения связных компонент переднего плана.

Работа метода демонстрируется на примере карт глубины, полученных с помощью сенсора Microsoft Kinect v2.

1. Определение переднего плана

В данной статье используется то же определение переднего плана, что и в предыдущей работе [1], основанное на понятии медиального представления.

1.1. Медиальное представление

Скелетом плоской фигуры называется множество всех центров максимальных вписанных в эту фигуру кругов. Радиальная функция ставит в соответствие каждой точке скелета радиус вписанного в фигуру круга с центром в этой точке. Совокупность скелета фигуры и ее радиальной функции называется медиальным представлением. Для многоугольных фигур скелет представим в виде плоского графа, узлы которого являются центрами некоторых вписанных кругов, а ребра являются отрезками прямых или парабол. Медиальное представление также может быть построено для бинарного изображения, в этом случае границы областей на изображении приближаются многоугольниками. Алгоритмы построения медиального представления для бинарных изображений описаны в книге [2].

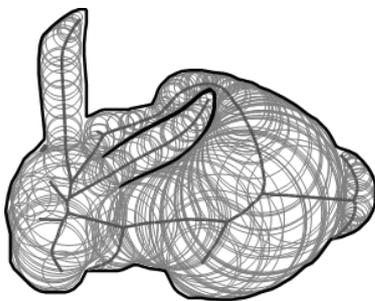


Рис. 1. Пример медиального представления фигуры.

Обозначим как $Z(p)$ глубину в точке p , заданную картой глубины.

Рассмотрим некоторый вписанный круг C узла скелета с центром в точке (x_c, y_c) и радиусом R . Построим второй круг C_ϵ , концентричный с кругом C и имеющий радиус $R + \epsilon$, $\epsilon > 0$. При таком построении границы фигуры разобьют C_ϵ на несколько фигур A_0, \dots, A_n : $\bigcup_{i=0}^n A_i = C_\epsilon$. При этом только одна из областей A_i будет полностью содержать в себе C , поскольку этот круг вписанный. Пусть это область A_0 : $C \subset A_0$.

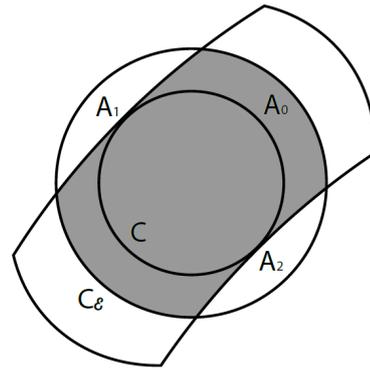


Рис. 2. Схема построений для определения переднего плана.

Определение 1. Будем говорить, что рассмотренный узел лежит на переднем плане с перепадом глубины d , если

$$\exists d > 0, \exists \epsilon > 0 : \max_{p \in C} Z(p) + d \leq \min_{p \in C_\epsilon \setminus A_0}$$

Это значит, что между вписанным кругом узла переднего плана и внешними областями, отсеченными границами фигуры, есть некоторый перепад глубины, причем круг находится ближе к камере.

2. Структура метода

Предложенный метод выделения переднего плана содержит три подзадачи и имеет следующую структуру:

- 1) Построение медиального представления сцены и разметка некоторых узлов:
 - выделение границ и бинаризация карты глубины;
 - построение медиального представления силуэтов сцены;
 - разметка части узлов медиального представления как передний или задний план.
- 2) Разметка оставшихся узлов медиального представления на передний и задний план с помощью минимального разреза.
- 3) Восстановление переднего плана в форме пиксельной маски.

Первые две подзадачи позволяют найти передний план в форме размеченного графа медиального представления, с помощью которого при необходимости можно вычислить пиксельную маску переднего плана для карты глубины.

3. Медиальное представление и первичная разметка

3.1. Бинаризация

На первом шаге на карте глубины вычисляются границы оператором Собеля, затем производится ее бинаризация по следующему правилу: черным (значение 0) помечаются все пиксели, лежащие на границе, а также

пиксели с неизвестной глубиной, остальные пиксели помечаются белым (значение 1). Это необходимо для разделения силуэтов объектов сцены.

Затем строится медиальное представление всех силуэтов сцены.

Порог, используемый при бинаризации границ, влияет на метод следующим образом: более низкие значения порога приводят к повышению детализации границы и медиального представления. При этом излишняя детализация может приводить к появлению шума и неустойчивости. Этот параметр можно подобрать эмпирически: таким образом, чтобы можно было различать близко находящиеся объекты при низком уровне шума.

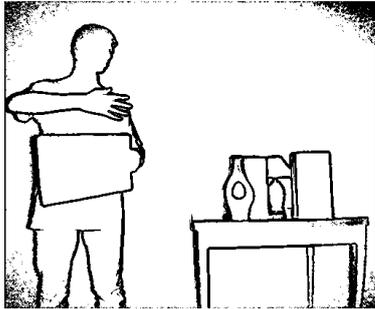


Рис. 3. Бинаризованная карта глубины.

3.2. Триангуляцией Делоне

называется триангуляция, для которой выполнено условие Делоне: круг, описанный вокруг любой ее грани, не содержит внутри себя точек триангуляции. Если никакие четыре точки множества не лежат на одной окружности, то триангуляция Делоне единственна. Для каждой точки ее ближайший сосед в евклидовой метрике будет соединен с ней некоторым ребром в триангуляции Делоне. Построение триангуляции Делоне на множестве всех узлов скелета позволяет задать систему соседства на множестве узлов скелета.

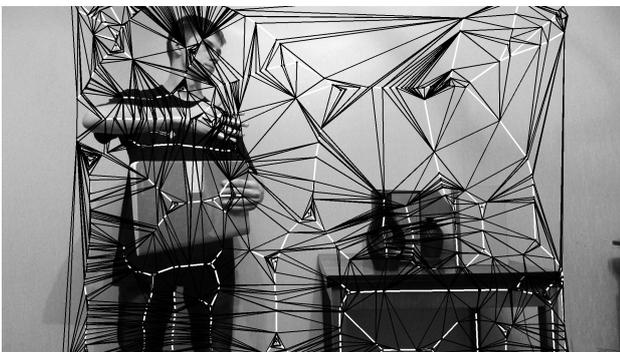


Рис. 4. Триангуляция Делоне на множестве узлов скелета.

3.3. Первичная разметка узлов

Рассмотрим два произвольных узла скелета C_1 , C_2 с центрами в точках p_1 , p_2 и радиусами R_1 , R_2 соот-

ветственно. Для этой пары узлов рассмотрим величины:

$\Delta Z = |Z(p_1) - Z(p_2)|$ — абсолютный перепад глубин между узлами; здесь $Z(p)$ — глубина точки p , задаваемая картой глубины;

$\Delta L = \|p_1 - p_2\| - R_1 - R_2$ — зазор между кругами (отрицательный, если круги пересекаются).

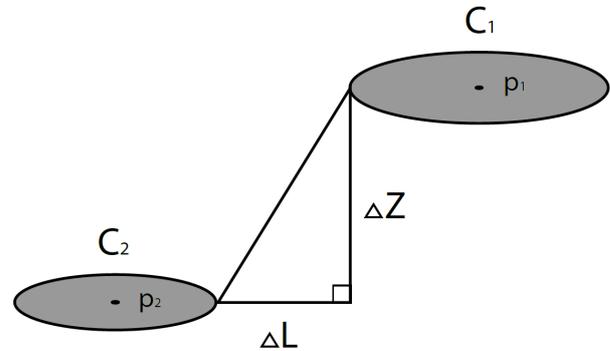


Рис. 5. Схема построений для первичной разметки узлов.

Определение 2. Будем говорить, что узел C_1 является передним планом относительно C_2 с параметрами $a > 0$, $b > 0$, если выполнены следующие условия:

$$\begin{cases} 0 < \Delta L \leq b \\ \frac{\Delta Z}{\Delta L} \geq a \\ Z(p_1) < Z(p_2) \end{cases}$$

При этом будем говорить, что узел C_2 является задним планом относительно C_1 . Если хотя бы одно из неравенств не выполнено, ни один из узлов не является передним/задним планом относительно другого.

Первое неравенство накладывает ограничения на зазор между вписанными кругами двух узлов: он должен быть положительным (круги не пересекаются) и не превосходящим параметра b . Необходимость анализировать круги, находящиеся на расстоянии друг от друга, возникает из-за возможности наличия пропусков (неопределенных значений) в карте глубины. Такие области неопределенных значений возникают из-за особенностей работы датчиков глубины и обычно располагаются по краям выступающих объектов — то есть, объектов переднего плана. Параметр b накладывает ограничения на максимальный зазор между кругами — ограничивает максимальную ширину области неопределенных значений вокруг объектов. Наблюдения показывают, что в датчиках глубины ширина таких областей действительно ограничена некоторым небольшим значением. Второе неравенство ограничивает отношение перепада глубин и зазора между кругами — то есть, регулирует крутизну перепада между ними. Если перепад достаточно крутой, круги будут являться передним/задним планом относительно друг друга. Третье неравенство означает, что узел C_1 находится ближе к камере, чем узел C_2 .

Определение 3. Будем говорить, что узлы C_1 и C_2 являются соседями, если они соединены некоторым ребром в построенной триангуляции Делоне, но при этом не соединены ребром в графе скелета сцены.

Если два узла соединены ребром в скелете сцены, то можно считать, что между ними нет значительного перепада глубины.

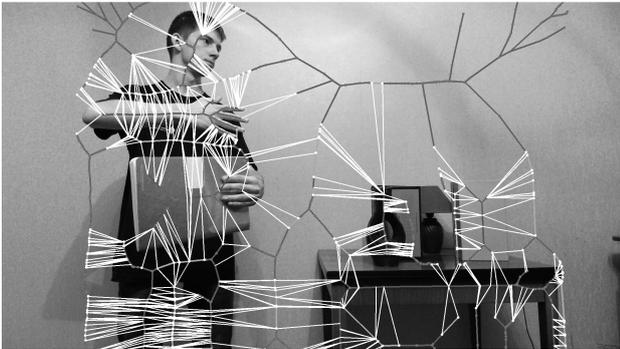


Рис. 6. Белым отмечены пары соседних узлов, являющихся передним/задним планом относительно друг друга.

Определение 4. Будем говорить, что узел C относится к переднему плану сцены, если он является передним планом относительно некоторых своих соседей и не является задним планом относительно всех своих соседей.

Будем говорить, что узел C относится к заднему плану сцены, если он является задним планом относительно некоторых своих соседей.

Используя это правило, узлы скелета сцены размечаются на передний план, задний план и узлы без меток.

4. Разметка оставшихся узлов

Каждый узел скелета соответствует некоторому вписанному в силуэт кругу — фрагменту объекта. На прошлом этапе некоторые фрагменты были помечены как передний план, а некоторые — как задний, но до этого момента не была использована свойство связности объектов сцены. При разметке часть узлов могла быть пропущена, поскольку все их соседи имеют маленький перепад глубины. Следующий шаг направлен на исправление разметки с использованием свойства связности объектов.

Построим потоковую сеть на основе скелета сцены: пусть граф $G_0 = (V_0, E_0)$ — скелет сцены (V_0 — вершины, E_0 — ребра). Построим на его основе орграф $G_1 = (V_1, E_1)$, где $V_1 = V_0$, а E_1 получается из E_0 заменой каждого ребра (u, v) на пару ориентированных ребер (дуг) (u, v) , (v, u) . Добавим вершины s, t — исток и сток соответственно. Далее все узлы переднего плана u_i соединим дугами (s, u_i) с истоком, а узлы заднего плана v_j соединим дугами (v_j, t) со стоком. Получим орграф (V, E) , $V = V_1 \cup \{s, t\}$, $E =$

$E \cup \{(s, u_i)|i\} \cup \{(v_j, t)|j\}$. Зададим пропускные способности всех дуг. Для дуг, идущих от стока или к стоку, определим пропускающую способность как некоторое большое число c_0 . Выбор пропускных способностей для остальных дуг будет влиять на выделяемый передний план. Если для дуг, инцидентных узлам переднего плана, задать пропускную способность c_1 , а для остальных дуг c_2 : $c_1 < c_2 \ll c_0$, то при поиске минимального разреза его граница будет проходить ближе к первичным узлам переднего плана, при этом объединяя некоторые узлы переднего плана.

Получим потоковую сеть $G = (V, E, s, t, c)$, где $c = \{c(e)|e \in E\}$ — заданные пропускные способности. Найдем минимальный разрез (V_s, V_t) в этой сети:

$$\begin{cases} \sum_{e=(u,v), u \in V_s, v \in V_t} c(e) \rightarrow \min_{V_s, V_t} \\ V_s \cup V_t = V, V_s \cap V_t = \emptyset, s \in V_s, t \in V_t \end{cases}$$

В данной работе для поиска минимального разреза используется алгоритм Бойкова-Колмогорова, описанный в [5].

Таким образом, множество узлов скелета будет разделено на два множества V_s, V_t . Все узлы, вошедшие в V_s , отнесем к переднему плану, остальные отнесем к заднему плану. Такой подход позволяет разметить все узлы скелета на передний/задний план с учетом связности объектов.

Будем относить ребро скелета к переднему плану, если оба инцидентных ему узла отнесены к переднему плану.

5. Пиксельная маска переднего плана

Скелетного представления переднего плана сцены уже достаточно для выделения отдельных объектов переднего плана, вычисления их координат и ограничивающих прямоугольников, но для некоторых задач, например, дополненной реальности, может понадобиться пиксельная маска переднего плана. При наличии размеченного скелета, маску переднего плана можно вычислить, разметив множества пикселей, покрываемых каждым ребром переднего плана. Подробно эта процедура описана в предыдущей работе [1].

6. Сложность алгоритма

Для карты глубины, содержащей N пикселей, при построении скелета и триангуляции Делоне порождается $O(N)$ элементов, а вычислительная сложность составляет $O(N \log N)$. Построенная потоковая сеть содержит $O(N)$ элементов, а сложность поиска минимального разреза составляет $O(N^4)$, согласно [5]. Таким образом, минимальный разрез является теоретическим узким местом. На практике основная вычислительная нагрузка приходится на построение скелета, а метод способен работать в режиме реального времени.

7. Реализация алгоритма

Метод был протестирован на картах глубины, полученных с помощью сенсора Microsoft Kinect v2. По

результатам экспериментов, метод способен адекватно оценивать передний план.

Метод работает в режиме реального времени: до 20 кадров в секунду (на процессоре Intel Core i3 2.3 ГГц, оперативная память 8 Гб). Размер карт глубины составляет 512×424 пикселей. По сравнению с методом, предложенным в предыдущей работе [1], данный метод более устойчив к наличию пропусков в картах глубины и дает более стабильный результат, хотя он также восстанавливает пиксельную маску переднего плана неточно. Метод адекватно работает с окклюзиями в сцене. Пример работы метода можно видеть на изображении 7.



Рис. 7. Передний план сложной сцены. Светлым обозначена пиксельная маска переднего плана.

8. Выводы

В данной статье развивается метод выделения переднего плана на картах глубины, предложенный в предыдущей работе [1]. Модифицированный метод сохраняет полезные свойства: работает в реальном времени, использует единственную карту глубины в качестве входных данных. В отличие от других известных подходов, метод не накладывает ограничений на структуру сцены и движение камеры, а также не требует обучения. Кроме того, модификация метода более устойчива к пропускам в данных и позволяет выделять объекты переднего плана в виде связанных компонент графа.

Данный метод может применяться в задачах отслеживания и распознавания объектов, дистанционного управления и дополненной реальности в качестве начального этапа для определения областей интереса, а также для определения порядка расположения объектов сцены.

Дальнейшая работа будет включать применение метода в задачах распознавания жестов и отслеживания объектов.

9. Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов 17-01-00917 и 16-57-45054.

10. Литература

- [1] Белобородов Д.С., Местецкий Л.М. Выделение переднего плана на картах глубины с помощью скелетного представления силуэтов объектов // Труды международной конференции ГрафиКон-2016, 2016. – с.378–383.
- [2] Местецкий Л. М. Непрерывная морфология бинарных изображений. Фигуры, скелеты, циркуляры // Litres, 2017.
- [3] Bokaris P., Muselet D., Trémeau A. 3D reconstruction of indoor scenes using a single RGB-D image // 12th International Conference on Computer Vision Theory and Applications (VISAPP 2017), 2017.
- [4] Bondi E., Seidenari L., Bagdanov A.D., Del Bimbo A. Real-time people counting from depth imagery of crowded environments // Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference, 2014. – pp.337–342.
- [5] Boykov Y., Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision // IEEE transactions on pattern analysis and machine intelligence, 2004. – Vol.26, No.9. – pp.1124–1137.
- [6] Fernandez-Sanchez E.J., Diaz J., Ros E. Background subtraction based on color and depth using active sensors // Sensors, 2013. – Vol.13, No.7. – pp.8895–8915.
- [7] Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM, 1981. – Vol.24, No.6. – pp.381–395.
- [8] Gupta S., Girshick R., Arbeláez P., Malik J. Learning rich features from RGB-D images for object detection and segmentation // European Conference on Computer Vision, 2014. – pp.345–360.
- [9] Hornacek M., Fitzgibbon A., Rother C. SphereFlow: 6 DoF scene flow from RGB-D pairs // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. – pp.3526–3533.
- [10] Jaimez M., Souiai M., Gonzalez-Jimenez J., Cremers D. A primal-dual framework for real-time dense RGB-D scene flow // Robotics and Automation (ICRA), 2015 IEEE International Conference, 2015. – pp.98–104.

Создание единой текстурной карты для объемной модели, полученной с помощью трехмерного сканера

Бобков Валерий¹, Кудряшов Алексей²
bobkov@dvo.ru | kudryashovA@dvo.ru

^{1,2} Институт автоматизации и процессов управления ДВО РАН, Владивосток, Россия;

В данной работе предложен метод генерации единой текстурной карты для компьютерной трехмерной модели, построенной по данным, полученным с разных точек наблюдения. В качестве источника для генерации текстуры служат фотоизображения самой модели. Большая часть существующих методов построения текстурной карты обладает очень высокой вычислительной сложностью и не всегда позволяет избежать заметных швов. Предложенный в данной работе подход предлагает решение этих проблем. Он основан на слиянии всех доступных текстур с учетом веса каждого отдельного пикселя в каждом треугольнике, учитывающий перекрытие и наблюдаемость этих треугольников. Для уменьшения времени работы предложены: алгоритм индексации треугольников в виде текстурной карты, который позволяет избежать полного перебора треугольников, алгоритм для быстрого расчета весов точек треугольника, а также используются многоядерные вычисления.

Ключевые слова: текстурирование, текстурная карта, воксельное пространство, трехмерная реконструкция, триангуляционная оболочка.

Creating a single texture map for the 3D model obtained by 3D scanner

Bobkov V.¹, Kudryashov A.²
bobkov@dvo.ru | kudryashovA@dvo.ru

^{1,2} Institute of Automation and Control Processes, Far Eastern Branch of RAS, Vladivostok, Russia

In this paper, we propose a method for generating a single texture map for 3D model reconstructed by 3D scanner. Simple image projection results in visible seams on the surface of the scans, and all the analogues of the current method considered are highly computationally complex. The proposed method of texturing offers a solution to these problems. The texturing algorithm is based on the blending of all available textures, taking into account the weight of each pixel in each triangle, as well the overlap and observability of these triangles. The algorithmic implementation of the method is optimized to improve computational performance of the method and the quality of the generated textured triangulation model.

Keywords: texturing, texture map, voxel space, 3d reconstruction, mesh.

1. Введение

Для создания реалистичных моделей объектов реального мира часто используются трехмерные сканеры, использующие стереокамеры. Они позволяют получить набор триангуляционных сеток (или карт высот), которые определяют форму объекта. Однако такие данные обладают большой избыточностью, поскольку сканирование ведется с большими перекрытиями. Поэтому существует задача объединения этих сеток. В данной работе используется воксельный подход [3]. Наряду с задачей объединения этих триангуляционных сеток, подразумевающей построение единой связной оболочки без дублирования участков поверхности, стоят задачи зашивки дыр и построение единой текстуры. Зашивка дыр происходит на этапе заполнения неявной функции в воксельном пространстве [4,6].

Существует довольно много подходов для текстурирования реконструированных объектов, однако для каждого метода реконструкции с его ограничениями необходимо подбирать свой метод создания текстуры. В [1, 2, 5] предложены несколько подходов, которые легли в основу предложенного метода.

Предлагаемый в настоящей статье метод построения единой целостной текстурной карты трехмерной модели основывается на более ранних работах авторов [7,8]. Работа выполнялась в контексте решения авторами более общей задачи по созданию трехмерных сцен на последовательностях изображений. Вклад авторов состоит в разработке оригинального алгоритма, в котором: а) производится построение гибридной весовой функции для слияния текстур, учитывающий следующие факторы: угол

наблюдения, тень/пересвет, близость к краю скана; б) обеспечивается оптимизация вычислений за счет эффективной структурной организации данных; в) достигается высокая скорость обработки данных за счет реализации параллельных вычислений на многоядерных процессорах.

2. Описание метода

Исходными данными являются трехмерная модель и набор ее фотоизображений с известной внешней и внутренней калибровкой камер. Построение самой модели выполняется модифицированным воксельным методом с зашивкой дыр, также разработанным авторами, но может выполняться и любым другим.

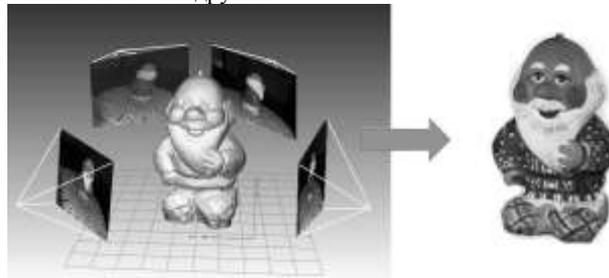


Рис. 1. Построение единой текстурной карты трехмерной модели по набору изображений, полученных с разных видов.

2.1 Предварительная фильтрация

Первая проверка – это проверка на наблюдаемость. Каждый треугольник построенной трехмерной модели проецируется на каждый из первоначальных видов. Для

каждого из них рассчитывается угол наклона от нормали треугольника к линии проходящей через центр каждой камеры. Очевидно, что если этот угол больше 90 градусов, то треугольник невидим и его можно отбросить. Для треугольников, у которых этот угол близок к 90 градусам, велика вероятность ошибки, но на случай, если нет другого более подходящего вида, эта проекция будет учтена. Вторая проверка – это проверка на видимость, то есть на то, что треугольник не перекрыт другими. В общем случае потребовался бы перебор всех треугольников модели, что заняло бы значительное время. Для оптимизации предлагается воспользоваться уже построенной индексной картой для первоначальной модели [7]. В случае использования модели, построенной другим методом объединения, такую индексную структуру надо дополнительно построить. Для всех углов треугольника объединённой модели получим значения индексов треугольников исходных сканов до объединения. Если расстояние между этими объектами меньше порогового значения, значит не существует объектов, которые бы загоразивали рассматриваемый треугольник. Особняком стоят треугольники, полученные в результате зашивки дыр – у них нет индексов на первоначальной индексной карте, поэтому их на видимость придется проверять либо с помощью полного перебора, либо, не использовать этот критерий, а использовать только критерий наблюдаемости. Логично, что для текстуры каждого треугольника нужно выбрать видимый участок, снятый с минимальным углом. Однако, на стыках текстур, полученных с разных видов, появляются заметные границы. Решению этой задачи посвящен следующий раздел статьи.

2.2 Попиксельное смешивание

Основной проблемой при нанесении текстуры на трехмерную модель с помощью исходных фотоизображений является различное освещение на разных видах. На рис. 2 видно, как один и тот же участок поверхности модели выглядит по-разному на разных видах.

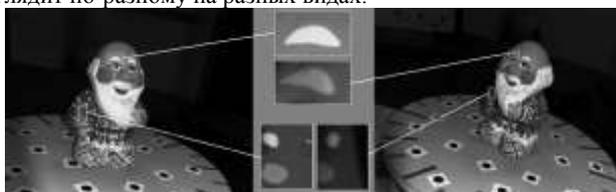


Рис. 2. Влияние освещения на текстуру.

Решить эту задачу можно путем смешивания значений цвета для каждой точки текстуры с разных видов. Однако равномерное смешивание текстур приводит к не самым идеальным результатам. Дело в том, что виды не равнозначны, на некоторых из них могут присутствовать сильные тени или пересветы, на некоторых, нужный нам треугольник, снят под очень малым углом, или находится на границе скана, где обычно могут возникать ошибки сканирования. Поэтому было предложено использовать гибридный вес для учета всех предложенных критериев выбора текстуры по следующей формуле:

$$W = W_{осв}(W_{уг} + W_{гр}), \quad (1)$$

где $W_{осв}$ – вес по освещенности, функция имеет трапециевидальную форму, принимает значение 0 – для теней (в которых яркость пикселя равна 0), затем линейно возрастает до 1 и также линейно падает до 0 для пересветов (в которых яркость пикселя равна 255), $w_{уг}$ – вес по углу наблюдения, линейно возрастает от 0 до 1, в зависимости от угла между вектором нормали к треугольнику и линией наблюдения, проходящей через камеру и центр треугольника, $w_{гр}$ – вес по расстоянию, линейно возрастает от 0 до 1, в зависимости от расстояния от центра треугольника до края скана.

Первые два параметра не вызывают трудностей с расчетом, то расчёт расстояния от каждого треугольника до границы скана является крайне ресурсоемким, поэтому была предложена следующая оптимизация. Уже существует построенная структура индексов для всех треугольников, которая в двухмерном виде содержит индексы треугольников и нули там, где треугольников нет. Необходимо, для каждого ненулевого значения вычислить расстояние до ближайшего нуля. Поскольку мы используем найденные значения только в качестве веса, то нет необходимости в вычислении абсолютных значений в 3d, достаточно вычисления дистанции на плоскости изображения. Однако, прямой перебор всех ненулевых значений и поиск ближайшего нулевого в заданном радиусе будет занимать продолжительное время. Поэтому была предложена еще одна оптимизация. Каждому ненулевому значению индекса присваиваем максимальный вес, затем проверяем есть ли рядом нулевое значение, если есть, добавляем в список и присваиваем значение 1. Затем будем производить поиск значений равных максимальному весу только в соседних ячейках элементов из этого списка. Такую итерацию производим необходимое число раз. Данная оптимизация позволяет сократить время вычисления веса в 6-8 раз по сравнению с полным перебором.

На рис. 3б показана построенная карта весов по дистанции от края скана для одного вида модели «Змея» (рис.3а), а также гибридная карта весов с учетом всех весов: по освещению, углу и расстоянию до края скана (рис. 3в), рассчитанная по формуле (1).



Рис. 3. а) Изображение модели «Змея», б) карта весов по дистанции от края, в) гибридная карта с учетом всех весов.

Рис. 3. а) Изображение модели «Змея», б) карта весов по дистанции от края, в) гибридная карта с учетом всех весов.

Рассчитав вес для каждого пикселя будем строить финальную текстуру. Для каждого пикселя каждого треугольника вычисляем цвет по формуле (2):

$$C = \sum_{i=0}^n W_i c_i, \quad (2)$$

где n – число видов, c_i – исходный цвет с i -го вида, вес W_i должен быть нормирован так, чтобы сумма всех весов для всех видов была равна 1.

2.3 Создание компактной текстурной карты

Заполненные цветом треугольники проецируются на пустые изображения, аналогичные по параметрам исходным видам. Очевидно, что полученная таким образом текстура будет занимать площадь, равную площади всех исходных изображений, что крайне избыточно. Для построения компактной текстуры будем использовать следующий алгоритм. Спроецированные треугольники будут образовывать цветные участки, которые с помощью волнового алгоритма выделяются и заносятся каждый в отдельный массив (рис. 4).

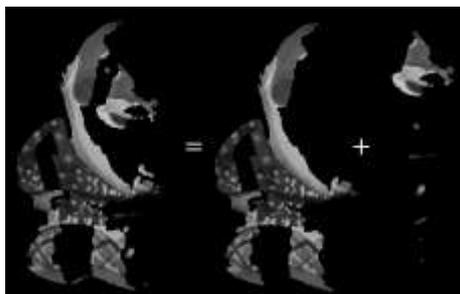


Рис. 4. Сегментация текстурных участков.

Туда же записываются индексы соответствующих треугольников. Теперь необходимо эти участки расположить максимально компактным образом. Представим эти участки в виде фигурок игры «Тетрис», только более сложной формы. Для скорости уменьшим их размер в несколько раз и представим в виде однобитовой карты (рис. 5).

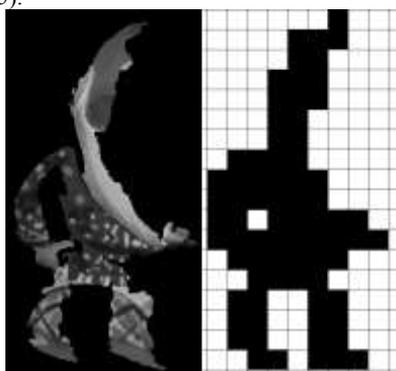


Рис. 5. Представление текстурного участка в виде однобитовой карты.

В первую очередь располагаем на финальной текстуре участки наибольшего размера, а в образовавшиеся промежутки укладываем все остальные. При перемещении необходимо пересчитать текстурные координаты.

3. Результаты

На следующих рисунках представлены несколько реконструированных моделей. Показаны несколько исходных изображений, сама модель и ее текстурная карта. В таблице 1 приведены параметры модели и время работы предложенного алгоритма. Там же, для сравнения, приведено время работы алгоритма UVAtlas [9], который также позволяет создать текстурную карту.

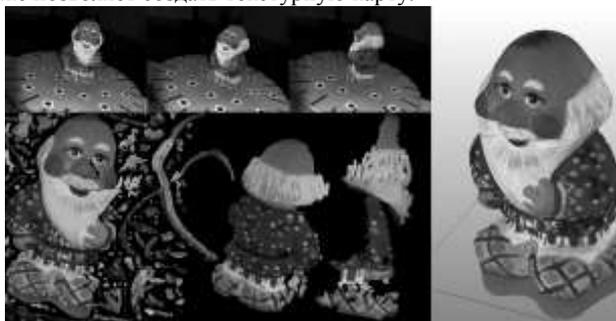


Рис. 6. Модель «Дед».

Отличие части текстурного участка от рис. 5 объясняется тем, что при создании компактной текстурной карты, часть треугольников можно перепроецировать на участки с большим количеством треугольников, для создания текстурных участков большей площади.



Рис. 7. Модель «Чайник».

Таблица 1. Сводная таблица результатов

Параметры \ Модель	Дед	Змея	Чайник
Треугольников в исходной модели, млн.	1,8	2,7	11,8
Количество видов	8	10	22
Время реконструкции, сек.	18	37	184
Время создания текстуры, сек.	6	12	99
Треугольников в финальной модели, млн.	0,3	0,6	4,8
Время работы UVAtlas, сек	84	213	>3800

В качестве недостатков метода стоит отметить то, что в областях, ненаблюдаемых ни на одном скане и в прилегающих к ним, могут создаваться артефакты. Также, помимо больших текстурных участков создается большое количество мелких, которыми закрашивается всего один или два треугольника. Дальнейшая работа будет направлена на решение указанных недостатков.

4. Заключение

В работе представлен метод создания единой текстурной карты для триангуляционной трехмерной компьютерной модели, по набору фотонизображения. Тестирование на реальных данных показало, что предложенные алгоритмические решения и структура данных с использованием октодеревя, индексации треугольников, вычисления гибридной весовой функции для каждой точки текстуры существенно повысили качество визуализации и производительность метода в сравнении с аналогами.

5. Благодарности

Работа выполнена при частичной финансовой поддержке РФФИ (грант № 15-07-00341, № 16-07-00350), Программы «Дальний Восток» (проект 15-I-4-011 о) и Программы Президиума РАН № I.33П «Фундаментальные проблемы математического моделирования».

6. Список литературы

- [1] Alshawabkeh Y., Naala N. Automatic multi-image phototexturing of complex 3D scenes. XVIII CIPA Int. Symposium, Torino. 2005. pp. 68-73.
- [2] Baumberg A. Blending Images for Texturing 3D Models. British Machine Vision Conference - BMVC, 2002. pp. 112-126.
- [3] Brian Curless, Marc Levoy. A Volumetric Method for Building Complex Models from Range Images. . In Proceedings of SIGGRAPH'96. 1996. P. 303-312

- [4] James Davis, Stephen R. Marschner, Matt Garr, Marc Levoy. Filling holes in complex surfaces using volumetric diffusion // First International Symposium on 3D Data Processing, Visualization, and Transmission, Padua, Italy. 2005. pp. 428–438.
- [5] R. Gal, Y. Wexler, E. Ofek, H. Hoppe, D. Cohen-Or. Seamless montage for texturing models Computer Graphics Forum 29 (2). 2009. pp. 479-486.
- [6] M. Kazhdan and H. Hoppe. Screened Poisson Surface Reconstruction. Transactions on Graphics, June 2013, Vol. 32, No. 3. pp. 29-42.
- [7] Бобков В.А., Кудряшов А.П. Воксельный метод построения триангуляционной поверхности по множеству видов // Информатика и системы управления, №2. 2012. С. 31-38.
- [8] Кудряшов А.П., Черкашин А.С. Построение единой триангуляционной поверхности по набору видов с зашивкой дыр // Информатика и системы управления. №1. 2015. С.36-40.
- [9] <https://uvatlas.codeplex.com/>

Об авторах

Бобков Валерий Александрович, д.т.н., заведующий лабораторией машинной графики института автоматки и процессов управления ДВО РАН.

Кудряшов Алексей Павлович, к.т.н., младший научный сотрудник лаборатории машинной графики института автоматки и процессов управления ДВО РАН.

Метод захвата движений как средство естественного интерфейса

Гайниyarов И.М.¹, Обабков И.Н.¹, Хлебников Н.А.¹

i.m.gainiarov@urfu.ru|i.n.obabkov@urfu.ru|na.khlebnikov@urfu.ru

¹Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, г. Екатеринбург, Россия

В данной работе представлена модульная архитектура мобильной системы захвата движений человека для работы в режиме реального времени. Техническую основу технологии составляет специальная сеть микроэлектромеханических датчиков (трехосевые акселерометры и гироскопы), закрепленных на теле человека. Каждый модуль датчиков подключается к одному микроконтроллеру для дальнейшей передачи данных на сервер хранения и обработки в клиент-серверной парадигме. Предполагаемая обобщенная конструкция имеет мобильное и свободное ношение датчиков в виде особого костюма. Такой подход помогает расширить сферу применения технологии захвата движений со стационарных подготовленных мест с видеокамерами к захвату движений конкретных людей без привязки к их местоположению. Так же это позволяет снизить хранимый поток информации, так как оцифровка каркаса движения производится датчиками на местах без передачи видеопотока. Все это делает возможным реализацию естественного интерфейса на основе костюма как для здоровых людей, так и для пациентов, нуждающихся в реабилитации.

Ключевые слова: Система захвата движений, Естественный интерфейс, Микроэлектромеханический гироскоп, Микроэлектромеханический акселерометр.

Motion capture method as a tool for natural interface

Gajniyarov I.M.¹, Obabkov I.N.¹, Khlebnikov N.A.¹

i.m.gainiarov@urfu.ru|i.n.obabkov@urfu.ru|na.khlebnikov@urfu.ru

¹Ural Federal University, Yekaterinburg, Russia

This paper presents a modular architecture of mobile motion capture system to work in real-time. The technical basis of technology is network of micro electro-mechanical sensors placed on human body. Each sensor module is connected to one microcontroller for further data transmission to server storage and processing. The design has free carrying sensors in a special suit. This approach helps to broaden the scope of movements capture technology from stationary prepared places with video cameras to capture without reference to their location. Also it allows to reduce the stored information, because digitization of the frame movement is performed by sensors. All this makes possible the realization of natural interface based on the suit as for healthy people and for patients who need rehabilitation.

Keywords: Motion capture, Wearable wireless system, MEMS sensor.

1. Введение

В настоящее время основной интерес технологии захвата движений находится в следующих областях: медицина, игровая индустрия, компьютерная анимация, охранные системы, робототехника [23].

В медицине это связано с повышением качества процесса реабилитационной терапии [5]. Например, в качестве съема показателей вне стационарных пунктов наблюдения при нарушениях опорно-двигательного или вестибулярного аппаратов. Или в качестве профессионального инструмента, как естественный интерфейс для проведения операции хирургом [1].

В игровой индустрии — это возможность получения естественных контроллеров для максимального погружения в виртуальную реальность.

В робототехнике — это управление промышленными манипуляторами, робототехническими средствами передвижения, экзоскелетами. Так же существует вариант с образовательными роботами. Один человек выступает в качестве эталона движения, с него только снимаются показания. Второй человек выступает в качестве приемника данных, с него не только снимаются показания, но и сравниваются с эталоном и сообщаются в виде стимулирующей обратной связи [3]. Большой интерес к антропоморфным робототехническим системам проявлен и в космической отрасли [8]. Например, стоят задачи дистанционного исполнения роботом сложных двигательных паттернов, скопированных с оператора [7].

В компьютерной анимации — это создание модели движения персонажей, мимики.

В охранных системах — это фиксация как самого факта движения в намеченных областях, так и распознавание паттерна движения отдельных людей в общем потоке.

Есть два принципиальных подхода к захвату движений [11]: анализ видеопотока [2] и подготовка данных о движении датчиками на местах крепления.

Плюсы первого подхода заключаются в простоте внедрения способа. Можно быстро подготовить пространство для получения сырых данных.

Минусами является повышающиеся требования к числу кадров в секунду у камеры для более детального захвата движений, большой объем хранения данных, сложность переноса камеры как оборудования для получения данных.

Плюсами второго подхода являются: мобильность, легкость оборудования и относительно малое число передаваемых данных.

Минусами является сложность интерпретации данных.

2. Постановка задачи

Основная цель создания костюма захвата движений — это создание единой платформы для создания разнообразных естественных интерфейсов. Это возможно по причине модульности системы. Получая весь объем данных можно выявлять зависимости отдельно на каждой руке или ноге, в зависимости между друг другом или во всем каркасе человека в целом [19].

Для этого необходимо создать электронную схему генерации необработанных данных, соответствующую следующим требованиям:

1. Автономное питание системы минимум на 24 часа бесперебойной работы.

2. Беспроводной способ передачи данных на сервер.

Соответствовать ГОСТ 12.2.025-76 ССБТ по классу электробезопасности.

4. Содержать датчики на большинстве суставных сочленениях человека (это более 45 MEMS датчиков).

5. Скрытое монтирование датчиков для предотвращения механическим повреждениям.

6. Размер площадки крепления каждого датчика не более 50 мм².

7. Предоставлять гарантию о расположении точек генерации данных через процедуру внутренней сертификации с помощью аппаратного криптоключа.

8. Состоять из 5 модулей [12] в базовой комплектации для всего тела человека.

Также необходимо разработать программный модуль для приема и обработки сырых данных. Работа для пользователя системы должна выглядеть как “черный ящик”, куда можно отправить данные и получить результат обработки. Такой подход с одной стороны решает вопрос безопасности интеллектуальной собственности, а с другой помогает работать системе на маломощных компьютерах.

3. Техническая реализация

3.1 Архитектура системы

В этом разделе особое внимание будет уделено вопросу мобильности, модульности и автономности системы.

Сбор данных осуществляется на плате NodeMCU на базе микроконтроллера ESP8266. Далее по сети WiFi укомплектованные данные [15] передаются на сервер, что позволяет не хранить данные на самом костюме. Общая схема представлена на Рисунке 1. Также возможна передача по Bluetooth и радиоканалу.

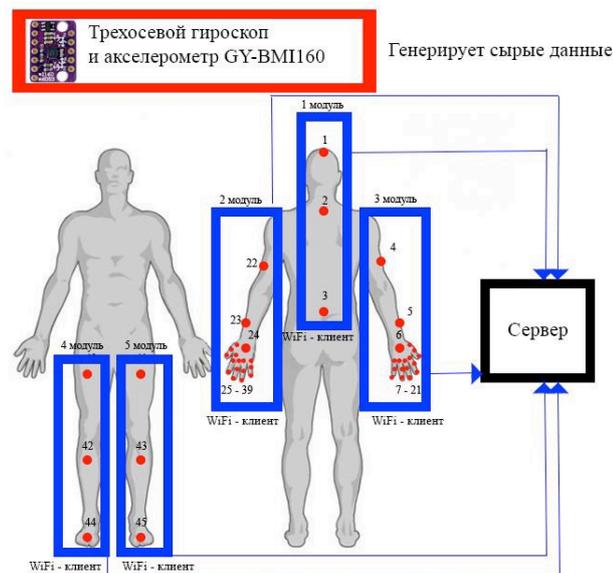


Рис. 1. Общая схема генерации и отправки данных.

Точками обозначены датчики. Прямоугольниками показано расположение модулей.

3.2 Архитектура модуля

Каждый модуль [10] в отдельности выполнен по схожей схеме, представленной на рисунке 2.

Все датчики подключены по SPI протоколу. Название и назначение каналов:

1. MISO – master input slave output. Сигнал от ведомого устройства ведущему.

2. MOSI – master output slave input. Сигнал от ведущего устройства ведомому.

3. SCL – канал синхронизации.

4. GDN – земля.

5. VCC – питание.

6. CS – chip select. Канал выбора устройства для снятия данных.

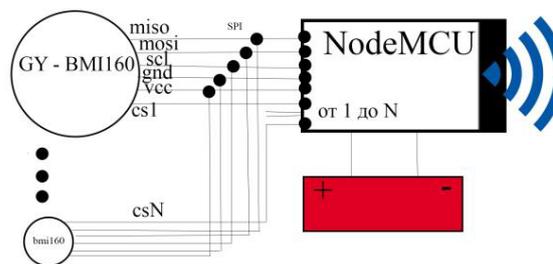


Рис. 2. Общая схема работы модуля. NodeMCU имеет одно питание. К нему по 5 общим шинам и N (по числу датчиков) подключены MEMS-устройства.

3.3 Программная реализация

Для проверки адекватности полученных данных реализованы две идеи. Первая заключается в начальных условиях, при которых после закрепления костюма на человеке и перед запуском питания необходимо, чтобы человек начал с исходной позиции. Таким образом происходит синхронизация последующих изменений в движениях модели и реального человека. Вторая заключается в создании визуального отладчика, который позволяет в режиме реального времени смотреть движения модели в соответствии с человеком.

Инструментом для реализации идеи послужил Blender 3D. Программный продукт позволяет создать не только необходимые модели, но и позволяет прописывать для них свою физическую модель на Python 3.

Были проведено тестирование для одиночных датчиков BMI160 на соответствие модели гироскопа реальным поворотам. Результаты положительные.

Также возник вопрос интерпретации полученных данных. Интегрированная калибровка датчика рассчитана на горизонтальное положение датчика во время запуска. Однако в условиях костюма это невозможно. В связи с этим необходимо при запуске генерировать матрицу корректирующих коэффициентов.

Сами показания датчиков нуждаются в учете их взаимосвязи [13]. Например, если рука поднимается – то данные со всех датчиков будут меняться, однако это будет линейная зависимость и означать она будет поднятие лишь предплечья, а не разнообразное движение руки.

На рисунке 3 представлено изменение положения руки модели человека через 1 секунду после запуска оборудования на реальном человеке. Генерируемые данные получаются корректными, однако существует проблема с нулевым дрейфом гироскопа. Эта проблема решается добавлением магнитометра к существующим акселерометру и гироскопу. Также для реализации автокалибровки данных тестируется возможность добавления датчика положения на кисть руки в единичном экземпляре и добавления датчиков Холла для фиксации прохождения руки через калибровочные магнитные точки на костюме во время движения человека.

Необходимо также решить вопрос зашумленности входящие данных. В данный момент рассматриваются следующие варианты постобработки на сервере (ввиду большой ресурсозатратности операций для микроконтроллера):

1. Фильтр Калмана.

2. Фильтр Маджвика [16].

3. Комплементарный фильтр [9].
4. Интегрирование данных по гироскопу [4].
5. Компенсация данных по акселерометру.
6. Дополнение данными с магнитометра.

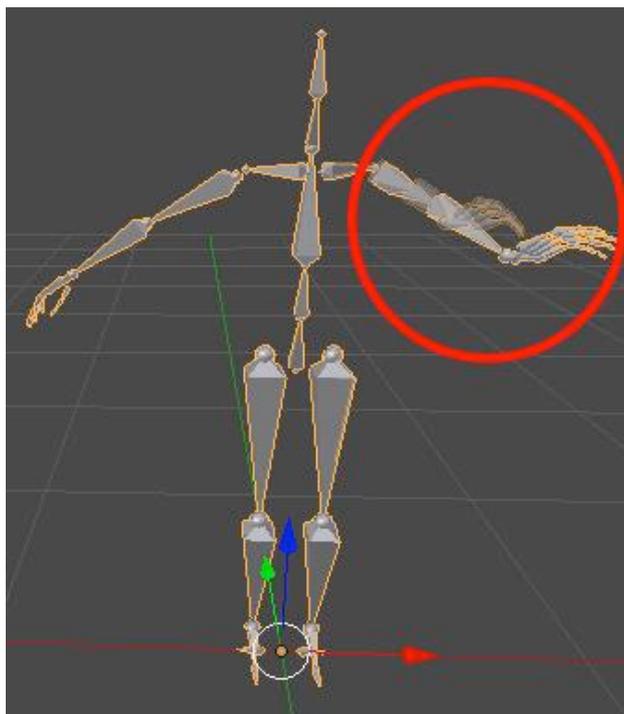


Рис. 3. Изменение положение руки модели с учетом поступивших с оборудования данных. Наложение двух снимков.

Еще одной программной задачей является выявление структуры в данных. Это необходимо решать с помощью инструментов машинного обучения.

Были проведены тесты быстрого действия системы. Сами датчики по документации способны работать на частоте до 3200 Гц для гироскопа и 1600 для акселерометра. Далее где идет узкое место – это протокол передачи. Используется стек TCP/IP для передачи по локальной сети через роутер. Если постоянно перед передачей пакета данных открывать соединение с сервером, а потом его закрывать, то частота составляет 1.8 Гц. Если при запуске оборудования открывать соединение один раз и закрывать по окончанию работы, то частота составляет от 54 до 76 Гц для 18 датчиков в одном модуле. Частота зависит от способа формирования пакета. С помощью POST запроса можно передать максимум 100 переменных за раз (это ограничение технологии). В дальнейшем на один модуль будет не более 16 датчиков данного типа, т.к. $6 \cdot 18 = 108$ значений, что больше 100. Поэтому реализовано два способа формирования пакета:

1. Склеивать значения одного датчика в одну строку, тогда передается пакет с 18 подразделами.

2. Склеивать значения пакетов только для двух последних датчиков, тогда число подразделов составляет 98.

Для первого случая частота ниже, т.к. необходимо работать с разбиением строки, второй способ быстрее ввиду меньшей вычислительной нагрузки.

4. Предварительный результат

В данный момент удалось создать модуль для захвата движений левой руки. Было реализовано два прототипа генерации данных. Первый включал Arduino MEGA 2560 в качестве микроконтроллера и Bluetooth HC-05 в качестве

передатчика. Прием осуществлялся по COM порту. Вторым прототип был спроектирован на базе NodeMCU с интегрированным микроконтроллером ESP8266, который поддерживает передачу по WiFi.

Предпочтение было отдано в пользу второго варианта, потому что в таком случае габариты устройства уменьшались в несколько раз, общая длина проводов также значительно сокращается, что приводит к меньшему энергопотреблению, снижению различного рода наводок на провода, больший радиус контакта с принимающим сервером увеличивает более, чем в два раза число датчиков. Помимо этого, WiFi протокол позволяет открыть возможность реализации множества других полезных функций в будущем.

В модуле левой руки работают 18 датчиков, одна литий-ионная батарея 2200 мАч и реализована беспроводная передача через Bluetooth или WiFi.

Со стороны сервера реализован прием данных и настроена среда для визуализации данных (Blender 2.78c и Unity 3D).

Обработка данных в данный момент реализуется с использованием фильтра Маджвика с магнитометром. Для проверки адекватности модели сначала были записаны исторические данные по простым движениям руки, пальцев, а также более сложные движения без перемещения всего тела в пространстве. После этого была натурная проверка на живом человеке. В данный момент еще модернизируется фильтр Маджвика для множественного использования гироскопов, акселерометров и магнитометров, т.к. классический вариант предполагает наличие только одного 9 (или 6) осевого датчика.

В дальнейшем планируется дальнейшее уменьшение носимого оборудования [14]. При малых существующих размерах собранное оборудование не носит промышленный характер. Возможна распайка датчиков в более компактную форму. Конечный вид микроконтроллера также будет претерпевать изменения. В конечном итоге при ношении костюма будут заметны только элементы крепления (перчатки, стяжки) и батарея питания. Однако эти вопросы относятся к области дизайна вещей и эргономичности.

В дальнейшем планируется расширить систему в плане разнообразия датчиков [17]. В ближайшее время возможно расширение за счет датчиков кожно-гальванической реакции и электроэнцефалографических датчиков (32 канальная система).

Помимо этого, планируется серия экспериментов для распознавания простейших движений человека для проверки базовой гипотезы о распознавании жестов. После этого будут эксперименты по распознаванию сложных движений вне рамок лабораторных условий [6].

Все это позволит в конечном итоге составить характерный портрет человека, основанный на индивидуальных особенностях движений человека. Что уже может применяться для количественной оценки изменений в человеке. Например, перед операцией человек носит подобный костюм в течение месяца. Сервер сохраняет все данные и выделяет характерные особенности человека. После проведения инвазивной операции можно будет оценить на сколько изменился человек и выработать индивидуальную программу по восстановлению в привычную форму.

Еще одной важной особенностью работы является возможность создания виртуального тела с обратной связью.

5. Благодарности

Работа была выполнена при поддержке ООО ИК "Энергософт" и ООО "Ведущая Утилизирующая Компания".

6. Ссылки

- [1] Авербух В.Л., Авербух Н.В., Стародубцев И.С., Тоболин Д.Ю. Использование жестовых интерфейсов при взаимодействии с объектами // Научная перспектива. № 10 (56) / 2014. Стр. 57-66.
- [2] Алфимцев А. Н., Демин Н. А. Захват и отслеживание удаленных объектов в видеопотоке //Инженерный журнал: наука и инновации. – 2013. – №. 11. – С. 23.
- [3] Гаврилова И. В. Перспективы применения робототехники в образовании //Новые информационные технологии в образовании и науке: НИТО-2017: материалы X международной научно-практической конференции. – РГППУ, 2017. – №. 10. – С. 458-460.
- [4] Иванов Д. С., Овчинников М. Ю. Использование одноосного гироскопа для определения ориентации макета в лабораторных условиях //Препринты Института прикладной математики им. МВ Келдыша РАН. – 2008. – №. 0. – С. 11-32.
- [5] Крак Ю. и др. Компьютерная система виртуального общения людей с проблемами слуха //16th International Conference «Knowledge-Dialogue-Solution» KDS. – 2008. – С. 161-165.
- [6] Кривonos Ю. Г., Крак Ю. В. Моделирование движений рук, мимики и артикуляции лица человека для синтеза и визуализации жестовой информации //Кибернетика и системный анализ. – 2011.
- [7] Крючков Б. И., Усов В. М. Создание моделей виртуальной реальности, как способ обучения космонавтов взаимодействию с роботом-помощником экипажа и как условие определение потенциальных областей его полезного применения //Труды международной научно-технической конференции «Экстремальная робототехника», СПб: Политехника-сервис. – 2013. – С. 230-244.
- [8] Крючков Б.И., Усов В.М., Михайлюк М.В. Применение методов захвата движений в космической робототехнике при инженерно-психологическом проектировании человеко-машинного взаимодействия // Пилотируемые полеты в космос. N 4 (21), 2016. Стр. 57-78.
- [9] Мансур М., Степанов О. А. Комплементарный фильтр в задачах комплексной обработки избыточных измерений //всероссийская научная конференция по проблемам управления в технических системах. – Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В.И. Ульянова (Ленина)(Санкт-Петербург), 2015. – №. 1. – С. 380-384.
- [10] Москвичёв Д. В., Суков С. Ф. Использование трехосевого mems гироскопа и акселерометра для задач определения пространственной ориентации объектов. – 2016.
- [11] Стародубцев И.С. Захват движения на основе иерархического гибкого скелета // Научно-технический вестник Поволжья. Направления: физико-математические науки, химические науки, технические науки. N 1. 2014. Стр. 159-162.
- [12] Стародубцев И.С. Проект «Виртуальная рука» // XV Международная конференция «Супервычисления и Математическое Моделирование». Тезисы. ФГУП «РФЯЦ ВНИИЭФ». Саров. 2014, стр. 124-125.
- [13] Шадрин, Д.В., Арышева, Г.В. Роботизированная рука с дистанционным управлением / Д.В. Шадрин, Г.В. Арышева

// Инженерия для освоения космоса: сб. науч. тр. / ТПУ. – Томск, 2016. – С. 278-281.

[14] Brigante C. M. N. et al. Towards miniaturization of a MEMS-based wearable motion capture system //IEEE Transactions on Industrial Electronics. – 2011. – Т. 58. – №. 8. – С. 3234-3241.

[15] Di Nisio A. et al. Design of a low cost multipurpose wireless sensor network //Measurements & Networking (M&N), 2015 IEEE International Workshop on. – IEEE, 2015. – С. 1-6.

[16] Ezekwe C. D., Geiger W., Ohms T. A 3-axis open-loop gyroscope with demodulation phase error correction //Solid-State Circuits Conference-(ISSCC), 2015 IEEE International. – IEEE, 2015. – С. 1-3.

[17] Makeig S. et al. Linking brain, mind and behavior //International Journal of Psychophysiology. – 2009. – Т. 73. – №. 2. – С. 95-100.

[18] Moeslund T. B., Granum E. A survey of computer vision-based human motion capture //Computer vision and image understanding. – 2001. – Т. 81. – №. 3. – С. 231-268.

[19] Mohammadzadeh F. F. et al. Feasibility of a wearable, sensor-based motion tracking system //Procedia Manufacturing. – 2015. – Т. 3. – С. 192-199.

Оценка погрешности измерений геометрических параметров, выполняемых с использованием призмочно-линзовых оптических систем

А.В. Гореvoy^{1,2,3}, А.С. Мачихин^{1,2}

gorevoy.a@gmail.com|aalexanderr@mail.ru

¹Научно-технологический центр уникального приборостроения РАН, Москва, Россия;

²НИУ «МЭИ», Москва, Россия;

³МГТУ имени Н.Э. Баумана, Москва, Россия

Использование эндоскопов со стереоскопическими призмочно-линзовыми оптическими системами для проведения визуально-измерительного контроля требует оперативной и надежной оценки погрешности измерений. Известные решения данной задачи основаны на серьезных допущениях, что делает невозможным их применение на практике без дополнительных испытаний. Основным методом оценки погрешности реальных измерений и аттестации приборов остается проведение серий контрольных измерений эталонных объектов. Предложен метод оценки погрешности геометрических измерений, выполненных стереоскопическим методом, подходящий для различных математических моделей устройств регистрации. Применимость метода подтверждена математическим моделированием для значений параметров и условий измерения, типичных для промышленных эндоскопов. Эксперименты с таким прибором показывают, что обычно используемые допущения о независимости погрешности определения соответствующих точек на изображениях приводят к заметному расхождению расчетных и экспериментальных данных. Результаты проведенных исследований могут быть полезны при проектировании измерительных систем на основе стереоскопического метода, а также разработке алгоритмов обработки изображений в подобных системах.

Ключевые слова: стереоскопический метод, трехмерные измерения, видеоэндоскопия, погрешность измерений, призмочно-линзовая оптическая система.

Uncertainty evaluation of geometric parameter measurements performed using prism-based optical system

A.V. Gorevoy^{1,2,3}, A.S. Machikhin^{1,2}

gorevoy.a@gmail.com|aalexanderr@mail.ru

¹Scientific and Technological Center of Unique Instrumentation of RAS, Moscow, Russia;

²National Research University "Moscow Power Engineering Institute", Moscow, Russia;

³Bauman Moscow State Technical University, Moscow, Russia

Remote visual inspection using video endoscopes equipped with stereoscopic prism-based optical systems requires operative and robust evaluation of measurement uncertainty. Known solutions are based on significant assumptions which hamper their implementation in practice without additional analysis. Currently, the main method for measurement accuracy estimation and video endoscope certification is multiple control measurements of the test chart parameters. We propose a method for estimating the error of geometric parameters measurements performed by the stereoscopic technique, suitable for different mathematical camera models. Applicability of the method has been confirmed by computer simulation for the parameter values and measurement conditions typical for industrial endoscopes. According to the experimental results obtained with an industrial endoscope, it has been shown that the commonly used assumptions about the independence of the error in determining the corresponding points on the images lead to a significant discrepancy between the calculated and experimental data. The results can be useful for the design of measurement systems based on the stereoscopic method, as well as the development of image processing algorithms for this type of systems.

Keywords: stereoscopic technique, 3D measurements, video endoscope, measurement uncertainty, prism-based optical system.

1. Введение

Эндоскопические системы являются одним из наиболее распространенных и универсальных средств визуально-измерительного контроля (ВИК) внутренних полостей промышленных объектов и сложного оборудования, не подлежащего разборке. К настоящему времени разработано несколько методов измерения трехмерных геометрических параметров выявляемых дефектов, аппаратно и программно реализованных в современных видеоэндоскопах: стереоскопический, многоточечный, теневой и фазовый [3, 5, 14, 15, 19]. Все эти методы основаны на принципе триангуляции. Стереоскопический метод является наиболее распространенным, поскольку при минимальных габаритах зонда позволяет проводить измерения для всех точек на изображении и не требует дополнительной системы подсветки. Реализация данного метода основана на

регистрации двух или более изображений объекта с разных ракурсов. Из-за малого диаметра зондов (как правило, 4-8 мм) размещение двух отдельных объективов и матричных приемников излучения (МПИ) затруднительно. Решением данной проблемы является использование сменных стереоскопических насадок, содержащих призмочно-линзовую оптическую систему (ОС), что позволяет зарегистрировать изображения с двух разных ракурсов на одном МПИ [4, 5, 13]. Калибровка такой системы и определение соответствующих точек на стереоскопических изображениях позволяют вычислять трехмерные координаты точек объекта и производить геометрические измерения [1, 2, 11, 17, 18] в соответствии со схемой, показанной на рис. 1.

Данная схема показывает основные элементы, которые рассматриваются при проектировании видеоэндоскопа с призмочной насадкой как измерительной стереоскопической системы. Поскольку алгоритм

вычисления трехмерных координат и алгоритм калибровки всегда содержат упрощенную математическую модель формирования изображения, то эту модель логично использовать для предварительных расчетов на начальном этапе проектирования и для формулировки требований к ОС и МПИ на следующих этапах [1, 2, 4]. Критерий качества измерительной системы формируется на основе величины погрешности измерения геометрических параметров. Таким образом, для расчетов необходимо иметь метод оценки погрешности, который позволяет рассчитать погрешность измерения трехмерных координат каждой точки рабочего объема при известной погрешности определения координат соответствующих точек на изображениях и погрешности параметров калибровки. Наличие такого метода также необходимо непосредственно в процессе ВИК для выдачи рекомендаций оператору и принятия им оперативных решений о результатах контроля.



Рис. 1. Структурно-функциональная схема видеэндоскопа с призматической насадкой.

Несмотря на активное развитие аппаратных и программных средств для реализации стереоскопического метода, задача определения погрешности реальных измерений остается недостаточно проработанной. Предложенные решения в области математического моделирования основаны на серьезных допущениях, что делает невозможным их применение на практике для ответственного оборудования ВИК без дополнительных испытаний. Основным методом оценки погрешности реальных измерений и аттестации приборов остается проведение с их помощью серий контрольных измерений эталонных объектов. В данной работе излагаются результаты теоретических и экспериментальных исследований по оценке погрешности измерений, выполненных при помощи эндоскопической системы с призматической насадкой, и анализируется влияние различных допущений на результат.

2. Метод оценки погрешности

Методы оценки погрешности измерения трехмерных координат можно разделить на две группы: использующие анализ интервалов и преобразование распределения вероятностей [1, 2, 9, 10, 16]. В первом случае границы

доверительных интервалов для координат точек на изображениях преобразуются в границы доверительного интервала для измеренных трехмерных координат. Недостатком таких методов являются завышенные значения погрешностей, невозможность учета корреляции исходных параметров и сложность реализации для измерений геометрических параметров, проводимых по нескольким точкам. Вторая группа методов направлена на определение распределения вероятностей для измеренных геометрических параметров по совместному распределению вероятностей для координат точек на изображениях. Из-за сложности вычисления преобразований распределения вероятностей для алгоритма триангуляции, как правило, считают исходное распределение нормальным и используют линейное приближение, что требует вычисления аналитических производных и может приводить к ошибочной оценке.

В данной работе мы используем УТ-метод (unscented transformation [8]), поскольку он позволяет универсальным способом рассчитать среднее значение и средноквадратическое отклонение (СКО) для различных математических моделей устройств регистрации и различных типов измерений (длина отрезка, площадь фигуры, расстояние от точки до плоскости и т. п.). Такой метод лишен недостатков, присущих методам линеаризации, не требует столь значительных вычислительных затрат, как метод со множеством случайных выборок (Monte-Carlo analysis), и может использоваться для обеих сформулированных выше задач.

Для описания метода примем, что модель формирования изображения записана в виде $\mathbf{p}_i = P_i(E_i(\mathbf{x}))$, где \mathbf{x} – координаты точек объекта в глобальной системе координат (ГСК), \mathbf{p}_i – координаты их изображений для i -й камеры, E_i – оператор преобразования координат из ГСК в систему координат (СК) i -й камеры, P_i – оператор преобразования, определяющего соответствие между координатами точек \mathbf{x}_i в СК камеры и координатами их изображений \mathbf{p}_i . Набор преобразований P_i и E_i параметризуется вектором \mathbf{k} [1, 2, 6]. В случае эндоскопической системы с призматической ОС под формированием изображения i -й камерой следует понимать формирование 1-й (левой) или 2-й (правой) части общего составного изображения [6, 7, 11, 17]. Исходными данными для метода при расчете погрешности измерения трехмерных координат для точки $\bar{\mathbf{x}}$ являются вектор координат изображений $\bar{\mathbf{p}} = (\bar{\mathbf{p}}_1^T, \bar{\mathbf{p}}_2^T)^T$ размерности $N = 4$, где $\bar{\mathbf{p}}_i = P_i(E_i(\bar{\mathbf{x}}))$, и соответствующая ему матрица ковариации Σ_p . Для проведения расчета зададим набор из $(2N + 1)$ сигма-точек $\tilde{\mathbf{p}}_j$ и весовых коэффициентов W_j [8]:

$$\begin{aligned} \tilde{\mathbf{p}}_0 &= \bar{\mathbf{p}}, & W_0 &= a/(N + a), \\ \tilde{\mathbf{p}}_j &= \bar{\mathbf{p}} + \left(\sqrt{(N + a)\Sigma_p}\right)_j, & W_j &= 1/2(N + a), \\ \tilde{\mathbf{p}}_{j+N} &= \bar{\mathbf{p}} - \left(\sqrt{(N + a)\Sigma_p}\right)_j, & W_{j+N} &= 1/2(N + a), \end{aligned}$$

где $j = 1 \dots N$, $\left(\sqrt{(N + a)\Sigma_p}\right)_j$ – j -я строка квадратного корня из матрицы $(N + a)\Sigma_p$, $a \in \mathfrak{R}$ – параметр. Далее вычислим координаты $\tilde{\mathbf{x}}_j$ для $j = 0 \dots 2N$ при помощи алгоритма триангуляции

$$\tilde{\mathbf{x}}_j = \underset{\tilde{\mathbf{x}}_j}{\operatorname{argmin}} (C(\tilde{\mathbf{x}}_j, \tilde{\mathbf{p}}_j, \mathbf{k})), \quad (1)$$

где C – оценочная функция [1, 2, 6]. В данной работе используется оценочная функция на основе расстояния в ГСК от точки $\tilde{\mathbf{x}}_j$ до лучей $\tilde{\mathbf{l}}_{i,j}$, полученных при помощи обратного преобразования $\mathbf{l}_i = E_i^{-1}(P_i^{-1}(\mathbf{p}_i))$. В итоге среднее значение измеренных трехмерных координат $\hat{\mathbf{x}}$ и соответствующая матрица ковариации $\Sigma_{\mathbf{x}}$ определяются как

$$\hat{\mathbf{x}} = \sum_{j=0}^{2N} W_j \tilde{\mathbf{x}}_j, \quad \Sigma_{\mathbf{x}} = \sum_{j=0}^{2N} W_j (\tilde{\mathbf{x}}_j - \hat{\mathbf{x}})(\tilde{\mathbf{x}}_j - \hat{\mathbf{x}})^T. \quad (2)$$

Данный метод не требует вычисления аналитических производных, позволяет анализировать математическую модель камеры и алгоритм триангуляции как «черный ящик», а также легко обобщается для вычисления погрешности любых геометрических параметров, измеряемых по M опорным точкам. Так, для расстояния r между двумя точками ($M = 2$) в качестве исходных данных будут выступать вектор $\bar{\mathbf{p}} = (\bar{\mathbf{p}}_{1,1}^T, \bar{\mathbf{p}}_{2,1}^T, \bar{\mathbf{p}}_{1,2}^T, \bar{\mathbf{p}}_{2,2}^T)^T$ размерности $N = 4M = 8$, где $\bar{\mathbf{p}}_{i,k} = P_i(E_i(\tilde{\mathbf{x}}_k))$, $k = 1 \dots M$, и матрица ковариации $\Sigma_{\mathbf{p}}$ размера $N \times N$, т. е. 8×8 . После того, как вычислены координаты $\tilde{\mathbf{x}}_{j,k}$ в соответствии с (1) и определены расстояния $\tilde{r}_j = \|\tilde{\mathbf{x}}_{j,1} - \tilde{\mathbf{x}}_{j,2}\|$, среднее значение измеренного расстояния \hat{r} и соответствующее СКО σ_r вычисляются аналогично (2). Таким же образом осуществляется модификация метода для вычисления площади треугольника ($M = 3$), расстояния от точки до плоскости ($M = 4$) и других геометрических параметров.

Отметим, что для данного метода, как и для других методов, использующих преобразование распределения вероятностей, требуется задать матрицы ковариации погрешности измерения координат на изображениях для всех точек, участвующих в рассматриваемом типе измерений. Обычно при предварительных расчетах допускают, что погрешность независима для всех координат, изображений и точек, и используют один параметр σ_p для ее описания [1, 2, 9, 10, 16], т. е. $\Sigma_{\mathbf{p}} = \sigma_p \mathbf{I}$, где \mathbf{I} – единичная матрица размера $N \times N$. Это допущение предполагает, что объект четко определен по обеим координатам (точка, угол, перекрестие, центр круга и т. д.) и основной причиной погрешности является дискретизация и равномерное размытие изображения при наличии некоррелированного шума.

Для выполнения исследований было разработано программное обеспечение (ПО), позволяющее проводить при помощи УТ-метода оценку погрешности различных типов измерений. Данное ПО также было использовано для компьютерного моделирования регистрации точек рабочего объема и получения синтетических данных. При этом модель формирования изображения $\bar{\mathbf{p}}_{i,k} = P_i(E_i(\tilde{\mathbf{x}}_k))$ используется для вычисления вектора $\bar{\mathbf{p}}$ «идеальных» координат набора из M опорных точек, к которым добавляется случайный вектор с заданной матрицей ковариации $\Sigma_{\mathbf{p}}$. Таким образом можно провести проверку правильности работы УТ-метода с помощью метода Монте-Карло для конкретных условий: заданных параметров системы регистрации, рабочего объема и матрицы ковариации погрешности измерения координат на изображениях.

3. Экспериментальные исследования

Для проведения экспериментов использовался видеозонд Mentor Visual iQ VideoProbe производства GE Inspection Technologies с зондом диаметром 6,1 мм и съемным стереоскопическим адаптером прямого обзора XLG3TM616060FG, обеспечивающим поле зрения каждого канала $60^\circ \times 60^\circ$ [5]. Сначала были зарегистрированы калибровочные серии изображений плоских тест-объектов в виде шахматной доски под различными углами к оси зонда на расстояниях 10..45 мм от него. Изображения проверочной серии были получены при линейном перемещении тест-объектов вдоль оси зонда в диапазонах 12..28 мм и 22..42 мм с шагом 1 мм. Определение координат узлов сетки тест-объекта на изображениях производилось автоматически, согласно алгоритму, описанному в [12]. Пример изображения из калибровочной серии с отмеченными координатами узлов показан на рис. 2.

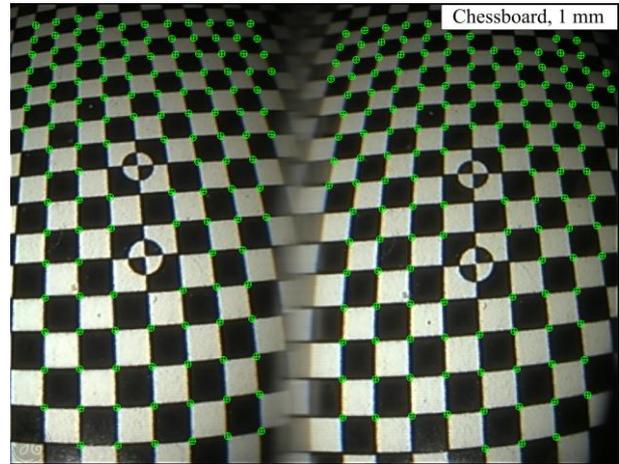


Рис. 2. Пример изображения тест-объекта с вычисленными координатами узлов.

По калибровочной серии изображений были определены параметры трассировочной модели формирования изображения. Как показано в [6, 7], такая модель более подходит для эндоскопа с призмочно-линзовой ОС, чем стандартная проективная модель с полиномиальной дисторсией, и не вносит заметной систематической погрешности. Затем изображения проверочных серий были использованы для вычисления трехмерных координат всех узлов при каждом положении тест-объекта, а также геометрических параметров: расстояния между соседними узлами (x - и y -отрезки, ориентация системы координат приведена на рис. 1), сдвига узла при линейном перемещении (z -отрезки) и расстояния от точки до плоскости. Далее, результаты калибровки, координаты узлов и вычисленные положения тест-объекта были использованы для оценки погрешности измерения тех же геометрических параметров УТ-методом при матрице ковариации $\Sigma_{\mathbf{p}} = \sigma_p \mathbf{I}$ для нескольких значений σ_p .

Полученные значения погрешности $\delta_r = r - \bar{r}$ (r – измеренное значение, \bar{r} – номинальное значение для данного типа измерений) демонстрируют сложную зависимость погрешности измерения от координат точки в пределах рабочего объема. Для проведения количественной оценки рабочий объем был разделен на зоны по z -координате и были вычислены для δ_r среднее значение Δ_r и СКО σ_r в пределах каждой зоны. Результаты для диапазона 12..28 мм показаны на рис. 3 и 4. Номинальные значения отрезков равны 1 мм для всех измерений. Для

сравнения с результатами оценки погрешности УТ-методом рассчитанные с его помощью значения \hat{r} и σ_r также были усреднены в пределах каждой зоны

$$\Delta_r = \frac{1}{L} \sum_{i=1}^L \hat{r}_i, \quad \sigma_r = \sqrt{\frac{1}{L} \left(\sum_{i=1}^L \sigma_{r_i}^2 + (\hat{r}_i - \Delta_r)^2 \right)},$$

где L – количество измерений в пределах данной зоны. Полученные значения показаны на рис. 3 и 4 для σ_p , равного 0,05 пикселей, 0,1 и 0,15 пикселей.

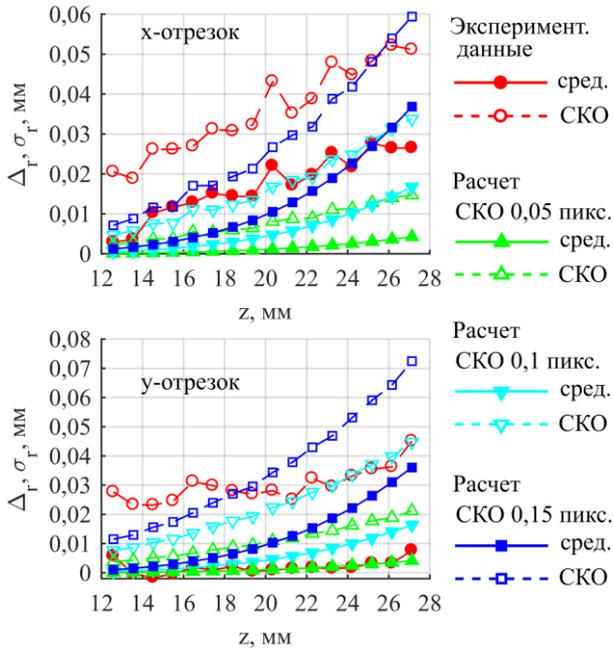


Рис. 3. Сравнение экспериментальных данных с результатами расчетов для измерения отрезков, перпендикулярных оси зонда.

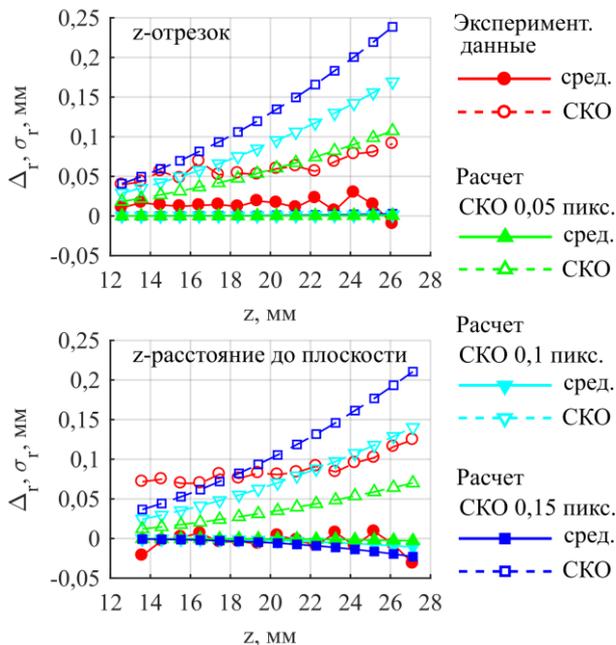


Рис. 4. Сравнение экспериментальных данных с результатами расчетов для измерения отрезков, параллельных оси зонда.

Как видно из показанных графиков, экспериментальные значения для z -отрезков соответствуют заметно меньшему значению σ_p по сравнению с другими измерениями.

Наиболее вероятной причиной является присутствие корреляции между погрешностями при определении координат одной точки на соседних изображениях серии. В таком случае матрицу ковариации Σ_p размером 8×8 можно представить в виде блочной матрицы

$$\Sigma_p = \begin{pmatrix} \Sigma'_p & \kappa_1 \Sigma'_p \\ \kappa_1 \Sigma'_p & \Sigma'_p \end{pmatrix},$$

где Σ'_p – матрица ковариации для одной точки (размером 4×4), κ_1 – коэффициент корреляции. В самом деле, если провести расчеты УТ-методом для z -отрезков при различных коэффициентах корреляции κ_1 и $\Sigma'_p = \sigma_p \mathbf{I}$, то становится очевидно, что увеличение коэффициента κ_1 приводит к тем же качественным изменениям, что и уменьшение σ_p . Аналогичный вывод следует из сравнения с результатами для измерения расстояния от точки до плоскости, где такая корреляция не может проявиться и результат гораздо ближе к результату для y -отрезков.

Также было проведено исследование влияния коэффициента корреляции погрешности проекций одной точки на двух частях изображения. При этом

$$\Sigma''_p = \begin{pmatrix} \Sigma''_p & \kappa_2 \Sigma''_p \\ \kappa_2 \Sigma''_p & \Sigma''_p \end{pmatrix},$$

где Σ''_p – матрица ковариации (размером 2×2) для одной проекции точки, κ_2 – коэффициент корреляции, κ_1 при этом равен нулю. Результаты расчетов также свидетельствуют о том, что увеличение коэффициента κ_2 аналогично уменьшению σ_p . Этот эффект одинаково проявляется для всех типов измерений. Таким образом, наличие данного типа корреляции не может служить причиной того, что экспериментальные значения для x -отрезков соответствуют большему значению σ_p . Причиной может быть систематическая погрешность, возникающая при калибровке, или другие не учтенные факторы. Кроме того, заметно, что экспериментальные значения для всех типов измерений содержат величину, не зависящую от расстояния, которая может соответствовать погрешности изготовления тест-объекта.

Для проверки используемого метода оценки погрешности используемого метода аналогичные расчеты были проведены для синтетических данных с заданной матрицей ковариации погрешности координат на изображениях по тем же результатам калибровки и вычисленным положениям тест-объекта. Применимость УТ-метода при данных параметрах была подтверждена для всех точек рабочего объема и всех использованных матриц ковариации.

4. Заключение

Проведенные исследования показали, что даже при использовании простого тест-объекта и автоматической обработки изображений полученные значения ошибки измерений не вполне совпадают с расчетными значениями при независимом распределении для всех координат, всех изображений и точек и одинаковом значении СКО для всех точек рабочего объема. Показано, что наличие корреляции, которую обычно не учитывают, способно заметно изменить результат. Тем не менее, положительное значение коэффициента корреляции уменьшает погрешность и одинаково влияет на разные измерения, а значит, расчет для нулевого коэффициента корреляции все равно можно использовать в качестве расчета для наихудшего случая

при сравнении разных вариантов и оптимизации на этапе предварительных расчетов. Из тех же соображений можно использовать такой расчет для выдачи оператору предупреждения о недостаточной точности измерений. Тем не менее, к абсолютным значениям погрешности, рассчитанным при данных допущениях, стоит относиться с осторожностью.

Достоверный учет других факторов, влияющих на вид матрицы ковариации погрешности координат точек на изображении для каждой точки рабочего объема, весьма затруднителен. Такими факторами являются текстура объекта, по-разному проявляющаяся при разном увеличении и освещении, асимметричное пятно рассеяния ОС, искажения, возникающие в электронном тракте. Поставить эксперимент так, чтобы разделить эти факторы, не представляется возможным. Более точные расчеты требуют детального моделирования процесса регистрации изображений с учетом освещенности и работы конкретного алгоритма определения соответствий. Такой подход не применим для предварительных вычислений, но может быть в дальнейшем использован для более подробных расчетов на следующих этапах проектирования совместно с разработкой алгоритмов обработки изображений.

5. Благодарности

Исследование выполнено за счет гранта Российского научного фонда (проект №17-19-01355). В части компьютерного моделирования исследование поддержано Российским фондом фундаментальных исследований (проект 17-29-03469).

6. Литература

- [1] Горевой, А.В., Колочкин, В.Я. Методы оценки погрешности измерения координат в комплексированных системах регистрации трехмерных образов объектов. *Инженерный журнал: наука и инновации* **9**: 45 (2013).
- [2] Горевой, А.В., Мачихин, А.С., Перфилов, А.М. Вычисление погрешности бесконтактного измерения площади дефектов на поверхностях сложной формы при видеоэндоскопическом контроле. *Научно-технический вестник информационных технологий, механики и оптики* **4(92)**: 140-148 (2014).
- [3] Мачихин, А.С. Измерительные возможности современных видеоэндоскопов. *Двигатель* **3**: 8-9 (2009).
- [4] Мачихин, А.С., Батшев, В.И. Проблемы создания стереоскопических объективов для видеоэндоскопов. *Инженерный журнал: наука и инновации* **7**: 1-6 (2013).
- [5] General Electric Company. Mentor Visual iQTM VideoProbeTM. https://www.gemeasurement.com/sites/gemc.dev/files/gea31352b_mentor_visual_iq_spect_sheet_r4.pdf
- [6] Gorevoy, A.V., Machikhin, A.S. Calibration of miniature prism-based stereoscopic imagers for precise spatial measurements. *Proc. SPIE* **9917**: 991707 (2016).
- [7] Gorevoy, A.V., Machikhin, A.S., Shurygin, A.V., Khokhlov, D.D., Naumov, A.A. 3D spatial measurements by means of prism-based endoscopic imaging systems. *ГРАФИКОН'2016 Труды 26-й Международной научной конференции*, 253-256 (2016).
- [8] Julier, S.J., Uhlmann, J.K. A consistent, unbiased method for converting between polar and cartesian coordinate systems. *The Proceedings of AeroSense: Acquisition, Tracking and Pointing XI* **3086**: 110-121 (1997).
- [9] Kee, W.L., Bai, Y., Lim, K.B. Parameter error analysis of singlelens prism-based stereovision system. *J. Opt. Soc. Am. A* **32**: 367-373 (2015).
- [10] Lia, Y., Zhang, J., Tian, J. Error Analysis in Stereo Vision for Location Measurement of 3D Point. *Proc. SPIE* **9813**: 981315 (2015).
- [11] Lim, K.B., Qian B. Biprism distortion modeling and calibration for a single-lens stereovision system. *J. Opt. Soc. Am. A* **33**: 2213-2224 (2016).
- [12] Mallon J., Whelan P.F.: Which pattern? Biasing aspects of planar calibration patterns and detection methods, *Pattern Recognition Letters* **28**: 921-930 (2007).
- [13] Nakano, S., Hori, F., Ogawa, K. Endoscopic device and endoscopic image distortion correction method. US Patent US20100004507A1 (2010).
- [14] Ng, A. Advances in 3D Video Borescope Measurement Technologies for Defect Characterization. *Proc. of NDT Malaysia* (2015). http://www.ndt.net/events/MINDTCE-15/app/content/Paper/33_Ng.pdf
- [15] Olympus Corporation. Последние достижения в области дистанционного визуального контроля и стереоизмерений. http://www.olympus-ims.com/en/downloads/download/?file=285216841&fl=ru_RU
- [16] Rodriguez, J.J., Aggarwal, J.K. Stochastic analysis of stereo quantization error. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**: 467-470 (1990).
- [17] Wu, L., Zhu, J., Xie, H. A modified virtual point model of the 3d dic technique using a single camera and a bi-prism. *Measurement Science and Technology* **25**: 115008 (2014).
- [18] Zhang, S. *Handbook of 3D Machine Vision: Optical Metrology and Imaging, Series in Optics and Optoelectronics*. CRC Press (2016).
- [19] Zhu, Y.-K., Tian, G.-Y., Lu, R.-S., Zhang, H. A review of optical NDT technologies. *Sensors* **11**: 7773-7798 (2011).

Об авторах

Горевой Алексей Владимирович — м.н.с. НТЦ УП РАН, инженер НИУ МЭИ, инженер МГТУ им. Н.Э. Баумана. Email: gorevoy.a@gmail.com.

Мачихин Александр Сергеевич — в.н.с. НТЦ УП РАН, доцент НИУ МЭИ. Email: aalexanderr@mail.ru.

Сравнение некоторых методов решения задачи детектирования лиц на изображениях

Е.А. Долотов¹, В.Д. Кустикова¹

dolotov.evgeniy@gmail.com|valentina.kustikova@itmm.unn.ru

¹ Институт информационных технологий, математики и механики

Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

Рассматривается задача детектирования лиц на изображениях. Проводится сравнение некоторых методов решения задачи, основанных на построении сверточных нейронных сетей. Сравнимые методы относятся к числу передовых, поскольку демонстрируют одни из лучших результатов на широко известном наборе данных Face Detection Data Set and Benchmark (FDDDB). Разрабатываются программные реализации методов, воспроизводятся результаты детектирования лиц на FDDDB. Эксперименты показывают сравнимые значения показателей качества с опубликованными на официальной странице FDDDB. Оценивается качество детектирования лиц на более сложных данных, входящих в состав набора WIDER FACE. Проводится анализ полученных результатов.

Ключевые слова: детектирование лиц, глубокое обучение, сверточные нейронные сети, Face Detection Data Set and Benchmark (FDDDB), WIDER FACE.

Comparison of some methods for solving the problem of face detection in images

E.A. Dolotov¹, V.D. Kustikova¹

dolotov.evgeniy@gmail.com|valentina.kustikova@itmm.unn.ru

¹ Institute of Information Technologies, Mathematics and Mechanics,

Lobachevsky State University of Nizhni Novgorod, Nizhni Novgorod, Russia

We consider the problem of face detection in images. We make a comparison of some methods, based on convolutional neural networks. The compared methods are among the advanced, as they demonstrate the best results on the widely known Face Detection Data Set and Benchmark (FDDDB) data set. We develop a software implementation of these methods and reproduce the results of face detection on FDDDB. The experiments show comparable values of the quality measurements with those published on the official FDDDB web page. We assess and analyse the quality of face detection on more complex data included in the WIDER FACE dataset.

Keywords: face detection, deep learning, convolutional neural networks, Face Detection Data Set and Benchmark (FDDDB), WIDER FACE.

1. Введение

Задача детектирования лиц на изображениях является одной из классических задач компьютерного зрения. Несмотря на простоту формулировки, задача является достаточно сложной в силу изменчивости визуального образа лиц. Изменчивость обусловлена вариативностью внешнего вида, масштаба и ракурса объекта, а также степенью его освещенности.

Цель настоящей работы состоит в том, чтобы оценить качество детектирования лиц с использованием известных методов [9, 12, 15, 18], основанных на обучении сверточных нейронных сетей, на наборе данных WIDER FACE [17], который содержит разнообразные изображения лиц, полученные в реальных условиях.

Работа построена следующим образом. Вначале дается общая схема работы каждого из рассматриваемых методов детектирования лиц. Приводится краткое описание программных реализаций. Выполняется сравнение качества на наборе данных Face Detection Data Set and Benchmark (FDDDB) [3]. Оценивается качество детектирования лиц на более сложных данных, входящих в состав набора WIDER FACE [17]. Проводится анализ полученных результатов.

2. Задача детектирования лиц

Задача детектирования лиц состоит в том, чтобы определить положение всех лиц людей на изображении. Положение определяется прямоугольником, окаймляющим границы лица.

3. Сравнимые методы

Существует большое количество различных методов детектирования лиц [1, 4, 8, 14, 20]. В настоящее время наилучшее качество показывают алгоритмы, основанные на применении сверточных нейронных сетей (Convolutional Neural Network, CNN). Различия алгоритмов данной группы, как правило, обнаруживаются в архитектурах сетей и параметрах их обучения [9, 14]. Наряду с этим, могут отличаться методы обработки признаков [4, 12], а также подходы к решению проблемы поиска лиц разного масштаба [8].

Множество методов, основанных на применении сверточных нейронных сетей, можно разделить на две группы:

– Методы, использующие выход предварительно обученной сверточной сети в качестве признакового описания изображения [12]. Построенное признаковое описание далее используется для обучения традиционных классификаторов.

– Методы, предоставляющие на выходе нейронной сети полную информацию о расположении лиц, областей или окаймляющих лица прямоугольников [9].

Рассмотрим более подробно четыре метода детектирования лиц, основанные на применении сверточных нейросетей и демонстрирующие одни из лучших результатов на широко известном наборе данных Face Detection Data Set and Benchmark (FDDDB) [3].

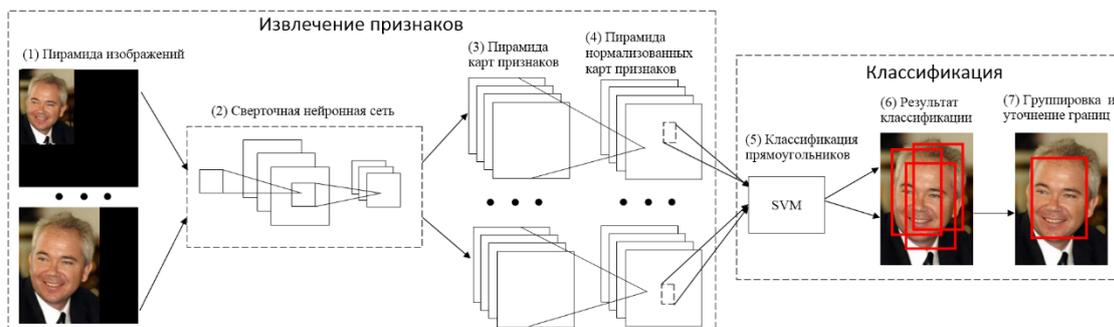


Рис. 1. Схема работы метода, основанного на построении модели деформируемых частей.

3.1 Метод, основанный на построении модели деформируемых частей

В алгоритме детектирования лиц [12] (A Deep Pyramid Deformable Part Model for Face Detection, DPMFD) для исходного изображения строится пирамида изображений, состоящая из разных масштабов одного изображения (рис. 1). Затем изображение с каждого уровня подается на вход сверточной нейронной сети, которая получена из сети AlexNet [10] посредством удаления последних полносвязных слоев. После обработки пирамиды изображений формируется соответствующая пирамида карт признаков, состоящая из такого же количества уровней, как и пирамида изображений. Каждый вектор признаков пирамиды признаков нормализуется. Далее осуществляется проход скользящим окном с единичным шагом по нормализованным картам признаков, извлекаются все прямоугольные области фиксированного размера. Полученные карты признаков преобразуются в вектора, которые затем классифицируются с помощью обученной линейной машины опорных векторов. При этом расстояние до гиперплоскости, разделяющей два класса, принимается за достоверность принадлежности к классу лиц. После этого этапа каждому лицу на изображении может соответствовать несколько прямоугольников. Поэтому необходимо объединить такие прямоугольники, а затем уточнить их границы, используя алгоритм, описанный в [5].

3.2 Метод, основанный на полностью сверточной сети, с малым количеством весов

В большинстве алгоритмов детектирования лиц применяются очень глубокие сверточные нейронные сети, которые имеют большое количество весов (AlexNet [10], VGG-16 [13], ResNet [6]). Метод [15] (FastCNN), описанный в данном разделе, доказывает, что эту задачу можно решить с помощью сверточной нейронной сети со значительно меньшим количеством слоев и весов.

Сначала обучается сеть, решающая задачу бинарной классификации «лицо/не лицо». Сеть принимает на вход трехканальное изображение в формате RGB с разрешением 32×32 пикселя. Затем изображение последовательно обрабатывается с помощью семи сверточных слоев. После каждого сверточного слоя применяется активационная функция PReLU [7], которая является кусочно-линейной, что позволяет делать нелинейные преобразования с выходом предыдущего слоя и в тоже время быстро вычислять производную. Выход сети после каждого сверточного слоя имеет меньший размер, это свойство позволяет решать задачу классификации без применения слоев другого типа. Выходом последнего сверточного слоя является пара чисел, каждое из которых определяет достоверность принадлежности к одному из двух классов. Сеть с такой архитектурой имеет примерно в 800 раз меньше весов, чем сеть AlexNet. Такое значительное уменьшение количества весов позволяет намного сократить

время работы детектора, а также сократить время, необходимое для обучения сети.

Так как описанная сеть является полностью сверточной, то она может быть применена к изображению произвольного размера. В результате всех вычислений на выходе сети получаются две матрицы, которые описывают достоверности принадлежности конкретной области изображения к одному из классов. Для уменьшения количества ложных срабатываний детектора каждая из этих матриц усредняется с окрестностью 3×3 . После чего по любой из этих двух матриц можно определить размеры и положения окаймляющих прямоугольников.

3.3 Метод, основанный на применении нестандартной функции потерь

При обучении нейронной сети для вычисления ошибки, как правило, используется метрика L_2 – евклидово расстояние между вектором, полученным на выходе нейронной сети, и вектором из обучающего набора данных. Этот подход имеет несколько недостатков. Во-первых, при использовании такой функции ошибки параметры оптимизируются независимо. Во-вторых, функция не является нормированной. Это приводит к тому, что ошибка на один пиксель на прямоугольнике большого размера обеспечивает такой же вклад, как и ошибка на прямоугольнике меньшего размера. В результате детектор обращает больше внимания на крупные объекты и игнорирует мелкие. Чтобы исправить эту ситуацию, необходимо использовать пирамиду из разных масштабов изображения, что негативно сказывается на времени работы детектора.

Авторы метода [18] (UnitBox), описанного в настоящем разделе, вводят новую функцию потерь, которая лишена этих недостатков. Данный подход основан на том, что для каждого пикселя изображения окаймляющий прямоугольник может быть описан с помощью четырехмерного вектора, где каждая компонента представляет собой расстояние от пикселя до верхней, нижней, левой и правой границ прямоугольника. Ошибку определения параметров прямоугольника предлагается вычислять с помощью метрики Intersection over Union (IoU), равной отношению площади пересечения прямоугольника, полученного в результате детектирования, и прямоугольника из разметки к площади их объединения. Использование такой функции ошибки позволяет оптимизировать все параметры прямоугольника одновременно. Также, несмотря на величину прямоугольника, значение величины IoU лежит в отрезке $[0, 1]$, что позволяет обучать нейронную сеть на объектах разного масштаба и выполнять детектирование объектов на изображении без дополнительного масштабирования.

Основой метода детектирования является широко известная глубокая сверточная нейронная сеть VGG-16 [13], из которой удалены все полносвязные слои и добавлены две новые ветки (рис. 2). Первая ветка служит для получения бинарной маски изображения

(одноканальное изображение, состоящие из нулей и единиц: 0 – фон, 1 – лицо), а вторая ветка отвечает за построение границ окаймляющих прямоугольников, которые представляются четырехмерными векторами.

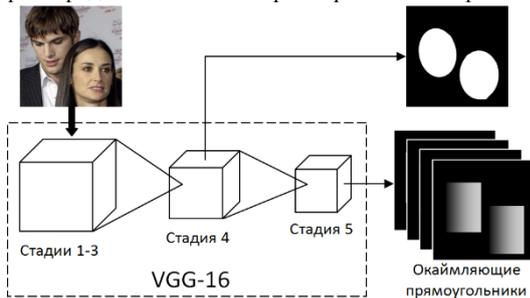


Рис. 2. Схема сверточной нейронной сети в методе UnitBox.

Для получения бинарной маски с помощью нейронной сети VGG-16 в конец четвертой стадии сети добавляются три слоя. Первые два – это сверточный слой с шагом в один пиксель и размером ядра $512 \times 3 \times 3 \times 1$ и слой линейной интерполяции для масштабирования выхода сверточного слоя до размера изображения, поступающего на вход сети. Третий слой обрезает изображение, полученное после линейной интерполяции, до размеров оригинального изображения. На основании полученного одноканального изображения вычисляется сигмоидальная кросс-энтропия, характеризующая ошибку построения бинарной маски. Для получения окаймляющих прямоугольников добавляются четыре дополнительных слоя, среди которых сверточный слой с шагом 1 и размером ядра $512 \times 3 \times 3 \times 4$, слой линейной интерполяции и слой, позволяющий обрезать выход после линейной интерполяции до размеров исходного изображения. На четвертом слое применяется функция ReLU, чтобы сделать выход сети неотрицательным. Затем вычисляется функция ошибки IoU. Общая ошибка сети является взвешенной суммой ошибок на обеих ветках.

К выходу первой ветки сети применяется пороговая функция для получения бинарной маски. На полученной бинарной маске находятся все компоненты связности. Из второй ветки сети извлекаются четырехмерные векторы, находящиеся на той же позиции, что и центры компонент связности. Четырехмерные векторы преобразуются в результирующие окаймляющие прямоугольники.

3.4 Метод, основанный на Faster R-CNN

Алгоритм [9] состоит из двух модулей: полностью сверточная нейронная сеть, которая используется для извлечения признаков из изображения и нахождения предполагаемых прямоугольников, и детектор на основе R-CNN [5], который использует эти прямоугольники. Указанные модули объединяются в одну нейронную сеть (Faster R-CNN, рис. 3).

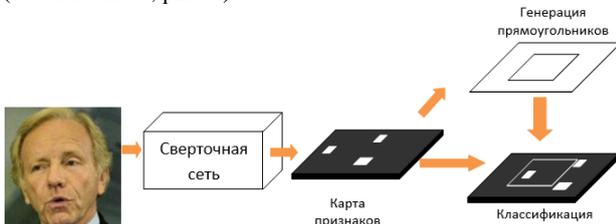


Рис. 3. Схема работы Faster R-CNN.

На первом этапе изображение произвольного размера подается на вход нейронной сети, которая обеспечивает выделение прямоугольников для дальнейшей классификации (Region Proposal Network, RPN). Эта сеть является полностью сверточной нейронной сетью, которая получена из сети VGG-16 путем удаления последних полностью связанных слоев и добавлением одного или нескольких полностью сверточных слоев. Добавленные сверточные слои используются непосредственно для

определения предполагаемых прямоугольников. Добавленные слои обходятся скользящим окном по карте признаков, полученной из сети VGG-16, и для каждой позиции окна извлекается вектор признаков малой размерности. Вектора признаков подаются на вход двум новым полностью связным слоям. Один из этих слоев используется для уточнения границ прямоугольника, а другой для классификации объекта, расположенного внутри этого прямоугольника.

В каждой позиции скользящего окна одновременно могут рассматриваться несколько прямоугольников, которые обладают разным размером или разным соотношением сторон. С помощью такого подхода на изображении одного масштаба можно найти лица разной величины.

После того как предполагаемые прямоугольники, содержащие лица, получены, их необходимо классифицировать. Обычно для этого используется другая сеть, которая обучается отдельно. Но в данном случае предлагается использовать для классификации часть слоев из сети RPN. Это позволяет значительно сократить время детектирования, а также упрощает процесс обучения модели, так как модель представляет собой единую нейронную сеть, которую можно обучать с помощью метода обратного распространения ошибки (Back Propagation). После применения сети на выходе получается набор прямоугольников. Каждому лицу на изображении могут соответствовать несколько различных прямоугольников. Поэтому завершающим этапом детектирования является объединение прямоугольников.

4. Программная реализация

Программная реализация описанных методов [20] выполнена с использованием языка программирования C++ на базе библиотек OpenCV и Caffe (рис. 4).

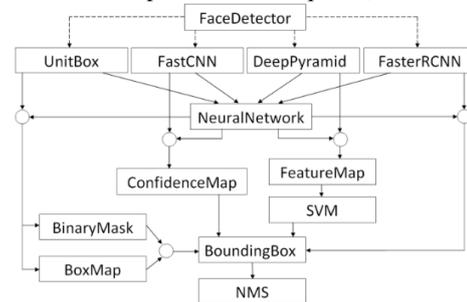


Рис. 4. Схема компонент системы.

1. *FaceDetector*. Базовый класс детектора лиц. *UnitBox*, *FastCNN*, *DeepPyramid*, *FasterRCNN* – наследники класса *FaceDetector*, каждый реализует соответствующий метод.
2. *NeuralNetwork*. Модуль для работы с нейронными сетями на базе библиотеки глубокого обучения Caffe [2].
3. *SVM*. Реализует метод скользящего окна и обеспечивает классификацию областей, накрываемых окном, с помощью метода опорных векторов. Использует реализацию метода из библиотеки компьютерного зрения OpenCV [11].
4. *BoxMap*, *FeatureMap*, *BinaryMask*. Содержат примитивы для представления окаймляющих прямоугольников через четырехмерные векторы, для представления пирамиды карт признаков и для представления бинарной маски.
5. *BoundingBox*. Класс, содержащий информацию об окаймляющих прямоугольниках.
6. *NMS*. Содержит реализацию нескольких стратегий объединения прямоугольников.

5. Критерий оценки качества детектирования

Для оценки качества детектирования с помощью разработанной программной реализации используется показатель, равный отношению площади пересечения прямоугольника, полученного в результате детектирования, и прямоугольника из разметки к площади их объединения (Intersection over Union, IoU). Таким образом, считается, что лицо обнаружено правильно, если данный показатель превышает некоторый порог. В противном случае принимается, что лицо не обнаружено. Срабатывание алгоритма детектирования на области, где лицо отсутствует, считается ложным срабатыванием.

На основании приведенного показателя осуществляется построение ROC-кривой, которая отражает зависимость количества ложных срабатываний алгоритма детектирования (false positive) от точности детектирования (true positive rate).

Построение ROC-кривых обеспечивается с помощью инструментов, предоставляемых разработчиками наборов данных, используемых в ходе апробации.

6. Тренировочные и тестовые данные

В процессе проведения экспериментов для обучения детекторов использовался набор данных WIDER FACE [17], который состоит из 32203 изображений, содержащих 393703 лица, максимальное и минимальное разрешение которых составляет 851×1295 и 1×1 пикселей соответственно.

Изображения разделены на 61 класс (спортивные мероприятия, демонстрации, интервью и т.д.). Набор содержит лица, достаточно сложные для детектирования – с большим перекрытием, в масках, с размытием, малого размера. Тестирование проводится на валидационном множестве набора WIDER FACE, который разделен на несколько частей с различными уровнями сложности, а также на наборе данных Face Detection Data Set and Benchmark (FDDB) [3], который состоит из 2845 изображений, содержащих 5171 лицо, максимальное и минимальное разрешение которых составляет 398×589 и 8×13 пикселей.

7. Результаты экспериментов

Результаты, полученные на данных FDDB, приведены на рисунке (рис. 5). Построенные ROC-кривые говорят о сравнимости результатов. При одинаковом количестве ложных срабатываний точность детектирования в среднем отличается не более чем на 1.5% для каждого из методов, что свидетельствует о корректности разработанных реализаций.

С использованием разработанных программных реализаций проводится оценка качества на наборе данных WIDER FACE. Результаты для четырех описанных методов, а также для методов [8, 20], демонстрирующих лучшие результаты, приведены на рис. 9.

На наборе WIDER FACE лучшие результаты показывает метод, основанный на применении Faster R-CNN. При этом на подмножестве с низким уровнем сложности методы Faster R-CNN и UnitBox имеют близкие результаты к лучшим результатам, опубликованным на официальной странице набора [16], в то время, как на наборах со средней и высокой сложностью они значительно отстают от этих методов. Данный факт объясняется тем, что они хуже распознают лица малой величины (рис. 6).

Методы FastCNN и DPMFD на наборах со средней и низкой сложностью показывают намного худшие результаты, но на наборе с высокой сложностью

результаты всех четырех методов близки, что во многом говорит об ориентированности методов на набор данных FDDB.

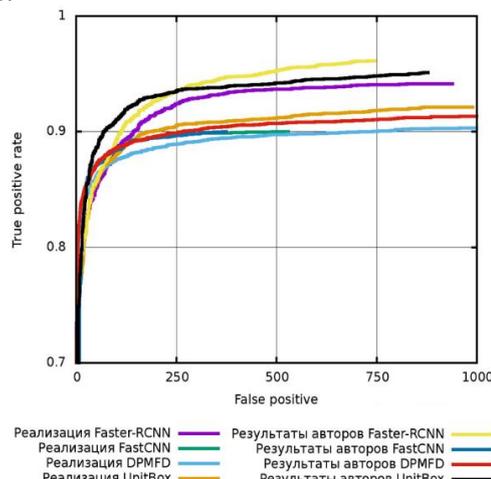


Рис. 5. Результаты работы на FDDB.



Рис. 6. Пример работы Faster R-CNN. Красным отмечены прямоугольники из разметки, зеленым результат поиска.

Тестирование выполняется на двух системах с различными конфигурациями. Конфигурация 1: Ubuntu, Intel(R) Core i5-2430M @ 2.4GHz, 6 GB. Конфигурация 2: CentOS, Intel(R) E5-2660 @ 2.2GHz, GPU: Nvidia Kepler K20, 64 GB. Средние времена работы методов на изображениях из базы WIDER FACE представлены в таблице 1.

Таблица 1. Время работы детекторов, с

	Конфигурация 1	Конфигурация 2
Faster R-CNN	4,5	0,39
FastCNN	5,3	0,42
DPMFD	15,4	5,2
UnitBox	3,9	0,36

Все четыре метода показывают хорошие результаты в довольно нестандартных случаях. Детекторы находят лица как на цветных изображениях, так и на изображениях в оттенках серого (рис. 7). Можно предположить, что это связано с тем, что в обучающем наборе изображений встречались изображения того и другого типа. Лица также находятся и на искусственно созданных изображениях (портреты, кадры из мультфильмов и т.д.). Исходя из этих результатов, можно сделать вывод, что с помощью глубоких сверточных сетей из изображения извлекаются признаки высокого уровня, такие, как границы лица, наличие глаз, носа, рта.

Очень часто на реальных изображениях лицо бывает перекрыто другими предметами. Все детекторы показывают возможность детектирования лиц в подобных случаях. Наилучший результат показывает метод,

основанный на Faster R-CNN. Даже при перекрытии около 60% данный метод достаточно точно определяет положение объекта и сохраняет верные пропорции лица (рис. 8). Метод UnitBox хуже справляется с этой задачей, так как определяет положение лица на основе бинарной маски, которая с данным случае получается искаженной.



Рис. 7. Пример работы DPMFD на изображениях с лицами различных цветов.



Рис. 8. Пример работы Faster R-CNN при перекрытии лица.

По результатам проведенных экспериментов можно сделать вывод, что среди описанных методов наилучшим качеством детектирования обладает метод, основанный на Faster R-CNN. Данный метод в отличие от других, рассмотренных и реализованных в работе, достаточно хорошо определяет лица небольшого масштаба. Методы UnitBox и Faster R-CNN используют только один масштаб изображения при детектировании, что дает им большой выигрыш во времени по сравнению с двумя другими методами.

8. Заключение

В данной работе рассмотрены четыре метода решения задачи детектирования лиц. Описанные методы показывают, что задачу детектирования лиц можно решать разными способами. Разработана программная реализация этих методов на базе библиотек компьютерного зрения OpenCV [11] и глубокого обучения Caffe [2]. Программная реализация выложена в открытый доступ [20]. Воспроизведены результаты детектирования лиц с использованием разработанной реализации на данных Fddb. Эксперименты показывают сравнимые значения показателей качества с опубликованными на официальной странице Fddb [3]. Проведена оценка качества детектирования на наборе данных WIDER FACE. Методы Faster R-CNN и UnitBox на наборе с низкой сложностью показывают результаты, близкие к наилучшим опубликованным, но на наборах со средней и высокой степень сложности они значительно им проигрывают, что

говорит об ориентированности указанных методов на более простой набор данных Fddb.

9. Литература

- [1] Barbu A., Lay N., Gramajo G. Face Detection with a 3D Model. – [https://arxiv.org/abs/1404.3596]. – 2015.
- [2] Caffe [http://caffe.berkeleyvision.org].
- [3] Face Detection Data Set and Benchmark Home [http://www.cs.umass.edu/fddb].
- [4] Farfadi S., Saberian M., Li L. Multi-view Face Detection Using Deep Convolutional Neural Networks. – [https://arxiv.org/abs/1502.02766]. – 2015.
- [5] Girshick R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation // In the Proc. of the Conf. on Computer Vision and Pattern Recognition. – 2014.
- [6] He K., et al. Deep Residual Learning for Image Recognition. – [https://arxiv.org/abs/1512.03385]. – 2015.
- [7] He K., et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. – [https://arxiv.org/abs/1502.01852]. – 2015.
- [8] Hu P., Ramanan D. Finding Tiny Faces. – [https://arxiv.org/abs/1612.04402]. – 2017.
- [9] Jiang H., Learned-Miller E. Face Detection with the Faster R-CNN. – [https://arxiv.org/abs/1606.03473]. – 2016.
- [10] Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. – 2012.
- [11] OpenCV [http://opencv.org].
- [12] Ranjan R., Patel V.M., Chellappa R. A Deep Pyramid Deformable Part Model for Face Detection. – [http://arxiv.org/abs/1508.04389]. – 2015.
- [13] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. – [https://arxiv.org/abs/1409.1556]. – 2015.
- [14] Sun X., Wu P., Hoi S. Face Detection using Deep Learning: An Improved Faster RCNN Approach. – [https://arxiv.org/abs/1701.08289]. – 2017.
- [15] Triantafyllidou D., Tefas A. A Fast Deep Convolutional Neural Network for Face Detection in Big Visual Data. – [http://arxiv.org/abs/1508.04389]. – 2015.
- [16] WIDER FACE: A Face Detection Benchmark Home [http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace].
- [17] Shuo Y., et al. WIDER FACE: A Face Detection

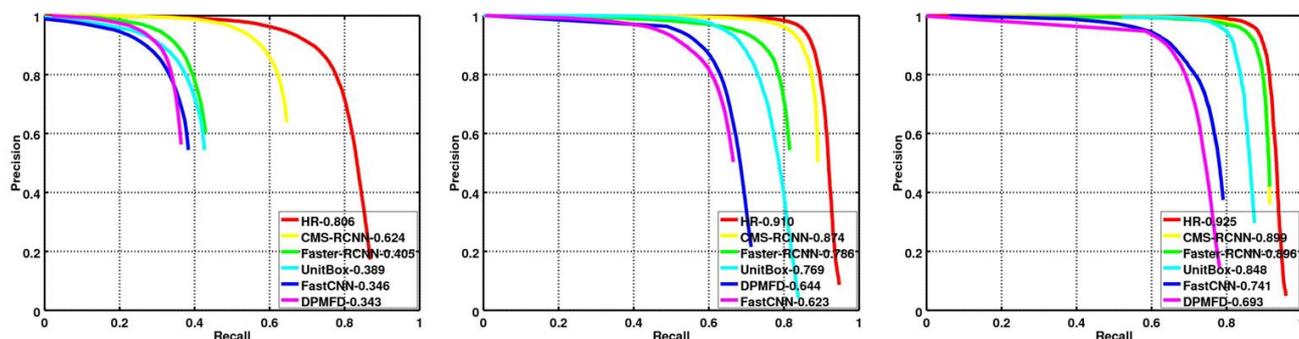


Рис. 9. Результаты работы на валидационной части набора WIDER FACE на изображениях высокой, средней и низкой сложности (слева направо). Показаны результаты методов: HR – Hybrid resolution [8], CMS-RCNN – Contextual Multi-Scale Region-based CNN [20], Faster RCNN – Faster Region-based CNN [9], UnitBox [18], FastCNN [15], DPMFD [12]. Для каждого из методов рядом с названием указано среднее значение точности.

- Benchmark. – [<https://arxiv.org/abs/1511.06523>]. – 2015.
- [18] Yu J., Jiang Y., Wang Z., Cao Z., Huang T.. UnitBox: An Advanced Object Detection Network. – INNS. – 2016.
- [19] Zhang K., Zhang Zh., Li Z. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. – [<http://arxiv.org/abs/1604.02878>]. – 2016.
- [20] Zhu C., et al. CMS-RCNN: Contextual Multi-Scale Region-based CNN for Unconstrained Face Detection. – [<https://arxiv.org/abs/1606.05413>]. – 2016.
- [21] Разработанная программная реализация [<https://github.com/DolotovEvgeniy/FaceDetection>].

Об авторах

Евгений Долотов, студент Института информационных технологий, математики и механики ННГУ им. Н.И. Лобачевского. Его e-mail dolotov.evgeniy@gmail.com.

Валентина Кустикова, к.т.н., старший преподаватель кафедры Математического обеспечения и суперкомпьютерных технологий Института информационных технологий, математики и механики ННГУ им. Н.И. Лобачевского. Ее e-mail valentina.kustikova@itmm.unn.ru.

Гибридный метод распознавания контуров на изображении на основе технологий компьютерного зрения

Р.А.Клестов¹, В.Ю.Столбов¹
klestovroman@gmail.com|valeriy.stolbov@gmail.com
¹ПНИПУ, Пермь, Россия

Рассматривается гибридный метод распознавания контуров, включающий несколько этапов обработки исходного изображения. Выделяются и описываются характерные особенности существующих методов распознавания контуров. Предлагается определенная последовательность применения методов. Особое внимание уделено методам, обеспечивающим работу со сложными входными данными, имеющими близкую к однородной структуру. Описывается математическая особенность работы каждого из методов. Приводятся примеры выходных данных при применении различных методов обработки изображения. На основе практического использования каждой из представленных технологий компьютерного зрения делаются соответствующие выводы и рекомендации. Проведено сравнение качества распознавания контуров на изображении в иллюстрациях.

Ключевые слова: компьютерная графика, компьютерное зрение, гибридный метод, цепной код Фримена, оператор Кэнни, Куwahara.

Complex recognizing contours method for an image based on computer vision technology

R.A.Klestov¹, V.Y.Stolbov¹
klestovroman@gmail.com|valeriy.stolbov@gmail.com
¹PNRPU, Perm, Russia

We consider a hybrid method for recognizing contours, which includes several stages of processing the original image. Identify and describe the characteristics of existing methods for recognizing contours. A certain sequence of methods is proposed. Particular attention is paid to methods that ensure work with complex input data having a close to homogeneous structure. A mathematical feature of the work of each of the methods is described. Examples of output data are given when applying various image processing methods. Based on the practical use of each of the presented computer vision technologies, appropriate conclusions and recommendations are made. The quality of contour detection in the image in the illustrations is compared.

Keywords: computer graphics, computer vision, complex method, Freeman Chain Code, Canny edge detector, Kuwahara.

1. Введение

Область компьютерного зрения является действительно инновационно привлекательной. Интерес к ней возник на заре попыток создания искусственного интеллекта [4]. В настоящее время количество новых решений и актуальных приложений для компьютерного зрения продолжает расти.

С увеличением количества предлагаемых способов и алгоритмов распознавания контуров внутри изображения становится неочевидным выбор какого-либо конкретного способа для определенных нужд. В данной статье рассматриваются основные используемые алгоритмы на примере их применимости к конкретной прикладной задаче выделения контуров на изображении.

Существует множество подходов к выделению границ, но практически все можно разделить на две категории [5]: методы, основанные на поиске максимумов, и методы, основанные на поиске нулей. Методы, основанные на поиске максимумов, выделяют границы с помощью вычисления «силы края», представленной в виде выражения градиента изображения, а затем поиска локальных максимумов силы края, используя предполагаемое направление границы, обычно перпендикулярного к вектору градиента. Методы, основанные на поиске нулей, ищут пересечения оси абсцисс выражения второй производной, которые обычно представляют собой нули лапласиана или нули нелинейного дифференциального уравнения. В качестве шага предобработки к выделению границ практически

всегда применяется сглаживание изображения, обычно фильтром Гаусса [5].

Хотя многие известные методы выделения границ основываются на вычислении градиента изображения, они отличаются типами фильтров, применяемых для вычисления градиентов в различном направлении. Предлагаемый в данной статье подход использует несколько уже существующих методов выделения границ с определенными параметрами поиска контуров для достижения более точного результата распознавания DICOM-изображения.

При этом обоснована наилучшая комбинация применяемых методов для решения задачи распознавания контура печени человека на примере DICOM изображения. Сложностью данной задачи является существенная неоднородность границ контура печени у больного человека, что требует использования различных групп методов поиска границ.

2. Поэтапное распознавание изображений

Условия и потребности при работе с алгоритмами компьютерного зрения могут быть самыми разными, поэтому постараемся максимально полно и поэтапно рассмотреть последовательность обработки входных данных.

Первым и необязательным этапом является необходимость размыть изображение. Выполняется эта операция из соображений более точного определения границ, так как шаг между градациями цветов будет выше, нежели на оригинальном изображении. Существует достаточно много фильтров, обеспечивающих сглаживание.

Наиболее точным в лабораторных условиях оказался фильтр Kuwahara, особенно в сером градиенте, что используется достаточно часто. Дополнительным «бонусом» является побочная чистка изображения от графического шума, что избавляет нас от необходимости дополнительной обработки входных данных. На рис.1 демонстрируется обработка входного изображения (а), фильтрами Гаусса (б), медианным фильтром (в) и фильтром Kuwahara (г).



Рис. 1. Обработка изображения: а) исходное; б) фильтр Гаусса; в) медианный; г) Kuwahara.

Фильтр Kuwahara выполняет нелинейную фильтрацию изображений с сохранением резких краев. После фильтрации изображение похоже на грубо нарисованную красками картину. Суть алгоритма состоит в следующем (рис. 2):

1. В окрестности каждого пикселя применить маску размером $(2r + 1) * (2r + 1)$;
2. Окрестность разделить на 4 области, указанные цифрами I, II, III, IV, размером $r * r$;
3. Вычислить среднюю интенсивность и дисперсию в каждой из областей I, II, III, IV;
4. Присвоить центральному пикселю X среднее значение интенсивности той области, дисперсия которой наименьшая.

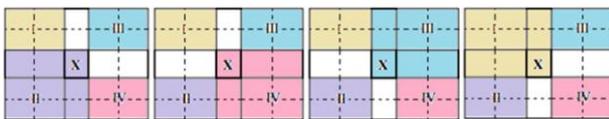


Рис. 2. Маска фильтра.

Применение выбранного фильтра освобождает от необходимости предварительного очищения изображения от шумов, забирая эту функцию на себя. Причём информация о границах объектов в изображении будет сохранена. Применяем к исходному изображению фильтр Kuwahara. На рис. 3 слева показано исходное изображение, а на рис. 3 справа – изображение, после применения фильтра Kuwahara.

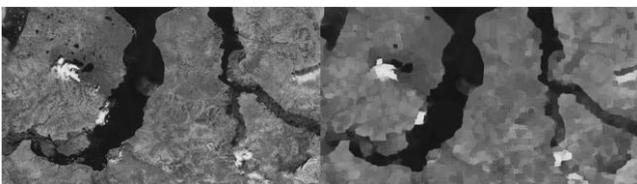


Рис. 3. Исходное изображение(слева) и рисунок, обработанный фильтром Kuwahara (справа).

Как наглядно видно, контуры изображения стали наиболее заметны и к данному рисунку требуется применить контурный фильтр. В данном случае будем использовать детектор границ Кэнни [1], использующий, в свою очередь, фильтр Собеля. Для функционирования оператора требуется две матрицы. По ним к входному изображению применяется свертка:

$$G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} * A \quad \text{и} \quad G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} * A$$

В каждой точке изображения приближенное значение градиента можно вычислить, используя полученные приближенные значения производных по следующей формуле:

$$G = \sqrt{G_x^2 + G_y^2}, \tag{1}$$

где A – исходное изображение, G_x и G_y – результат (число) операции свертки по выбранным матрицам, а знак $*$ обозначает саму двумерную операцию свертки. Координата x возрастает слева направо, а y – сверху вниз.

Действие алгоритма сводится к тому, что вначале будет произведен поиск градиентов, т.е. границы будут отмечаться там, где градиент изображения приобретает максимальное значение. Только локальные максимумы будут отмечены как границы, а потенциальные границы будут определены порогами. Итоговые границы определяются путем подавления всех краев, несвязанных с определенными (сильными) границами. Принцип такого подавления проиллюстрирован на рис.4. Так из исходного изображения будет получен его контурный вариант.

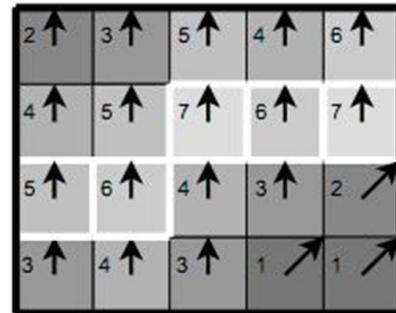


Рис. 4. Принцип подавления границ.

На рис.5 показан конечный результат изображения. Полученные контуры продемонстрированы на снимке печени, как пример применения алгоритмов. Следует обратить внимание что в данном случае также применен фильтр по минимальному радиусу и позиционированию контура, для отсекация ложных выделений. Но только получить данный контур недостаточно. Его нужно представить в виде кода, который можно было бы хранить.

Один из наиболее используемых методов кодирования контуров – цепной код Фримена [5]. Цепные коды применяются для конвертации границы в последовательность прямых определенной длины и указателя направления. В основе этого лежит 2х или 3х-битная решетка на 4 и 8 направления соответственно. Длина определяется разрешением решетки, а направление – указателем из решетки. (рис. 6).

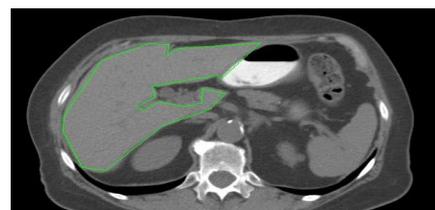


Рис. 5. Пример работы оператора Кэнни.

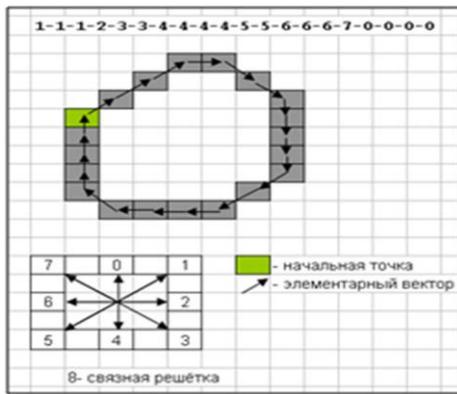


Рис. 6. Цепной код Фримена.

На контуре фиксируется точка, которая называется начальной точкой. Затем, контур обходится, и каждый вектор смещения записывается комплексным числом $a+ib$, где a – смещение точки по оси X, а b – смещение по оси Y. Смещение берется относительно предыдущей точки. При таком подходе контур определяется как совокупность элементарных векторов, представленных своими двумерными координатами. Изменение начальной точки приведет к циклическому сдвигу вектор-контра, а изменение масштаба изображения можно рассматривать как умножение элементарного вектора на масштабный коэффициент.

Скалярным произведением контуров $N1$ и $N2$ называется такое комплексное число [2]:

$$\eta = (N1, N2) = \sum_{n=0}^{k-1} (\gamma_n, v_n), \quad (2)$$

где k – размерность вектор-контра, γ_n – n -й элементарный вектор контра $N2$, v_n – n -й элементарный вектор контра $N1$, (γ_n, v_n) – скалярное произведение комплексных чисел, вычисляемых как $(a+ib, c+id) = (a+ib)(c-id) = ac+bd+i(bc-ad)$. (3)

Действительная часть скалярного произведения комплексных чисел совпадает со скалярным произведением соответствующих векторов, т.е. включает в себя векторное скалярное произведение. Чем больше значение скалярного произведения, тем «ближе» вектора друг к другу. Для перпендикулярных векторов оно будет равно нулю, станет отрицательным для векторов, направленных в разные стороны, а для коллинеарных – будет иметь максимальное значение.

Нормированное скалярное произведение (НСП), определяемое формулой (4) в пространстве комплексных чисел, также является комплексным числом. При этом, единица – это максимально возможное значение модуля НСП и она достигается только при выполнении условия (6).

$$\eta = \frac{(N2, N1)}{|N2| |N1|}, \quad (4)$$

где $|N1|$ и $|N2|$ – нормы (длины) контуров. Нормы контуров вычисляются как

$$|N2| = \left(\sum_{n=0}^{k-1} |\gamma_n|^2 \right)^{\frac{1}{2}}, \quad (5)$$

$$N2 = \mu N1, \quad (6)$$

где μ – произвольное комплексное число.

Исходя из физического смысла умножения комплексных чисел, получается вывод, что модуль НСП достигнет максимального значения – единицы – только если контур $N2$ является тем же контуром $N1$, даже если он

будет повернут на некоторый угол и промасштабированный на определенный коэффициент.

Но равенство в (6) достигается только тогда, когда начальные точки контуров совпадают. В противном случае, даже если контуры одинаковы, модуль НСП таких контуров не будет равен единице. Для решения этой проблемы можно воспользоваться взаимокорреляционной функцией (ВКФ) двух контуров, определяемой формулой (7), значения которой показывают степень похожести контуров, если сдвинуть начальную точку N на m позиций.

$$\tau(m) = (N2, N1(m)), \quad (7)$$

где $N1(m)$ – контур, полученный из $N1$ путем циклического сдвига его элементарного вектора на m элементов. ВКФ определена на всем множестве целых чисел, но поскольку циклический сдвиг на k приведет к исходному контуру, то ВКФ имеет период k и нужно рассмотреть значения этой функции только в пределах от 0 до $k-1$.

Максимальный модуль среди значений ВКФ в (8) показывает степень похожести контуров (достигает единицы для одинаковых контуров), а аргумент $\arg(\tau_{\max})$ дает угол поворота одного контра относительно другого.

$$\tau_{\max} = \max \left(\frac{\tau(m)}{|N2| |N1|} \right), \quad (8)$$

где $\tau(m)$ – модуль ВКФ, $|N1|$ и $|N2|$ – нормы (длины) контуров, определяемые формулами (5) и (6).

3. Заключение

Предложен гибридный метод распознавания контуров на изображении, включающий три основных этапа: размытие, выделение границ, сохранение границ. Данный подход был использован при решении задачи выделения границ печени человека и распознавании контуров фаз и зерен на изображении микроструктуры стали [3], где условия работы были наиболее сложными ввиду наличия малоразличимых градаций серого на снимках и сильной неоднородности границ контуров распознаваемых объектов. Однако при подобном подходе все еще остается сложность определения того, что именно выделять, в случае если нам нужен только один контур, как в примере поиска органа человека. В случае если существует два контура этого же органа на снимке, данный подход не сработает ввиду ограничения только на один контур из-за схожего контра селезенки. Так или иначе эта проблема осталась открытой и требует дальнейшего рассмотрения.

4. Поддержка

Исследование проводилось при финансовой поддержке Министерства образования и науки РФ в рамках проекта RFMEFI58617X0055. Название совместного проекта с ЕС в рамках программы "Горизонт 2020": "Fracture across Scales and Materials, Processes and Disciplines" Акроним проекта: "FRAMED".

5. Литература

- [1] Гонсалес Р., Вудс Р. Цифровая обработка изображений // Техносфера. Москва. 2005
- [2] Ванюкова Д.И., Попов С.С., Соколов П.А. Совмещение цифрового картографического изображения местности с радиолокационным изображением. [Электронный ресурс] В кн.: Материалы XVI конференции молодых ученых «Навигация и управление движением», Санкт-Петербург, 2014, Март. URL: <http://www.elektropribor.spb.ru/kmu2014/refs> (дата обращения 10.07.2017).
- [3] Ключев А.В., Столбов В.Ю., Шарыбин С.И.

Визуализация сложных зеренных структур металлов и сплавов при идентификации их параметров // Научная визуализация. 2016. Т.8. №3. С. 95-101.

- [4] Ясницкий Л.Н. Введение в искусственный интеллект: учеб. пособие для студ. высш. учеб. заведений. — М. : Издательский центр «Академия», 2010. — 176 с.
- [5] Shih Frank. Image processing and pattern recognition: fundamentals and techniques. IEEE Press, 2010. – 537 p.

Об авторах

Клестов Роман Андреевич, магистр Пермского национального исследовательского политехнического университета. Его e-mail: klestovroman@gmail.com

Столбов Валерий Юрьевич, д.т.н., профессор кафедры вычислительной математики и механики Пермского национального исследовательского политехнического университета. Его e-mail: valeriy.stolbov@gmail.com.

Автономная навигация в закрытых помещениях с использованием адаптивного описания эталонных изображений

Л.И. Лебедев, Ю.Г. Васин, М.П. Осипов¹
lebedev@pmk.unn.ru|pmk@unn.ac.ru|osipovmp@mail.ru

¹Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

В работе описываются методы коррекции результатов позиционирования посредством автоматического распознавания фрагментов графических изображений, инвариантного относительно аффинных преобразований. Распознавание осуществляется на базе корреляционно-экстремального контурного метода. В целях увеличения быстродействия и качества распознавания предлагается метод адаптивного формирования описаний эталонных изображений в зависимости от параметров проецирования. Приводится оценка сложности предлагаемого метода распознавания.

Ключевые слова: распознавание, коррекция, аффинное преобразование, эталонные изображения, адаптивное описание, видео поток, КЭЖМ.

Autonomous navigation in enclosed spaces using an adaptive description of standard images

L.I. Lebedev, Yu.G. Vasin, M.P. Osipov¹
lebedev@pmk.unn.ru|pmk@unn.ac.ru|osipovmp@mail.ru

¹N.I. Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

The paper describes methods for correcting the results of positioning by automatically recognizing fragments of graphic images that are invariant with respect to affine transformations. Recognition is carried out on the basis of the correlation-extreme contour method. In order to increase the speed and quality of recognition, a method is proposed for the adaptive formation of descriptions of standard images, depending on the projection parameters. The complexity of the proposed recognition method is estimated.

Keywords: recognition, correction, affine transformation, standard image, adaptive description, video stream, CECM.

1. Введение

В настоящее время актуальной темой исследовательских работ является поиск доступных технологических решений в задаче навигации внутри зданий и помещений.

Классический подход к навигации с использованием технологий спутниковых радионавигационных систем на основе GPS и ГЛОНАСС малоэффективен внутри помещений поскольку, стены и перекрытия выступают помехой для прохождения спутникового сигнала. Поэтому на данный момент предлагаются различные подходы решения задачи навигации в закрытых помещениях, использующие: сети передачи данных (Bluetooth, Wi-Fi, FM-радиоволны, UWB), магнитное поле Земли, микромеханические инерционные датчики (акселерометры, датчики угловой скорости), QR-коды, RFID-метки и т.д. [7]. Но каждый подход в отдельности имеет существенные ограничения по точности, надежности, стоимости и сложности реализации.

Традиционным способом навигации внутри помещений на данный момент остается визуальная навигация, представляющая собой систему указателей и именных табличек, размещенных в здании и служащих путеводителем по нему. Несмотря на простоту реализации, существенный недостаток такого подхода состоит в большой информационной нагрузке на человека и сложности анализа полученной информации. В последнее время в ряде работ было предложено использовать методы компьютерного зрения как для анализа движения [4], так и для анализа информационных объектов помещения (указателей, именных табличек и т.д.) [9], что позволяет автоматизировать процесс навигации. На стенах и дверях зданий часто можно встретить плоские объекты,

указывающего и предупреждающего характера. Это названия и номера помещений, знаки пожарной безопасности, путей эвакуации, указатели и т.д. Такие информационные объекты имеют яркую окантовку, что легко выделяет их из общего фона. Если, используя методы компьютерного зрения, распознать такие плоские объекты в кадре с камеры мобильной навигационной системы, то зная точное расположение этих объектов можно восстановить информацию о местоположении пользователя. Для этого предварительно создается база данных изображений плоских информационных объектов помещения (картины, названия и номера помещений, планы эвакуации, знаки пожарной безопасности и т.д.) с координатной привязкой их расположения. Таким образом, в момент, когда необходима корректировка своего местоположения, пользователь производит видеосъемку близлежащего плоского информационного объекта с помощью мобильного устройства со встроенной камерой. Далее, используя методы поиска изображения по содержанию (англ. Content-based image retrieval (CBIR)), происходит сравнение элементов изображения полученных при видеосъемке с базой данных изображений информационных объектов. Если искомое изображение найдено, то, зная информацию о расположении изображаемого объекта, хранящуюся в базе данных, можно определить расположение и ориентацию камеры, производящей съемку, а, следовательно, расположение и ориентацию пользователя. Недостатком такого подхода является не уникальность информационных объектов, а следовательно неоднозначность определения местоположения пользователя.

В предыдущей работе авторов [3] для решения задачи навигации внутри здания был применен комбинированный подход, в основе которого лежат методы инерциальной

навигации [5] и компьютерного зрения [12] при использовании поэтажных планов здания, технологий моделирования и 3D визуализации. Использование комбинированного подхода позволяет компенсировать недостатки каждого метода и повысить надежность позиционирования. Для распознавания информационных объектов помещения использовался метод SIFT поиска локальных особенностей для определения соответствий на изображениях [6]. Этот метод традиционно применяется исследователями при использовании методов компьютерного зрения в задаче навигации [9,12]. Но метод SIFT (SURF) имеет высокую вычислительную сложность и плохо работает для объектов простой формы и без ярко выраженной текстуры какими, как правило, являются информационные объекты помещения. В работе [8] для распознавания текстовой информации в табличках на стенах помещения в задаче навигации используется метод Optical Character Recognition (OCR). Но данный метод не является инвариантным относительно ортогональных преобразований и масштабирования, что значительно сокращает возможности его применения в задаче навигации. В свою очередь, информационные объекты помещения имеют четкие границы и легко выделяются из окружающего фона, поэтому поиск такого класса объектов можно эффективно производить методами контурного анализа. В качестве метода контурного анализа предлагается использовать корреляционно-экстремальный контурный метод (КЭКМ) [2].

2. Постановка задачи

На базе эталонных изображений уникальных информационных объектов должна решаться задача автоматического распознавания элементов текущего кадра из видеопотока с последующей его идентификацией. Под уникальностью информационных объектов здесь понимается как содержащееся в них информация, так и возможность ее распознавания предлагаемыми методами. Кроме того, для обеспечения быстродействия и качества распознавания, предлагаемые методы должны быть инвариантными относительно аффинных преобразований, а также к адаптивному описанию фрагментов изображения в зависимости от угла сканирования информационного объекта. Коррекция результатов позиционирования должна осуществляться на основе данных о местоположении эталонного изображения и параметров сканирования, полученных в результате распознавания информационного объекта.

3. Методы решения

Распознавание изображений информационных объектов будет осуществляться методами сравнения с эталонами посредством обнаружения и идентификации фрагментов, сходных по описанию с элементами эталонных изображений. Эталонное изображение формируется на основе описаний фрагментов изображения плоского информационного объекта, являющегося центральной фронтальной перспективой. Совокупность элементов эталонного изображения должна быть достаточной для качественного автоматического распознавания информационных объектов с одной стороны и минимальной по составу для обеспечения необходимого быстродействия с другой стороны.

Распознавание фрагментов изображения будем осуществлять корреляционно-экстремальным контурным методом (КЭКМ) [2]. В основу КЭКМ положено вычисление оценки близости ϵ_m по описаниям $\mathbf{w} = (x, y)^T$ фрагмента изображения \mathbf{O} и $\mathbf{w}^e = (x^e, y^e)^T$

эталона \mathbf{E} , которые задаются посредством векторной (контурной) модели представления информации:

$$\epsilon_m = \mathbf{Dw}^e - R^2 / \mathbf{Dw}, \tag{1}$$

где \mathbf{Dw}^e , \mathbf{Dw} - дисперсии контуров эталона и фрагмента изображения соответственно, а R - величина, вычисленная по значениям смешанных корреляционных моментов:

$$R^2 = S_n^2 + C_s^2, \\ S_n = \text{cov}(\mathbf{x}^e, \mathbf{y}) - \text{cov}(\mathbf{y}^e, \mathbf{x}), \\ C_s = \text{cov}(\mathbf{x}^e, \mathbf{x}) + \text{cov}(\mathbf{y}^e, \mathbf{y}). \tag{2}$$

Начальные моменты первого и второго порядка находятся по формулам:

$$\bar{\mathbf{w}}^e = \mathbf{Mw}^e = \frac{1}{2 \cdot S^e} \cdot \sum_{i=1}^{m-1} (\mathbf{w}_i^e + \mathbf{w}_{i+1}^e) \cdot S_i^e \\ \mathbf{My}^2 = \frac{1}{3 \cdot S} \cdot \sum_{i=1}^{n-1} (y_i^2 + y_{i+1}^2 + y_i \cdot y_{i+1}) \cdot S_i \tag{3}$$

где $\mathbf{w}^e = \{\mathbf{w}_1^e, \mathbf{w}_2^e, \dots, \mathbf{w}_m^e\}$, $\mathbf{w} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$, $\mathbf{w} = (x, y)^T$ - исходные описания контуров эталона и распознаваемого фрагмента соответственно, а S_i^e , S_i - длины ребер контуров. Вычисление смешанных моментов второго порядка осуществляются по вспомогательным описаниям эталона $\hat{\mathbf{w}}_j^e = (\hat{x}_j^e, \hat{y}_j^e)^T$ и фрагмента $\hat{\mathbf{w}}_j = (\hat{x}_j, \hat{y}_j)^T$, $j = 1, 2, \dots, l \leq m + n - 2$ соответственно [8]:

$$\mathbf{Mx}^e \mathbf{y} = \frac{1}{6 \cdot S^e} \cdot \sum_{i=1}^{l-1} (2 \cdot \hat{x}_i^e \cdot \hat{y}_i + 2 \cdot \hat{x}_{i+1}^e \cdot \hat{y}_{i+1} + \hat{x}_i^e \cdot \hat{y}_{i+1} + \hat{x}_{i+1}^e \cdot \hat{y}_i) \cdot \hat{S}_i^e \tag{4}$$

Оценка близости ϵ_m корректна только для согласованных описаний контуров. Для согласования описаний контуров применяются эффективные по быстродействию методы относительных смещений и метод парабол [2, 11]. Поэтому, КЭКМ инвариантен не только относительно ортогональных преобразований и масштабирования, но также и к циклическим описаниям контуров, что обеспечивает наиболее полный состав элементов в классах эквивалентностей.

В работе [10] было доказано, что любое аффинное преобразование может быть представлено совокупностью операций, включающих проецирование контура на плоскость, полученную последовательным вращением плоскости XOY вокруг осей OX и OY на углы β и γ соответственно, ортогональное преобразование и масштабирование проецированного контура. Для обеспечения инвариантности относительно аффинных преобразований вначале формируются описания проекций контура по формулам

$$\begin{cases} x_{pr} = x \cdot \cos \gamma + y \cdot \sin \beta \cdot \sin \gamma \\ y_{pr} = y \cdot \cos \beta \end{cases}, \tag{5}$$

а затем решается оптимизационная задача вида

$$\epsilon_m^A = \min_{(\beta, \gamma) \in D^o} \epsilon_m(\beta, \gamma), \tag{6}$$

где оценка сходства (близости) $\epsilon_m(\beta, \gamma)$ для заданных значений углов β и γ находится с помощью КЭКМ в соответствии с формулами (1)-(4). Решение

оптимизационной задачи (6) осуществляется с использованием метода сеток. Область $D^\circ(\beta, \gamma)$ изменения параметров β и γ может быть изначально детализирована при распознавании контуров, описывающих специфические фрагменты изображений (рамки, условные знаки и т.д.) при непосредственном вычислении матрицы аффинного преобразования A° . Область D° в процессе распознавания фрагментов информационных объектов может корректироваться. Необходимо также отметить, что нахождение области D° позволяет значительно увеличить быстродействие алгоритма автоматического распознавания.

4. Полученные результаты

В результате исследований различных структур описания эталонных изображений была выбрана модель представления в формате КМР, адаптированная для решения задачи распознавания информационных объектов. Предлагаемая структура описания эталонного изображения включает три поля: заголовок, параметрическое описание каждого из эталонов и их метрические описания путем задания последовательностей координат точек контуров. Заголовочная часть содержит сведения об эталонном изображении в целом: идентификатор эталонного изображения, количество эталонов N_1 , представляющих изображение, координаты привязки изображения (x, y, H) , разрешение и ряд других сведений об изображении. Атрибутивное описание эталонов состоит из записей постоянной длины и включает данные о количестве точек в метрическом представлении эталона, его габаритах, местоположении эталона на изображении информационного объекта, ссылке на адрес начала метрического описания эталона, типа и кода эталона, раскраска контура. Также, вычисляются и запоминаются дисперсионные характеристики эталона и ряд других сведений, необходимых для обеспечения быстродействия при нахождении согласованных описаний. Кроме того, для каждого контура эталонного изображения задаются условия его активизации и значимости. Это позволяет в зависимости от условий сканирования информационного объекта изменять необходимый набор эталонов, обязательных для опознавания полученного изображения.

Для получения контурной модели представления изображения были реализованы несколько алгоритмов, которые последовательно обеспечивают перевод его в оттенки «серого», затем, осуществляют бинаризацию с порогом Th , полученным с использованием метода Оцу, далее решают задачу получения штриховой модели изображения и, наконец, построения самой контурной модели в соответствии с методом формирования КМР [1]. В совокупности все представленное алгоритмическое и программное обеспечение построения контурной модели обладает необходимым для решения задачи распознавания изображений информационных объектов быстродействием.

Формирование эталонного изображения осуществляется на основе полученной контурной модели базового информационного объекта в интерактивном режиме. Пользователю предоставляется возможность из множества контуров, описывающих изображение в формате КМР выбрать необходимую из них совокупность, достаточную для идентификации изображения этого информационного объекта, получаемого в других условиях сканирования. Уменьшение угла сканирования информационного объекта влечет изменение состава контуров в описании изображения в формате КМР в

результате их слипания. Поэтому, при формировании эталонного изображения предусмотрена возможность получения описаний слипшихся контуров в базовых изображениях, чтобы использовать их в качестве эталонов, что способствует повышению качества распознавания. Параметры, характеризующие ранговые предпочтения контуров и их условия использования в обязательном порядке задаются пользователем также интерактивным режиме. Остальные атрибуты описания контура формируются автоматически.

Для распознавания информационных объектов был реализован алгоритм на базе КЭЖМ. Информационный объект будет считаться распознанным, если на нем будут идентифицированы все контура, соответствующие элементам эталонного изображения с учетом их представления. Особенностью данного алгоритма распознавания является то, что он может использоваться как прямое, так и обратное аффинное преобразование. Прямое аффинное преобразование, обеспечивающее переход от базового описания изображения к текущему, необходимо для локализации области поиска контуров, соответствующих местоположению рассматриваемого эталона. Это исключает необходимость распознавания всех контуров на изображении текущего информационного объекта. Обратное аффинное преобразование дает возможность при вычислении оценки сходства использовать методы относительных смещений и парабол для получения согласованных описаний контуров. Это также направлено на увеличение быстродействия алгоритма распознавания.

На примере распознавания конкретного информационного объекта, кадр с изображением которого представлен на рис. 1, представим все этапы, необходимые для идентификации и проведения коррекции результатов позиционирования. Отметим, что это полученное с камеры изображение является боковой фронтальной перспективой.



Рис. 1. Кадр с изображением информационного объекта.

Вначале для данного информационного объекта сформируем эталонное изображение. Очевидно, что заранее невозможно определить угол, под которым будет получено распознаваемое изображение информационного объекта. Поэтому формирование эталонов будет осуществляться на основе элементов изображения, представленного на рис.2 и являющегося центральной фронтальной перспективой того же объекта. Изображение данного базового информационного объекта в формате контурной модели представлено на рис. 3. Представленная контурная модель изображения получена при бинаризации с порогом $Th = 120$. Как было отмечено выше, при формировании эталонного изображения необходимо использовать такие фрагменты базового изображения, чтобы они однозначно определяли данный информационный объект. Эксклюзивными элементами

данного информационного объекта являются символы, идентифицирующие ответственного за противопожарную безопасность и текст, характеризующий предметную область распознаваемого объекта. Поэтому, эталонное изображение должно включать, во-первых, описания рукописных букв 2-х «Л» (2 контура), 3-х «е» (2), «б» (1), «в» (1), «И» (1), а во-вторых, символы из фрагментов текста, например, «помещения», «пожаре» и «или». Метрические описания одних и тех же символов представляются одним контуром и при распознавании КЭМ заимствуются, что позволяет увеличить быстродействие алгоритма распознавания отдельных букв. При вращении базового изображения относительно оси OY на угол более 65° происходит слипание символов «д» и «е» в фамилии ответственного, а при угле более 77° - слипание символов «Л» и «е», «б» и «е». При превышении этих углов в первом случае слипания буква «е» перестает быть значимым эталоном, а во втором, эталонное изображение для более качественного распознавания можно пополнить вновь образовавшимися контурами. В результате получим эталонное изображение с адаптивным для распознавания составом эталонов.



Рис. 2. Базовое изображение информационного объекта.

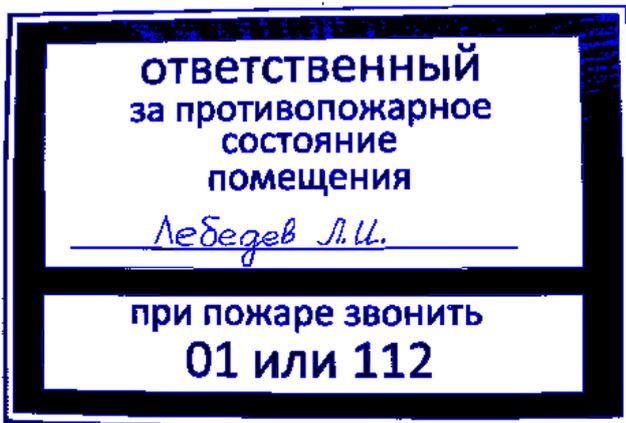


Рис. 3. Базовый информационный объект в формате КМР.

Изображение информационного объекта в формате КМР включает описание двух внутренних контуров, задающих границы рамок. Включение любого из этих контуров в описание эталонного изображения позволило легко получить матрицу аффинного преобразования

$$A^\circ = \begin{pmatrix} 0.3396 & -0.0776 \\ 0.0542 & 0.9846 \end{pmatrix},$$

по ее компонентам найти углы проецирования $\beta \approx 4.9^\circ$ и $\gamma \approx 69.6^\circ$, определить границы области $D^\circ(\beta, \gamma)$ и целенаправленно отбирать элементы информационного

объекта для распознавания, используя относительные координаты положения эталонов на базовом изображении.

На рис. 4 приведен фрагмент изображения текущего информационного объекта. Все символы данного изображения, соответствующие эталонам базового информационного объекта, были распознаны с вероятностью, значительно превосходящей пороговое значение. Исключение составляет третья буква «е» в фамилии, однако ее распознавание не осуществлялось алгоритмом из-за отсутствия соответствующего эталона согласно его области использования.

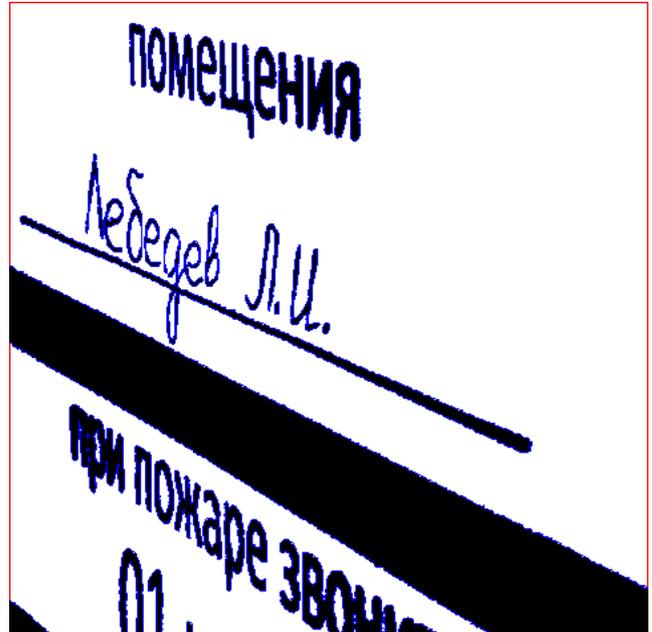


Рис. 4. Фрагмент КМР- изображения текущего кадра.

Таким образом, данное изображение было успешно распознано, а координаты его привязки использованы для коррекции результатов позиционирования.

Оценим вычислительную сложность предложенного подхода коррекции текущих координат позиционирования, которая полностью определяется сложностью распознавания изображения информационного объекта. Наиболее сложным в вычислительном плане является получение начальной матрицы аффинного преобразования A° . Для ее получения в общем случае необходимо вычислить оценки сходства эталона со всеми N_2 контурами на изображении информационного объекта. Таким образом, сложность вычисления матрицы A° оценивается величиной $\Xi_1 = \lambda \cdot N_2 \cdot O(m + \bar{n})$, $\lambda < 100$, где \bar{n} - среднее количество точек в описаниях используемых контуров изображения информационного объекта. Отметим, что за счет логической фильтрации количество контуров N_2 можно значительно уменьшить.

После получения матрицы аффинного преобразования A° вычисление оценок сходства осуществляется только для эталонов и контуров на изображении информационного объекта, соответствующих им по местоположению. Сложность вычисления этих оценок сходств с учетом количества уточнений ϑ матрицы аффинного преобразования оценивается величиной $\Xi_2 = N_1 \cdot \vartheta \cdot O(\bar{m} + \bar{n})$, ($\vartheta \leq 6$). Таким образом, сложность распознавания оценивается величиной $\Xi = \Xi_1 + \Xi_2$.

5. Заключение

Предлагаемый подход коррекции результатов позиционирования, ориентированный на использование алгоритма распознавания фрагментов графических изображений, реализованного на базе КЭМ и инвариантного относительно аффинных преобразований, увеличил возможности осуществления автономной навигации в закрытых помещениях. Дальнейшего увеличения качества автоматического распознавания можно достичь путем уменьшения перспективных искажений изображения.

6. Благодарности

Работа выполнена при поддержке Российского Фонда Фундаментальных исследований, проект № 16-07-01214 А и проект № 16-07-01198.

7. Литература

- [1] Васин Ю.Г., Башкиров О.А., Чудинович Б.М., Гервагин Ю.Г. Подсистема растрового ввода-вывода и обработки графической информации РАСТР I // Автоматизация обработки сложной графической информации. Межвузовский сборник / Горьковский государственный университет, г. Горький, 1984.
- [2] Лебедев Л.И. Корреляционно – экстремальные контурные методы распознавания. Теоретические основы: Учебное пособие./ Нижний Новгород, Изд-во Нижегородского государственного университета, 2013. 113 с. – ISBN 978-5-91326-308-7.
- [3] М.П. Осипов, А.О. Патрушев Методы корректировки местоположения в задаче автономной навигации в закрытых помещениях / Труды 26-ой Международной конференции по компьютерной графике и зрению, Нижний Новгород. 2016, С. 417-419.
- [4] S. Hilsenbeck, A. Moller, R. Huitl, G. Schroth, M. Kranz, and E. Steinbach, "Scale-preserving long-term visual odometry for indoor navigation," in International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sydney, Australia, Nov. 2012.
- [5] Jeong Won Kim, A Step, Stride and Heading Determination for the Pedestrian Navigation System / Jeong Won Kim, Han Jin Jang, Dong-Hwan Hwang, Chansik Park // In: Positioning, Vol.1, No.8, December 2004 – pp.273-279.
- [6] Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. pp. 1150–1157.
- [7] Rainer Mautz, Indoor Positioning Technologies / Rainer Mautz -Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, February 2012 – с.127.
- [8] Sami M., Ayaz Y., Jamil M., Gillani S. O., Muhammad N. Text detection and recognition for semantic mapping in indoor navigation. International Conference on IT Convergence and Security. IEEE, 2015, pp. 1–4.
- [9] Serrão, M., Shahrabadi, S., Moreno, M., Josre, J. T., Rodrigues, J. I. Rodrigues, J. M. F., du Buf, J. M. H., Computer vision and GIS for the navigation of blind persons in buildings. Universal Access in the Information Society, 14(1), pp. 67-80, 2014.
- [10] Vasin Ju.G., Lebedev L.I. Recognition of base of similarity estimations, invariant with respect to affinities. //8th International conference on Pattern recognition and image analysis: new information technologies"/ Yoshkar-Ola, v.2, 2007, pp. 57-60.
- [11] Vasin Yu.G., Lebedev L.I. The problem of obtaining coherent contour descriptions in the calculation of similarity estimates. //8th Open German-Russian Workshop "Pattern recognition and image understanding" (OGRW-8-2011): Workshop Proceedings. 2011. P. 324-327.
- [12] Werner M., Kessel M., and Marouane C., Indoor positioning using smartphone camera, in Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '11), 6, p. 1, September 2011.

Об авторах

Лебедев Леонид Иванович, к.ф.-м.н., ст.н.с., ведущий научный сотрудник института информационных технологий, математики и механики Нижегородского государственного университета им. Н.И. Лобачевского. Его e-mail: lebedev@pmk.unn.ru.

Васин Юрий Григорьевич, д.т.н., главный научный сотрудник института информационных технологий, математики и механики Нижегородского государственного университета им. Н.И. Лобачевского. Его e-mail: pmk@unn.ac.ru.

Осипов Михаил Павлович, к.т.н., ведущий научный сотрудник института информационных технологий, математики и механики Нижегородского государственного университета им. Н.И. Лобачевского. Его e-mail: osipovmp@mail.ru.

Глубинные двоичные дескрипторы изображения человека для его повторной идентификации и сопровождения в видео

В. С. Лютов¹, А. С. Конушин^{1,2}

vladimir.lutov@graphics.cs.msu.ru|anton.konushin@graphics.cs.msu.ru

¹МГУ, Москва, Россия;

²ВШЭ, Москва, Россия

В работе рассматривается задача повторной идентификации человека – определения личности неизвестного по фотографии с камеры видеонаблюдения. В работе предлагается базовый метод, основанный на модификации нейросетевого алгоритма классификации VGG16. Экспериментальное сравнение алгоритма продемонстрировало большую точность в сравнении с аналогами. Также предлагается модификация алгоритма, строящая бинарные дескрипторы изображения сравнимой точности с исходным дескриптором.

Ключевые слова: компьютерное зрение, повторная идентификация, глубинное обучение.

Human image deep binary descriptor for person reidentification and tracking in video

V. S. Liutov¹, A. S. Konushin^{1,2}

vladimir.lutov@graphics.cs.msu.ru|anton.konushin@graphics.cs.msu.ru

¹MSU, Moscow, Russia;

²HSE, Moscow, Russia

The work focuses on human reidentification, i.e. identifying an unknown person using a photo from a surveillance camera. A base method that involves modifying the VGG16 neural network algorithm is proposed. Experimental comparison of the algorithm proved to be highly accurate compared to similar approaches. Furthermore, an algorithm modification, which creates binary visual descriptors that have comparable accuracy with the original descriptor, is presented.

Keywords: computer vision, reidentification, deep learning.

1. Введение

В работе рассматривается задача построения дескрипторов изображения человека для повторной идентификации и сопровождения. Повторная идентификация - автоматическое сопоставление обнаружений людей с их личностями. Иначе говоря, требуется определить, видели ли мы раньше обнаруженного человека, и, если видели, то где и когда именно.

Например, в сценариях видеонаблюдения, решение этой задачи позволит выяснить откуда пришел, куда ушел и с кем общался участник некоторого инцидента. Данная задача является актуальной задачей компьютерного зрения, так как для работы в реальных условиях требуется высокая скорость поиска изображений и высокая переносимость алгоритма, подготовленного на одних данных на данные, полученные из независимого источника.

Быстрая повторная идентификация в первую очередь может использоваться в сопровождении, для дальнейшего повышения качества ассоциации фрагментов траекторий одного и того же человека, например, когда человек в промежутке между этими фрагментами был скрыт каким-то элементом сцены [9].

На вход алгоритму повторной идентификации подается результат работы алгоритма обнаружения людей – обнаружения. Обнаружения – выделенные из полученных со статических камер изображения прямоугольники, содержащие изображения людей. Требует-

ся определить, как обнаружения соотносятся с личностями людей.

Эта задача является частным случаем задачи идентификации человека. Она традиционно решается за два этапа: построение дескриптора изображения человека и поиск по базе дескрипторов [8]. Пример работы приведен на рисунке 1. В данной работе предлагается решать подзадачу построения дескриптора.



Запрос 5 самых близких кадров

Рис. 1. Пример работы предложенного алгоритма.

Чем левее изображения-кандидаты, тем ближе их дескрипторы к дескриптору-запросу. Зеленым отмечены верные изображения, красным - неверные.

2. Обзор существующих методов

Построение бинарного дескриптора состоит из двух этапов: сначала строится вещественный дескриптор достаточно высокого качества, затем на основе алгоритма построения вещественного дескриптора строится бинарный дескриптор.

2.1. Вещественные

Методы построения вещественных дескрипторов можно разбить на три множества:

1. **Методы без машинного обучения** не используют информацию о том, какие изображения находятся в коллекции изображений и работают для всех изображений одинаково. Исследования [8] [21] показывают, что чем больше мы принимаем во внимание коллекцию изображений, с которой работаем, тем большую точность работы получаем.
2. **Методы с машинным неглубинным обучением** используют информацию о том, какие изображения находятся в коллекции изображений. Чаще всего они представляют собой комбинацию классических методов обработки изображений и методов машинного обучения, также они используют математическую статистику. По результатам обзора [8] метод “Иерархический гауссовый дескриптор для повторной идентификации человека” GOG [14] показывает наилучшую точность в данном классе.
3. **Методы с глубинным обучением** показывают наилучшие результаты для этой задачи [5] [19]. Данные методы широко применяются в различных задачах компьютерного зрения и показывают все лучшие результаты, например в задаче выделения движущихся объектов со статичной камеры видеонаблюдения [15].

Нейросетевые методы повторной идентификации весьма популярны. На данный момент изучаются как независимые методы повторной идентификации [18], так и комплексные подходы, использующие решение смежных задач, таких как анализ цвета одежды человека [22] или поиск изображения по его текстовому описанию [10].

2.2. Бинарные

В данной работе акцентируется внимание на построении бинарных дескрипторов произвольной длины. В работах, посвященных повторной идентификации, подобные исследования ранее не проводились.

Однако, в других работах по компьютерному зрению подобные методы рассматривались. Построение бинарных дескрипторов (хешей) небольшой размерности можно сделать разными способами. В ходе этой работы были экспериментально проверены следующие частные методы:

1. Наивная бинаризация - поэлементное сравнение исходного дескриптора с 0. Значениям больше 0 соответствует 1, остальным – 0.
2. Выделение наиболее значимых элементов исходного дескриптора с помощью метода машинного обучения. Например, с помощью алгоритма “random forest” [1].
3. Преобразование в пространство меньшей размерности с помощью метода главных компонент [4], затем бинаризация.
4. Нейросетевые методы построения бинарных дескрипторов. Были проверены два таких алгоритма:

алгоритм сигмоиды [12] и его модификация DBE [13].

Алгоритм сигмоиды заключается в добавлении к сети, строящей вещественные дескрипторы, полносвязного слоя с сигмоидальной функцией активации.

В алгоритме DBE вместо этого слоя используется последовательность из нескольких слоев:

$$f_{DBE}(X) = \tanh(\text{ReLU}(\text{BN}(XW_{DBE} + b_{DBE}))),$$

где f_{DBE} – алгоритм DBE, X – вещественный дескриптор, $\tanh(Z)$ – поэлементный гиперболический тангенс, $\text{ReLU}(Z)$ – поэлементный $\max(0, z_i)$, BN – слой “batch normalization” [7], W_{DBE} , b_{DBE} – оптимизируемые веса алгоритма, матрица и вектор, соответственно.

3. Предложенный метод

По результатам проведенного обзора в качестве базового алгоритма был выбран алгоритм VGG16, предобученный на задаче ImageNet. Он состоит из 5ти блоков свертки+max-pooling и 3 полносвязных слоев. На вход он получает изображение размером 224x224x3, на выходе – вероятность принадлежности каждой из картинок к тому или иному классу из 1000 классов. Предложенная модификация изображена на рисунке 2 и подробно описана ниже. Исходный код доступен по ссылке <https://github.com/vslutov/reidentification>.



Рис. 2. Схема предложенного алгоритма.

3.1. Преобработка данных

Случайным образом разделим тренировочные данные на обучающую и валидационную выборки. Отношение количества примеров в обучающей и валидационной выборке равно 9/1. В процессе обучения нейросеть будет видеть только обучающую выборку, а на валидационной выборке мы будем проверять насколько хорошо она обучилась.

Во время обучения цвета на картинках нормализовались – каждый канал каждого пикселя приводился линейным преобразованием к такому виду, чтобы выборочное мат. ожидание каждого отдельного канала отдельного пикселя по всей выборке было равно нулю, а выборочная дисперсия – единице. Однако в оригинальной нейросетевой модели VGG16, обученной на ImageNet [17], предобученные веса были получены на картинках с другой преобработкой. В работе [17] использовалось вычитание среднего RGB значения по каждому пикселю по обучающей выборке, иначе говоря приведение выборочного мат. ожидания к нулю.

Необходимо отметить, что предполагается использовать предложенный алгоритм для потоковой обработки видео с камер видеонаблюдения. В разное время суток кадры такого видео имеют разный уровень яркости и контрастности. Нормализация выборочных мат. ожидания и дисперсии в рамках небольшой выборки позволяет сгладить данное различие.

После нормализации картинки горизонтально зеркально отражались с вероятностью $\frac{1}{2}$ и поворачивались на случайный угол до 20 градусов.

3.2. Адаптация под задачу повторной идентификации человека

Так как мы собираемся обрабатывать изображения людей, изображения будут вертикально вытянуты. Предобученные на ImageNet полносвязные слои пригодны только для картинок фиксированного размера (как в ImageNet, 224x224x3), поэтому использовать их в данной задаче нецелесообразно и придется их убрать. Сверточные слои оставим без изменений.

Также в ходе экспериментального тестирования выяснилось, что при удалении самого глубокого блока сверток и max-pooling слоя точность алгоритма увеличивается. Это связано с тем, что в данной задаче для алгоритма важнее низкоуровневые свойства такие как цвет и текстура одежды.

Добавляем в конец сети GlobalAveragePooling слой. Это самый простой способ получить низкоуровневое представление информации о картинке.

Добавляем в конец сети слой "batch normalization" [7]. Этот слой делает предсказуемым распределение каждого элемента в дескрипторе – мат. ожидание равно 0, дисперсия 1. Эта модификация увеличивает точность работы алгоритма и позволит нам сделать наивную бинаризацию простым сравнением с нулем, длина хеша при такой бинаризации равна числу выходов слоя "batch normalization" – 512.

Добавим в конец сети один полносвязный слой с функцией активации softmax и количеством нейронов равному числу классов в обучающей выборке – для Market1501 [20] это 751. Назовем этот слой классифицирующим, он используется только во время обучения. Во время обучения в качестве функции ошибки у нас будет выступать ошибка идентификации – многоклассовая перекрестная энтропия выходов классифицирующего слоя и ожидаемого результата. Вещественный дескриптор – выход слоя "batch normalization", он имеет размерность 512 вещественных чисел.

3.3. Обучение нейронной сети

Используем оптимизатор nadam [2]. Размер батча – 128 изображений. Если в течение 4х эпох ошибка на валидационной выборке не уменьшается, то уменьшим скорость обучения в 10 раз. Если в течение 10 эпох ошибка на валидационном наборе не уменьшается, то мы достигли локального минимума, остановим обучение.

1. Отключаем обучение всех слоев, кроме последнего полносвязного. Обучаем 50 эпох на обучающей выборке.

2. Включаем обучение всех сверточных слоев. Еще раз проводим 50 эпох обучения на обучающей выборке. Дескриптор – выходы предпоследнего слоя (нормализации по батчу).

Этот метод обучения показал точность rank-1 85%, что сравнимо с наилучшими из известных методов решения этой задачи. Из чего можно сделать вывод, что в этой задаче низкоуровневые признаки сохраняют основную необходимую информацию и дальнейшее усложнение архитектуры нецелесообразно.

3.4. Построение бинарных выходов

Введем параметр hash_size – количество бит в выходном бинарном дескрипторе. Добавим к полученной на предыдущем этапе сети еще один слой – полносвязный слой с hash_size нейронами и функцией активации сигмоида, назовем этот слой бинаризирующим. Расположим его после слоя "batch normalization", но перед классифицирующим слоем. При этом классифицирующий слой требуется заново инициализировать и обучить всю сеть еще раз по предложенной выше схеме, задав в качестве начального приближения полученные на предыдущем этапе веса. Бинарный дескриптор или хеш – результат сравнения выходов бинаризирующего слоя с 0.5, он имеет размерность hash_size бит. Путем изменения этого параметра в данной работе проверены хеши длиной 128, 256 и 512 бит соответственно.

Предложенный метод бинаризации повторяет алгоритм сигмоиды из статьи [12] за исключением того, что в предложенном методе в качестве базовой архитектуры нейронной сети вместо алгоритма из статьи используется модифицированный алгоритм VGG16.

4. Экспериментальное исследование

Таблица 1. Сравнение эталонных коллекций.

Коллекция данных	VIPeR [3]	PRID [6]	CUHK 03 [11]	Market 1501 [20]
Число личностей	632	385	1467	1501
Примеров на человека	1	1	2-10	≈ 15
Размер кадров	128 × 48	128 × 64	≈ 160 × 60	128 × 64
Число камер	2	2	2	6

Был проведен обзор открытых коллекций по результатам которого составлена таблица 1.

Для экспериментальной оценки выбрана коллекция данных Market1501, состоящая из 1501 набора изображений людей. Изображения цветные и имеют разрешение 128x64, они разделены на 3 непересекающихся множества:

1. Обучающая выборка, состоящая из 12936 изображений 751 человека. Для каждого изображения есть метка, какой человек на нем изображен.

2. Тестовая выборка, состоящая из 19732 изображений 750 человек и два класса отвлекающих изображений. Все фотографии подписаны. Личности в обучающей и тестовой выборке не пересекаются.
3. Выборка запросов, состоящая из 3368 изображений 750 человек. Здесь находятся другие фотографии тех же личностей, что и в тестовой выборке. Фотографии подписаны.

4.1. Протокол тестирования

Рассмотрим два протокола тестирования, представленные на выбранной коллекции: запрос по **одному** и **нескольким** изображениям.

Запрос по **одному** изображению:

1. Обучаем алгоритм извлечения дескриптора изображения человека исключительно на обучающей выборке.
2. Строим базу дескрипторов – каждой картинке ставим в соответствие дескриптор с помощью тестируемого алгоритма. В тестовой выборке фотографии 750 личностей и 2 класса с объектами, не являющимися людьми. Изображения распределены по классом примерно равномерно.
3. Берем изображения-запросы из выборки запросов, для них строим дескрипторы-запросы. Всего 3368 запросов. Картинки из выборки запросов до этого никак не использовались.
4. Для каждого дескриптора-запроса находим k ближайших соседей среди дескрипторов из базы дескрипторов, отсортированных по близости (по метрике L_2).
5. Определяем классы этих дескрипторов.

Запрос по **нескольким** изображениям отличается тем, что запросы состоят из набора изображений. Эти наборы состоят из всех изображений из `query_set`, соответствующих конкретному человеку, снятому с конкретной камеры, чаще всего это 2-6 изображений. Действия проводятся такие же, за исключением пункта 3.

3. Сначала для каждого изображения в запросе строится дескриптор, затем для построения дескриптора-запроса дескрипторы в наборе объединяются с помощью какого-то алгоритма – чаще всего для каждого элемента считается среднее или максимум, но возможны более сложные алгоритмы.

Для предложенного алгоритма в данной работе считалось среднее по каждому элементу.

4.2. Используемые метрики качества

Для определения схожести изображений использовалось сравнение вещественных и бинарных дескрипторов по метрикам L_2 и L_1 соответственно.

$rank - k$ – процент запросов из выборки запросов, для которых изображение нужного класса содержится среди первых k примеров, отсортированных по схожести.

Также в этой эталонной коллекции для большинства реализаций используется мера точности Mean Average Precision (mAP) [16].

На выбранной эталонной коллекции данных уже сделано несколько реализаций алгоритмов повторной идентификации. Предложенное решение сравним с следующими реализациями:

1. Лучшим на данный момент подходом для этой задачи не использующим нейросети GOG [8].
2. Лучшими на данный момент подходами для этой задачи, основанными на нейронных сетях ResNet [5] и MobileNet [19].

Для существующих реализаций мы будем использовать точность, описанную в оригинальных работах, а реализацию предложенного алгоритма протестируем по протоколу, описанному в обзоре существующих методов. Отметим, что авторы статьи [8] не проводили тестирования алгоритма GOG на запросах, состоящих из нескольких изображений, поэтому результаты для этого протокола не добавлены в соответствующую таблицу, однако сравнительно низкая точность на запросах из одного изображения, позволяет предположить, что он уступает нейросетевым алгоритмам построения дескрипторов изображения человека. В качестве меры точности будем использовать метрики mAP и rank-1, rank-5. Для проверки методов бинаризации и сравнения качества бинаризации при разных размерах бинарного дескриптора мы будем использовать тот же тестовый протокол, но будем использовать только основную меру качества для данной эталонной коллекции – rank-1.

Для проверки переносимости базового алгоритма с коллекции Market1501 на коллекцию CUNK03 [11] и обратно был реализован такой же протокол тестирования на CUNK03. Все личности были случайным образом разбиты на два непересекающихся множества – тренировочный и тестовые наборы, затем из тестового набора были выделены случайные 15% изображений в качестве множества запросов, эти изображения больше не присутствуют в тестовом наборе. Именно такое отношение изображений-запросов ко всем изображениям для тестирования используется в Market1501.

5. Результаты

Таблица 2. Сравнение методов построения вещественных дескрипторов, бинаризация не проводилась, запрос из **одного** изображения.

Алгоритм	rank-1, %	rank-5, %	mAP, %
GOG [8]	58.6	79.4	–
Предложенный метод	85.33	99.55	51.42
ResNet+ TripletLoss [5]	86.67	93.38	81.07
MobileNet+ DML [19]	87.73	–	68.83

Таблица 3. Сравнение методов построения вещественных дескрипторов, бинаризация не проводилась, запрос из **нескольких** изображений.

Алгоритм	rank-1, %	rank-5, %	mAP, %
ResNet+ TripletLoss [5]	91.75	95.78	87.18
MobileNet+ DML [19]	91.66	–	77.14
Предложенный метод	92.91	99.73	62.10

Экспериментальное сравнение с лучшими реализациями на данный момент (таблицы 2 и 3) показало, что предложенная реализация имеет сравнимую с наилучшей точность по мере качества rank-1 в задаче повторной идентификации с запросом из одного изображения. А с запросом из нескольких изображений предложенный алгоритм справился лучше известных аналогов, что демонстрирует его применимость в обработке видеопоследовательностей, где и используется несколько изображений.

Таблица 4. Сравнение методов бинаризации. Изображена точность rank-1, больше – лучше.

Алгоритм	512 бит	256 бит	128 бит
Наивная бинаризация	83.72	–	–
Выделение главных компонент [4]	78.50	75.15	68.34
”Random forest” [1]	83.72	77.46	65.17
Алгоритм DBE [13]	83.90	78.50	71.79
Предложенный метод	83.40	79.15	75.20

Экспериментальная проверка (таблица 4) показала, что наилучшим решением для дескриптора длиной 512 бит является алгоритм DBE, но наивный подход ненамного ему уступает. При уменьшении дескриптора наилучший результат показывает метод сигмоиды.

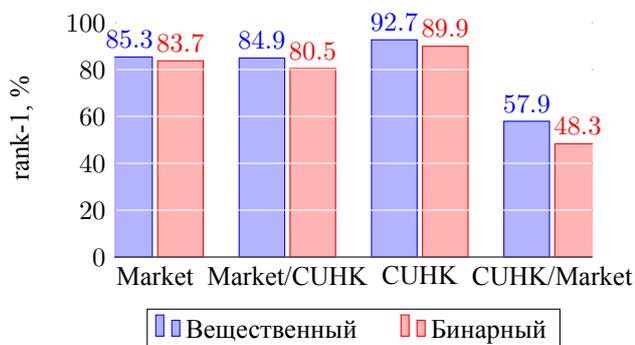


Рис. 3. Оценка возможности переноса на другие независимые данные

Проведенная экспериментальная оценка (рис. 3) показала, что алгоритм, подготовленный на данных из Market1501, применим на других полученных из независимого источника данных. Точность rank-1 уменьшилась менее, чем на 1%. Однако алгоритм, подготовленный на данных из CUHK03 теряет в точности при переходе на Market1501, что показывает, что эталонная коллекция CUHK03 недостаточно разнообразна, на ней происходит переобучение.

Нужно заметить, изображения в разных эталонных коллекциях имеют различное разрешение, а люди на них занимают разный процент кадра. Вопрос зависимости качества обучения от этих дополнительных параметров эталонной коллекции требует дополнительного исследования в будущем. На данный момент он не может быть решен ввиду малого количества коллекций достаточного для обучения нейросетевых алгоритмов размера.

6. Заключение

В работе предложен алгоритм построения дескриптора изображения человека для повторной идентификации и сопровождения в видео. С задачей поиска по нескольким изображениям он справился лучше аналогов по мере качества rank-1, rank-5 и mAP на эталонной коллекции Market1501.

Бинарная модификация уступила базовому алгоритму по точности (83.90% против 85.33% по метрике rank-1), но полученные дескрипторы занимают на 1-2 порядка меньше памяти. Проверка переносимости показала, что предложенный алгоритм можно использовать на данных, полученных из независимого источника.

7. Дальнейшие исследования

С помощью разветвленных нейросетевых архитектур можно добиться совместной обработки разноуровневых признаков, что может привести к росту качества получаемого дескриптора.

Было бы интересно переработать тестовый протокол, чтобы увидеть результаты работы предложенного алгоритма и конкурентов для случая, когда 750 людей из тестовой выборки представлены каждый всего одной или двумя фотографиями. Эта задача возникает, если нужно сравнить кадр с видеокамеры с одной-двумя фотографиями преступников.

8. Литература

- [1] Breiman L. Random Forests // Machine learning. – 2001. – Т. 45. – №. 1. – С. 5-32.
- [2] Dozat T. Incorporating Nesterov momentum into Adam. – 2015. – <http://cs229.stanford.edu/proj2015/054report.pdf>, 2015.
- [3] Gray D., Brennan S., Tao H. Evaluating appearance models for recognition, reacquisition, and tracking // Proc. IEEE International Workshop on Performance

- Evaluation for Tracking and Surveillance (PETS). – 2007. – Т. 3. – №. 5.
- [4] Halko N., Martinsson P. G., Tropp J. A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions // SIAM review. – 2011. – Т. 53. – №. 2. – С. 217-288.
- [5] Hermans A., Beyer L., Leibe B. In Defense of the Triplet Loss for Person Re-Identification // arXiv preprint arXiv:1703.07737. – 2017.
- [6] Hirzer M. et al. Person re-identification by descriptive and discriminative classification // Scandinavian conference on Image analysis. – Springer Berlin Heidelberg, 2011. – С. 91-102.
- [7] Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift // International Conference on Machine Learning. – 2015. – С. 448-456.
- [8] Karanam S. et al. A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets // arXiv preprint arXiv:1605.09653. – 2016.
- [9] Kuplyakov D., Shalnov E., Konushin A. Further Improvement on an MCMC-based Video Tracking Algorithm. // ГРАФИКОН'2016 Труды 26-й Международной научной конференции. 2016. С. 440-444.
- [10] Li S. et al. Person search with natural language description // arXiv preprint arXiv:1702.05729. – 2017.
- [11] Li W. et al. Deepreid: Deep filter pairing neural network for person re-identification // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2014. – С. 152-159.
- [12] Lin K. et al. Deep learning of binary hash codes for fast image retrieval // Proceedings of the IEEE conference on computer vision and pattern recognition workshops. – 2015. – С. 27-35.
- [13] Liu L. et al. End-to-end Binary Representation Learning via Direct Binary Embedding // arXiv preprint arXiv:1703.04960. – 2017.
- [14] Matsukawa T. et al. Hierarchical gaussian descriptor for person re-identification // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – С. 1363-1372.
- [15] Morozov F., Konushin A. Background subtraction using a convolutional neural network. // ГРАФИКОН'2016 Труды 26-й Международной научной конференции. 2016. С. 445-447.
- [16] Serepce S. et al. Information retrieval – 2017. – https://en.wikipedia.org/wiki/Information_retrieval.
- [17] Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // arXiv preprint arXiv:1409.1556. – 2014.
- [18] Ulu A., Ekenel H. K. Convolutional neural network-based representation for person re-identification // Signal Processing and Communication Application Conference (SIU), 2016 24th. – IEEE, 2016. – С. 945-948.
- [19] Zhang Y. et al. Deep Mutual Learning // arXiv preprint arXiv:1706.00384. – 2017.
- [20] Zheng L. et al. Scalable person re-identification: A benchmark // Proceedings of the IEEE International Conference on Computer Vision. – 2015. – С. 1116-1124.
- [21] Zheng L., Yang Y., Hauptmann A. G. Person Re-identification: Past, Present and Future // arXiv preprint arXiv:1610.02984. – 2016.
- [22] Cheng Z., Li X., Loy C. C. Pedestrian Color Naming via Convolutional Neural Network // Asian Conference on Computer Vision. – Springer, Cham, 2016. – С. 35-51.

Способ бесконтактного измерения размеров крупногабаритных объектов

А.А. Михайличенко¹, А.Б. Клещенко^{2,3}
alexey.a.mikh@gmail.com|aktech@inbox.ru

¹Институт математики, механики и комп. наук, Южный Федеральный университет, Ростов-на-Дону, Россия;

²Физический факультет, Южный Федеральный университет, Ростов-на-Дону, Россия;

³ООО «ИЗОСКАН», г. Ростов-на-Дону, Россия

В работе представлен метод бесконтактного измерения размеров крупных объектов на принципе лазерной триангуляции. Предложенный метод отличается возможностью применения в условиях сильной посторонней за- светки, высокой скоростью измерений и низкой стоимостью оборудования.

Ключевые слова: Компьютерное зрение, Бесконтактное измерение, Лазерное сканирование, Трехмерная рекон- струкция.

An Approach to Non-Contact Measurement of Large-Sized Objects

A.A. Mikhaylichenko¹, A.B. Kleschenkov^{2,3}
alexey.a.mikh@gmail.com|aktech@inbox.ru

¹Institute of Mathematics, Mechanics and Computer Science, Southern Federal University, Rostov-on-Don, Russia;

²Faculty of Physics, Southern Federal University, Rostov-on-Don, Russia; ³ISOSCAN, Ltd, Rostov-on-Don, Russia

In this paper we present an approach to non-contact measurement of large-sized objects based on laser triangulation principle. The proposed method is characterized by application in conditions of strong external illumination, high measurement rate and low hardware cost.

Keywords: Computer Vision, Non-Contact Measurement, Laser Scanning, 3D Reconstruction.

1. Введение

Во многих задачах промышленной автоматизации, логистики и управления движением требуется измерение размеров крупногабаритных объектов бесконтактным способом. Одним из распространенных подходов является построение пространственной модели объекта с последующим определением геометрических параметров. Для пространственной реконструкции можно использовать технологии с применением сканирующих лазерных дальномеров [1], лидаров, различные алгоритмы пассивного и активного стереовидения [2, 3] и т.п.

В работе предлагается метод построения трехмерного облака точек измеряемого объекта методом лазерной триангуляции. Суть метода заключается в высокоскоростном измерении профилей объекта с одновременным движением объекта. Новизна метода заключается в применении в качестве подсветки линии из точек, в отличие от классического подхода, при котором используется непрерывная линия. Это позволяет поднять отношение сигнал/шум без увеличения мощности подсветки, а также за счет дополнительной обработки данных уменьшить вероятность ошибок измерений при возникновении бликов. Полученный набор профилей и измеренная скорость движения объекта в моменты измерения профилей позволяют построить пространственную модель объекта.

2. Обзор

В настоящее время существуют различные подходы и технические решения для реконструкции трехмерных моделей объектов. У каждой из технологий свои условия применения, требования к оборудованию, преимуществам и недостаткам.

3D сенсоры вроде Microsoft Kinect или Intel RealSense имеют ограниченную область применения. Устройства, использующие инфракрасную подсветку или инфракрасные датчики глубины и показывающие хорошие результаты в помещении, в уличных условиях из-за солнечного света неприменимы. Современные версии RealSense, которые помимо инфракрасных датчиков используют алгоритмы стереозрения, в уличных условиях демонстрируют лучшие результаты, однако в случае бликов, тени или просто обширных однотонных участков на изображении зачастую дают неверные результаты.

Одними из наиболее популярных для промышленных применений являются системы на основе сканирующих лазерных дальномеров. Во многих приложениях измеряемый объект движется (например, груз на конвейере или автомобиль на дороге). Если скорость движения известна, применяется система из двух сканирующих дальномеров, которые используются для измерения поперечного профиля объекта одновременно с его движением через измерительную зону. Если же скорость движения объекта неизвестна, то для ее измерения система дополняется третьим сканирующим дальномером. Такой подход позволяет получить вы-

сокую точность измерений при низкой скорости движения объекта. Однако, при увеличении скорости до 10 – 15 м/с из-за ограниченной скорости сканирования (как правило, 50–100 изм./сек) возрастает ошибка измерения длины. Другим важным недостатком является высокая стоимость оборудования, которая доходит до 15–20 тыс. долл. США.

Снизить стоимость решения можно с помощью технологии стереовидения [2, 3], которая позволяет получить трехмерную модель объекта по данным с двух и более камер. Технология реконструкции трехмерной сцены методом стереовидения в настоящее время набирает популярность в различных приложениях, однако существующие методы либо обладают большой вычислительной сложностью, что усложняет их применение в задачах реального времени, либо имеют низкую точность реконструкции. Более того, точность реконструкции снижается для объектов с хорошо отражающими поверхностями из-за бликов на изображении.

Еще одна технология трехмерного сканирования заключается в измерении расстояний до точек на поверхности объекта методом лазерной триангуляции [4, 5]. Лазерная триангуляция основана на способности лазерного луча распространяться в хорошо коллимированной форме на большие расстояния. Для измерения расстояния до определенной точки происходит подсветка этой точки лазером. Отраженный свет от этой точки контролируется детектором, который устанавливается на некотором расстоянии источника от лазерной подсветки, таким, что источник излучения, объект и детектор образуют треугольник. Положение яркого пятна на фоточувствительной матрице детектора показывает направление входящего света, т.е. угол между лазерным лучом и отраженным светом, откуда может быть рассчитано расстояние.

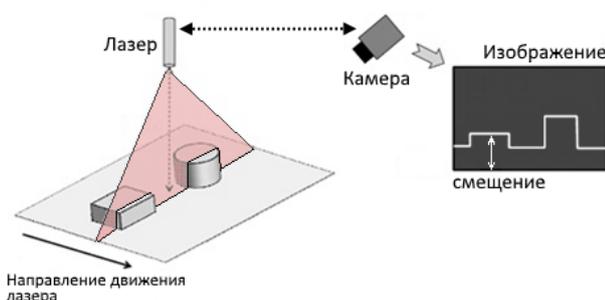


Рис. 1. Измерение профиля объекта методом лазерной триангуляции.

Для измерения профиля объекта вместо одного луча используют линию [6]. Эту линию, как правило, получают из луча с помощью призмы. В этом случае, на приемной матрице фоточувствительного элемента формируется уже не одна точка, а отраженное от объекта изображение линии на его поверхности. Форма изображения кривой позволяет получить профиль объекта (рис. 1). В работе предложен способ измерения

профиля на основе лазерной триангуляции и включающий ряд решений для повышения точности измерений.

3. Описание

Исходя из условий применения к системе измерения профилей предъявляются противоречивые требования — эксплуатация на открытом пространстве при наличии прямых солнечных лучей и соблюдение требований по лазерной безопасности, что накладывает ограничения на мощность лазерного излучения. В [5] показано, что без применения специальных приемов метод активной лазерной триангуляции в условиях прямого солнечного света практически не работает. Поэтому было применено несколько технических решений, позволяющих максимизировать отношение полезного сигнала к уровню естественного освещения.

Во-первых, в качестве профиля лазерной подсветки используется не сплошная линия, а линия, состоящая из отдельных точек — для формирования такой структурированной подсветки применены специализированные дифракционные оптические элементы. Такой прием позволяет снизить мощность источника подсветки пропорционально увеличению расстояния между изображением точек на поверхности объекта при сохранении яркости в самой точке. Во-вторых, для оптической фильтрации сигнала подсветки и минимизации уровня шумовой фоновой засветки, возникающей от посторонних источников света, солнечного света и бликов, применены узкополосные оптические фильтры на длину волны лазерной подсветки. И в-третьих, используется минимально возможное время экспозиции измерительной камеры, при этом момент экспозиции синхронизирован с моментом включения подсветки. Эти три приема позволили существенно увеличить отношение сигнал/фон не превышая предельно допустимую разрешенную интегральную энергию лазерного излучения.



Рис. 2. Схема измерительного комплекса.

В каждый момент времени система получает изображение объекта, подсвеченного подсветкой. Одной из главных задач детектора является задача обнаружении точек, формируемых на поверхности объекта подсветкой. После того, как на кадре найдены необходимые пятна (или подтверждено их отсутствие), по смещению пятна определяется расстояние от камеры до объекта.

На следующем этапе смещение преобразуется в реальные координаты соответствующей точки, набор которых для одного кадра и образует профиль объекта. По набору таких профилей для разных моментов времени, зная мгновенную скорость объекта для каждого профиля, можно получить трехмерную модель объекта.

3.1. Поиск пятен на изображении

На рис. 3, представлено типовое изображение, получаемое измерительной камерой. Видно, что точки от лазерной подсветки представлены в виде «пятен» различного радиуса и яркости. В компьютерном зрении существуют различные подходы к обнаружению подобных «пятен» — дифференциальные детекторы, которые основаны на вычислении производных [7], методы, основанные на поиске локального максимума интенсивности [8] или других особенностей изображения (SURF, SIFT) и т.п.



Рис. 3. Пример изображения, полученного с измерительной камеры.

В задачах реального времени важным критерием при выборе алгоритма является быстрдействие, так как это определяет частоту измерений, и, соответственно — степень детализации реконструкции. В качестве основного был выбран алгоритм, основанный на свертке изображения с лапласианом гауссиана (Laplacian of Gaussian, LoG).

Еще одной важной особенностью метода, позволяющей понизить вероятность ложного обнаружения пятен является то, что положение точки на изображении для каждого из лучей при изменении дальности до объекта изменяется вдоль определенной кривой — траектории точки. Для каждого из лучей есть своя траектория (рис. 4), которая не пересекается с траекториями других точек. Таким образом, поиск пятен на изображении достаточно проводить только вдоль заранее известных траекторий, что снижает вероятность обнаружения ложных пятен и существенно ускоряет обработку одного кадра за счет уменьшения областей поиска.

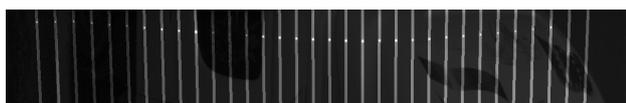


Рис. 4. Траектории точек, формируемых лазерной подсветкой на поверхности объекта.

Дополнительно, чтобы избежать срабатывания алгоритма по ложным пятнам (бликам) или по шуму, обнаруженные пятна подвергаются фильтрации по уровню энергии. Критериями здесь являются отношение яркости предполагаемого пятна к следующему по яр-

кости пятну и превышение яркости пятна над среднеквадратичной яркостью остальной области поиска данного пятна. Пороги достаточно подобрать один раз при калибровке всего алгоритма на реальных данных.

3.2. Калибровка

Измерение дальности до объекта для каждого луча с учетом известного углового расстояния между лучами позволяет построить профиль объекта, заданный некоторым количеством точек на плоскости. При этом, дальность является функцией пиксельного смещения образа луча на изображении. Таким образом, для точной количественной оценки расстояния необходима информация о соответствии между пиксельным смещением точки на изображении и дальностью до объекта. В идеальных условиях эту зависимость можно определить аналитически, однако, из-за оптических искажений объектива, температурной зависимости размеров оптических элементов и т.п. эта задача сильно усложняется.

Для получения однозначного соответствия между дальностью и смещением проще провести ряд измерений, которые, по сути, являются калибровкой системы. Калибровка позволяет получить требуемые зависимости для всех лучей, и состоит из следующих этапов:

- получение изображений точек (пятен) для каждого из лучей при различных дальностях из рабочего диапазона и локализация этих точек;
 - аппроксимация полученной зависимости дальность/смещение для каждого из лучей — наилучший результат был получен при аппроксимации иррациональным полиномом 2-го порядка вида
- $$y = \frac{a}{x + b};$$
- определение траекторий перемещения изображения точек каждого из лучей при изменении дальности (аппроксимация полиномом 2-го порядка);
 - измерение углов между лучами.

Одной из важнейших характеристик системы является точность реконструкции, которая определяется точностью измерения дальности и угловыми расстояниями между лучами. В свою очередь, точность измерения дальности определяется точностью проведения калибровки и разрешающей способностью по дальности. Разрешающая способность h измерения дальности методом лазерной триангуляции зависит от дальности до объекта D , разности между лазером и камерой B , фокусного расстояния f и физического размера пикселя p :

$$h(D) = \frac{D^2 p}{fB} \quad (1)$$

Как видно, разрешение метода квадратично падает с увеличением дальности до объекта. Для повыше-

ния разрешения возможно определение координат пятна с субпиксельной точностью. Для этого используется аппроксимация окрестности пятна гауссианом, с помощью которого и определяется центр этого пятна. На рис. 5 представлена зависимость разрешения метода от дальности до объекта, измеренная экспериментально.

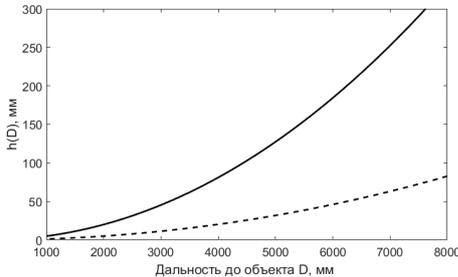


Рис. 5. Зависимость разрешающей способности измерения дальности от расстояния до объекта без уточнения (сплошная линия) и с субпиксельным уточнением в 4 раза (пунктирная линия).

Разрабатываемая система предназначена для измерения габаритных размеров транспортных средств (ТС), поэтому для возможности полной реконструкции измеряемых ТС необходимо определение положение измерителей профиля относительно дороги. Для этого рассмотрим систему координат (СК) с центром в точке размещения камеры (рис. 6, а). Ось y направлена вдоль луча с номером 1. Точки на поверхности земли в этой СК имеют координаты:

$$\begin{aligned} p_1 &= (0, -d_1), \\ p_2 &= (d_2 \sin \varphi_2, -d_2 \cos \varphi_2), \\ &\dots \\ p_n &= (d_n \sin \varphi_n, -d_n \cos \varphi_n). \end{aligned} \quad (2)$$

Используя координаты этих точек, находим при помощи метода наименьших квадратов уравнение прямой L в СК Ox_y . Прямую L далее будем называть «горизонтом». Чтобы определить высоту расположения камеры H , достаточно найти расстояние до этой прямой от точки $(0, 0)$.

Таким образом, в результате калибровки определяются данные о точной зависимости пиксельного смещения и дальности до поверхности объекта для каждого из лучей, траектории перемещения этих пятен (для оптимизации их поиска), углы между лучами, а также положение измерительной камеры относительно плоскости отсчета. Эта информация позволяет однозначно определить пространственное положение точек профиля объекта в абсолютных координатах.

Построение пространственной модели объекта. Рассмотрим рис. 6, б. Если углы лучей заданы относительно первого луча, для дальнейших преобразований координат необходимо определить его угол. К этому моменту уже должны быть известны точки горизонта и его уравнение в виде

$$y = k_H x + b_H. \quad (3)$$

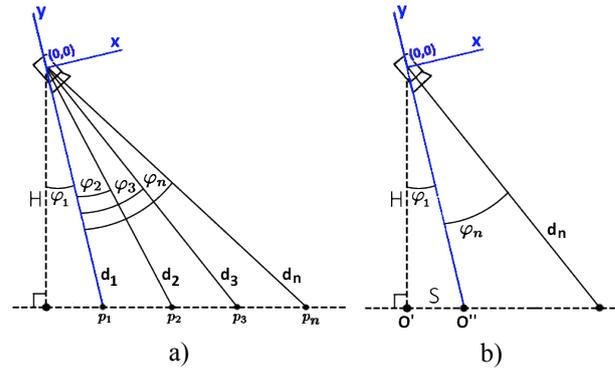


Рис. 6. Иллюстрации к преобразованиям координат.

O' — точка пересечения горизонта и нормали, проведенной к горизонту через точку $(0, 0)$, O'' — точка пересечения горизонта и оси y , где $O''_x = 0$, а O''_y находим из уравнения (3) при $x = 0$. Обозначим отрезок $O'O'' = S$, и тогда

$$\varphi_1 = \arctg \left(\frac{S}{H} \right). \quad (4)$$

Для перехода из СК центра камеры и первого луча в качестве оси y к СК с центром в точке O' можно воспользоваться следующими формулами:

$$\begin{cases} x_n = d_n \sin (\varphi_n + \varphi_1), \\ y_n = H - d_n \cos (\varphi_n + \varphi_1), \end{cases} \quad (5)$$

где x_n, y_n — координаты точки p_n в СК с центром в O' , d_n — дальность для n -го луча.

Реконструкция отдельных поперечных профилей объекта производится при помощи формул (5). Для определения длины объекта достаточно знать временные метки каждого профиля и скорость объекта для каждой такой временной метки.

Несмотря на довольно точную реконструкцию, которая производится описанным выше методом, на завершающей стадии обработки для гарантированного исключения выбросов, возникающих от шума и бликов на поверхности объекта, выполняется фильтрация облака точек, удаляющая такие выбросы. Описание примененного алгоритма фильтрации выбросов можно найти в [9].

4. Эксперимент

Описанный подход был апробирован на задаче измерения габаритных размеров движущихся ТС. Для этого применено два измерителя профиля, размещенных над обочинами полосы движения. Результаты реконструкции представлены на рис. 7.

Измерения мгновенной скорости движения ТС производится оптическим методом с помощью растрового измерителя поперечной скорости, обеспечивающего точность измерения $0,2 - 0,5\%$.

Программное обеспечение измерительного комплекса разработано на языке C++. При описанном спо-

собе обработки изображения размером 200×1280 с целью поиска на нем пятен от подсветки время обработки одного изображения на процессоре Intel Core i7 не превышает 3–4 мс, при этом одновременно могут обрабатываться изображения от нескольких измерителей профиля.

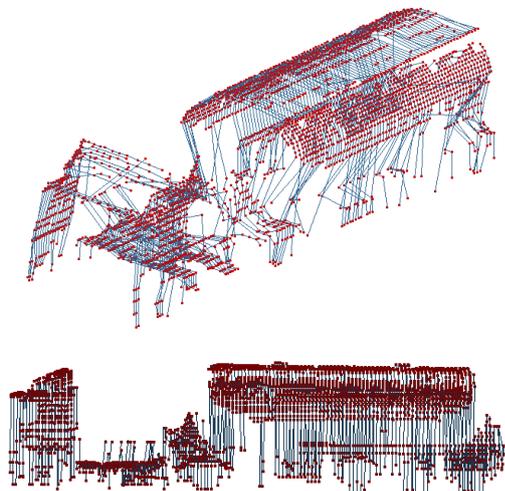


Рис. 7. Пример реконструкции.

Погрешность измерения поперечных размеров зависит от места установки измерителей профиля. При высоте их установке порядка 6 м над уровнем дороги и таком же расстоянии между ними и с применением алгоритма субпиксельной локализации она не превышает 3 – 3,5 см при высоте объекта 1,5 м и уменьшается при увеличении размера ТС (т.к. уменьшается измеряемая дальность). При использовании двух измерителей профиля с частотой 100 изм./сек, разрешающая способность по определению длины ТС, движущегося на скорости 40 км/ч, составляет около 10 см и уменьшается при повышении частоты кадров либо при снижении скорости движения объекта.

5. Выводы

В работе представлен метод, позволяющий с высокой точностью проводить реконструкцию крупногабаритных движущихся объектов. Результаты реконструкции могут быть использованы для определения линейных размеров объекта, анализа его формы и т.п. Погрешность измерения ширины и высоты ТС составляет не более 3-3,5 см, а измерения длины — не более 10 см при скорости до 40 км/ч.

Актуальным остается вопрос о методе измерения скорости. В настоящее время для этого применен растровый измеритель скорости, но ведутся исследования по применению альтернативных методов — например, использования оптического потока.

Полученные результаты позволяют сделать вывод о правильности выбранного подхода для измерения размеров движущихся объектов и целесообразности дальнейших исследований.

6. Благодарности

Работа выполнена при частичной финансовой поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере (проект №1639ГС1/24342).

7. Литература

- [1] Christie D., Jiang C., Paudel D., Démonceaux C. 3D Reconstruction of Dynamic Vehicles using Sparse 3D-Laser-Scanner and 2D Image // International Conference on Informatics and Computing, 2016.
- [2] Hirschmuller H. Stereo Processing by Semi-Global Matching and Mutual Information // Pattern Analysis and Machine Intelligence, IEEE transactions on, 2008.
- [3] Hu F.J., Zhao Y. Comparative research of matching algorithms for stereo vision // Journal of Computational Information Systems 9(13).
- [4] Ashok W. R., Panse M. S., Apte H. Laser Triangulation Based Object Height Measurement // International journal for research in emerging science and technology, 2015. – Volume: 2, No.3.
- [5] Ilstrup D., Manduchi R. Active Triangulation in the Outdoors: A Photometric Analysis // Proceedings of the 5th International Symposium on 3D Data Processing, Visualization, and Transmission, 2010.
- [6] Jiang C., Yang J., Zhang L., Liu Z. Information extraction and three-dimensional reconstruction for a small laser scanning detector // International Conference on Optoelectronics and Microelectronics, 2012.
- [7] Kong H., Akakin H.C., Sarma S.E. A Generalized Laplacian of Gaussian Filter for Blob Detection and Its Applications // IEEE Transactions on Cybernetics, 2013. – Volume: 43, No.6. – pp.1719-1733.
- [8] Vincent L., Soille P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations // IEEE Trans. Pattern Anal. Mach. Intell., 1991. – Volume. 13, No. 6, – pp. 583–598.
- [9] Posio J., Leiviskä K., Ruuska J., Ruha P. Outlier Detection for 2D Temperature Data // Proceedings of the 17th World Congress The International Federation of Automatic Control Seoul, Korea, July 6–11, 2008.

Выделение отрезков на аэроснимках в рамках задачи обнаружения строений

Л.В. Новоторцев¹, А.Г. Волобой¹

torets13@gmail.com|voloboy@gin.keldysh.ru

¹Институт Прикладной Математики им. Келдыша РАН, Москва, Россия

В рамках фотограмметрии часто возникает задача обнаружения строений на аэрофотоснимках для последующей обработки. Один из эффективных способов обнаружения строений - выделение отрезков на снимке. Существующие способы, опирающиеся на данный подход, очень чувствительны к точности нахождения отрезков. Поэтому необходимо предоставить максимально точный метод обнаружения отрезков.

Ключевые слова: аэроснимки, выделение прямых, обнаружение строений, компьютерное зрение.

Extracting straight line segments from aeroimages in the context of building detection task

L.V. Novotortsev¹, A.G. Voloboy²

torets13@gmail.com|voloboy@gin.keldysh.ru

¹Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

One of the tasks of photogrammetry is building detection for further processing. An efficient way to approach this task is straight line detection based algorithms. But existent methods which use that scheme are sensitive to lost segment errors. That is way it is important to propose method for detecting most of straight line segments.

Keywords: aeroimages, straight line detection, building detection, computer vision.

1. Введение

Задача распознавания и реконструкции строений на снимках, полученных при помощи аэро- или космосъемки, является довольно распространенной в рамках фотограмметрии. Одна из сфер, где активно используются результаты этой задачи, - автоматическое создание или обновление электронных карт, где наличие трёхмерных моделей значительно увеличивает наглядность продукта. Другая сфера - анализ застройки как в коммерческих (нахождение выгодных и пригодных мест для новой застройки), так и в природоохранных (обнаружение нелегальной застройки) целях.

При решении этого типа задач используются три вида источников изображений: космосъемка, аэросъемка и съемка БПЛА (Беспилотный Летательный Аппарат). Космосъемка позволяет получить наиболее качественные снимки, более того, часто полученные снимки имеют большее количество каналов, чем 3 цветовых (RGB). Однако космоснимки очень дорогие, что ограничивает их применение. Снимки, полученные с самолетов и с БПЛА, отличаются качеством снимающей камеры (на БПЛА обычно ставят более простые и дешевые) и калибровкой камеры. Кроме того, аэросъемка близка к надирной (иными словами, направление съемки близко к перпендикуляру к земле).

В данной статье рассматривается обработка данных, полученных с помощью аэросъемки, поскольку такой тип данных наиболее распространен на текущий момент из-за оптимального соотношения стоимости данных и их качества и доступности.

Одним из распространенных подходов к решению задачи обнаружения и трехмерной реконструкции строений является выделение отрезков на снимке и последующий их анализ [1, 4, 8]. В отличие от космосъемки, рассматриваемые снимки не обладают такой высокой четкостью, поэтому необходимо обеспечить наибольшую точность и полноту полученных результатов. Существующие методы нахождения строений и их

реконструкции, опирающиеся на выделенные отрезки, довольно сильно чувствительны к найденным отрезкам и хорошо справляются с шумовыми "ложными" сигналами (однако они значительно замедляют скорость работы этих методов).

По этой причине в данной статье предлагается использовать алгоритм, основанный на максимально чувствительном выделении отрезков в сочетании с последующей фильтрацией шума. При этом акцент ставится на минимизацию найденных отрезков и уже второстепенно на минимизацию шума.

2. Обзор существующих методов

Основными подходами к решению задачи выделения отрезков на изображениях являются метод Хафа [11], алгоритм LSD [10] и метод EDLines[7].

Метод Хафа опирается на результаты, полученные после применения двоичного фильтра нахождения границ. Наиболее распространенным примером такого фильтра является фильтр Кэнни [5] и его модификации [4, 8].

Метод Хафа основан на методе голосования. Общий подход заключается в том, что для каждого пикселя, относящегося к грани, строятся всевозможные прямые, проходящие через этот пиксель. Использование оригинального метода Хафа в общем случае не годится из-за "глобальности" этого алгоритма (ищутся прямые, а не отрезки). Поэтому чаще используются модификации метода, такие как иерархический [13] или вероятностные подходы [14]. Помимо этого существует большое количество модификаций, направленных на уменьшение ошибок дискретизации, подбору критерия определения выделения прямой по таблице и скорости работы метода в целом [9, 15, 16].

Метод LSD является более новым алгоритмом, который обеспечивает большую скорость работы и для большинства случаев более качественными результатами [10]. Этот алгоритм работает с градиентным полем (получаемым при помощи фильтра Собеля [12]), а не с двоичной картой границ, что дополнительно увеличивает скорость работы.

Основным недостатком метода LSD в рамках поставленной задачи является пропуск некоторых отрезков, в том числе и явных на больших снимках. Этот недостаток можно компенсировать, применяя LSD к сегментам изображения и объединяя полученные отрезки исходя из информации о градиентах.

Метод EDLines является еще одним распространенным алгоритмом нахождения отрезков на изображениях. Идея данного подхода схожа с методом LSD, только гораздо большее внимание уделяется скорости алгоритма (позиционируется как алгоритм реального времени) и неразрывности нахождения прямых [7].

К недостаткам этого алгоритма относится большое количество близлежащих параллельных прямых. К примеру, прямые, образованные коньком крыши. В рамках задачи обнаружения строений, данный недостаток является существенным, поскольку осложняет обнаружение и реконструкцию крыш.

3. Основная идея

Чувствительность метода LSD увеличивается при уменьшении размера изображения, поэтому в данной работе предлагается использовать следующий алгоритм:

- 1) разбить изображение на сегменты и применить метод LSD к каждому сегменту;
- 2) выполнить объединение отрезков между сегментами;
- 3) выполнить более тщательное слияние отрезков на основе карты градиентов.

Рассмотрим каждый из этих пунктов подробнее.

4. Разбиение изображения на сегменты

Как уже было сказано выше, чувствительность LSD увеличивается при уменьшении изображения [10]. Но при этом есть ограничение на размер сегмента, связанное с погрешностью обнаружения угла прямой в сегменте.

На таблице 1 приведены соответствие размера сегмента и ошибки расчёта угла прямой в нём [2].

$\Delta\theta$	1	2	3	4	5	6	7	8	9	10
L	82	41	28	21	17	14	12	11	10	9

Таблица 1. Погрешность определения угла и минимальные размеры сегмента для обеспечения этой погрешности, где $\Delta\theta$ - погрешность расчета угла, L - размер сегмента.

В этой работе был выбран размер 100x100 пикселей на сегмент для достижения высокой точности определения угла и чувствительности алгоритма.

5. Объединение отрезков между сегментами

После получения отрезков в каждой области необходимо выполнить объединение отрезков.

В первую очередь все найденные отрезки помечаем как необработанные. Далее последовательно проходим по всем сегментам, на которые было разбито изображение. В каждом сегменте обрабатываем полученные в нём отрезки, которые всё ещё помечены как необработанные.

Для каждого отрезка выполняем проверку на наличие его продолжений в соседних областях. При выполнении проверки необходимо учитывать, что концы отрезков и углы их наклона могут не совпадать. Поэтому объединение выполняется при условии, если концы находятся не далее, чем на 10 пикселей и углов наклона отрезков отличаются не более чем на 1 градус. Величины подобраны исходя из

выбранного размера сегмента и погрешности определения угла (таблица 1). В этом случае выполняется слияние отрезков.

Слияние проводится в несколько этапов.

- 1) Усредняются параметры прямых на которых лежат отрезки. При усреднении в качестве весов используются длины объединяемых отрезков. Это делается для избегания существенного искажения, вызванного отрезками малой длины, которые имеют большую погрешность определения угла.
- 2) Концы отрезков определяются на основе дальних концов объединяемых отрезков и их проекции на полученную в результате усреднения прямую.

6. Детальное объединение отрезков

Несмотря на уже выполненное объединение отрезков между областями, результат может быть неудовлетворительным, поскольку отрезки в некоторых областях могут быть не найдены из-за низкого контраста или особенностей алгоритма LSD.

Для устранения подобного рода ошибок необходимо произвести дополнительные проверки на возможные объединения прямых.

Для каждого необработанного отрезка проверяется наличие отрезков на заданном расстоянии (в данной работе предлагается ставить это расстояние равным средним размерам строений на снимке). Расстояние между отрезками определяется минимальным расстоянием между их концами. Для того чтобы выполнилось объединение необходимо, чтобы было выполнены следующие условия.

- 1) Углы прямых, определяемых отрезками, должны отличаться не более чем на допустимую погрешность. Допустимая погрешность определяется при помощи таблицы 1 по длине наименьшего отрезка.
- 2) Выполняется проверка, насколько изменится угол объединенной прямой по сравнению с прямой, образованной большим из отрезков. Если этот угол превышает значение из таблицы 1 для соответствующей длины, то объединение не выполняется.
- 3) Между ближайшими концами рассматриваемых отрезков выполняется проверка поля градиентов. Слияние выполняется, если в рассматриваемой области находится градиент имеющий направление, совпадающее с нормалью рассматриваемой прямой. Отрезки, для которых выполнены эти три условия объединяются по тому же принципу, описанному в разделе 5.

7. Результаты

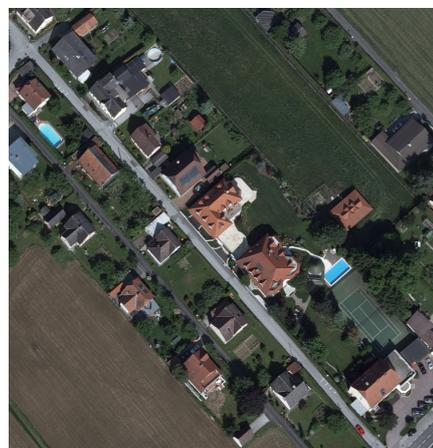


Рис 1. Исходное изображение.

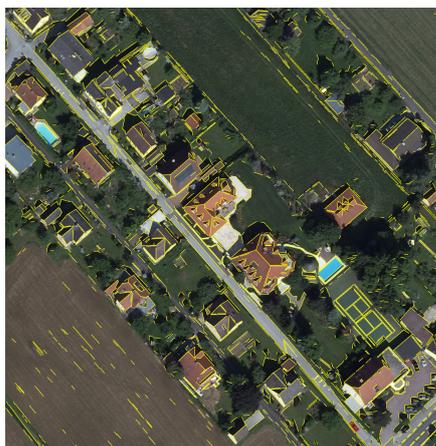


Рис 2. Результаты, полученные алгоритмом LSD.



Рис 3. Результаты, полученные алгоритмом EDLines.

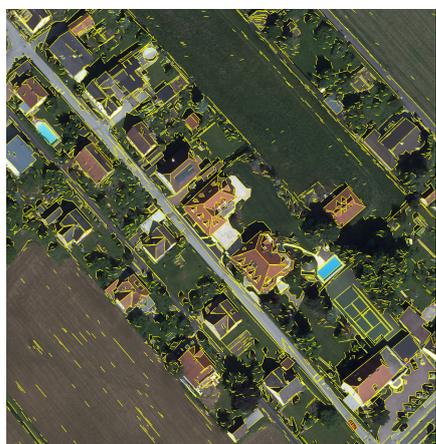


Рис 4. Результаты, полученные рассматриваемым алгоритмом.

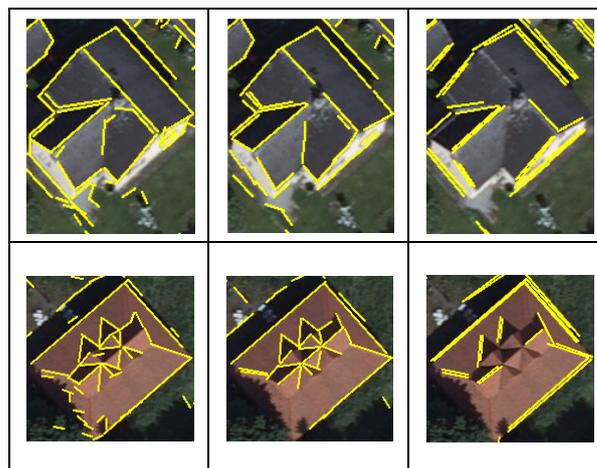


Таблица 2. Сравнение результатов выделения отрезков на участках изображения.

	LSD	EDLines	Предложенный метод	
			Получение отрезков	Объединение отрезков
T	2.68	1.56 c	2.79	0.67
N	2511	1467	4817	4289

Таблица 3. Сравнение методов на скорость работы и на количество найденных отрезков (T – время, затраченное на выполнение, N – Количество найденных отрезков).

На рисунке 1 приведено исходное изображение, на рисунке 2 изображение с линиями, найденными алгоритмом LSD, на рисунке 3 представлен алгоритм EDLines, а на рисунке 4 приведён результат применения описываемого метода. Более подробное сравнение участков изображений приведено в таблице 2. Также в таблице 3 приведены основные численные данные рассматриваемых методов.

На рисунке 3 видно, что по сравнению с результатами алгоритма LSD получено гораздо больше шума, но при этом некоторые прямые не найденные методом LSD были успешно выделены предложенным алгоритмом. Именно это и являлось целью данного исследования, поскольку дальнейшая обработка при обнаружении и реконструкции строений легко отбракует лишние шумы. Так же представленный алгоритм позволяет избежать дублирования обнаружения одних и тех же граней дома. Предложенный алгоритм работает медленнее обоих методов, представленных в сравнении, однако выигрывает в обнаружении отрезков, как это видно на рисунках в таблице 3.

8. Дальнейшие исследования

Данный алгоритм также можно модифицировать для нахождения пропущенных отрезков. Для этого необходимо искать отрезки между концами найденных отрезков, аналогично исследуя градиенты.

Предложенный метод	LSD	EDLines

9. Список литературы

- [1] Дьяконова С. В., Зайченко Ю. П. Подход к решению задачи автоматизированного обнаружения зданий на спутниковых изображениях. Вісник Національного технічного університету України КІП. Інформатика, управління та обчислювальна техніка, 58, С 51-55, 2013.
- [2] Новоторцев Л., Волобой А. Улучшенный алгоритм обнаружения отрезков для реконструкции строений по аэрофотоснимкам. Приволжский научный журнал, № 4, 2014. С. 49-53.
- [3] Baillard C., Zisserman A. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. International Archives of Photogrammetry and Remote Sensing, 33(B2) part 2, pp 56-62, 2000.
- [4] Bao P., Zhang D., Wu X. Canny edge detection enhancement by scale multiplication. IEEE transactions on pattern analysis and machine intelligence, 27(9), pp. 1485-1490, 2005.
- [5] Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, pp. 679-698, 1986.
- [6] Cohen J. P. et al. Rapid building detection using machine learning. arXiv preprint arXiv:1603.04392. – 2016.
- [7] Cuneyt Akinlar, Cihan Topal, EDLines: A real-time line segment detector with a false detection control, Pattern Recognition Letters, 32(13), pp. 1633-1642, 2011.
- [8] Dollár P., Zitnick C. L. Fast edge detection using structured forests. Pattern Analysis and Machine Intelligence, IEEE Transactions, 37(8), pp. 1558-1570, 2015.
- [9] Du, S., van Wyk, B. J., Tu, C., & Zhang, X. An improved Hough transform neighborhood map for straight line segments. IEEE Transactions on Image Processing, 19(3), pp. 573-585, 2010.
- [10] von Gioi R. G. et al. LSD: a line segment detector //Image Processing On Line, 2, 35-55, 2012.
- [11] Hough P. V. C. Method and Means for Recognizing Complex Patterns, U.S. Patent 3069654, Dec. 18, 1962.
- [12] I. Sobel and G. Feldman. A 3×3 Isotropic Gradient Operator for Image Processing. R. Duda and P. Hart (Eds.), Pattern Classification and Scene Analysis, pp. 271–272, 1973.
- [13] Li H., Lavin M. A., Le Master R. J. Fast Hough transform: A hierarchical approach. Computer Vision, Graphics, and Image Processing, 36(2), pp. 139-161, 1986.
- [14] Matas, J. and Galambos, C. and Kittler, J.V., Robust Detection of Lines Using the Progressive Probabilistic Hough Transform. CVIU 78 1, pp 119-137, 2000.
- [15] Nguyen T. T., Pham X. D, and Jeon J. An improvement of the Standard Hough Transform to detect line segments. IEEE International Conference on Industrial Technology, pp. 1-6, 2008.
- [16] O'Rourke J. Dynamically Quantized Spaces for Focusing the Hough Transform. IJCAI, 81, pp. 24-28, 1981.

Кластеризация пикселей цветового изображения методом Уорда

И.Г. Ханьков, М.В. Харинов
{igk, khar}@iias.spb.su

Санкт-Петербургский Институт информатики и автоматизации Российской академии наук (СПИИРАН),
Санкт-Петербург, Россия

В статье рассматривается кластеризация пикселей, предназначенная для сцен любого содержания. Для эффективного выделения объектов на изображениях применяется метод Уорда и способ тестирования алгоритмов на правдоподобие с использованием стереопар. Обсуждаются особенности применения метода Уорда кластеризации пикселей, и предлагается модель применения метода Уорда. Выполняется сравнение с аналогичными решениями.

Ключевые слова: сегментация изображения, кластеризация пикселей, суммарная квадратичная ошибка, минимизация, иерархия приближений, выпуклая последовательность.

Ward's pixel clustering for color image

I.G. Khanykov, M.V. Kharinov
{igk, khar}@iias.spb.su

St.Peterburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS),
St.Petersburg, Russia

In the paper pixel clustering, intended for scenes of any content is considered. To effectively detect objects, Ward's method and the method for testing the believability of algorithms using stereo-pairs are proposed. The features of the Ward's pixel clustering are discussed, and a model of the Ward's method application is developed. Comparison with similar solutions is performed.

Keywords: image segmentation, pixel clustering, total squared error, minimization, hierarchy of approximations, convex sequence.

1. Введение

На сегодняшний день, для расширяющейся предметной области обработки изображений разработано множество алгоритмов выделения объектов и продолжают разрабатываться новые алгоритмы для новых изображений и возникающих задач. Общим недостатком большинства алгоритмов является неустойчивость выделения объектов на изображениях заранее неизвестного содержания при изменении условий съемки. Число релевантных алгоритмов можно существенно сократить, если потребовать, чтобы они давали сходные результаты выделения объектов хотя бы при обработке стереопар, при которой каждое из двух изображений одной и той же сцены обрабатывается по отдельности, а также когда оба изображения для выделения объектов объединяются в одно (рис. 1). Тем самым, в первом случае (рис.1, сверху), моделируется выделение одних и тех же объектов на различных изображениях, а во втором случае (рис.1, внизу), моделируется выделение одинаковых объектов на одном изображении.

2. Двухмасштабная модель изображения

В работах [3,4,6,7] выдвигается идея двухэтапной сегментации изображения, при которой «объекты интереса» сначала локализуются в некотором фрагменте изображения, а затем детектируются в масштабе предварительно выделенного фрагмента. Согласно фундаментальной работе [6,7] изображение представляется в *двухмасштабной* модели суммой кусочно-гладкой компоненты, задающей протяженные объекты с резкими границами между ними, и дополнительной компоненты, которая включает в себя информацию о текстуре, мелких деталях, шумовую компоненту и пр.. В качестве *объектов*, разделенных контурными границами, рассматривается объединение покрывающих изображение связанных сегментов, которые различаются по цвету и др. признакам. В вычислениях существенно используются априорная информация об «объектах интереса».



Рис.1. Стандартные стереоизображения «Teddy» для тестирования алгоритмов сегментации изображений произвольной сцены.

3. Метод Уорда и двухмасштабная модель дихотомической иерархии объектов

Модель дихотомической иерархии объектов развивается с начала 90-х годов [8] и фактически является моделью применения метода Уорда [13,16] для кластеризации пикселей изображения, что при современных изображениях из миллионов пикселей является сложной научно-технической проблемой. Метод Уорда – это классический метод кластеризации [1,5], который в обработке изображений используется крайне редко [13] из-за большой вычислительной сложности, квадратично возрастающей с ростом числа пикселей N в изображении.

Для изображения из N пикселей в модели дихотомической иерархии объектов на выходе получается последовательность из N кусочно-постоянных приближений изображения с последовательно

возрастающим числом $g = 1, 2, \dots, N$ кластеров пикселей. При этом вычисляется $2N - 1$ различных множеств пикселей. Основной особенностью указанной последовательности приближений является то, что она описывается выпуклой последовательностью $E_1, E_2, \dots, E_N \equiv 0$ значений суммарной квадратичной ошибки E :

$$E_g \leq \frac{E_{g-1} + E_{g+1}}{2}, \quad g = 2, 3, \dots, N - 1,$$

или монотонно возрастающей последовательностью $\Delta E_2 \leq \Delta E_3 \leq \dots \leq \Delta E_{N-1} \leq 0$ неположительных приращений E . Характерно, что по этому признаку обсуждаемая последовательность иерархических приближений имитирует последовательность оптимальных приближений. Последняя отвечает минимально возможным значениям E и среднеквадратичного отклонения σ приближения от изображения ($E \equiv 3N\sigma^2$), но в общем случае не является иерархической.

Кластеризация пикселей по Уорду выполняется посредством итеративного слияния кластеров 1, 2, а именно тех, что обеспечивают минимальное приращение суммарной квадратичной ошибки ΔE_{merge} :

$$1, 2: \Delta E_{merge}(1, 2) = \min,$$

где ΔE_{merge} выражается через число пикселей n_1, n_2 и средние яркости I_1, I_2 в кластерах 1, 2 в виде:

$$\Delta E_{merge}(1, 2) = \frac{n_1 n_2}{n_1 + n_2} \|I_1 - I_2\|^2 \geq 0.$$

Если при минимизации приращения ΔE_{merge} вместо всевозможных пар кластеров анализируются только пары укрупняемых смежных множеств пикселей, то кластеризация переходит в сегментацию, как в модели Мамфорда-Шаха [2,14,15], позволяющей повысить скорость вычислений по Уорду за счет предварительного уменьшения числа обрабатываемых множеств пикселей.

Для операции *merge* при вычислениях поддерживается обратная операция *split* разделения кластера, в частности, сегмента на два, слиянием которых данный кластер или сегмент был получен. При этом приращение ΔE_{split} суммарной квадратичной ошибки E получается с обратным знаком:

$$\Delta E_{split}(1 \cup 2) = -\Delta E_{merge}(1, 2) \leq 0.$$

В модели предусмотрено также преобразование приближения изображения посредством операции *correct* реклассификации пикселей. В этом случае, при переносе n_3 пикселей кластера $3 \subset 1$, имеющих среднюю яркость I_3 , из кластера 1 в кластер 2 приращение $\Delta E_{correct}$ суммарной квадратичной ошибки E выражается в виде:

$$\Delta E_{correct} = \frac{n_2 n_3}{n_2 + n_3} \|I_3 - I_2\|^2 - \frac{n_1 n_3}{n_1 - n_3} \|I_3 - I_1\|^2.$$

Немаловажной особенностью дихотомической иерархии, получаемой методом Уорда или в модели Мамфорда-Шаха, является ее неоднозначность, которая связана с выраженной повторяемостью минимальных значений приращения суммарной квадратичной ошибки, особенно на начальных шагах укрупнения кластеров (или сегментов). При этом результат итеративных вычислений определяется выбором на каждом шаге очередной пары объединяемых множеств пикселей из набора равноправных вариантов и зависит от порядка их слияния, а изображение аппроксимируется многочисленными иерархическими последовательностями квазиоптимальных приближений,

которые описываются переплетающимися выпуклыми кривыми (рис. 2).

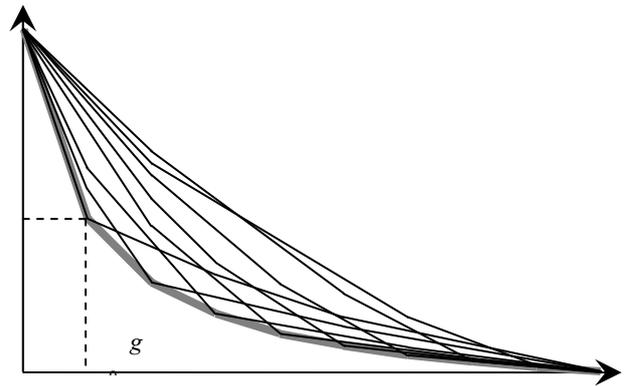


Рис.2. Неоднозначная аппроксимация последовательности оптимальных приближений (границная серая кривая) выпуклыми последовательностями квазиоптимальных приближений (верхние кривые).

На рис.2 изображены выпуклые графики суммарной квадратичной ошибки E в зависимости от числа кластеров g . Из-за неоднозначной аппроксимации оптимальных приближений изображения иерархическими квазиоптимальными приближениями возникает параметризация последних значением числа кластеров g_0 , при которой кривая для квазиоптимальных приближений максимально приближается к кривой для оптимальных приближений. Приближение с g_0 кластерами пикселей разделяет иерархию объектов с числом кластеров от 1 до $g_0 - 1$ от иерархии частей внутри объектов с числом кластеров от g_0 до N . Тем самым изображение иерархически структурируется в двух масштабах — объектов и частей объектов.

Разделение изображения на g_0 изображений объектов, задаваемых кластерами пикселей, определяется тем, что:

- 1) последовательность приближений внутри каждого объекта как самостоятельного изображения описывается выпуклой последовательностью значений суммарной квадратичной ошибки E ;
- 2) максимальное падение $|\Delta E_{split}|$ суммарной квадратичной ошибки при разделении надвое любого объекта не превышает минимального приращения ΔE_{merge} при слиянии любой пары объектов в один объект: $\min \Delta E_{merge} \geq \max |\Delta E_{split}|$;
- 3) реклассификация любой части объекта из одного объекта в другой не повышает суммарной квадратичной ошибки E : $\min \Delta E_{correct} \geq 0$.

Выполнение первого условия достигается расчетом иерархии частей по Уорду и обеспечивает максимальное падение $|\Delta E_{split}|$ суммарной квадратичной ошибки при разделении каждого объекта надвое.

Выполнение второго условия поддерживается встречным применением операций *split* и *merge* в так называемом *SI*-методе [10] и обеспечивает кластеризацию объектов при достраивании иерархии объектов до полной метод Уорда без нарушения условия выпуклости при начальном числе кластеров g_0 .

Третье условие обеспечивается итеративным применением операции *correct*, обеспечивающей

максимальное падение суммарной квадратичной ошибки $|\Delta E_{split}|$. Указанный метод является уточненной версией метода K -средних [1,5] и его обобщением на обработку иерархически структурированных кластеров пикселей.

4. Экспериментальные результаты

Рисунки 3-6 показывают результаты обработки тестовых изображений рис. 1 посредством итеративного слияния множеств пикселей.

На рис. 3-4 приведены результаты сегментации в модели сегментации Мамфорда-Шаха для результирующих изображений из $g=2$ и $g=7$ сегментов различных цветов.

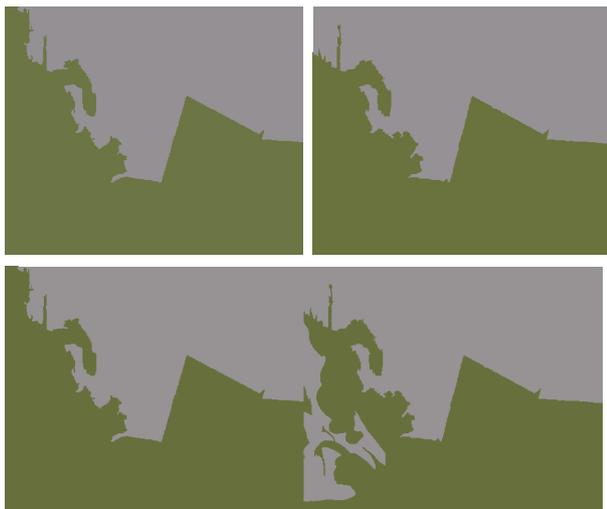


Рис. 3. Сегментация в модели Мамфорда-Шаха при $g=2$.

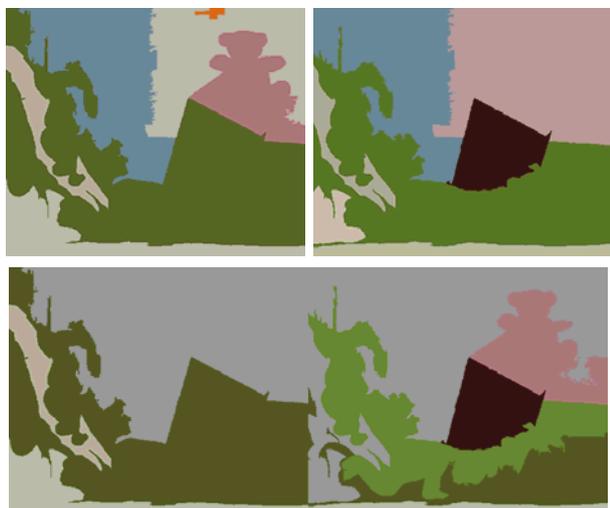


Рис. 4. Сегментация в модели Мамфорда-Шаха при $g=7$.

Приближения изображений одинаковых объектов при $g=2$ сегментах недостаточно детальны (рис. 3), а при $g=7$ сегментах выделение объектов плохо предсказуемо (рис.4).

На рис. 5-6 приведены результаты кластеризации по Уорду для приближений с кластерами пикселей $g_0=2$ и $g_0=7$ различных цветов. Под приближениями указаны достигнутые значения среднеквадратичного отклонения σ , связанного с E соотношением $E \approx 3N\sigma^2$.

По сравнению с моделью сегментации Мамфорда-Шаха, метод Уорда кластеризации пикселей обеспечивает единообразное детектирование гораздо большего числа объектов. Устойчивое выделение объектов достигается при различных E на одном и том же изображении (приближения внизу на рис. 5, 6). Это проверено в экспериментах с многочисленными совмещенными изображениями стереопар.

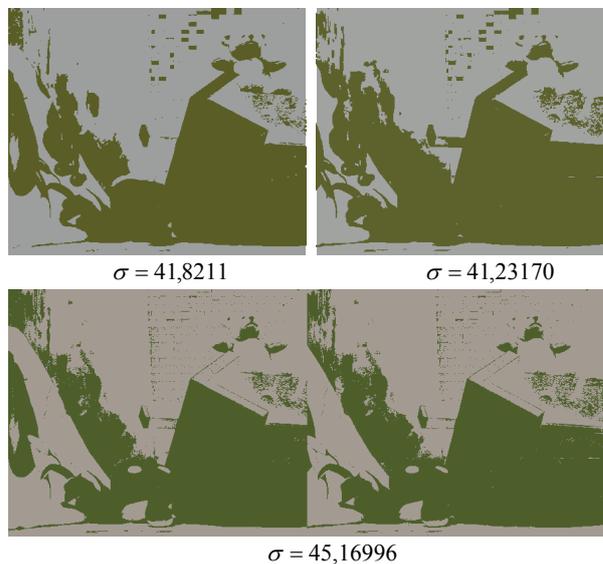


Рис. 5. Результаты кластеризации по Уорду при $g_0=2$.

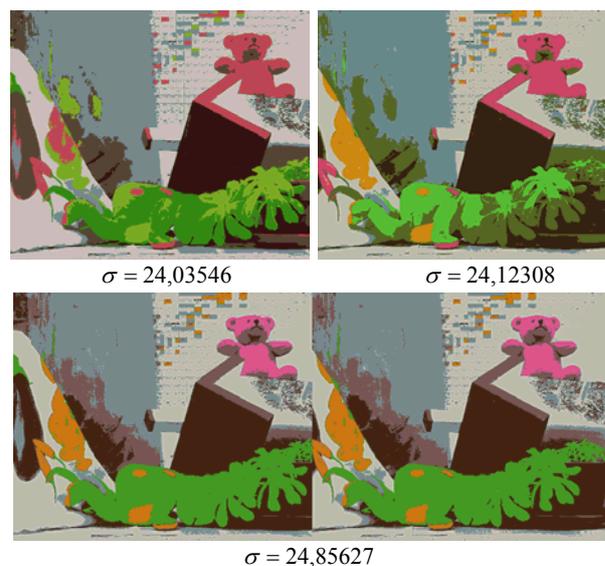


Рис. 6. Результаты кластеризации по Уорду при $g_0=7$.

Для улучшения выделения объектов на отдельных изображениях за счет получения близких по E приближений (вверху на рис. 5, 6), необходимо развитие программной реализации минимизации E по пунктам 1)-3) предыдущего раздела.

5. Заключение

Судя по нашему опыту, метод Уорда обеспечивает устойчивое выделение на изображении объектов в виде близких по усредненным цветам кластеров пикселей, которые «видит» компьютер. При этом внедрение метода Уорда в практику обработки изображений приводит к двухмасштабной модели дихотомической иерархии

объектов [8-11], которая поддерживает скоростные вычисления с кластерами пикселей и сегментами изображения, а также обеспечивает эффект запоминания миллионов приближений изображения в RAM.

В качестве прототипа модели дихотомической иерархии объектов рассматривается двухмасштабная модель изображения [6,7], основными ограничениями которой являются:

- 1) на выходе вычисляется единственное разбиение исходного изображения;
- 2) объекты описываются посредством связанных сегментов, не объединяемых в кластеры;
- 3) анализ множеств пикселей выполняется в паре квадратных окошек различного масштаба.

Указанные ограничения преодолеваются в нашей модели [8-11], в которой:

- 1) на выходе вычисляются варианты иерархических последовательностей квазиоптимальных приближений с числом кластеров от 1 до N ;
- 2) объекты описываются посредством кластеров пикселей и сегментов изображения;
- 3) анализ изображения выполняется в адаптивных алгоритмах слияния/разделения множеств пикселей.

Смысл модели [8-11] прозрачен и сводится к тому, что исходное изображение разбивается на самостоятельные иерархически структурированные изображения объектов классическим методом Уорда [13,16], но с учетом особенностей изображения. При этом нетривиальный формализм модели состоит в определении разбиения изображения на g_0 объектов и организации расчетов в терминах специальной структуры данных динамической сети, которая «набрасывается» на пиксели для ускорения вычислений и экономии памяти [11]. Для эффективного внедрения двухмасштабной модели дихотомической иерархии объектов в практику обработки изображений, вычисления в терминах динамической сети предполагается оформить в виде стандартных утилит обработки, чтобы ими пользовались программисты-инженеры, не вдаваясь в детали математического описания.

Для современной теории и практики обработки изображений проблема выделения объектов особенно актуальна. Возможно, в обозримом будущем ее удастся решить на основе отечественных моделей [3,4], [6,7] и [8-11].

6. Литература

- [1] Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. — 607 с.
- [2] Бугаев А.С., Хельвас А.В. Поисковые исследования и разработка методов и средств анализа и автоматического распознавания потоковой информации в глобальных информационных системах. Шифр «Лацкан» // Отчет по НИР. М.: Изд-во МФТИ, 2001. Т. 1, — 140 с.
- [3] Визильтер Ю.В., Желтов С.Ю. Проблемы технического зрения в современных авиационных системах // Механика, управление и информатика, 2011. №6. — С. 11–44.
- [4] Луцив В.Р. Объектно-независимый подход к структурному анализу изображений / Дис. докт. технич. наук. СПб.: ГУАП, 2011. — 318 с.
- [5] Мандель И.Д. Кластерный анализ // М.: Финансы и статистика. 1988. — 176 с.

[6] Чочиа П.А. Некоторые алгоритмы обнаружения объектов на основе двухмасштабной модели изображения // Информационные процессы, 2014. Т.14, №2, — С. 117–136.

[7] Чочиа П.А. Теория и методы обработки видеoinформации на основе двухмасштабной модели изображения / Дис. докт. технич. наук. М.: ИППИ РАН, 2016. — 302 с.

[8] Харинов М.В. Разработка динамических структур данных системы автоматизированного распознавания изображений / руков. В.В. Александров / Автореф. Дис. канд. технич. наук. — С.П. 1993. — 20с.

[9] Харинов М.В., Ханьков И. Г. Применение метода Уорда для кластеризации пикселей цифрового изображения // Вестник Бурятского государственного университета.

[10] Харинов М.В. Обратимое слияние структурированных кластеров пикселей // Компьютерная графика и зрение (Графикон'2016) / Труды 26-й международной конференции по компьютерной графике и зрению: Москва, МГУ им. М. В. Ломоносова, 19 – 23 сентября 2016 г., Нижний Новгород, — С. 298–302.

[11] Харинов М.В. Основы модели квазиоптимальных приближений изображения // Вестник Бурятского государственного университета, Математика и информатика, Улан-Удэ: Изд-во Бурятского госуниверситета, 2016. №1 — С. 60-72.

[12] Bar L., Chan T.F., Chung G., Jung M., Vese L.A., Kiryati N., Sochen N. Mumford and Shah Model and Its Applications to Image Segmentation and Image Restoration. Handbook of Mathematical Methods in Imaging, 2015. — pp. 1539-1597.

[13] Jain A. K., Murty M.N., Flynn P.J. Data clustering: a review // ACM computing surveys (CSUR). – 1999. Vol. 31. №3. — pp. 264-323.

[14] Jenatton R., Gramfort A., Michel V., Obozinski G., Eger E., Bach F., Thirion B. Multiscale mining of fMRI data with hierarchical structured sparsity // SIAM Journal on Imaging Sciences. 2012. Vol. 5. №. 3. — pp. 35-856.

[15] Mumford D., Shah J. Boundary detection by minimizing functionals, I // Proc. IEEE Comput. Vision Patt. Recogn. Conf., San Francisco, 1985. — pp. 22–26.

[16] Ward J.H., Jr. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 1963. Vol. 58, Issue 301, — pp. 236-244.

Об авторах

Ханьков Игорь Георгиевич, аспирант, м.н.с. Лаборатории прикладной информатики СПИИРАН Его e-mail igk@iias.spb.su.

Харинов Михаил Вячеславович, доцент, с.н.с. Лаборатории прикладной информатики СПИИРАН. Его e-mail khar@iias.spb.su.



Обработка и анализ
биомедицинских изображений

Biomedical Image Processing
and Analysis

GraphiCon 2017

24–28 сентября 2017
Пермь, Россия

Classification methods on different brain imaging modalities for Alzheimer disease studies

A. Khvostikov^{1,2}, J. Benois-Pineau¹, A. Krylov², G. Catheline³

xubiker@gmail.com|jenny.benois-pineau@u-bordeaux.fr

kryl@cs.msu.ru|gwenaelle.catheline@u-bordeaux2.fr

¹LaBRI UMR 5800, University of Bordeaux, Bordeaux, France;

²CMC Depart., Lomonosov Moscow State University, Moscow, Russia;

³INICIA UMR 5287, University Victor Segalen Bordeaux 2, Bordeaux, France

Computer-aided early diagnosis of Alzheimer's Disease (AD) and its prodromal form, Mild Cognitive Impairment (MCI), has been the subject of extensive research in recent years. Some recent studies have shown promising results in the AD and MCI determination using structural and functional Magnetic Resonance Imaging (sMRI, fMRI), Positron Emission Tomography (PET) and Diffusion Tensor Imaging (DTI) modalities.

This paper reviews the major trends in automatic classification methods such as feature extraction based methods as well as deep learning approaches in medical image analysis applied to the field of Alzheimer's Disease diagnostics. Different fusion methodologies to combine heterogeneous image modalities to improve classification scores are also considered.

Keywords: *Medical imaging, Alzheimer's Disease, Mild cognitive impairment, Machine Learning, Deep learning, Convolutional neural networks, Fusion, Review*

1. Introduction

Alzheimer's Disease (AD) is the most common type of dementia. Dementia refers to diseases that are characterized by a loss of memory or other cognitive impairments, and is caused by nerve cells degeneration in the brain.

Structural Magnetic Resonance Imaging (sMRI) and amyloid-Positron Emission Tomography (amyloid-PET) have been added in the newly proposed criteria for the predementia phase of the disease [1]. Other modalities as diffusion MRI (DTI) and functional MRI measured at rest (fMRI) are not yet used in these definitions. However, AD subjects also present modification on these two modalities as well [2, 3].

Since AD-related neurodegeneration is associated with gray matter atrophy, most previous works were focused on volumetric approaches that are based on comparison of anatomical brain structures assuming one-to-one correspondence between subjects. The wide-spread voxel-based morphometry (VBM) [4] is an automatic volumetric method for studying the differences in local concentrations of white and gray matter and comparison of brain structures of the subjects to test with reference normal control (NC) brains. Tensor-based morphometry (TBM) [5] was proposed to identify local structural changes from the gradients of deformations fields when matching tested brain and the reference healthy NC. Object-based morphometry (OBM) [6] was introduced for shape analysis of anatomical structures.

In general, the automatic classification on brain images of different modalities can be applied to the whole brain [7–10], or performed using the domain knowledge on specific regions of interest (ROIs). Concerning AD, hippocampal structural changes are strongly correlated to the severity of disease [11]. The

changes in such regions are considered as AD biomarkers.

Advances in computer vision and content-based image retrieval research make penetrate the so-called feature-based methods into classification approaches for (AD) detection [12–14]. The reason for this is in inter-subject variability, which is difficult to handle in VBM. On the contrary, the quantity of local features which can be extracted from the brain scans together with captured particularities of the image signal allowed an efficient classification with lower computational workload [14]. The obtained feature vectors are classified using machine learning algorithms.

Lately with the development of neural networks the feature-based approach became less popular and is gradually replaced with convolutional neural networks of different architectures.

The goal of the present paper is to give a substantial overview of these recent trends in classification of different brain imaging modalities in the problem of computer-aided diagnostics of Alzheimer disease and its prodromal stage, i.e. mild cognition impairment (MCI).

The paper is organized as follows. In Section 2 we will overview the main feature-based approaches and in Section 3 we will compare different approaches based on using neural networks. Particular attention will be paid in each case to fusion of modalities. All reviewed approaches are compared in Table 1. Section 4 concludes our review.

2. Feature-based classification

Feature-based classification can be performed on images of different modalities. Here we compare and discuss the usage of sMRI, DTI and sMRI fusion with other modalities.

sMRI In previous joint work [13], Ahmed et al. computed local features on sMRI scans in hippocampus and posterior cingulate cortex (PCC) structures of the brain. The originality of the work consisted in the usage of Gauss-Laguerre Circular Harmonic Functions (GL-CHFs) instead of traditional SIFT and SURF descriptors [15, 16]. CHFs perform image decomposition on the orthonormal functional basis, which allows capturing local directions of the image signal and intermediate frequencies. It is similar to Fourier decomposition, but is more appropriate in case of smooth contrasts of MRI modality. For each projection of each ROI a signature vector was calculated using a bag-of-visual-words model (BoVWM) with a low-dimensional dictionary with 300 clusters. This led to the total signature length of 1800 per image. Principal component analysis was then applied to reduce the signature length to 278. The signatures then were classified using SVM with RBF kernel and 10-fold cross-validation and reached the accuracy level of 0.838, 0.695, 0.621 for AD/NC, NC/MCI and AD/MCI binary classification problems accordingly.

DTI This modality is probably the most recent to be used for AD classification tasks. Both Mean Diffusivity (MD) and Fractional Anisotropy (FA) maps are being explored for this purpose. In [17] the authors acquired DTI images of 15 AD patients, 15 MCI patients, and 15 healthy volunteers (NC). After the preprocessing steps the FA map, which is an indicator of brain connectivity, was calculated. The authors considered 41 Brodmann areas, calculated the connectivity matrices for this areas and generated a connectivity graph with corresponding 41 nodes. Two nodes corresponding to Brodmann areas are marked with an edge if there is at least one fiber connecting them. Then the graph is described with the vector of features, calculated for each node and characterizing the connectivity of the node neighborhood. Totally each patient is characterized by 451 feature. The vectors were reduced to the size of 430 and 110 using ANOVA-based feature selection approach. All vectors were classified with the ensemble of classifiers (Logistic regression, Random Forest, Gaussian native Bayes, 1-nearest neighbor, SVM) using 5-fold cross-validation. The authors have achieved the 0.8, 0.833, 0.7 accuracy levels for AD/NC, AD/MCI and MCI/NC accordingly on their custom database.

Another methodology is described in [18]. The authors use the fractional anisotropy (FA) and mode of anisotropy (MO) values of DTI scans of 50 patients from the LONI Image Data Archive (<https://ida.loni.usc.edu>). After non-linear registration to the standard FA map, the authors calculate the skeleton of the mean FA image as well as MO and perform the second step of registration. After that a Relief feature algorithm is performed on all voxels of the image, relevant ones are used for 10-cross valida-

tion training the SVM classifier with RBF kernel. The declared accuracy is 0.986 and 0.977 for classification AD/MCI, AD/NC accordingly.

Data fusion In [19] authors use a fusion of sMRI and PET images together with canonical correlation analysis (CCA). After preprocessing and aligning images of 2 modalities given the covariance data of sMRI and PET image they find the projection matrices by maximizing the correlation between projected features. Here

$$X_1 \in R^{d \times n}, X_2 \in R^{d \times n}$$

are the d -dimensional sMRI and PET features of n samples,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

is a covariance matrix,

$$(B_1, B_2) = \arg \max_{(B_1, B_2)} \frac{B_1^T \Sigma_{12} B_2}{\sqrt{B_1^T \Sigma_{11} B_1} \sqrt{B_2^T \Sigma_{22} B_2}}$$

are the projection matrices and

$$Z_1 = B_1^T X_1, Z_2 = B_2^T X_2$$

are the resulting projections. The authors construct the united data representation for each patient:

$$F = [X_1; X_2; Z_1; Z_2] \in R^{4d \times n}$$

and calculate SIFT descriptors. This descriptors are used to form the BoVW model, the classification is performed using SVM. The achieved accuracy is 0.969, 0.866 and for classifying AD/NC and MCI/NC accordingly.

Ahmed et al. in [14] demonstrated the efficiency of using the amount of cerebrospinal fluid (CSF) in the hippocampal area calculated by an adaptive Otsu's thresholding method as an additional feature for AD diagnostics. In [12] they further improved the result of [13] by combining visual features derived from sMRI and DTI MD maps with a multiple kernel learning scheme (MKL). Similar to [13] they selected hippocampus ROIs on the axial, saggital and coronal projections and described them using Gauss-Laguerre Harmonic Functions (GL-CHFs). These features are clustered into 250 and 150 clusters for sMRI and MD DTI modalities and encoded using the BoVW model. Thus they got three sets of features: BoVW histogram for sMRI, BoVW histogram for MD DTI and CSF features. The obtained vectors are classified using MKL approach based on SVM. The achieved accuracy is 0.902, 0.794, 0.766 for AD/NC, MCI/NC and AD/MCI classification.

3. Classification with neural networks

Deep neural networks (DNN) and specifically convolutional NN have become popular now due to their good generalization capacity and available GPU Hardware needed for parameter optimization. Their main drawback for AD classification is the small amount of available training data and also a low resolution of input images when the ROIs are considered. This problem can be eliminated in several ways: i) by using

shallow networks with relatively small number of neurons, ii) applying transfer learning from an existing trained network or iii) pretraining some of the layers of the network.

Forming shallow networks kills the idea of deep learning to recognize structures at different scales and reduces the generalization ability of the network, so this methodology has not often been used since recently, despite it has shown decent results [20].

One way to enlarge the dataset is to use domain-dependent data augmentation. In case of medical images this often comes down to mirror flipping, small-magnitude translations and weak gaussian blurring [20].

Autoencoders The idea of pretraining some of the layers in the network is easily implemented with autoencoders (AE) or in image processing tasks more often with convolutional autoencoders (CAE). Autoencoder consists of an input layer, hidden layer and an output layer, where the input and output layers have the same number of units (Fig.1). Given the input vector $x \in \mathbb{R}^n$ autoencoder maps it to the hidden representation h :

$$h = f(Wx + b),$$

where $W \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, n is the number of input units, p is the number of hidden units, f is an encoder function e.g. sigmoid. After that the hidden representation h is mapped back to $\hat{x} \in \mathbb{R}^n$:

$$\hat{x} = g(\widehat{W}h + \widehat{b}),$$

where $\widehat{W} \in \mathbb{R}^{n \times p}$, $\widehat{b} \in \mathbb{R}^n$, g is the identity function. The weights and biases are found by gradient methods to minimize the cost function:

$$J(W, b) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \|\hat{x}^{(i)} - x^{(i)}\|^2,$$

where N is the number of inputs.

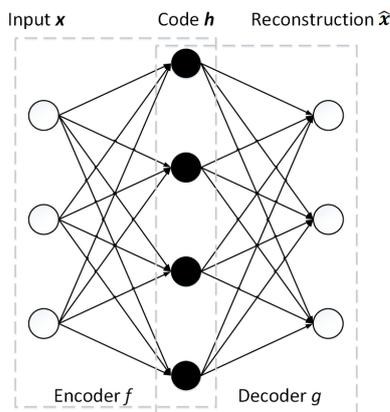


Figure 1. Architecture of autoencoder

The overcompleted hidden layer is used to make the autoencoder extracting features.

Introducing spatial constraints with convolutions easily aligns the model of autoencoder to the convolutional autoencoder (CAE) and 3D convolutional autoencoder (3D-CAE).

In [8] authors added a sparsity constraint to prevent hidden layers of autoencoder from learning the identity function. They use 3D convolutions on the both sMRI and PET modalities and train the autoencoder on random $5 \times 5 \times 5$ image patches. Maxpooling, fully-connected and softmax layers were applied after autoencoding. Mixing data of sMRI and PET modalities is performed at FC layer. The use of autoencoders allowed the authors to increase the classification accuracy by 4-6% and leads to the level of 0.91 for AD/NC classification.

Nearly the same approach with sparse 3D autoencoder was used in [9] to classify sMRI images into 3 categories (AD/MCI/NC). The proposed network architecture is shown in Fig.2. Larger obtained dataset and more accurate network parameters configuration allowed the authors to reach the accuracy of 0.954, 0.868 and 0.921 in AD/NC, AD/MCI and NC/MCI determination accordingly.

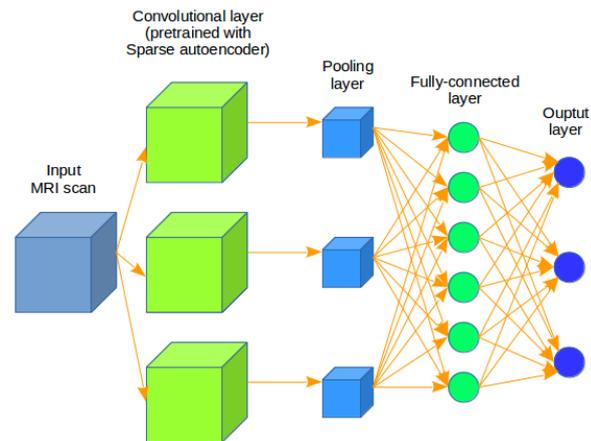


Figure 2. Typical CNN architecture with CAE pretraining.

The authors of [7] extended the idea of applying autoencoders. They proposed using three stacked 3D convolutional autoencoders instead of only one. Two fully-connected layers before the softmax were used for a progressive dimension reduction. The usage of stacked 3D CAE allowed the authors to achieve one of the best accuracy level: 0.993, 1, 0.942 for AD/NC, AD/MCI and MCI/NC classification on sMRI images only.

Transfer Learning Transfer learning is considered as the transfer of knowledge from one learned task to a new task in machine learning. In the context of neural networks, it is transferring learned features of a pre-trained network to a new problem. Glozman and Liba in [21] used the widely known AlexNet [22], pretrained

on the ImageNet benchmark and fine-tuned the last 3 fully-connected layers (Fig.3). The main problem of transfer learning is the necessity to transform the available data so that it corresponds to the network input. In [21] the authors created several 3-channel 2D images from the 3D input of sMRI and PET images by choosing central and nearby slices from axial, coronal and sagittal projections. They then interpolated the slices to the size 227×227 compatible with AlexNet. Naturally one network was used for each projection. To augment the source data only mirror flipping was applied. This transfer learning based approach allowed the authors to reach 0.665 and 0.488 accuracy on 2-way (AD/NC) and 3-way (AD/MCI/NC) classifications accordingly on a subset of ADNI database.

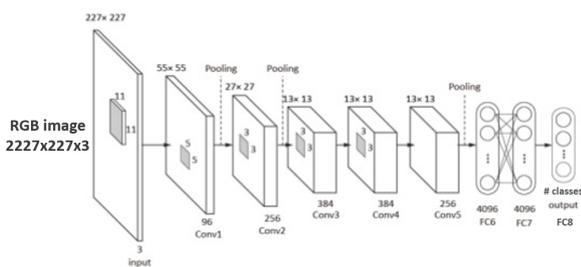


Figure 3. AlexNet architecture. Includes 5 convolutional layers and 3 fully-connected layers.

2D convolutional neural networks In [23], [24], [25] the authors compared the classification of structural and functional MRI images using the LeNet-5 architecture by transforming the source 3D and 4D (in case of MRI) data to a batch of 2D images. LeNet-5 consists of two convolutional and two fully-connected layers. The reached level of accuracy for 2-class classification (AD/NC) was 0.988 for sMRI and 0.999 for fMRI images.

Billones et al. proposed in [10] to use a modified 16-layered VGG network [26] to classify sMRI images. The key feature of this paper was to use 2D convolutional network to classify each slice of source data separately. The authors selected 20 central slices for each image and the final score was calculated as the output of the last softmax layer of the network. The accuracy of each slice among all images was also studied, 17 slices were selected as representative, 3 slices (the first and two last slices in the image sequence) demonstrated lower level of accuracy. All in all authors reached a very good accuracy level: 0.983, 0.939, 0.917 for AD/NC, AD/MCI and MCI/NC classification.

In [20] Aderghal et al. used 3 central slices in each projection of a hippocampal ROI. The network architecture represented three 2D convolutional networks (one network per projection) that were joined in the last fully-connected layer. The reached accuracy

for AD/NC, AD/MCI and MCI/NC classification is 0.914, 0.695 and 0.656 accordingly.

Other networks A new approach was proposed in [27]. Shi et al. used a deep polynomial network to analyze sMRI and PET images. It differs from classical CNNs by non-linearity of operations. The building block of the architecture is shown in Fig.4. Here, n^i represents a layer of nodes, (+) means a layer of nodes that calculate the weighted sum $n(z) = \sum_i w_i z_i$, all other nodes compute $n(z_1, z_2) = \sum_i w_i (z_1)_i (z_2)_i$. These blocks were combined into a deep network, the input layers were fed with the average intensity of the 93 ROIs selected on sMRI and PET brain images.

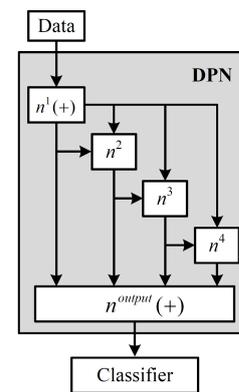


Figure 4. An example of a DPN module.

This architecture allowed the authors to reach very good level of accuracy: 0.971, 0.872 for AD/NC, MCI/NC classification. The used algorithm also demonstrated a good level of accuracy (0.789) for MCI-C/MCI-NC determination, where MCI-C stands for MCI patients that lately converted to AD and MCI-NC stands for MCI patient that were not converted.

In [28] the authors compared the residual and plain 3D convolutional neural networks for sMRI image classification. Here the authors examined the four binary classification tasks AD/LMCI/EMCI/NC, where LMCI and EMCI stands for the late and early MCI stages accordingly. Both networks demonstrated nearly the same performance level, the best figures being obtained for AD/NC classification 0.79-0.8.

In [29] Suk et al. try to combine two different methods: sparse regression and convolutional neural networks. The authors got different sparse representations of the 93 ROIs of the sMRI data by varying the sparse control parameter, which allowed them to produce different sets of selected features. Each representation is a vector, so the result of generating multiple representations can be treated as a matrix. This matrix is then fed to the convolutional neural network with 2 convolutional layers and 2 fully-connected layers. This approach led to the classification accuracy

Algorithm	Methodology	Modalities	Content	Data (size)	Accuracy		
					AD/NC	AD/MCI	MCI/NC
Magnin et al. [6]	Volumetric	sMRI	Full brain	custom (38)	0.945	-	-
Ahmed et al. [13]	Feature-based	sMRI	2 ROIs	ADNI (509)	0.838	0.695	0.621
Ebadi et al. [17]	Feature-based	DTI	Full brain	custom (34)	0.8	0.833	0.7
Lee et al. [18]	Feature-based	DTI	Full brain	LONI (141)	0.977	0.977	-
Lei et al. [19]	Feature-based	sMRI + PET	Full brain	ADNI (398)	0.969	-	0.866
Ahmed et al. [14]	Feature-based	sMRI + DTI	1 ROI	ADNI (203)	0.902	0.766	0.794
Vu et al. [8]	NN-based	sMRI + PET	Full brain	ADNI (203)	0.91	-	-
Payan and Montana [9]	NN-based	sMRI	Full brain	ADNI (2265)	0.993	1	0.942
Glozman and Liba [21]	NN-based	sMRI + PET	Full brain	ADNI (1370)	0.665	-	-
Sarraf et al. [23]	NN-based	sMRI, fMRI	Full brain	ADNI (302)	0.988, 0.999	-	-
Billones et al. [10]	NN-based	sMRI	Full brain	ADNI (900)	0.983	0.939	0.917
Aderghal et al. [20]	NN-based	sMRI	1 ROI	ADNI (815)	0.914	0.695	0.656
Shi et al. [27]	NN-based	sMRI + PET	Full brain	ADNI (202)	0.971	-	0.872
Korolev et al. [28]	NN-based	sMRI	Full brain	ADNI (231)	0.79-0.8	-	-
Suk et al. [29]	NN-based	sMRI	93 ROIs	ADNI (805)	0.903	-	0.742

Table 1. Comparison of different classification methods.

level of 0.903 and 0.742 for AD/NC and MCI/NC classification.

4. Discussion and Conclusion

As it can be seen from the Table 1 relatively new feature-based and neural network-based methods demonstrate very good level of performance compared to the classical volumetric methods that are performed manually by medical experts.

It should be mentioned, that the direct comparison of the reviewed algorithms for Alzheimer's disease diagnostics is impossible. The proposed results were obtained using images from several databases and in different quantities (see Table 1). Moreover different classification problems were challenged: although most papers focus on the 3-class AD/MCI/NC binary classification, some of them consider only 2-class AD/NC classification [8, 23–25] and even 4-class AD/eMCI/lMCI/NC classification [28]. Also [10, 29] deserve special attention as the authors try to solve a problem in demand of prediction of Alzheimer converters.

This review allows the community of researchers working on AD classification problems to position their approach and design more efficient classification schemes.

5. Acknowledgements

This research was supported by Ostrogradsky scholarship grant established by French Embassy in Russia. We thank Pierrick Coupé from LABRI UMR 5800 University of Bordeaux/CNRS/Bordeaux-IPN who provided insight and expertise that greatly assisted the research.

6. References

- [1] Jeffrey L Cummings, Bruno Dubois, et al. International work group criteria for the diagnosis of alzheimer disease. *Medical Clinics of North America*, 97(3):363–368, 2013.
- [2] Massimo Filippi and F Agosta. Imaging biomarkers in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, 31(4):770–788, 2010.
- [3] Julio Acosta-Cabronero and Peter J Nestor. Diffusion tensor imaging in alzheimer's disease: insights into the limbic-diencephalic network and methodological considerations. *Frontiers in aging neuroscience*, 6, 2014.
- [4] J. Ashburner and K. Friston. Voxel-based morphometry—the methods. *Neuroimage*, 11(6):805–821, 2000.
- [5] C. Studholme et al. Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change. *IEEE transactions on Medical Imaging*, 25(5):626–639, 2006.
- [6] B. Magnin et al. Support vector machine-based classification of alzheimer's disease from whole-brain anatomical mri. *Neuroradiology*, 51(2):73–83, 2009.
- [7] E. Hosseini-Asl et al. Alzheimer's disease diagnostics by a deeply supervised adaptable 3d convolutional network. *arXiv preprint arXiv:1607.00556*, 2016.
- [8] Tien Duong Vu et al. Multimodal learning using convolution neural network and sparse autoencoder. In *Big Data and Smart Computing (Big-Comp), 2017 IEEE International Conference on*, pages 309–312. IEEE, 2017.
- [9] A. Payan and G. Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *arXiv preprint arXiv:1502.02506*, 2015.
- [10] C. Billones et al. Demnet: A convolutional neural network for the detection of alzheimer's disease and mild cognitive impairment. In *Region 10 Conference (TENCON), 2016 IEEE*, pages 3724–3727. IEEE, 2016.
- [11] V. Planche et al. Hippocampal microstructural damage correlates with memory impairment in clinically isolated syndrome suggestive of mul-

- tiple sclerosis. *Multiple Sclerosis Journal*, page 1352458516675750, 2016.
- [12] Olfa Ben Ahmed et al. Recognition of alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning. *Neurocomputing*, 220:98–110, 2017.
- [13] Olfa Ben Ahmed et al. Alzheimer's disease diagnosis on structural mr images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. *Computerized Medical Imaging and Graphics*, 44:13–25, 2015.
- [14] Olfa Ben Ahmed et al. Classification of alzheimer's disease subjects from mri using hippocampal visual features. *Multimedia Tools and Applications*, 74(4):1249–1266, 2015.
- [15] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.
- [16] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer vision—ECCV 2006*, pages 404–417, 2006.
- [17] A. Ebadi et al. Ensemble classification of alzheimer's disease and mild cognitive impairment based on complex graph measures from diffusion tensor images. *Frontiers in Neuroscience*, 11, 2017.
- [18] W. Lee, B. Park, and K. Han. Svm-based classification of diffusion tensor imaging data for diagnosing alzheimer's disease and mild cognitive impairment. In *International Conference on Intelligent Computing*, pages 489–499. Springer, 2015.
- [19] B. Lei et al. Discriminative learning for alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion. *Frontiers in aging neuroscience*, 8, 2016.
- [20] K. Aderghal et al. Fuseme: Classification of smri images by fusion of deep cnns in 2d+eps projections. In *International Workshop on Content-Based Multimedia Indexing*, pages 1–7, 2017.
- [21] T. Glozman and O. Liba. Hidden cues: Deep learning for alzheimer's disease classification cs331b project final report.
- [22] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [23] S. Sarraf and G. Tofghi. Classification of alzheimer's disease structural mri data by deep learning convolutional neural networks. *arXiv preprint arXiv:1607.06583*, 2016.
- [24] S. Sarraf and G. Tofghi. Deep learning-based pipeline to recognize alzheimer's disease using fmri data. *bioRxiv*, page 066910, 2016.
- [25] S. Sarraf et al. Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri. *bioRxiv*, page 070441, 2016.
- [26] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [27] Jun Shi et al. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE journal of biomedical and health informatics*, 2017.
- [28] S. Korolev et al. Residual and plain convolutional neural networks for 3d brain mri classification. *arXiv preprint: 1701.06643*, 2017.
- [29] Heung-Il Suk et al. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical image analysis*, 37:101–113, 2017.

About the authors

Alexander Khvostikov is a PhD student, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia. His contact e-mail is xubiker@gmail.com.

Jenny Benois-Pineau is a full professor of Computer Science at the University Bordeaux, Bordeaux, France. She is a head of video and indexing research group at LABRI UMR 5800. Her contact email is jenny.benois-pineau@u-bordeaux.fr.

Andrey Krylov is a professor, head of the Laboratory of Mathematical Methods of Image Processing, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia. His contact e-mail is kryl@cs.msu.ru.

Gwenaelle Catheline is an assistant professor at the University Victor Segalen Bordeaux 2, Bordeaux, France. She is a researcher at INCIA UMR 5287. Her contact email is gwenaelle.catheline@u-bordeaux.fr.

Метод синтеза реалистичных данных крио-электронной микроскопии одной частицы и крио-электронной томографии

Н.А. Аношина¹, Т.Б. Сагиндыков¹, Д.В. Сорокин^{1,2}

anshinan@yandex.ru|tamerlan.sagindykov@gmail.com|dsorokin@cs.msu.ru

¹Лаборатория математических методов обработки изображений, факультет вычислительной математики и кибернетики, Московский Государственный Университет имени М.В.Ломоносова, Москва, Россия;

²Centre for Biomedical Image Analysis, Faculty of Informatics, Masaryk University, Brno, Czech Republic

Крио-электронная микроскопия широко используется в области структурной биологии для построения трехмерных моделей изучаемых частиц с разрешением в несколько ангстрем. Наиболее распространенными методами для получения моделей столь высокого разрешения являются крио-электронная микроскопия одной частицы и крио-электронная томография. В обеих техниках, ключевым шагом алгоритмов обработки данных является совмещение изображений, двумерных для крио-электронной микроскопии одной частицы, и трехмерных для крио-электронной томографии. Существует некоторое количество методов совмещения изображений для совмещения двумерных и трехмерных изображений крио-электронной микроскопии. Однако, оценить качество таких методов на реальных данных не представляется возможным, в силу отсутствия "ground truth". Также, оценка качества методов совмещения на реальных данных осложнена высоким уровнем шума. В данной работе предложены методы генерации наборов двумерных и трехмерных реалистичных данных крио-электронной микроскопии одной частицы и крио-электронной томографии. Предложенные алгоритмы моделируют искажения, возникающие при получении данных в реальном электронном микроскопе, а именно типичный для данного вида микроскопии шум, эффект потерянного клина, двумерное проецирование, функцию переноса контраста. В работе приведены примеры сгенерированных синтетических двумерных и трехмерных изображений.

Ключевые слова: синтез изображений, крио-электронная микроскопия, крио-электронная томография.

1. Введение

Крио-электронная микроскопия является одной из основных развивающихся технологий изучения строения клеток, вирусов и белковых комплексов в молекулярном разрешении [7]. Основной принцип работы крио-ЭМ заключается в визуализации радиационно-чувствительных образцов в просвечивающем электронном микроскопе при криогенных условиях. Существуют различные методы крио-ЭМ: крио-электронная томография [3], [11], крио-электронная микроскопия одной частицы [1] и электронная кристаллография, — все они успешно используются для анализа биологических структур в различных ситуациях.

Основная проблема получения атомарного разрешения в электронном микроскопе — повреждение образца в результате взаимодействия электронов с органическим веществом [2]. Для получения высокого разрешения в крио-электронной микроскопии используют образцы, замороженные до температуры жидкого азота или гелия, поскольку криогенные температуры уменьшают повреждение образца в результате облучения. Для последующего получения моделей структур используют подходы, основанные на обработке полученных изображений.

Одним из широко распространенных подходов является крио-электронная микроскопия одной частицы [9]. В данном методе снимают большое количество изображений высокого разрешения ($< 4\text{\AA}$), каждое из которых содержит сотни одинаковых, но ориентированных случайным образом частиц. Затем на каждом из полученных изображений локализуются все проекции частиц (Рис. 1). Для этого, происходит классифика-

ция проекций на несколько классов, содержащих проекции, соответствующие одинаковым ориентациям, в трехмерном пространстве. В рамках каждого класса изображения проекций могут иметь различную ориентацию в двумерной плоскости проекции, а также небольшое смещение. С помощью методов совмещения изображений проекции частиц в рамках каждого класса преобразуются к одной ориентации в двумерной плоскости проекции, а затем происходит итеративное усреднение. Таким образом получают проекции частиц, отношение сигнала к шуму у которых значительно выше, чем у исходных объектов.

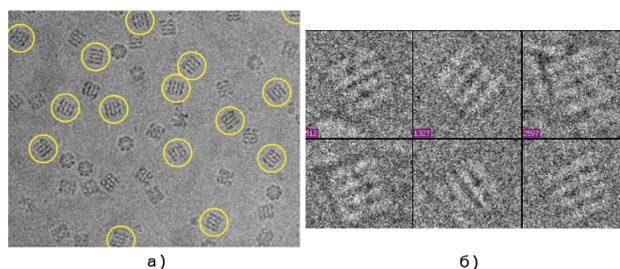


Рис. 1. а) Пример изображения крио-электронной микроскопии одной частицы, желтыми кругами отмечены локализованные проекции частиц одного класса. б) Увеличенные изображения частиц одного класса с различной ориентацией (изображения взяты из Gabriel Lander, Thesis Defense presentation).

В крио-электронной томографии для получения данных, образцы, как и в крио-электронной микроскопии, подвергают заморозке, предохраняя их от повреждений, возникающих в результате облучения. Пла-

стинку, с замороженным образцом, содержащую большое количество изучаемых частиц, поворачивают под разными углами относительно электронного луча с шагом в 1 – 2 градуса в диапазоне от -60° до $+60^\circ$ и просвечивают в электронном микроскопе, получая двумерную проекцию образца для каждого угла. Полученную серию изображений используют для трехмерной реконструкции, которая как правило, осуществляется с использованием взвешенного обратного проецирования [6]. В результате получается трехмерное томографическое изображение (Рис. 2), содержащее десятки одинаковых частиц, ориентированных различным образом. Чтобы получить трехмерное изображение частицы, где отношение сигнала к шуму значительно выше, в полученном трехмерном томографическом изображении локализируются все частицы. Далее, с помощью методов совмещения изображений субтомограммы, содержащие найденные частицы, преобразуются к одной ориентации в пространстве, а затем итеративно усредняются.

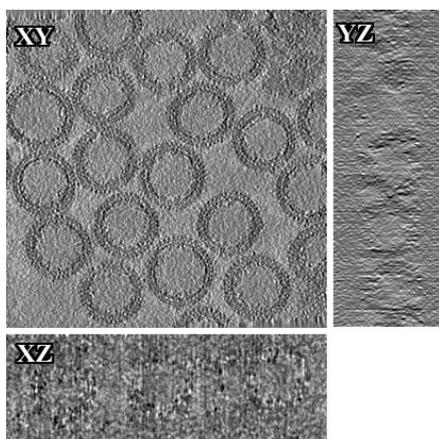


Рис. 2. Центральные срезы томографического изображения, содержащего одинаковые частицы (изображение из депозитария Protein Data Bank in Europe, набор данных EMD-4018).

Для того чтобы оценить качество работы метода совмещения изображений, нужны синтетические данные, где ориентация объектов задается контролируемо. В тоже время, такие данные должны быть максимально схожи с реальными изображениями крио-электронной микроскопии и томографии. В литературе описано использование синтетических данных для тестирования алгоритмов обработки данных крио-электронной микроскопии [4], [5], однако детали методов, используемых для генерации данных авторами не приводятся. Целью данной работы является создание наборов двумерных и трехмерных синтетических изображений крио-электронной микроскопии и томографии. Для этого предложен алгоритм генерации трехмерной модели объекта, схожего с реальными молекулярными комплексами, изучаемыми с помощью методов крио-электронной микроскопии и томографии. Далее, для синтезированных изображений происходит моделиро-

вание процессов, происходящих при получении данных в электронном микроскопе, а именно: моделирование шума и функции переноса контраста, а также построение двумерных проекций для крио-электронной микроскопии одной частицы и моделирование эффекта отсутствующего клина для крио-электронной томографии. Созданные с помощью предложенного метода данные используются для тестирования алгоритма совмещения трехмерных биомедицинских изображений.

2. Метод генерации данных

2.1. Метод построения 3D объекта

В данном подразделе рассмотрен пример алгоритма построения 3D объекта. В качестве реального образца для построения объекта может быть использована любая структура, исходя из особенностей решаемой задачи. В данной работе для построения синтетического объекта был взят белковый комплекс, имеющий цилиндрическую форму. Предложенный метод строит модель объекта, описывая его уравнением цилиндрической поверхности (Рис. 3а). За основание цилиндрического объекта возьмем окружность и добавим колебания по поверхности объекта:

$$C = \rho + C_1 \sin(C_2 \phi),$$

где C – радиус основания цилиндра, C_1, C_2 – некоторые константы. Для моделирования интенсивности объекта используется функция, зависящая от радиуса ρ и высоты h в цилиндрической системе координат, имеющая более одного максимума. В данной работе для задания интенсивности использовалась функция

$$f(\rho, h) = f_1(\rho)f_2(h),$$

где $f_1(\rho)$ – полиномиальная функция четвертого порядка с коэффициентами $p_0 = 3, p_1 = 4.4341, p_2 = -0.9049, p_3 = 0.0685, p_4 = -0.0017$, зависящая от радиуса ρ , а $f_2(h)$ – тригонометрическая функция, зависящая от высоты цилиндрического объекта h .

Стоит отметить, что высота цилиндра, функция $f_2(h)$ и прочие параметры алгоритма построения 3D объекта подбираются эмпирически для получения объекта необходимой формы и интенсивности. После построения и моделирования интенсивности объекта применим к нему функцию переноса контраста и добавим пуассоновский шум.

2.2. Виды искажений крио-ЭМ

Шум. Появление пуассоновского шума обусловлено статистической природой электромагнитных волн, таких как рентгеновское излучение, видимый свет и гамма-лучи. Источники рентгеновских лучей излучают определенное число фотонов в единицу времени. Такие источники имеют случайные колебания числа испускаемых фотонов. В результате полученное изображение имеет пространственную и временную случайность, соответствующую распределению

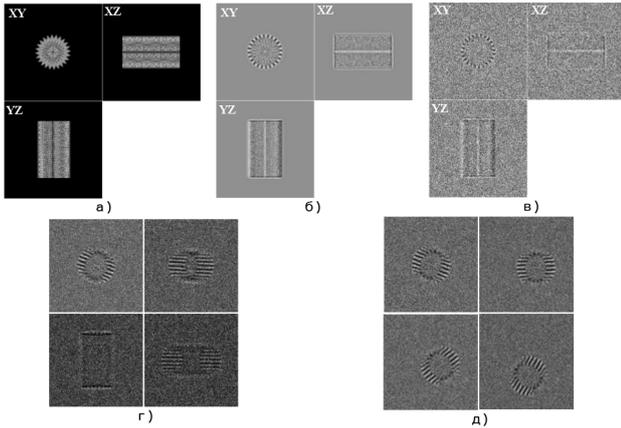


Рис. 3. а) Построенный 3D объект. б) Результат применения функции переноса контраста. в) Добавление пуассоновского шума. г) Примеры различных проекций трехмерного объекта, построенные с помощью предложенного метода с различным уровнем шума. д) Проекция одинаковых частиц, отличающихся смещением и поворотом.

Пуассона (Рис. 3в):

$$P(\lambda, k) = \frac{\lambda^k e^{-\lambda}}{k!}, \lambda > 0, k \in \mathbb{N}.$$

В электронном микроскопе в результате взаимодействия электронов с органическим веществом образец постепенно сгорает, поэтому полученные изображения обладают низким отношением сигнала к шуму (SNR) [12]. Также разрешение полученных изображений зависит от функции переноса контраста.

Функция переноса контраста. Для большинства образцов крио-электронной микроскопии не существует амплитудного значения контраста, поскольку электроны очень слабо взаимодействуют с легкими атомами образцов. В таких условиях можно сгенерировать фазовую контрастность изображений с помощью дефокусировки микроскопа: дефокусировка объектива генерирует низкочастотную контрастность за счет снижения контрастности на более высоких частотах. Это полезно, однако передача информации из микроскопа происходит необычным образом. Перемещение сигнала от образца до плоскости изображения в электронном микроскопе описывается *функцией переноса контраста* [12] (Рис. 3б): это осциллирующая функция в пространственной частотной области описывает, как аберрации в просвечивающем электронном микроскопе изменяют изображение образца:

$$CTF(f) = A(\sin(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s)) + B \cos(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s))),$$

где f – пространственная частота, Δz – дефокусировка, c_s – сферическая аберрация, λ – длина волны электрона, A – зависящая от дефокусировки огибающая функция, соответствующая затуханию сигнала для заданной

угловой апертуры, B – доля амплитудного значения контраста [13].

2.3. Метод построения двумерных проекций

Данный метод состоит в построении проекции 3D объекта на плоскость, которая задается нормалью \vec{n} . Для схожести с реальными данными, к координатам нормали в сферической системе координат (ρ, ϕ, θ) для построения проекции 3D объекта на плоскость добавлялась погрешность $\delta \in [0^\circ, 2^\circ]$ к углам ϕ и θ соответственно, которая выбиралась случайным образом (Рис. 3г).

2.4. Эффект потерянного клина (missing wedge)

В крио-электронной томографии эффект *потерянного клина* возникает вследствие того, что двумерные проекции образца, из которых восстанавливается томограмма, получают в ограниченном диапазоне углов (обычно $\pm 60^\circ$). Это происходит из-за физических ограничений при наклоне тонкой пластинки с образцом (при наклоне на большие углы, толщина образца становится слишком большой). Согласно теореме о центральном сечении [10], двумерный Фурье-образ проекции равен сечению трехмерного Фурье-образа исходного объекта, проходящему через начало координат в частотной области под тем же углом, под которым была получена проекция. Таким образом, эффект *потерянного клина* моделируется с помощью наложения клиновидной маски на исходное изображение в частотной области:

$$I_{mw} = \mathfrak{F}^{-1}(\hat{I} \cdot M),$$

где I_{mw} – результат моделирования, \hat{I} – преобразование Фурье исходного изображения, M – маска для удаления клина (Рис. 4).

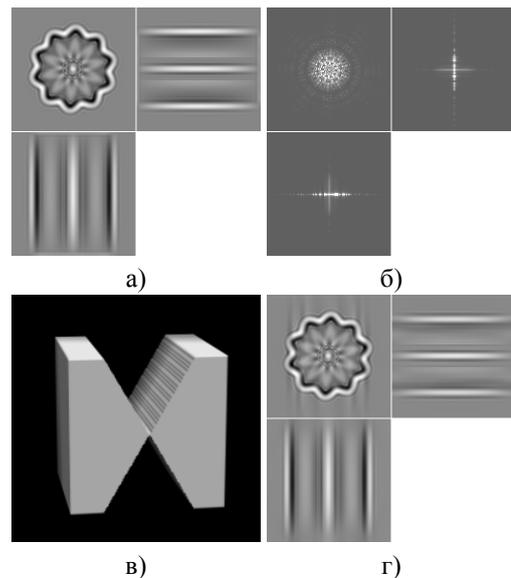


Рис. 4. а) Модельное 3D изображение. б) Фурье-образ изображения (а). в) Маска потерянного клина. г) Результат удаления клина в частотной области.

3. Алгоритм получения двумерных данных

С помощью предложенного подхода можно сгенерировать двумерные данные (Рис. 3), для этого нужно:

1. Сгенерировать трехмерное модельное изображение с заданием параметров C_1 и C_2 и функции моделирования интенсивности.
2. Для созданного изображения применить функцию переноса контраста и добавить пуассоновский шум.
3. Для каждой заданной нормали \vec{n} спроецировать объект на плоскость, заданную \vec{n} .
4. Полученные проекции повернуть на заданные углы и переместить на заданные векторы смещения.

Результатом является набор двумерных проекций трехмерного объекта с различными искажениями, характерными для крио-ЭМ одной частицы. Параметры данных искажений, а также углы проецирования, поворота и смещения задаются пользователем, что позволяет получать различные наборы данных для тестирования алгоритмов совмещения (Рис. 3д).

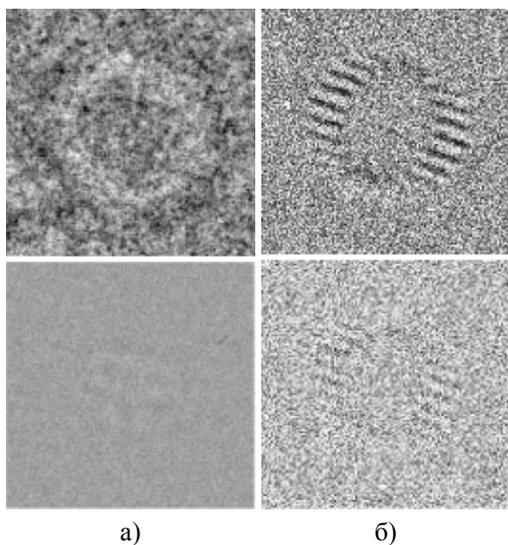


Рис. 5. Сравнение синтетических и реальных данных.
а) Реальные данные (набор данных EMPIAR-10026 [8]). б) Синтетические данные.

В сравнении с реальными данными, объект на синтетических данных виден более явно, однако это позволяет подобрать необходимые параметры и наладить работу метода. При изменении различных параметров можно добиться наибольшего сходства с различными реальными данными (Рис. 5).

4. Алгоритм получения трехмерных данных

Для создания набора трехмерных данных необходимо выполнить следующие шаги:

1. Сгенерировать трехмерное модельное изображение с заданием параметров C_1 и C_2 и функции моделирования интенсивности.
2. Получить несколько трехмерных изображений из полученного в предыдущем пункте, выполнив по-

вороты на заданные углы вокруг всех трех осей координат и сдвинув на заданные векторы смещения.

3. Смоделировать эффект *потерянного клина* на полученных изображениях.
4. Для полученных изображений применить функцию переноса контраста и добавить пуассоновский шум.

Результатом является набор трехмерных изображений объекта, ориентированного различным образом, с применением искажений характерных для крио-электронной томографии (Рис. 6). Параметры данных искажений, а также углы поворота и векторы смещения задаются пользователем, что позволяет получать различные наборы данных для тестирования алгоритмов совмещения трехмерных изображений крио-электронной томографии.

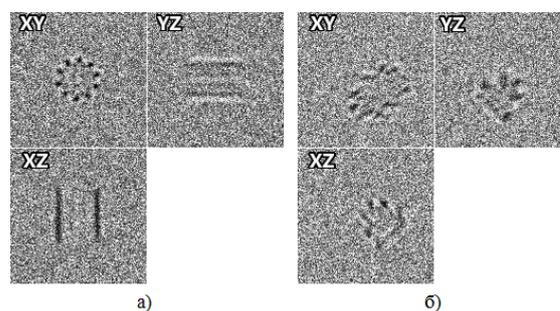


Рис. 6. Пример созданных трехмерных изображения одного объекта с разных ракурсов: изображение (б) отличается от (а) поворотом вокруг оси-Х на $\pi/6$, вокруг оси-У на $-\pi/4$ и сдвигом на вектор $[10, 15, 1]$.

Сравнение с реальными данными показывает, что в результате синтеза изображений с помощью предложенного алгоритма можно получить синтетические данные, близкие по структуре к данным полученным при крио-электронной томографии, а моделируемые эффект *потерянного клина*, функция переноса контраста и пуассоновский шум воссоздают искажения, получаемые при съемке в электронном микроскопе (Рис. 7).

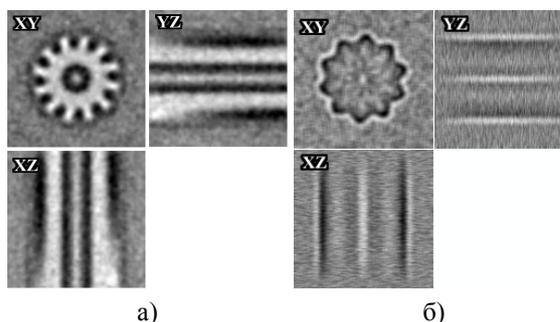


Рис. 7. а) Изображение реального объекта крио-электронной томографии. б) Результат моделирования трехмерного изображения.

5. Заключение

В данной работе предложен метод генерации реалистичных данных крио-электронной микроскопии.

Разработанные алгоритмы позволяют получить наборы данных с контролируемыми параметрами искажений, а также контролируемыми углами поворота и смещения для тестирования алгоритмов совмещения изображений. Получены наборы двумерных изображений для тестирования алгоритмов совмещения изображений, используемых в крио-электронной микроскопии одной частицы. Также получены наборы трехмерных изображений для тестирования алгоритмов совмещения изображений, используемых в крио-электронной томографии. Стоит отметить, что данная работа посвящена моделированию искажений, возникающих при получении данных в реальном электронном микроскопе. При этом, при практическом применении предложенного метода, геометрическая форма синтезируемого объекта может быть заменена на любую другую, исходя из особенностей решаемой задачи, для которой генерируются синтетические данные.

6. Благодарности

Работа выполнена при финансовой поддержке РФФИ, грант 17-11-01279.

7. Литература

- [1] Al-Amoudi, A., Chang, J.-J., Leforestier, A., McDowall, A., Salamin, L. M., Norlén, L. P., Richter, K., Blanc, N. S., Studer, D., and Dubochet, J. Cryo-electron microscopy of vitreous sections. *The EMBO journal* 23, 18 (2004), 3583–3588.
- [2] Glaeser, R. M. Limitations to significant information in biological electron microscopy as a result of radiation damage. *Journal of ultrastructure research* 36, 3 (1971), 466–482.
- [3] Hoenger, A., and Bouchet-Marquis, C. Cellular tomography. *Adv Protein Chem Struct Biol* 82 (2011), 67–90.
- [4] Lander, G. C., Stagg, S. M., Voss, N. R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.-W., et al. Appion: an integrated, database-driven pipeline to facilitate em image processing. *Journal of structural biology* 166, 1 (2009), 95–102.
- [5] Lyumkis, D., Brilot, A. F., Theobald, D. L., and Grigorieff, N. Likelihood-based classification of cryo-em images using frealign. *Journal of structural biology* 183, 3 (2013), 377–388.
- [6] McEwen, B., Radermacher, M., Rieder, C., and Frank, J. Tomographic three-dimensional reconstruction of cilia ultrastructure from thick sections. *Proceedings of the National Academy of Sciences* 83, 23 (1986), 9040–9044.
- [7] Milne, J. L., Borgnia, M. J., Bartesaghi, A., Tran, E. E., Earl, L. A., Schauder, D. M., Lengyel, J., Pierson, J., Patwardhan, A., and Subramaniam, S. Cryo-electron microscopy—a primer for the non-microscopist. *FEBS Journal* 280, 1 (2013), 28–45.
- [8] Russo, C. J., and Passmore, L. A. Ultrastable gold substrates for electron cryomicroscopy. *Science* 346, 6215 (2014), 1377–1380.
- [9] Schmid, M. F. Single-particle electron cryotomography (cryoet). *Adv Protein Chem Struct Biol* 82 (2011), 37–65.
- [10] Sigworth, F. J. Principles of cryo-em single-particle image processing. *Microscopy* 65, 1 (2016), 57–67.
- [11] Tocheva, E. I., Li, Z., and Jensen, G. J. Electron cryotomography. *Cold Spring Harbor perspectives in biology* 2, 6 (2010), a003442.
- [12] Vulovic, M. *Modeling of Image Formation in Cryo-Electron Microscopy*. TU Delft, Delft University of Technology, 2013.
- [13] Wade, R. The phase contrast characteristics in bright field electron microscopy. *Ultramicroscopy* 3 (1978), 329–334.

Компьютерный синтез эхокардиографических изображений сердечной мышцы человека: проблемы и их решения

Т.В. Голубятников¹, С.В. Поршневу¹, Н.И. Клиорин²
golubyatnikov.tv@gmail.com|s.v.porshnev@urfu.ru|nat@bgu.ac.il

¹Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, г. Екатеринбург, Россия;

²Университет им. Бен-Гуриона в Негеве, г. Бер-Шева, Израиль

В статье обсуждаются проблемы, возникающие при создании акустических 3D-моделей сердца, используемых далее для синтеза эхокардиографических изображений сердца на основе результатов расчета распределений акустических полей в программных симуляторах акустических полей Field II [6], Creanuis [5] и FOCUS [7]. Предложен способ создания акустических 3D-моделей сердца, основанный на использовании существующих геометрических 3D-моделей сердца человека. Представлены результаты синтеза изображений сердечной мышцы человека, подтверждающие его работоспособность.

Ключевые слова: алгоритмы синтеза, эхокардиографические исследования, ультразвуковые изображения, сердечная мышца, научная визуализация, акустическая модель, акустическое поле, обработка сигналов, Field II, Creanuis, FOCUS.

Computer synthesis of human heart echocardiographic images: problems and solutions

T.V. Golubyatnikov¹, S.V. Porshnev¹, N.I. Klorin²
golubyatnikov.tv@gmail.com|s.v.porshnev@urfu.ru|nat@bgu.ac.il

¹Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, Russian Federation;

³ Ben-Gurion University of the Negev, Beer-Sheva, Israel

The article discusses the problems that arise in the development of acoustic 3D-heart models, which are used later for the synthesis of echographic heart images. The synthesis is based on the results of acoustic field distributions calculation in Field II [6], Creanuis [5] and FOCUS [7] software. An approach for creating acoustic 3D-models of the heart have been proposed. The method is based on the use of existing geometric 3D-models of the human heart. The results that confirming method performance have been presented.

Keywords: synthesis algorithms, echocardiographic studies, ultrasound imaging, cardiac (heart) muscle, scientific visualization, acoustic model, acoustic field, signal processing, Field II, Creanuis, FOCUS.

1. Введение

Сегодня в кардиологической практике для проведения ультразвуковых исследований (УЗИ) сердца человека широко используются УЗИ-сканеры, представляющие собой программно-аппаратные комплексы (ПАК). Информация, получаемая в реальном времени с помощью ПАК, представляет собой последовательные во времени мгновенные эхокардиографические изображения сердца (УЗИ-кадры). Далее, для получения объективной оценки состояния сердечной мышцы и выявления (при наличии таковых) имеющихся у пациента патологий в функционировании сердечной мышцы, врачами-кардиологами проводится анализ УЗИ-видеофильмов. Данный анализ состоит в измерении на УЗИ-кадрах геометрических размеров тех или иных областей сердца и вычислении на их основе значений ряда количественных показателей, характеризующих состояние сердечной мышцы: площадей желудочков сердца и предсердий, фракцию выброса левого желудочка, время сокращения левого желудочка, объем сердца и других [17].

На первом этапе анализа УЗИ-кадров решается задача оконтуривания: на каждом кадре областей сердца, состояние которых определяет функциональное состояние сердечной мышцы. В связи с тем, что УЗИ-изображения сердечной мышцы оказываются низкоконтрастными, на них отсутствуют четко идентифицируемые границы некоторых областей сердца и, одновременно, присутствуют различные артефакты, обусловленные высоким уровнем спекл-шумов. Данные особенности УЗИ-изображений не позволяют и по сей день создать автоматические алгоритмы оконтуривания областей сердца на УЗИ-кадрах [22, 23]. В этой ситуации

врач-кардиолог, на основании собственных представлений о положении границ областей сердца производит, как правило, сегментацию УЗИ-изображений вручную. При этом обобщить опыт квалифицированных экспертов в виде некоторого набора правил оказывается невозможным. В результате достоверность диагноза оказывается зависящей от квалификации и опыта конкретного эксперта, проводившего анализ УЗИ-видеофильма.

Априори понятно, что можно решить перечисленные выше проблемы, препятствующие созданию автоматизированных алгоритмов выделения на УЗИ-кадрах областей сердца и их программных реализаций, если иметь в наличии изображения сердечной мышцы с известными формой и геометрическими размерами ее областей.

В статье обсуждаются проблемы, возникающие при создании акустических 3D-моделей сердца, а также предложен способ создания данных моделей на основе использования существующих геометрических 3D-моделей сердца человека.

2. Алгоритмы синтеза УЗИ-изображений

Как известно [18], синтез УЗИ-изображений внутренних органов человека, реализуемый в УЗИ-сканерах, основан на анализе параметров акустических полей, создаваемых акустическими импульсами, рассеянными биологическими тканями. Следовательно, общая задача компьютерного синтеза УЗИ-изображений (и УЗИ-изображений сердца, в том числе) может быть декомпозирована на следующие подзадачи:

1. Синтез геометрических 3D-моделей и соответствующих им акустических 3D-моделей.

- Решение задачи дифракции импульсных акустических полей, рассеиваемых внутренними органами человека.
- Синтез на основе рассчитанных распределений акустических полей внутри организма человека УЗИ-изображений внутренних органов.

Сегодня, используемые для моделирования акустических полей методы, можно разделить на две группы:

- основанные на вычислении линейной свертки [2, 8, 13, 14];
- основанные на решении уравнений акустики.

Использование методов линейной свертки из группы №1 позволяет существенно снизить требования к вычислительным мощностям [9], однако, точность получаемых здесь результатов существенно уступает аналогичным результатам, получаемым с использованием методов из группы №2, в которых производится решение волнового уравнения.

Методы, которые относятся к группе №2 традиционно используют одно из следующих уравнений: линейное волновое уравнение, предложенное Л.А. Черновым [3], уравнение Хохлова-Заболотской-Кузнецова (ХЗК) [10, 21], либо полное волновое уравнение, описывающее распространение ультразвука через биологические ткани с учетом поглощения, нелинейности, дифракции, а также множественного рассеяния, отражений и рефракции [15].

3. Алгоритм обработки сигналов и формирования УЗИ-изображений

На основании расчета акустического поля вдоль каждой из линий сканирования и с учетом положения активных акустических элементов приемопередатчика во временной области вычисляют значения давления рассеянного звукового поля. В результате проведенных вычислений получают распределение отраженного акустического поля в полярной системе координат. Для каждой из линий сканирования сигналы, зарегистрированные на элементах приемопередатчика, объединяются с учетом функции аподизации в единый сигнал (RF signal) [19]. Дальнейшая обработка RF сигналов, зарегистрированных по j направлениям сканирования, осуществляется путем последовательного выполнения следующих действий:

- Компенсация временного поглощения. Для компенсации эффекта поглощения обычно применяют следующую коррекцию $p_j = p_0 e^{-\alpha f d_j}$, где α – коэффициент поглощения, f – частота генерируемого сигнала и d_j – расстояние, пройденное акустической волной вдоль j -го луча.
- Фильтрация RF сигналов. Применяется для уменьшения шума вне диапазона передающей частоты и выделения, спектральной составляющей, соответствующей биологической ткани, – тканевой гармоники (tissue harmonic imaging). Традиционно, в качестве частотного фильтра используется фильтр Гаусса.
- Выделение с помощью демодуляции высокочастотной составляющей или использования преобразования Гильберта огибающей сигналов $h(t)_j$.
- Логарифмическая компрессия сигналов $h(t)_j$:

$$h_c(t)_j = \log_{10}(1 + a \cdot h(t)_j) / \log_{10}(1 + a),$$

где a – коэффициент компрессии, обеспечивающий приведение динамического диапазона сигнала к динамическому диапазону, воспринимаемому человеческим глазом.

- Преобразование функций $h_c(t)_j$ акустической волны из полярной в декартову систему координат и интерполяция распределения преобразованного акустического поля.

Более подробная информация о каждом шаге постобработки RF сигналов описана в [18].

Проиллюстрируем процесс синтеза УЗИ-изображения следующим простым примером (рис. 1). Здесь сканируемая сцена состоит из трех рассеивающих объектов цилиндрической формы с радиусами поперечного сечения $2 \cdot 10^{-3}$ м, центры которых находятся в точках с координатами $(0,0,0.01)$, $(0,0,0.015)$ и $(0,0,0.02)$ м, соответственно. Секторный приемопередатчик, расположенный в точке $(0,0,0)$, производит последовательное сканирование в трех угловых направлениях -15° , 0° и 15° , где углы отсчитываются от оси oZ в плоскости $Y=0$. Направления сканирования представлены линиями, исходящими из точки $(0,0,0)$. Серым цветом выделена область, видная на синтезируемом изображении. Для каждой из линий сканирования выполнялась обработка RF сигнала в соответствие с описанным выше алгоритмом.

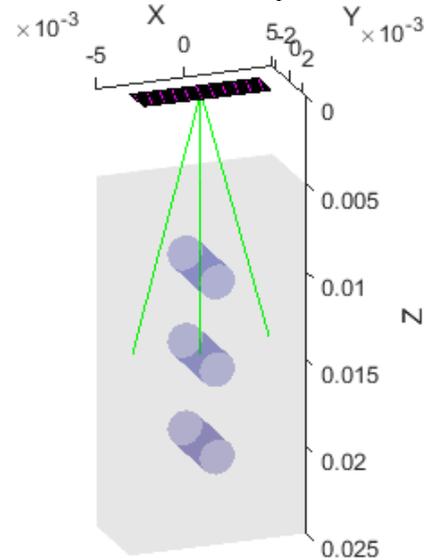


Рис. 1. Модель визуализируемой сцены.

Для каждой из линий сканирования выполняется поэтапная обработка радиочастотного сигнала в соответствии с описанным выше алгоритмом. Результаты выполнения этапов данного алгоритма для линии сканирования 0° представлены на рисунке 2б. Отметим, что на данном рисунке вдоль оси абсцисс отложено не время, но расстояние от источника акустической волны до рассеивателя, вычисляемое по формуле:

$$l_i = \frac{1}{2} c t_i = \frac{1}{2} \frac{c}{F_d} i, \quad i = \overline{0, N-1},$$

где F_d – частота дискретизации RF сигнала, а N – количество отсчетов дискретного RF сигнала. Логарифмически сжатые сигналы, вдоль каждой из линий сканирования, представлены на рисунке 2а. Здесь сигналы представлены в декартовой системе координат, а на рисунке 2б изображены те же сигналы, но уже в декартовой системе координат. Увеличивая количество линий сканирования можно улучшить качество синтезированного изображения (Рис 2с).

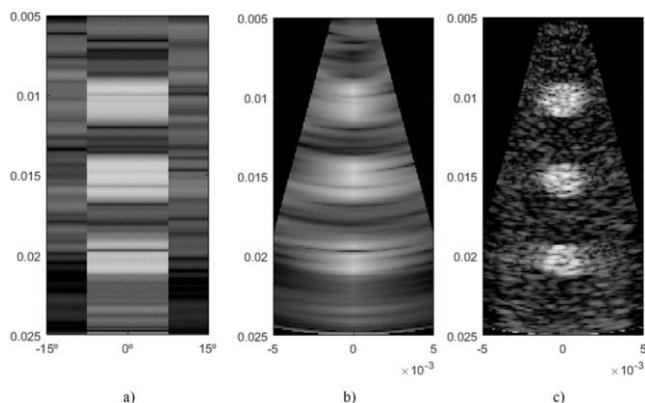


Рис. 2. Синтезированное УЗИ-изображение (а, b – три линии сканирования в угловых направлениях -15° , 0° и 15° , представленные в полярной и декартовой системах координат соответственно, и с – 60 линий сканирования).

Из рисунка 2 видно, что синтезированное УЗИ-изображений, действительно, соответствует исходной сцене.

Проанализировав ряд программных инструментариев (ПИ) [4, 5, 6, 7, 11], предназначенных для моделирования акустических полей и синтеза УЗИ-изображений было выявлено, что наиболее часто используется модель рассеивающей среды (объектов), состоящая из набора точечных рассеивателей. В этой связи в данной статье предлагается новый способ, позволяющий создавать рассеивающие среды с точечными рассеивателями.

4. Способ формирования акустических 3D-моделей областей сердца

В основу нового способа формирования моделей рассеивающих объектов была положена гипотеза о том, что можно создать адекватную 3D-акустическую модель сердца на основе использования каскадных геометрических 3D-моделей. Далее заполнить случайным образом некоторую область пространства правильной формы, заведомо большую отражающего объекта, и классифицировать точки по их принадлежности к соответствующим поверхностям каскадной геометрической 3D-модели, при этом точечные рассеиватели, не прошедшие классификацию считать лежащими вне границ сердца. В конечном итоге для каждого из точечных рассеивателей задать соответствующие значения акустических характеристик.

В процессе создания каскадной геометрической 3D-модели объекта требуется решать следующие задачи:

1. Импорт и экспорт модели и её отдельных частей в каком-либо стандартном формате 3D-моделей.
2. Замыкание контуров и поверхностей внутренних областей 3D-модели и проверка их на замкнутость.
3. Создание общих для нескольких фигур контуров и поверхностей, что требует наличия механизма ссылок на элементы 3D-модели.
4. Отображение сечений фигуры для возможности визуального контроля и представления результата синтеза каскадной геометрической 3D-модели объекта.

В соответствии с вышеуказанными требованиями нами был выбран редактор 3ds Max компании Autodesk, а для постобработки геометрической 3D-модели объекта, построенной в графическом редакторе 3ds Max, был выбран пакет MATLAB, имеющий необходимый инструментарий для работы с 3D-моделями.

Для проверки приведенной выше гипотезы была выбрана каркасная модель сердца [1], которая была сконвертирована в редакторе 3ds Max в тип Editable Poly.

Далее были выделены поверхности, представляющие из себя основные интересные объекты: клапаны легочного ствола и аорты, митральный клапан, правый предсердно-желудочковый клапан, внутренние поверхности, включающие левый и правый желудочки и предсердия, а также внешнюю поверхность границы сердца. Для последующей классификации точечных рассеивателей были выделены замкнутые объемы, состоящие из набора поверхностей исходной каркасной модели и дополнительных поверхностей, которые были добавлены для замыкания поверхностей некоторых отделов сердца. Наборы поверхностей, ограничивающих перечисленные выше отделы сердца, были выделены в отдельные слои (Layers). В том случае, когда замкнутые 3D-фигуры имели смежные поверхности, они выделялись в отдельные ссылочные поверхности (Reference) и для их идентификации использовался модификатор (Normal), задающий направление нормали внутрь соответствующей замкнутой 3D-фигуры. Для контроля замкнутости объемов был использован инструмент Open Edges. В результате был получен набор слоев, представляющих собой замкнутые объемы всех интересных областей сердца.

На следующем этапе для визуализации соответствующих плоскостей сканирования 3D-модели сердца была произведена настройка дополнительного окна просмотра (Viewport), была добавлена камера (Camera) и проведена конфигурация параметров срезов по соответствующим осям. В предположении, что сканирование будет производиться в плоскости XoZ , а центр приемопередатчика будет расположен на оси oZ , модель была позиционирована и повернута относительно центра координат таким образом, чтобы получить четырехкамерную проекцию сердца вдоль главной оси. Сечение 3D-модели сердца с названиями рассеченных объемов представлено на рисунке 3.

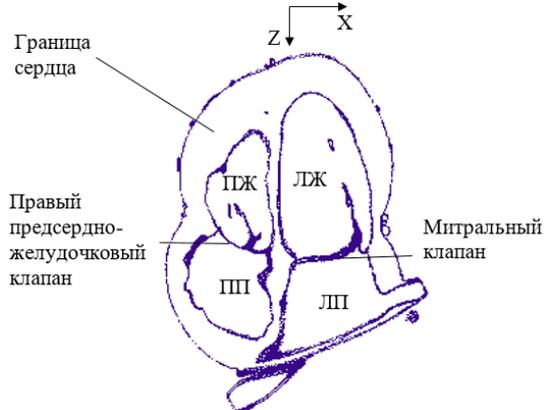


Рис. 3. Сечение геометрической 3D-модели сердца в плоскости $Y=0$ (проекция по главной оси): ПЖ – правый желудочек, ЛЖ – левый желудочек, ПП – правое предсердие, ЛП – левое предсердие.

Для обработки данных в MATLAB каждая из полученных 3D-областей была записана в отдельный файл стандартного графического формата .obj. Ввиду того, что встроенный механизм экспорта 3ds Max не позволяет выполнить экспорт выделенных объектов в отдельные файлы, для решения этой задачи был использован скриптовый язык MaxScript и авторами написан скрипт экспорта, код которого может быть найден в открытом доступе [16].

После загрузки геометрий 3D-моделей областей сердца в MATLAB были вычислены координаты описанного вокруг геометрий параллелепипеда. Далее были сгенерированы $0.4 \cdot 10^6$ 3D-координат рассеивателей в соответствии с равномерным случайным распределением внутри обсуждаемого параллелепипеда. При меньшем числе

рассеивателей качество синтезированного УЗИ-изображения по экспертным оценкам было признано не удовлетворительным. Априори понятно, что увеличение числа рассеивателей будет приводить к увеличению похожести реальных и синтезированных УЗИ-изображений при одновременном прямо пропорциональном увеличении времени расчетов, поэтому анализ синтезированных УЗИ-изображений при большем числе рассеивателей является направлением дальнейших исследований в ходе которых планируется оптимизировать используемые алгоритмы и применить более производительные вычислительные средства. Для проверки принадлежности каждого рассеивателя к одной из областей сердца была использована соответствующая m -функция, находящаяся в свободном доступе [12]. Также проводился специальный контроль за положением рассеивателей и их отнесения к соответствующей 3D-области, обусловленный тем, что некоторые из геометрических 3D-моделей перечисленных выше областей сердца оказываются вложенными друг в друга. Иллюстрация результатов выполненной классификации рассеивателей к митральному клапану и сухожильным хордам приведена на рисунке 4.

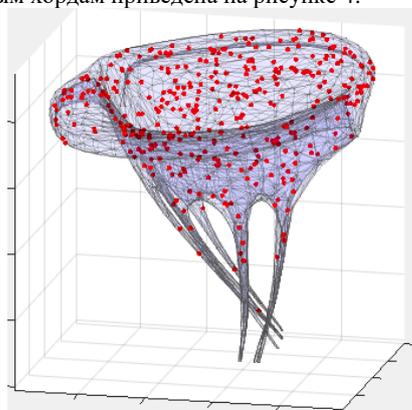


Рис. 4. Визуализация рассеивателей, отнесенных к митральному клапану и сухожильным хордам.

Отметим, что вследствие сложной структуры каркасной геометрической 3D-модели не удается избежать нежелательных пересечений выделенных 3D-областей. Это, в свою очередь, может приводить к возникновению погрешности при идентификации рассеивателей. Однако, визуальные оценки количества некорректно классифицированных рассеивателей показали, что их число не превосходит 1%, а потому проблемой можно пренебречь.

Далее для каждого рассеивателя, отнесенного в результате классификации к какой-либо из областей сердца, были определены эмпирически диапазоны их коэффициентов рассеяния. Для этого проводился синтез УЗИ-изображений при различных диапазонах значений коэффициентов рассеяния выбранных областей сердца. Далее проводилась независимая экспертная оценка степени близости синтезированных УЗИ-изображений к реальным УЗИ-изображениям сердца. Диапазоны значений коэффициентов рассеяния для выбранного экспертами синтезированного УЗИ-изображения представлены в таблице 1.

Таблица 1. Расчетные формулы коэффициентов рассеяния для классифицированных в замкнутые объемы рассеивателей. Функция $rnd()$ – генератор случайных чисел в диапазоне $[0, 1]$

№	Название объема	Диапазоны значений коэффициентов рассеяния
1	Граница сердца	$[0.2, 0.9]$
2	Правый желудочек и предсердие	$[0.01, 0.04]$

3	Левый желудочек и предсердие	$[0.01, 0.04]$
4	Митральный клапан	$[0.2, 0.95]$
5	Клапан легочного ствола	$[0.2, 0.95]$
6	Клапан аорты	$[0.2, 1]$
7	Правый предсердно-желудочковый клапан	$[0.2, 1]$

Так в пакете MATLAB была создана акустическая 3D-модель, использованная для расчета акустических полей и синтеза УЗИ-изображения сердца с помощью Field II, которую можно также использовать и в других ПИ с аналогичным типом модели рассеивающей среды (Creanuis, FOCUS).

5. Результаты синтеза УЗИ-изображения сердца

Пример УЗИ-изображения, синтезированного в ПИ Field II представлен на рисунке 5. Параметры среды и используемого для синтеза секторного приемопередатчика, основанного на фазированной решетке, представлены в таблице 2.

Таблица 2. Параметры рассеивающей среды и конфигурация секторного приемопередатчика

Параметр	Значение	Описание
F_0	$5 \cdot 10^6$	Центральная частота, Hz
F_s	$100 \cdot 10^6$	Частота сэмпирования, Hz
C_0	1540	Скорость распространения звука, м/с
ρ_0	1000	Плотность биологической ткани, кг/м ³
E_w	$1.54 \cdot 10^{-4}$ $= C_0 / F_0 / 2$	Ширина приемопередающего элемента, м
$Kerf$	$2.5 \cdot 10^{-6}$	Расстояние между элементами апертуры, м
E_h	0.007	Высота приемопередающего элемента, м
N_e	64	Количество приемопередающих элементов
$angle$	90	Размер сектора, °
N_i	121	Количество равноудаленных между собой линий сканирования

Из рисунка 5b видно, что синтезированное УЗИ-изображение оказывается подобным реальному УЗИ-кадру с изображением сердца в четырехкамерной проекции по главной оси (рис 5a). При этом формы и размеры областей сердца на синтезированном УЗИ-изображении оказываются близкими к соответствующим формам и размерам областей акустической 3D-модели, что можно наблюдать на рисунке 5b, где контуром отмечена граница пересечения 3D-модели с плоскостью, в которой производится сканирование и формируется синтезированный УЗИ-кадр. Различия же в изображении связаны в основном с тем, что коэффициенты рассеяния точечных рассеивателей на синтезированном УЗИ-изображении выбраны условно постоянными (см. таблицу 2), тогда как в реальных биологических тканях сердца значения данной характеристики варьируются, по-видимому, в более широком диапазоне значений.

6. Заключение

Предложен способ создания акустических 3D-моделей рассеивающих сред и получены результаты, подтверждающие его работоспособность. Данный способ позволяет на основе геометрической 3D-модели, представленной в стандартном графическом формате,

создавать акустическую модель рассеивающей среды и, используя существующие программные инструменты для моделирования распределения акустических полей, синтезировать УЗИ-изображения.

Продемонстрирована возможность автоматизации процесса создания акустических 3D-моделей с помощью пакета MATLAB на основе использования результатов предварительной обработки геометрических 3D-моделей в графическом редакторе.

Дальнейшие направления развития данного подхода состоят в разработке алгоритмов постобработки синтезированных УЗИ-изображений с целью приведения их текстур к текстурам реальных УЗИ-снимков, а также использования более реалистичных (не «точечных», но «сплошных» – volumetric data models (VDM) [20]) акустических 3D-моделей рассеивающих объектов. Представляется возможным использовать для их создания данные МРТ или КТ сердца. Их анализ позволит создать VDM сердечной мышцы, в которых для каждой точки сердца можно будет задать их акустические свойства. Исследование данного подхода является предметом последующих публикаций.

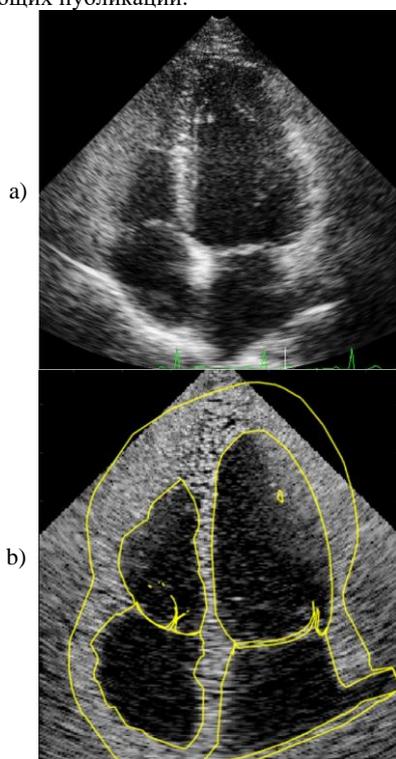


Рис. 5. Пример реального а) и синтезированного оконтуренного б) УЗИ-изображений сердца.

7. Литература

- [1] 3D model of the heart, <https://free3d.com/3d-model/human-heart-2-79840.html>
- [2] Bamber J.C. and Dickinson R.J. Ultrasonic B-scanning: a computer simulation, *Phys. Med. Biol.*, 25, 463-479, 1980.
- [3] Chernov L. A. Propagation of Waves in a Medium Having Random Inhomogeneities, M., Izd-vo AN SSSR., p. 156-157, 1958.
- [4] Comsol Multiphysics Acoustics module, <https://www.comsol.com/acoustics-module>
- [5] Creanuis software. <https://www.creatis.insa-lyon.fr/site7/fr/CREANUIS>
- [6] Field II software. <http://field-ii.dk/>
- [7] FOCUS software, <http://www.egr.msu.edu/~fultras-web/>
- [8] Gao H. et al. A fast convolution-based methodology to simulate 2-Dd/3-D cardiac ultrasound images, in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 2, pp. 404-409, February 2009.
- [9] Gao T. and Hergum, H. Torp, and D'hooge J. Comparison of the performance of different tools for fast simulation of ultrasound data, *Ultrasonics*, vol. 52, no. 5, pp. 573-577, 2012.
- [10] Kuznetsov V. P., Equation of nonlinear acoustics, *Sov. Phys. Acoust.* – 1971. – Т. 16. – №. 4. – P. 467-470.
- [11] k-Wave software, <http://www.k-wave.org/>
- [12] MATLAB function to check for 3D coordinates in a 3D object, <https://www.mathworks.com/matlabcentral/fileexchange/37856-inpolyhedron-are-points-inside-a-triangulated-volume->
- [13] Meunier J. and Bertrand M. Echographic image mean gray level changes with tissue dynamics: a system-based model study, in *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 4, pp. 403-410, April 1995.
- [14] Meunier J., Bertrand M. and Mailloux G. A model for dynamic texture analysis in two-dimensional echocardiograms of the myocardium. In: *SPI 0768*, pp. 193-200, 1987.
- [15] Pinton G., Dahl J., Rosenzweig S. and Trahey G. A heterogeneous nonlinear attenuating full-wave model of ultrasound, *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 56, no. 3, pp. 474-488, 2009.
- [16] Script for exporting layers to individual files, <https://github.com/golubyatnikov-tv/3dsMax/blob/master/ExportVolumes.ms>
- [17] Shiller N., Osipov M.A. *Clinical echocardiography*. – M.: Praktika, 2005.
- [18] Szabo T.L. *Diagnostic Ultrasound Imaging: Inside Out*, *Biomedical Engineering Series, Elsevier, Academic Press*, 2004.
- [19] Varray F. Simulation in nonlinear ultrasound : application to nonlinear parameter imaging in echo mode configuration, *Other. Universit'e Claude Bernard - Lyon I*, 2011.
- [20] Volumetric data visualization, <http://ps-medtech.com/volumetric-imaging/>
- [21] Zabolotskaya E. A, Khokhlov R. V. Quasi-plane waves in the nonlinear acoustics of confined beams, *Sov. Phys. Acoust.* – 1969. – Т. 15. – №. 1. – С. 35-40.
- [22] Zyuzin V.V., Porshnev S.V. and Bobkova A.O., Contouring left ventricle from echocardiographic image sequence in long-axis view, *Proceedings of 2nd International Conference on Information and Communication Technologies and Applications (ICTA 2012)*, 2012. P. 213-216.
- [23] Zyuzin V.V., Porshnev S.V., Bobkova A.O. and Bobkov V.V., The method of semiautomatic contouring of the human heart left ventricle on echocardiographic images, *Fundamental studies №8 Academy of Natural Sciences*, 2013, P. 44-48.

Об авторах

Поршнеv Сергей Владимирович, д.т.н., профессор института радиоэлектроники и информационных технологий Уральского федерального университета. Его e-mail s.v.porshnev@urfu.ru.

Клиорин Натан Иосифович, д.т.н., профессор университета им. Бен-Гуриона в Негеве. Его e-mail nat@bgu.ac.il.

Голубятников Тимофей Валерьевич, аспирант института радиоэлектроники и информационных технологий Уральского федерального университета. Его e-mail golubyatnikov.tv@gmail.com.

Оптимизация корреляционных методов регистрации биомедицинских изображений

Ф.А. Гурьянов, А.С. Крылов

fegur108@gmail.com|kryl@cs.msu.ru

Лаборатория математических методов обработки изображений,

Факультет вычислительной математики и кибернетики,

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

Предложен алгоритм поиска оптимального размера уменьшения изображений для ускорения корреляционных методов регистрации изображений. Алгоритм основан на приведении сигнала к одномерному виду и его декомпозиции адаптивным методом эмпирических мод. После выбора наиболее репрезентативной внутренней моды мы можем определить минимальное расстояние между экстремумами этой эмпирической моды. Таким образом мы можем оценить, во сколько раз можно уменьшить изображение на первом шаге сопоставления так, чтобы не потерять точность регистрации. В данной работе рассматривается задача сопоставления биомедицинских изображений с использованием метода максимизации взаимной информации. Проведено тестирование алгоритма на изображениях клеток HeLa.

Ключевые слова: корреляционный метод регистрации изображений, максимизация взаимной информации, декомпозиция адаптивным методом эмпирических мод, биомедицинские изображения, клетки HeLa.

Optimization of correlation methods for biomedical image registration

F.A. Guryanov, A.S. Krylov

fegur108@gmail.com|kryl@cs.msu.ru

Laboratory of Mathematical Methods of Image Processing,

Faculty of Computational Mathematics and Cybernetics,

Lomonosov Moscow State University, Moscow, Russia

We propose an algorithm for achieving the optimal image reduction to accelerate the correlative registration methods. Our approach is based on reduction of the signal to the one-dimensional form and its adaptive empirical mode decomposition. After selecting the most informative intrinsic mode function we can calculate the minimal distance between the extremes of the mode. Thus, we can estimate the maximum image reduction rate that allows to maintain the accuracy of registration. This paper investigates the case of biomedical image registration using the method of mutual information maximization. We test the algorithm on the images of HeLa cells.

Keywords: correlative image registration method, mutual information maximization, adaptive empirical mode decomposition, biomedical images, HeLa cells.

1. Введение

Регистрация изображений является важной задачей, возникающей при обработке биомедицинских изображений [7]. В данной работе будут рассматриваться снимки клеток HeLa, полученные с помощью микроскопа [10]. Пример таких изображений приведён на рис. 1. Эти клетки широко используются в различных исследованиях по фармакологии и биологии [11]. Для дальнейшего анализа нам необходимо сначала совместить изображения этих клеток. Существует два типа регистрации изображений: rigid и nonrigid [3]. Жёсткая (rigid) регистрация предполагает совмещение изображений с использованием только аффинных преобразований (сдвиг, вращение, масштаб), в то время как эластичная (nonrigid) регистрация допускает деформацию объекта для совмещения. Перед тем, как производить эластичную регистрацию обычно необходимо совместить объекты при помощи только аффинных преобразований. В данной статье мы рассматриваем задачу жёсткой регистрации изображений.

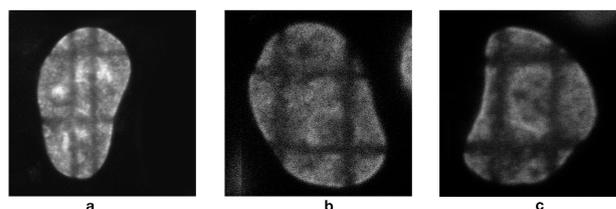


Рис. 1. Пример изображений клеток HeLa для регистрации.

Существуют следующие основные типы жёстких методов совмещения биомедицинских изображений [3, 4]:

1. Основанные на совмещении каких-либо особых точек.
2. Основанные на аппроксимации изображений простыми геометрическими фигурами (круг, овал, прямоугольник и другие) [1].
3. Методы, в которых используются обратное преобразование Фурье и логарифмическое полярное преобразование [8].
4. Корреляционные методы.

2. Корреляционные методы

Наиболее устойчивыми к различным типам и модальностям изображений являются корреляционные методы. В таких методах вводится понятие похожести изображений (этим и отличаются корреляционные методы друг от друга). И далее для регистрируемого изображения происходит подбор лучшего варианта преобразования. Такого, при котором преобразованное и целевое изображение максимально похожи. В данной работе в качестве метрики похожести мы используем взаимную информацию (mutual information) [6]. Метод максимизации взаимной информации (ММИ) является одним из самых распространённых энтропийных корреляционных методов. Энтропийные методы не зависят от характера изображения, так как работают в основном с гистограммами изображений.

Основным недостатком корреляционных методов, в частности ММИ, является низкая скорость работы алгоритма, так как необходимо перебрать все возможные варианты трансформации изображения. Поэтому обычно, вместо полного перебора вариантов, используют какую-нибудь оптимизированную версию поиска максимума значения метрики [9]. Один из самых очевидных и действенных способов оптимизации - искать на сетке приблизительный максимум, а потом уточнять значение полным перебором, но уже в значительно меньшей окрестности. Или уменьшить изображения в несколько раз и применить алгоритм для них, а после регистрации уменьшенных копий уточнить результат на оригинальных изображениях.

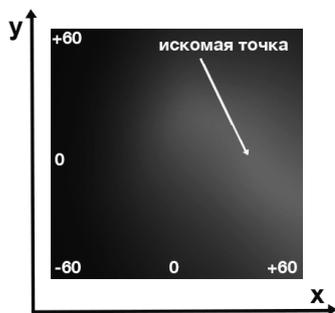


Рис. 2. График зависимости взаимной информации от смещения изображения по осям x и y . Чем светлее точка, тем больше взаимная информация при соответствующем сдвиге.

Подобные оптимизации можно производить, так как взаимная информация увеличивается плавно по мере приближения к искомой точке [6]. Пример изменения МИ, при смещении изображения по осям x и y , приведен на рис. 2.

Таким образом, производить подобную оптимизацию возможно, но остаётся открытым вопрос - во сколько раз можно уменьшить изображение, чтобы не потерять точность регистрации? Или, с каким шагом можно производить поиск по сетке так, чтобы максимальное значение МИ на сетке соответствовало области

с максимальным значением МИ на всей рассматриваемой области? В данной работе нами предложен и протестирован метод такой оценки, основанный на методе эмпирических мод.

3. Метод эмпирических мод

Метод эмпирических мод (Empirical Mode Decomposition или EMD) был разработан в 1998 году и нашёл своё применение во многих исследованиях [2, 5]. Он позволяет анализировать нелинейные и нестационарные данные и получать частотное распределение данных по времени для одномерных сигналов, путём разложения сигнала на функции с разной частотой – эмпирические (или внутренние) моды (Intrinsic Mode Functions или IMF).

Определение 1. Эмпирическая мода это функция, обладающая следующими свойствами:

1. Количество экстремумов не более чем на единицу отличается от количества пересечений нуля этой функцией на рассматриваемом промежутке.
2. Значение полусуммы верхней и нижней огибающих данной функции близко к нулю.

В результате разложения сигнала $f(t)$ на эмпирические моды мы получаем следующую сумму:

$$f(t) = r(t) + \sum_{i=1}^N \phi_i(t), \quad (1)$$

где $\phi_i(t)$ - эмпирические моды, а $r(t)$ - остаток. Первые моды содержат высокочастотные характеристики сигнала, а последние и остаток - низкочастотные.

3.1. Быстрый алгоритм декомпозиции

Существуют разные методы декомпозиции сигнала на эмпирические моды. В данной работе мы используем быстрый адаптивный метод декомпозиции на эмпирические моды [2]. Алгоритм декомпозиции для одномерного сигнала I состоит из следующих шагов:

1. Задаём изначальный размер окна $w = 3$.
2. Находим в сигнале I локальные экстремумы с размером окна поиска равным w . Локальные максимумы p должны удовлетворять следующим условиям:

$$I(p) > I(q), \forall q \in W_w(p), \quad (2)$$

где $W_w(p)$ обозначает окно с центром в точке p и размером w . Тогда локальные минимумы q должны удовлетворять условиям:

$$I(p) < I(q), \forall q \in W_w(p). \quad (3)$$

3. Для каждого локального максимума находим расстояние до другого ближайшего максимума. Обозначим это расстояние как d_{\max} . Для каждого локального минимума находим расстояние до другого ближайшего минимума и обозначаем его d_{\min} . Находим минимальное из этих значений $d = \min(d_{\max}, d_{\min})$.

4. Обновляем размер окна $w = 2 \lceil \frac{d}{2} \rceil + 1$.
5. Вычисляем верхнюю U и нижнюю L огибающую с обновлённым размером окна w :

$$U(p) = \max_{q \in W_w(p)} I(q), L(p) = \min_{q \in W_w(p)} I(q). \quad (4)$$

6. Вычисляем среднюю огибающую R как среднее арифметическое верхней и нижней огибающих, размытых с размером окна размытия w :

$$R(p) = \frac{1}{w} \sum_{q \in W_w(p)} \frac{U(q) + L(q)}{2}. \quad (5)$$

7. Для сигнала I вычисляем эмпирическую моду $M = I - R$. Получаем высокочастотную эмпирическую моду M и низкочастотный остаток R .
8. Далее считаем $I = R$ и повторяем шаги 2-7 (при каждом повторе получаем очередную внутреннюю моду) до тех пор пока R станет невозможно дальше разложить (то есть когда у R будет меньше двух максимумов или минимумов).

3.2. Размер окна

В описанном выше алгоритме декомпозиции важную роль играет размер окна w . Размер окна на каждом шаге это расстояние между экстремумами одного типа (минимумами или максимумами). Чем выше номер эмпирической моды, тем более низкочастотным сигналом она является. А значит у каждой следующей внутренней моды количество экстремумов будет уменьшаться, а расстояние между ними увеличиваться. То есть размер окна для каждой эмпирической моды тоже будет увеличиваться. При этом w является достаточно важной характеристикой сигнала, так как уменьшив частоту сигнала в w раз мы не потеряем локальных экстремумов.

4. Метод максимизации взаимной информации (ММИ)

Метод максимизации взаимной информации это корреляционный метод регистрации изображений, который описывается следующей формулой:

$$t^* = \arg \max_t \sum_{u,v} H_{I_t J}(u, v) \log \frac{H_{I_t J}(u, v)}{H_{I_t}(u)H_J(v)}, \quad (6)$$

где H_I и H_J нормализованные гистограммы сигналов I и J соответственно:

$$H_I(u) = \frac{|\{p \in \text{dom}I : I(p) = u\}|}{N_I}, \quad (7)$$

$$H_J(v) = \frac{|\{p \in \text{dom}J : J(p) = v\}|}{N_J}, \quad (8)$$

а $H_{I_t J}$ нормализованная совместная гистограмма сигналов I_t и J :

$$H_{I_t J}(u, v) = \frac{|\{p \in \text{dom}J : I_t(p) = u, J(p) = v\}|}{N_J}. \quad (9)$$

Этот метод широко используется для совмещения изображений [6, 7].

5. Оптимизация метода ММИ

На примере ММИ рассмотрим как с помощью описанного выше подхода декомпозиции сигнала методом эмпирических мод можно оптимизировать корреляционные методы регистрации биомедицинских изображений клеток.

5.1. Сведение к одномерному сигналу и его декомпозиция

Как можно видеть из примеров изображений (рис. 1), важными характеристиками рассматриваемых изображений являются границы клеток. Выделим границы клеток и, на основе характеристик границ, сделаем вывод о том, во сколько раз можно уменьшать изображения для применения алгоритма так, чтобы не потерять точность регистрации.

Сначала по изображению клетки построим бинарную маску клетки. Маска строится с помощью выделения по порогу яркости с дополнительным морфологическим анализом для удаления пропусков внутри маски и лишних элементов вне клетки. Данные были взяты из работы [10]. Пример клетки и маски для этой клетки изображён на рис. 3.

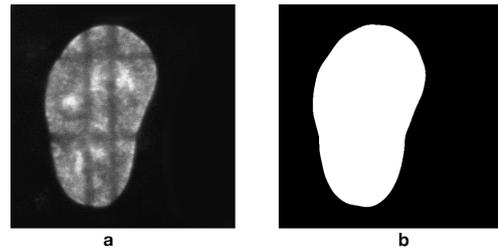


Рис. 3. Пример изображения клетки HeLa (a) и маски (b) для этого изображения.

Далее мы переводим эти маски в полярную систему координат с точкой отсчёта в центре масс маски клетки. Таким образом мы переходим от 2D картинке к одномерному сигналу - граница клетки в полярной системе координат. Результат такого перехода изображён на рис. 4.

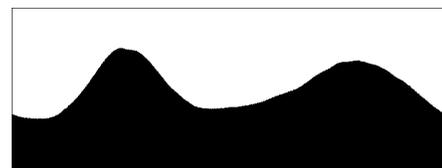


Рис. 4. Инвертированная маска клетки HeLa (смотри рис. 3) в полярной системе координат. По оси x расположен угол α , по оси y - радиус R .

Теперь производим декомпозицию полученного одномерного сигнала с использованием быстрого адаптивного метода эмпирических мод. Результат декомпозиции изображения a из рис. 1 можно увидеть на рис.

5. Нас интересует размер окна w получающийся после подсчёта каждой эмпирической моды.

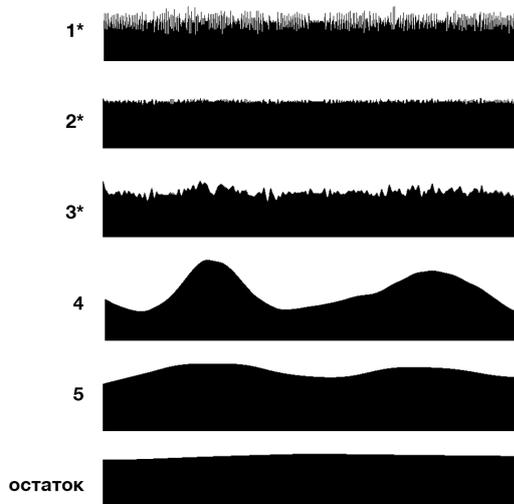


Рис. 5. Результат декомпозиции методом эмпирических мод сигнала на рис. 4. Внутренние моды 1 - 5 и остаток. Отмеченные звёздочкой (*) моды масштабированы для наглядности.

5.2. Вычисление коэффициента уменьшения изображения

Таблица 1 отражает важные свойства использованного нами алгоритма [5] - в определённый момент происходит значительный скачок размера окна. Шаг на котором произошёл скачок размера окна (шаг под номером k) мы определяем по следующей формуле:

$$k \rightarrow \min: \frac{\ln(w_{k+1})}{\ln(w_k)} \geq \frac{\ln(w_{i+1})}{\ln(w_i)}, \forall i. \quad (10)$$

изображение \ мода	мода						
	1	2	3	4	5	6	7
a	3	3	8	181	213	-	-
b	3	3	6	7	31	203	334
c	3	3	4	8	54	202	338

Таблица 1. Зависимость размера окна от номера внутренней моды для трёх тестовых изображений.

Изображения (a, b, c) можно увидеть на рис. 1. Выделены размеры окна, отвечающие уравнению 10 (прямо перед скачком размера окна).

Например в таблице 1 показано, что для изображения a из рис. 1 (первая строчка таблицы) скачок размера окна произошёл между третьей и четвёртой модой (мы в таком случае выделяем третью моду). Изменение характера эмпирических мод для данного изображения при переходе от третьей моды к четвёртой можно увидеть на рис. 5.

Следующие после скачка размера окна внутренние моды содержат самые низкочастотные характеристики

изображения. Для целей уменьшения регистрируемых изображений мы выберем эмпирическую моду прямо перед скачком размера окна. В этой эмпирической моде содержатся самые важные высокочастотные характеристики, так как шум и прочие случайные экстремумы отсеялись при вычислении первых эмпирических мод. Поэтому в качестве основного параметра для вычисления коэффициента уменьшения изображения мы возьмём размер окна w соответствующий этой эмпирической моде.

Теперь используя размер окна w мы вычисляем коэффициент k уменьшения изображения следующим образом:

$$k = \frac{\pi w (width + height)}{4N}, \quad (11)$$

где $width$ и $height$ это соответственно ширина и высота изображения, а N - количество шагов по параметру угла α при переходе к одномерному сигналу из полярной системы координат. Здесь мы по сути переводим w из полярной системы координат обратно к декартовой системе координат и пикселям. Поэтому значение зависит и от размеров изображения и от количества шагов по параметру угла при переходе к одномерному сигналу.

Полученный параметр k используется для оптимизации подбора преобразования, в данной работе - для максимизации взаимной информации.

6. Результаты

Мы протестировали предложенный метод на наборе изображений клеток HeLa [10]. Считалось, что наш метод работает правильно, если результат работы оптимизированного метода совпадал с результатом работы при полном переборе вариантов. Оценим коэффициент ускорения алгоритма. Вместо полного перебора вариантов трансформации угла и сдвигов по осям x и y , мы ищем максимум взаимной информации сначала по сетке с шагом k . Это будет требовать в k^3 раз меньше вычислений (так как сетка по трём параметрам). И затем будут необходимы дополнительные вычисления для уточнения результата.

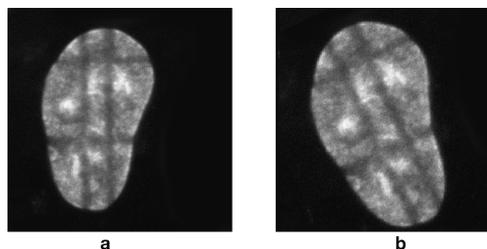


Рис. 6. Пример изображения (a) для регистрации (с изображением b на рис. 1) и результата регистрации (b).

Характерные результаты работы алгоритма для различных пар изображений приведены в таблице 2. Можно видеть, что при применении оптимизации резуль-

тат регистрации изображений не изменяется. При этом значительно уменьшается время работы алгоритма. Можно увидеть, что наша оценка коэффициента ускорения была адекватной. Например, для изображений 1, нахождение максимума на сетке будет происходить приблизительно в 7^3 , то есть примерно в 340 раз быстрее, чем при полном переборе. С учётом дальнейшего уточнения результата и полного перебора вокруг найденного узла сетки, получившийся коэффициент ускорения ($\frac{5156}{22}$) около 230 выглядит достаточно существенным. Таким образом можно утверждать, что полученная оценка коэффициента оптимизации является удачной, так как мы значительно ускоряем процесс регистрации и при этом не ухудшаем качество работы метода максимизации взаимной информации.

	x	y	angle	scale	coefficient	time (s)
изображения 1	40	-10	-25	1.15	-	5156
изображения 1*	40	-10	-25	1.15	7	22
изображения 2	4	18	-8	1	-	4957
изображения 2*	4	18	-8	1	7	24
изображения 3	-40	-11	25	0.85	-	4892
изображения 3*	-40	-11	25	0.85	8	19

Таблица 2. Результаты тестов для наборов изображений из таблицы 1: пары изображений (a;b), (b;c), (a;c) соответственно. Показаны вычисленные при регистрации значения сдвига по осям x и y, угла поворота, увеличения. Указано время, которое потребовалось для регистрации. Звёздочкой (*) отмечены строки, где применялась предлагаемая оптимизация (чётные строки). Для этих строк указано значение коэффициента оптимизации, вычисленное описанным выше способом.

7. Выводы

Предложен алгоритм оценки возможного уменьшения регистрируемых изображений для ускорения корреляционных методов регистрации изображений. Алгоритм протестирован для оптимизации метода максимизации взаимной информации. Тестирование показало хорошие результаты на наборе биомедицинских изображений клеток HeLa. Мы смогли существенно улучшить скорость работы алгоритма сопоставления, не ухудшив качество регистрации.

Дальнейшая работа будет направлена на более детальное исследование различных жёстких корреляционных методов и оптимизацию нежёстких методов регистрации биомедицинских изображений.

8. Благодарности

Исследование выполнено за счет гранта Российского научного фонда (проект №17-11-01279).

9. Литература

- [1] J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage* 38, 95–113, 2007.
- [2] S. M. A. Bhuiyan, R. R. Adhami, J. F. Khan. A novel approach of fast and adaptive bidimensional empirical mode decomposition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1313–1316, 2008.
- [3] L. G. Brown. A Survey of Image Registration Techniques. Department of Computer Science Columbia University NY 10027, 1–60, 1991.
- [4] J. Flusser and T. Suk. A moment-based approach to registration of images with affine geometric distortion. *IEEE Trans. on Geoscience and Remote Sensing*, 32(2): 382–387, 1994.
- [5] F. Guryanov, A. S. Krylov. Fast medical image registration using bidirectional empirical mode decomposition. In *Signal Processing: Image Communication.*, 1–6, 2017.
- [6] F. Maes, D. Vandermeulen, P. Suetens. Medical Image Registration Using Mutual Information. *Proceedings of the IEEE*, 91(10): 1699–1722, 2003.
- [7] Francisco, P. M. Oliveira and J. M. R. S. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2): 73–94, 2014.
- [8] B. S. Reddy and B. N. Chatterji. An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration. *IEEE Trans. on Image Processing*, 5(8): 1266–1271, 1996.
- [9] I. Sitdikov, F. Guryanov, A. S. Krylov. Accelerated Mutual Entropy Maximization for Biomedical Image Registration. In *Image Processing Theory, Tools and Applications*, 337–340, 2015.
- [10] D. V. Sorokin, J. Suchankova, E. Bartova, P. Matula. Visualizing Stable Features in Live Cell Nucleus for Evaluation of the Cell Global Motion Compensation. *Folia Biologica*, 60: 45–49, 2014.
- [11] Y. Zhao, R. Yao, L. Ouyang, H. Ding, T. Zhang, K. Zhang, S. Cheng, W. Sun. Three-Dimensional Printing of Hela Cells for Cervical Tumor Model in Vitro. *Biofabrication*, 6: 1–10, 2014.

Метод поиска эффективной последовательности алгоритмов обработки медицинских данных при сегментации анатомических структур сердца

В.В. Данилов¹, И.П. Скирневский¹, О.М. Гергет¹
viacheslav.v.danilov@gmail.com | skirnevskiy@tpu.ru | gerget@tpu.ru
¹Томский Политехнический Университет, Томск, Россия

В работе рассматривается способ определения эффективной последовательности операций при обработке медицинских изображений, в частности, операций, связанных с задачей контурирования анатомических структур сердца. Во-первых, мы представляем новую метрику Индекс оптимальности, которая является средневзвешенным коэффициентом и метрик точности, а также среднего времени выполнения каждого алгоритма. Данный индекс позволяет оценить, насколько быстро и точно выполняет обработку каждый алгоритм. Значение индекса оптимальности нормализовано и варьируется от 0 до 1, что облегчает сравнение различных подходов. Во-вторых, в работе сравниваются наиболее известные методы фильтрации, повышения контраста и сегментации, а все полученные результаты анализируются с использованием введенного индекса оптимальности. На этапе удаления цифрового шума были использованы такие фильтры как: медианный фильтр, фильтр полной вариации, билатеральный фильтр, фильтр потока кривизны, фильтр нелокального среднего и фильтр сдвига среднего. В качестве методов сегментации были применены следующие алгоритмы: метод сегментации водоразделами, кластеризация, сегментация выращиванием регионов, морфологическая сегментация, сегментация на основе графов и сегментация изолиниями. Все исследования проводились на статическом кадре из серии изображений, полученных в рамках эхокардиографического исследования сердца.

Ключевые слова: эхокардиография, УЗИ сердца, фильтрация, сердце, сегментация.

A method for effective sequence searching of algorithms processing medical data during segmentation of anatomical structures of the heart

V.V. Danilov¹, I.P. Skirnevskiy¹, O.M. Gerget¹
viacheslav.v.danilov@gmail.com | skirnevskiy@tpu.ru | gerget@tpu.ru
¹ Tomsk Polytechnic University, Tomsk, Russia

The paper considers a method for determining the effective sequence of steps during the processing of medical images, in particular, related to the problem of delineation the anatomical structures of the heart. First, we proposed the metric called the Optimality Index (OI) that is a weighted average of several accuracy coefficients, indices, and mean processing time. The metric allows to estimate how fast and accurate each image processing algorithm is. Moreover, the Optimality Index varies from 0 to 1, which facilitates the comparison of different approaches. The second thing is concerned with comparison of filtering, sharpening, and segmentation technique. All the obtained results are analysed and interpreted by dint of the new metric (OI). During noise reduction step, we compared Median filter, Total Variation filter, Bilateral filter, Curvature Flow filter, Nonlocal Means filter, and Mean Shift filter. To clarify the borders of anatomical structures we used Linear sharpen. Lastly, we applied Watershed segmentation, Clusterization, Region growing segmentation, Morphological segmentation, Image Foresting segmentation, and Isoline delineation as segmentation techniques. All the research studies were performed on a static frame from a series of images obtained as part of an echocardiographic study of the heart.

Keywords: echocardiography, cardiac ultrasound, filtering, heart, segmentation.

1. Введение

В настоящее время сердечно-сосудистые заболевания (ССЗ) являются одной из самых распространенных причин смерти в мире. По данным Всемирной организации здравоохранения, в 2012 году из-за расстройств, связанных с сердцем, умерло около 17,5 миллионов человек, что составляет 31% всех смертей [24]. На данный момент превентивные меры, позволяющие спрогнозировать риск появления ССЗ на основе не инвазивного исследования, продолжают набирать популярность. Очевидным преимуществом прогнозирования или выявления заболевания на ранних стадиях является относительная простота лечения, отсутствие необходимости оперативного вмешательства и значительное снижение рисков осложнения после лечения заболевания на поздних стадиях.

Сегментация медицинских изображений, в частности сегментация отделов сердца, играет важную роль, как в качественном, так и в количественном анализе, а её основной целью является извлечение определенных регионов интереса и оценка их характеристик.

В настоящее время одним из наиболее распространенных методов диагностики сердца является ультразвуковое исследование (УЗИ), обеспечивающее визуализацию сердечных структур в разных плоскостях. Сегментация анатомических структур сердца на основе данных ЭхоКГ является решающим этапом расчета основных клинических показателей, таких как толщина стенок, фракция выброса, аномалии движения стенки, масса сердца, объем предсердий и желудочков и т.д.

Сегодня ручное получение контуров, сделанное клиницистом, является самой популярной техникой при выполнении сегментации сердца. Однако, по мимо того, что ручная разметка очень трудоемкая и монотонная задача, ключевой проблемой является то, что результаты контурирования оцениваются только визуально. Подобная субъективная оценка, с учетом большой физической и умственной нагрузки, может привести к серьезным ошибкам и проблемам на последующих стадиях исследования.

Таким образом, существует потребность внедрения автоматизированных систем, позволяющих не только выполнять сегментацию анатомических структур сердца в автоматизированном или полу автоматическом режимах, но

и проводить объективную оценку полученных результатов. В связи с этим, основной целью данной работы был поиск метода для объективной оценки качества работы алгоритмов обработки цифровых изображений. В исследовании также освещаются последовательности методов фильтрации и сегментации, дающие в связке наилучшие результаты при обработке данных эхокардиографии.

2. Материалы и методы

Основным предположением, лежащим в основе данной работы, является то, что сегментация камер сердца будет более точной, надежной и быстрой, только в том случае, если поток обработки изображений (workflow) будет состоять из оптимальных этапов фильтрации, повышения резкости и сегментации.

Материалами данного исследования служат наборы данных 2D ЭхоКГ, которые были получены на свином сердце с использованием преобразователя X7-2, эхокардиографа Philips iE33 и программного обеспечения PSS5.1 Ultrasound. Обработка данных осуществлялась на компьютере со следующими характеристиками: Intel Core i7-4820K 3.7GHz CPU и NVIDIA GeForce 960 GT.

Основной поток обработки состоит из нескольких шагов, а именно: чтение данных, фильтрация, повышение резкости, сегментация, пост-процессинг и оценка точности. Главные шаги потока обработки указаны ниже.

1. Фильтрация. В работе использовались наиболее популярные алгоритмы удаления цифрового шума при обработке изображений:
 - a. Медианный фильтр [25, 26].
 - b. Фильтр полной вариации [5, 17].
 - c. Билатеральный фильтр [13, 22].
 - d. Фильтр потока кривизны [14].
 - e. Фильтр нелокального среднего [3, 4].
 - f. Фильтр сдвига среднего [6].
2. Повышение резкости. Данная техника используется для повышения визуальной чёткости изображения. В работе рассматривается линейный алгоритм увеличения резкости с помощью метода нерезкого маскирования (unsharp masking) [7]. Также был рассмотрен метод деконволюции, который позволяет проводить более качественную обработку за счёт широкого спектра параметров и определения типа и «формы» шума. В работе используется только линейный алгоритм, так как этап повышения резкости не является ключевым, а используется как дополнительный шаг в целях улучшения визуального отображения.
3. Сегментация. Для сравнения было выбрано 6 различных техник сегментации:
 - a. Сегментация водоразделами [1, 20].
 - b. Кластеризация [12].
 - c. Сегментация выращиванием регионов [2, 16].
 - d. Морфологическая сегментация
 - e. Сегментация на основе графов [10].
 - f. Сегментация на основе изолиний.
4. Пост-процессинг. Этап включает процедуры, которые обрабатывают связанные области после сегментации.

На этапе пост-процессинга мы использовали математическую морфологию, которая была первоначально предложена Ж. Серра [19]. Для оценки точности работы алгоритмов в качестве дополнительных показателей, были оценены: общее количество пикселей и доля, охватываемая каждым регионом; индекс структурного сходства [23]. Для сравнения областей автоматического и ручного сегментирования были вычислены три метрики, такие как индекс Сёрнсена-Дайса [21] [8], индекс Жаккара [11] и

коэффициент корреляции. Для оценки точности и расстояния между контурами была применена модифицированная дистанция Хаусдорфа [9].

Основная концепция работы основана на использовании последовательного применения методов из основного потока обработки изображений, указанного на рис. 1.



Рис. 1. Рабочий поток обработки изображений.

Одной из главных сложностей исследования с помощью такого подхода является проблематичность оценки всех вариантов рабочего процесса. Например, для выбранных нами алгоритмов общее число вариантов рабочего потока равно 72. Чтобы объективно оценить наилучший путь рабочего процесса, мы построили ориентированный граф, вершинами которого являются процедуры обработки изображений, а именно: фильтрации, повышения резкости и сегментации. При прохождении любого пути оцениваются время выполнения алгоритма на каждом блоке и ошибки сегментации. Эти две метрики позволяют судить о том, насколько точно и быстро выполняется выбранная комбинация алгоритмов.

3. Результаты

3.1 Фильтрация

При решении задачи удаления цифрового шума с изображения следует учитывать, что каждый фильтр имеет различное поведение и параметры, такие как: размер ядра, связанность, регуляризация, пространственное распространение и др. Более того, разные фильтры корректно работают только с определенным типом цифрового шума. В нашем исследовании параметры каждого фильтра были выявлены эмпирически. Под оптимальными принимались такие параметры, при которых метод удалял цифровой шум и в тоже время не изменял структуру границ объектов. На рис. 2 приведен регион правого сердца после этапа фильтрации.

Как можно увидеть из рис. 2, большинство фильтров при корректной настройке параметров справляются с задачей удаления цифрового шума без значительного размытия

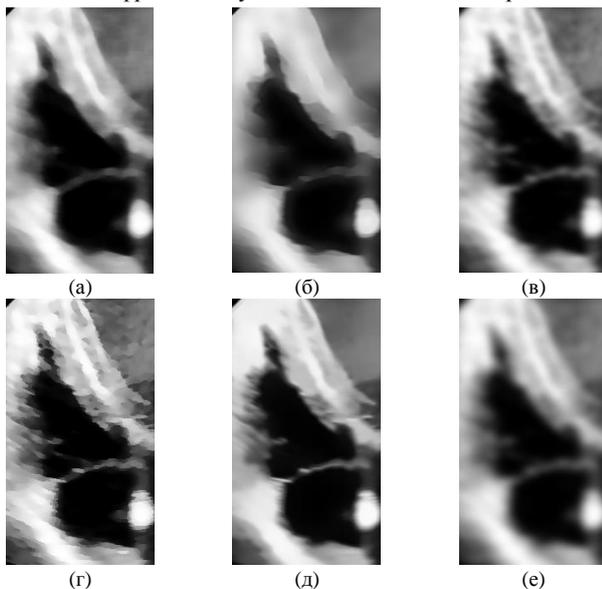


Рис. 2. Фильтрация региона правого сердца: (а) Медианный фильтр, (б) Фильтр полной вариации, (в) Билатеральный фильтр, (г) Фильтр потока кривизны, (д) Фильтр нелокального среднего, (е) Фильтр сдвига среднего. Тем не менее, фильтр потока кривизны разрывает и повреждает границы объектов, что может оказать отрицательное влияние на последующих этапах обработки.

3.2 Увеличение контраста

Линейные алгоритмы увеличения контраста, основанные на производных первого и второго порядка (в частности, Лапласиан) могут использоваться наряду с нелинейными вариантами на основе градиента или нечетких множеств. В нашей работе использованы линейные алгоритмы контрастирования. Стоит отметить, что радиус окна варьировался в процессе эксперимента от 15 до 60 пикселей. На рис. 3 показан фрагмент эхокардиографии, на котором изображена верхушка правого желудочка после применения алгоритма увеличения контраста для различных фильтров удаления цифрового шума.

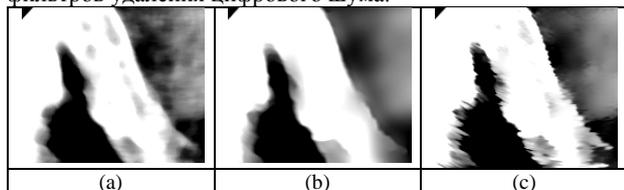


Рис. 3. Повышение контраста изображения после использования: (а) Медианного фильтра, (б) Фильтра полной вариации, (в) Билатерального фильтра.

После двух шагов, а именно, удаления цифрового шума и увеличения контраста фильтр потока кривизны значительно повреждает границы изображения. Причиной подобного рода искажения может быть либо неспособность корректной работы данного фильтра с изображениями эхокардиографии в силу особенностей структуры, разрешения и типа шума, либо некорректно подобранные параметры алгоритма.

Очевидно, что комбинация алгоритмов фильтрации может быть более эффективной при решении задач удаления цифрового шума без искажения входных данных. Данные исследования проводились в работах [15, 18]. Так, комбинация фильтра Гаусса и Лапласиана или алгоритм Марра-Хильдрета могут быть эффективной связке.

Сочетание билатерального фильтра и фильтра Гаусса также значительно снижает среднеквадратическую ошибку и увеличивает значение индекса подобия. Однако, в работе не освещаются комбинации алгоритмов так как это выходит за рамки предметной области, а именно исследования механизмов объективной оценки алгоритмов обработки изображений.

3.3 Сегментация

Чтобы правильно оценить точность автоматических методов сегментации в исследовании использовались вручную размеченные изображения. Контурирование было выполнено опытным врачом-кардиологом (рис. 4).

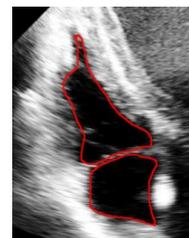


Рис. 4. Ручное выделение регионов, выполненное врачом-кардиологом.

Суть анализа точности алгоритмов автоматической сегментации заключалась в сравнении полученных результатов для каждого из методов с контуром, полученным вручную.

Результаты работы алгоритмов сегментации представлены на рис. 5.

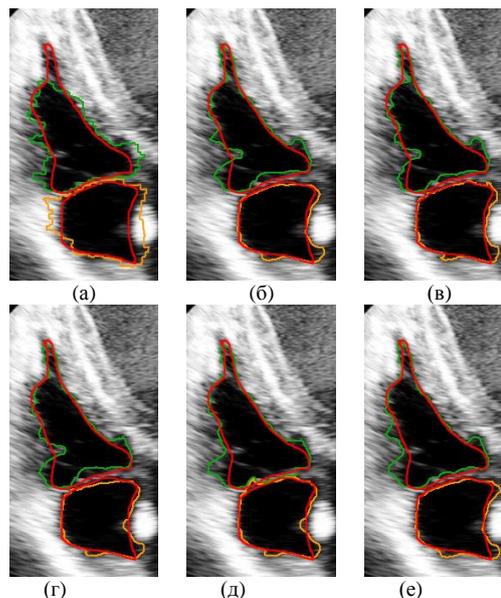


Рис. 5. Контур региона правого сердца: (а) Сегментация водорозделами, (б) Кластеризация, (в) Сегментация выращивание регионов, (г) Морфологическая сегментация, (д) Сегментация на основе графов, (е) Изолинии.

Зеленые и желтые контуры, описывающие правый желудочек и правое предсердие соответственно, были получены с помощью морфологических преобразований. Каждый контур представляет из себя разность между изображениями после эрозии и дилатации (1).

$$g(x, y) = e(x, y) - d(x, y), \quad (1)$$

где $g(x, y)$ – выходное изображение с контурами,

$e(x, y)$, $d(x, y)$ – изображение после эрозии, и дилатации соответственно.

В нашей работе мы получили толщину контура в 2 пикселя. Как можно заметить из рис. 5, автоматическое выделение не такое гладкое, как ручное, что свидетельствует о наличии ошибок. Однако, искажения

подобного рода можно нейтрализовать или свести к минимуму за счет морфологических операций или легкой фильтрацией.

4. Количественная оценка

4.1 Оценка методов фильтрации и контрастирования

Используемыми в повсеместной практике метриками для количественной оценки и сравнения качества работы алгоритмов удаления цифрового шума и методов увеличения контраста являются: среднеквадратическая ошибка (СКО, MSE), отношение сигнал/шум (ОСШ, SNR), пиковое отношение сигнала к шуму (ПОСШ, PSNR) и индекс структурного сходства (ИСС, SSIM). Мы вычислили все метрики для различных стадий. На рис. 6 приведены значения MSE после фильтрации и контрастирования входного изображения.

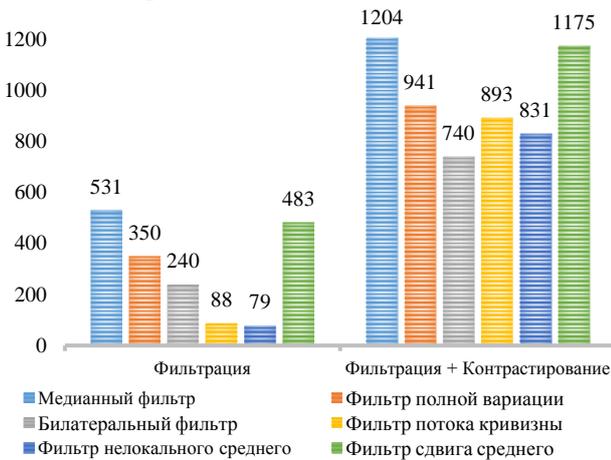


Рис. 6. Среднеквадратическая ошибка между фильтрованным, контрастированным и исходным изображениями для различных фильтров.

Как представлено на рис. 6, на этапе удаления цифрового шума наиболее эффективными являются Фильтр потока кривизны и нелокального среднего. Однако, в связке с алгоритмом увеличения контраста лучший результат показал Билатеральный фильтр. Следует отметить, что медианный фильтр и Фильтр сдвига среднего значительно ухудшают изображение.

Для метрики ОСШ и ПОСШ, согласно полученным данным, фильтр Нелокального среднего и Фильтр потока кривизны более устойчивы к шуму, в то время как Медианный фильтр и Фильтр сдвига среднего имеют самые низкие значения ОСШ, что говорит о низкой эффективности фильтров при решении задач подавления цифрового шума.

Очевидно, не имеет никакого смысла сравнивать между собой СКО и ОСШ, ПОСШ и ИСС, так как они имеют разные масштабы, а некоторые из них оперируют только абсолютными значениями. Для решения этой проблемы в работе была введена новая метрика комплексной оценки качества изображений после этапов фильтрации – Коэффициент оптимальности (3). Метрика позволяет анализировать данные разных типов и позволяет идентифицировать лучший алгоритм фильтрации цифровых изображений. Аргументы данной метрики нормализованы в масштабе от 0 до 1. В работе использовался линейный метод нормализации (2):

$$y_n = \frac{y_i - y_{min}}{y_{max} - y_{min}}, \quad (2)$$

где y_n – нормализованная метрика,

y_{min} , y_{max} – минимальное и максимальное значение диапазона параметра соответственно,
 y_i – значение параметра.

Стоит отметить, что параметры MSE, SNR и PSNR зачастую отражают одинаковую картинку и имеют прямую зависимость, однако во время экспериментов возникали аномальные выбросы: при относительно небольших значениях PSNR, уровень шума (MSE) был достаточно большим. Поэтому для полноценного описания и охвата всех случаев все три параметра MSE, SNR и PSNR были включены в Коэффициент оптимальности.

$$OC_F = \frac{a * MSE_n + b * SNR_n + c * PSNR_n + d * SSIM_n + e * PTime_n}{N}, \quad (3)$$

где N – количество используемых метрик,

a, b, c, d, e – параметры взвешенного усреднения, позволяющие варьировать важность и вклад параметров (в данном случае все константы использованы полноценно, т.е. равны 1),

$PTime$ – время работы алгоритма.

4.2 Оценка алгоритмов сегментации

Оценка методов сегментации является одной из важных задач, так как выполнение визуальной верификации перед внедрением алгоритма в клиническую практику может стать причиной не выявленных ошибок на этапе тестирования. Однако, помимо визуальной оценки полученных контуров, существует ряд метрик, позволяющих произвести численную оценку качества работы алгоритмов. Как было указано в разделе 2, мы использовали следующие метрики: индекс структурного сходства, индекс Сёрнсена-Дайса, индекс Жаккара и коэффициент корреляции. Результаты оценки с использованием данных метрик приведены на рис. 7.

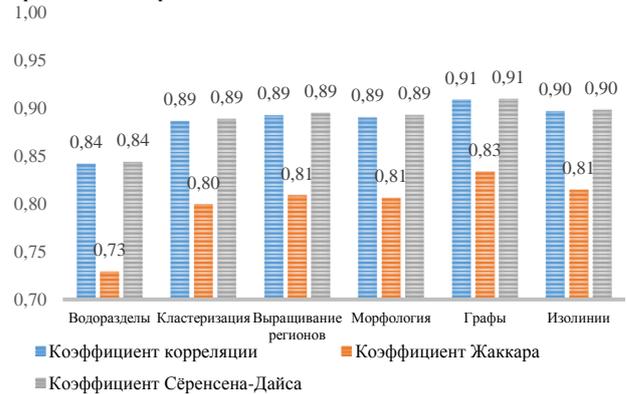


Рис. 7. Индексы подобия и корреляции для области правого желудка.

Как можно заметить из рис. 7, все три метрики дают примерно одинаковое значение для всех алгоритмов сегментации, однако метод сегментации водоразделами уступает в точности. Причиной является захват алгоритмом лишних областей.

Также, как и в случае с оценкой качества фильтрации, метрики оценки качества сегментации имеют разные размерности, и их невозможно сравнивать между собой. Внедрение модификации Коэффициента оптимальности для оценки методов сегментации может позволить решить эту проблему.

$$OC_S = \frac{a * DI_n + b * JI_n + c * CC_n + d * HD_n + e * MHD_n + f * DD_n + g * JD_n + h * PTime_n}{N}, \quad (4)$$

где N – количество используемых метрик,

a, b, c, d, e, f, g, h – параметры взвешенного усреднения, позволяющие варьировать важность параметров (в данном случае все константы равны 1),

DI – коэффициент Сёрнсена-Дайса, JI – коэффициент Жаккара, CC – коэффициент корреляции, HD – дистанция Хаусдорфа, MHD – модифицированная дистанция Хаусдорфа, DD – различие Дайса, JD – различие Жаккара, $PTime$ – время работы алгоритма.

В таблице 1 представлены усредненные значения параметров для методов фильтрации и сегментации, а также итоговый коэффициент оптимальности (КО). Параметры методов были подобраны эмпирическим путём и могут быть не оптимальными для различных входных данных. В данной работе выбранные параметры в минимальной степени искажают исходное изображение, о чём свидетельствует максимальное значение коэффициента оптимальности.

Таблица 1. Коэффициент оптимальности для различных методов обработки цифровых изображений.

Метод	Параметры	КО
Фильтрация		
Медианный фильтр	Размер окна = 5	0.58
Фильтр полной вариации	Параметр регуляризации = 0.3 Количество итераций = 25	0.79
Билатеральный фильтр	Разброс пикселей = 0.3 Пространственный размах = 1 Количество итераций = 5	0.96
Фильтр потока кривизны	Проводимость = 0.5 Время пропускания = 3	0.82
Фильтр нелокального среднего	Размер окна = 7	0.39
Фильтр сдвига среднего	Диапазон = 3 Дистанция = 3 Количество итераций = 5	0.60
Сегментация		
Сегментация водоразделами	Порог = 0.3	0.62
Кластеризация	Количество кластеров = 4	0.74
Сегментация выращиванием регионов	Порог = 0.15	0.84
Морфологическая сегментация	Порог = 0.06	0.69
Сегментация на основе графов	Радиус соединения = 0.7	0.75
Сегментация изолиниями	–	0.90

Таким образом, комбинация метода фильтрации и сегментации с минимальным значением коэффициента оптимальности может позволить получить оптимальную связку алгоритмов при решении различных задач обработки медицинских данных.

5. Заключение

На сегодняшний день существует значительное число алгоритмов фильтрации и сегментации изображений. При оценке результатов работы методов обработки изображений, основным критерием зачастую является субъективное визуальное восприятие итоговых данных исследователем или клиницистом. Настоящее исследование демонстрирует вариант поиска подходящего алгоритма при обработке ультразвуковых изображений, опираясь на полученные значения Коэффициента оптимальности, как объективного индикатору эффективности выбранного метода. Коэффициент оптимальности рассчитывается как средневзвешенное значение нескольких коэффициентов и показателей. Кроме того, метрика нормализована в диапазоне от 0 до 1, что даёт более объективные результаты, которые можно легко интерпретировать и сравнивать.

Таким образом, в данной работе не предлагается унифицированное решение для любого класса задач обработки медицинских данных, однако, вводится метрика, которая может быть полезна исследователям при оценке качества работы их алгоритмов.

6. Благодарности

Данная работа выполнена при поддержке Министерства Образования и Науки РФ, а также государственного задания «Наука» №12.8205.2017/БЧ.

7. Литература

- [1] Arbelaez P. [и др.]. Contour detection and hierarchical image segmentation. // IEEE transactions on pattern analysis and machine intelligence. 2011. № 5 (33). С. 898–916.
- [2] Brice C.R., Fennema C.L. Scene analysis using regions // Artificial Intelligence. 1970. № 3–4 (1). С. 205–226.
- [3] Buades A., Coll B., Morel J.-M. Non-Local Means Denoising // Image Processing On Line. 2011. (1). С. 490–530.
- [4] Buades A., Coll B., Morel J.-M. A non-local algorithm for image denoising // Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. 2005. № 0 (2). С. 60–65 т. 2.
- [5] Chambolle A. [и др.]. An introduction to total variation for image analysis // Theoretical foundations and numerical methods for sparse recovery. 2010. (9). С. 263–340.
- [6] Comaniciu D., Meer P. Mean shift: A robust approach toward feature space analysis // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. № 5 (24). С. 603–619.
- [7] Deng G. A generalized unsharp masking algorithm // IEEE Transactions on Image Processing. 2011. № 5 (20). С. 1249–1261.
- [8] Dice L.R. Measures of the Amount of Ecologic Association Between Species // Ecology. 1945. № 3 (26). С. 297–302.
- [9] Dubuisson M.-P., Jain A.K. A modified Hausdorff distance for object matching // Proceedings of 12th International Conference on Pattern Recognition. 1994. № 1 (1). С. 566–568.
- [10] Falcão A.X., Stolfi J., Alencar Lotufo R. de The image foresting transform: theory, algorithms, and applications. // IEEE transactions on pattern analysis and machine intelligence. 2004. № 1 (26). С. 19–29.
- [11] Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura // Bulletin del la Société Vaudoise des Sciences Naturelles. 1901. № January 1901 (37). С. 547–579.
- [12] Kanungo T. [и др.]. An efficient k-means clustering algorithm: analysis and implementation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. № 7 (24). С. 881–892.
- [13] Kwon K., Kim M.-S., Shin B.-S. A fast 3D adaptive bilateral filter for ultrasound volume visualization // Computer Methods and Programs in Biomedicine. 2016. (133). С. 25–34.
- [14] Malladi R., Sethian J. a. Image processing via level set curvature flow. // Proceedings of the National Academy of Sciences of the United States of America. 1995. № 15 (92). С. 7046–50.
- [15] Marr D., Hildreth E. Theory of Edge Detection // Proceedings of the Royal Society B: Biological Sciences. 1980. № 1167 (207). С. 187–217.
- [16] Pratt W.K. Digital Image Processing / W.K. Pratt, 3-е изд., John Wiley & Sons, 2001. 735 с.
- [17] Rudin L.I., Osher S., Fatemi E. Nonlinear total variation based noise removal algorithms // Physica D: Nonlinear Phenomena. 1992. № 1–4 (60). С. 259–268.
- [18] Scott E. Umbaugh Digital image processing and computer vision / Scott E. Umbaugh, 1990. 254 с.
- [19] Serra J. Image Analysis and Mathematical Morphology / J. Serra, London - New York: Academic Press, 1982. 610 с.
- [20] Smistad E. [и др.]. Medical image segmentation on GPUs – A comprehensive review // Medical Image Analysis. 2015. № 1 (20). С. 1–18.
- [21] Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and

- its application to analyses of the vegetation on Danish commons // Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter. 1948. № 4 (5). С. 1–34.
- [22] Tomasi C., Manduchi R. Bilateral Filtering for Gray and Color Images // International Conference on Computer Vision. 1998. С. 839–846.
- [23] Wang Z. [и др.]. Image quality assessment: From error visibility to structural similarity // IEEE Transactions on Image Processing. 2004. № 4 (13). С. 600–612.
- [24] World Health Organization Cardiovascular diseases // Fact sheet on CVDs [Электронный ресурс]. URL: <http://www.who.int/mediacentre/factsheets/fs317/en/> (дата обращения: 16.12.2016).
- [25] Zeng H. [и др.]. An Improved Algorithm for Impulse Noise by Median Filter // AASRI Procedia. 2012. (1). С. 68–73.
- [26] Zhu Y., Huang C. An Improved Median Filtering Algorithm for Image Noise Reduction // Physics Procedia. 2012. (25). С. 609–616.

Об авторах

Данилов Вячеслав Владимирович, инженер лаборатории дизайна медицинских изделий, RASA-центр, Томского политехнического университета.
E-mail: viacheslav.v.danilov@gmail.com.

Игорь Петрович Скирневский, инженер лаборатории дизайна медицинских изделий, RASA-центр, Томского политехнического университета. E-mail: skirnevskiy@tpu.ru.

Ольга Михайловна Гергет, к.т.н., научный сотрудник лаборатории дизайна медицинских изделий, RASA-центр, Томского политехнического университета. E-mail: gerget@tpu.ru.

Алгоритм встраивания информации в пары блоков ДКП-коэффициентов сжатого цифрового изображения

О.О. Евсютин¹, Р.В. Мещеряков¹, А.В. Ращупкина¹
ooo@keva.tusur.ru|mrv@keva.tusur.ru|angelinara@mail.ru

¹Томский государственный университет систем управления и радиоэлектроники, Томск, Россия

Один из известных способов защиты конфиденциальной информации основан на использовании стеганографического кодирования, когда защищенная передача данных реализуется с сокрытием самого факта передачи. В данной работе предлагается новый стеганографический алгоритм, работающий со сжатыми цифровыми изображениями. Отличительной особенностью данного алгоритма является использование пар схожих блоков ДКП-коэффициентов JPEG-изображения для встраивания частей сообщения. Операция встраивания отдельных битов сообщения состоит в задании определенной разности между соответствующими коэффициентами двух блоков. Для повышения эффективности стеганографического встраивания использована модификация генетического алгоритма, основанная на троичной логике. С помощью генетического алгоритма определяется наилучший вариант распределения изменений между соответствующими элементами каждой пары блоков, выбранной при встраивании. Целевой функцией является величина PSNR. Полученный алгоритм обеспечивает высокое качество встраивания и позволяет избежать существенных искажений модели исходного изображения при встраивании секретного сообщения за счет использования пар схожих блоков.

Ключевые слова: информационная безопасность, защита данных, цифровая стеганография, сокрытие данных, JPEG.

Algorithm of the information embedding into the pairs of blocks of the compressed digital image DCT-coefficients

O.O. Evsutin¹, R.V. Meshcheryakov¹, A.V. Rashchupkina¹
ooo@keva.tusur.ru|mrv@keva.tusur.ru|angelinara@mail.ru

¹Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia

One of the known methods of the confidential information protection is based on the steganographic coding use when the protected data transfer is implemented with concealment of the fact of transmission. The new steganographic algorithm working with the compressed digital images is offered in this paper. The difference of this algorithm is use of pairs of the similar blocks of the JPEG image DCT-coefficients for the message parts embedding. Operation of the message separate bits embedding consists in specifying of a certain difference between the appropriate coefficients of two blocks. The modification of the genetic algorithm based on ternary logic is used for increase in the steganographic embedding efficiency. By means of the genetic algorithm is defined the best version of distribution of changes between the appropriate elements of each blocks pair selected in the embedding process. Value of PSNR is the object function. The received algorithm provides the high embedding quality and allows to avoid essential distortions of the initial image model in the process of the confidential message embedding due to use of pairs of the similar blocks.

Keywords: information security, data protection, digital steganography, data hiding, JPEG.

1. Введение

Одно из современных направлений защищенной передачи данных в информационных системах основано на использовании методов цифровой стеганографии, реализующих встраивание в цифровые объекты скрытых информационных последовательностей различного назначения.

Стеганографические методы защиты информации позволяют решать такие задачи, как обеспечение конфиденциальности информации и обеспечение аутентификации цифровых объектов [6].

В настоящем исследовании речь идет о встраивании информации в сжатые JPEG-изображения. При работе с такими изображениями встраивание осуществляется посредством внесения изменений в квантованные коэффициенты дискретного косинусного преобразования (ДКП-коэффициенты). Соответствующие алгоритмы можно разделить на два класса: алгоритмы, оперирующие отдельными ДКП-коэффициентами, и алгоритмы, оперирующие группами ДКП-коэффициентов. Во втором случае для встраивания одного элемента данных секретного сообщения используется два и более ДКП-коэффициента, и встраивание заключается в установлении определенного соотношения между данными

коэффициентами в зависимости от значения встраиваемого элемента данных.

Основной целью данной работы является получение алгоритма встраивания информации в сжатые JPEG-изображения, обеспечивающего высокое качество стегоизображения за счет адаптации встраивания к изображению-контейнеру при использовании для встраивания элементов сообщения групп ДКП-коэффициентов.

2. Предшествующие работы

Большинство алгоритмов, работающих с JPEG-изображениями, относится к первому классу, когда элементы сообщения независимым образом распределяются по отдельным ДКП-коэффициентам. В качестве примеров можно отметить алгоритмы, основанные на методе PM1 [1, 12], различные LSB-подобные алгоритмы [2, 9], алгоритмы, основанные на методе QIM [3, 10] и т.д.

Классическим примером алгоритма, реализующего встраивание элементов сообщения в группы ДКП-коэффициентов, является алгоритм, представленный в работе [13]. Он встраивает один бит сообщения (цифрового водяного знака) в пару среднечастотных ДКП-коэффициентов. Встраивание заключается в изменении данных коэффициентов таким образом, чтобы модуль их

разности был больше или меньше некоторой фиксированной величины в зависимости от значения встраиваемого бита.

Схожий алгоритм, встраивающий два бита сообщения в пару ДКП-коэффициентов, представлен в работе [7]. В данном случае встраивание изменяет соотношение между выбранными коэффициентами и знак их среднего арифметического значения.

В настоящее время в рассматриваемом направлении цифровой стеганографии появляются работы, в которых встраивание сообщения осуществляется более чем в одно изображение.

В работах [4, 5] предлагается встраивать сообщение в несколько изображений одной сцены, например, в несколько последовательно полученных фотоизображений. В [5] предлагается использовать два изображения, в более позднем исследовании [4] – произвольное количество изображений.

В [11] предлагается алгоритм встраивания информации в схожие между собой блоки ДКП-коэффициентов стереопар. Схожесть блоков определяется сопоставлением низкочастотных областей ДКП-спектра, а встраивание осуществляется в области средних частот.

Необходимо отметить, что в данных исследованиях не идет речь о схемах разделения секрета, когда секретное сообщение распределяется между некоторым количеством различных изображений. В рассматриваемом случае предполагается передача всех использованных для встраивания цифровых изображений одному получателю.

Преимущество использования нескольких изображений для встраивания одного сообщения заключается в уменьшении количества изменений, приходящихся на одно изображение, за счет чего повышается устойчивость перед стегоанализом.

3. Методы исследования

В данном разделе представлены основные положения, лежащие в основе проведенного исследования.

3.1 Встраивание информации в пары блоков ДКП-коэффициентов

Рассмотрим алгоритм встраивания, представленный в статье [11]. Кроме отмеченного преимущества подход, реализуемый данным алгоритмом, обладает также и недостатком, связанным с нетипичностью использования стереоизображений в обычной переписке пользователей, привлекающей нежелательное внимание к стегоконтейнеру. Однако данный подход может быть легко перенесен на встраивание информации в одиночные изображения.

Встраивание сообщения в стереопару в соответствии с [11] осуществляется следующим образом.

В блоках ДКП-коэффициентов выделяется область поиска, состоящая из 6 низкочастотных элементов, и область встраивания, состоящая из 21 среднечастотного элемента. Для каждого блока левого изображения, входящего в стереопару, находится наиболее похожий на него блок в правом изображении. Поиск ведется в квадратной области, состоящей из $K \times K$ блоков. Критерием схожести является минимальность суммы квадратов разностей соответствующих ДКП-коэффициентов, рассчитанной для областей поиска пары блоков.

После нахождения парного блока вычисляются разности между парами соответствующих ДКП-коэффициентов, принадлежащих областям встраивания. Пары равных друг другу ДКП-коэффициентов используются далее для встраивания элементов сообщения, а все прочие пары ДКП-коэффициентов изменяются таким

образом, чтобы разности между ними увеличились на единицу по абсолютному значению.

Перед встраиванием каждые три бита сообщения представляются в виде двух чисел из множества $\{-1, 0, 1\}$. Каждое из таких чисел встраивается в пару соответствующих ДКП-коэффициентов двух схожих блоков стереопары. Для этого между данными коэффициентами устанавливается разность, соответствующая встраиваемому числу.

Необходимо отметить, что область поиска не изменяется при встраивании сообщения, поэтому поиск схожих блоков может быть безошибочно повторен на этапе извлечения встроеного сообщения.

Описанный подход к встраиванию сообщения в стереопару может быть перенесен на встраивание в одиночные изображения достаточно простым образом, если вести поиск парных блоков ДКП-коэффициентов по единственному имеющемуся изображению.

3.2 Оптимизация встраивания с помощью генетического алгоритма

Для повышения эффективности работы стеганографических алгоритмов во многих исследованиях ставятся задачи оптимизации, решаемые с помощью метаэвристик, как это можно увидеть, например, из обзорной работы [8]. Достаточно часто для этого применяется генетический алгоритм.

В рассматриваемом подходе к встраиванию информации в сжатые JPEG-изображения возможность для оптимизации появляется при выборе конкретного способа установления необходимого разностного соотношения между парой ДКП-коэффициентов.

Пример представлен на рис. 1 для случая, когда встраивание производится в пары коэффициентов с произвольным исходным значением разности.

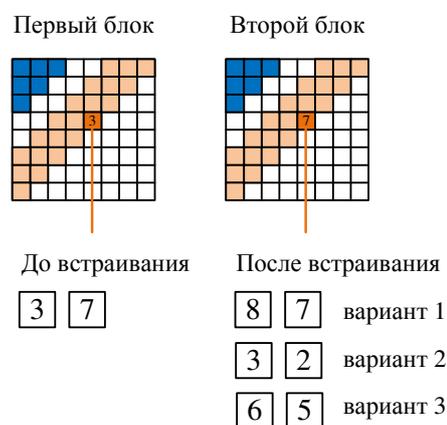


Рис. 1. Выбор способа изменения разности между парами ДКП-коэффициентов

Для решения соответствующей задачи оптимизации в настоящем исследовании используется модификация генетического алгоритма, оперирующая векторами-особями, записанными в троичном алфавите. Длина вектора для данной пары ДКП-блоков соответствует количеству пар ДКП-коэффициентов, используемых для встраивания, а значения, принимаемые отдельными элементами, имеют следующий смысл:

- 1) 0 – изменяется коэффициент первого блока;
- 2) 1 – изменяется коэффициент второго (парного) блока;
- 3) 2 – равные изменения вносятся в оба коэффициента.

В качестве целевой функции принимается величина PSNR.

4. Предлагаемый алгоритм

В данном разделе представлен алгоритм, реализующий встраивание информации в пары блоков ДКП-коэффициентов сжатого цифрового изображения с использованием генетического алгоритма.

Вход:

сообщение $M = m_1 m_2 \dots m_L$, $m_i \in \{0, 1\}$, $i = \overline{1, L}$;
пустой стегоконтейнер — цифровое изображение, сжатое по методу JPEG; параметр глубины поиска K ; порог схожести блоков T ; размер области встраивания u ; параметры генетического алгоритма.

Выход:

заполненный стегоконтейнер.

Шаг 1. Преобразовать сообщение M , записанное в двоичном алфавите $\{0, 1\}$, в сообщение M' , записанное в алфавите $\{-1, 0, 1\}$.

Шаг 2. Восстановить из JPEG-файла блоки квантованных ДКП-коэффициентов для трех компонент цветового пространства YCbCr.

Шаг 3. Создать пустой список S для хранения номеров заполненных ДКП-блоков.

Шаг 4. Для $i = 1, N$, где N — количество ДКП-блоков, выполнить следующее:

Шаг 4.1. Выбрать очередной ДКП-блок C_i .

Шаг 4.2. Если номер выбранного блока содержится в списке S , пропустить данный блок. В противном случае перейти к следующему шагу.

Шаг 4.3. В окрестности Мура порядка K выбрать первый найденный ДКП-блок C_j , $i \neq j$, такой, что

$$d(C_i, C_j) < T, \text{ где } d(C_i, C_j) = \sum_{l=0}^5 (c_i^l - c_j^l)^2 \text{ и номер}$$

данного блока отсутствует в списке S . Если блок, удовлетворяющий данному условию, отсутствует в окрестности блока C_i , то принять в качестве C_j блок с наименьшим значением $d(C_i, C_j)$.

Шаг 4.4. Сгенерировать популяцию генетического алгоритма, состоящую из r особей вида $p^k = (p_1^k, p_2^k, \dots, p_u^k)$, $p_l^k \in \{0, 1, 2\}$.

Шаг 4.5. Развивать популяцию генетического алгоритма в течение τ итераций, приняв за целевую функцию значение PSNR, рассчитываемое для блоков пикселей, восстанавливаемых из ДКП-блоков C_i и C_j .

Шаг 4.6. Выбрать особь с наибольшим значением целевой функции $p^{\text{best}} = (p_1^{\text{best}}, p_2^{\text{best}}, \dots, p_u^{\text{best}})$.

Шаг 4.7. Встроить часть последовательности M' в пару ДКП-блоков C_i и C_j следующим образом:

$$\forall l = \overline{1, u},$$

если $p_l^{\text{best}} = 0$, то

$$\tilde{c}_i^l = c_i^l + \Delta,$$

$$\tilde{c}_j^l = c_j^l,$$

если $p_l^{\text{best}} = 1$, то

$$\tilde{c}_i^l = c_i^l,$$

$$\tilde{c}_j^l = c_j^l + \Delta,$$

если $p_l^{\text{best}} = 2$, то

$$\tilde{c}_i^l = c_i^l + \frac{\Delta}{2},$$

$$\tilde{c}_j^l = c_j^l + \frac{\Delta}{2},$$

где Δ — это величина изменений, необходимых для установления заданной разности между коэффициентами c_i^l и c_j^l , зависящая от значения встраиваемого элемента последовательности M' .

Шаг 4.8. Добавить номера блоков C_i и C_j в список S .

Шаг 5. Осуществить статистическое кодирование ДКП-коэффициентов и завершить алгоритм.

Можно увидеть, что в представленном алгоритме встраивание осуществляется в пары ДКП-коэффициентов независимо от исходного значения разности между ними. Это позволяет более точно оценивать емкость изображения-контейнера до начала встраивания.

Алгоритм извлечения является достаточно очевидным и в настоящей статье не приводится.

5. Вычислительные эксперименты и их обсуждение

Вычислительные эксперименты с полученным алгоритмом проводились на выборке из 9 полноцветных тестовых JPEG-изображений разрешением 512×512 пикселей: «Airplane», «Baboon», «Earth», «House», «Lenna», «Peppers», «Sailboat», «Splash», «Tiffany» [14]. Встраиваемые сообщения представляли собой тексты на русском языке, сжатые с помощью программы-архиватора.

Примеры тестовых изображений приведены на рис. 2.

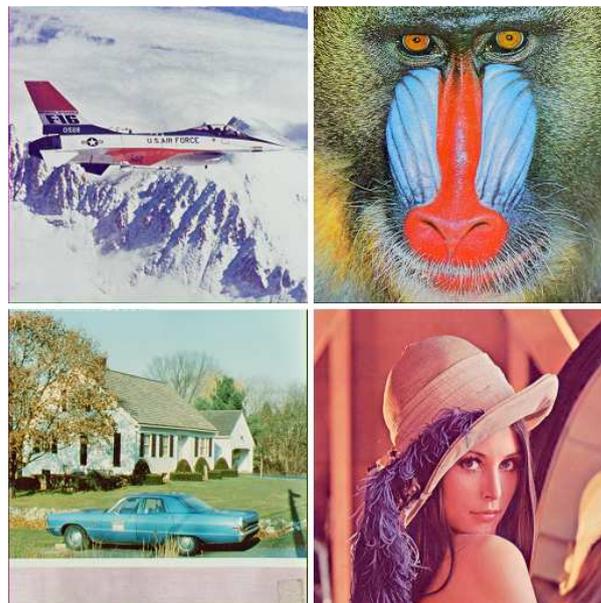


Рис. 2. Примеры тестовых изображений.

На рис. 3 представлена зависимость значения PSNR от длины встраиваемого сообщения, полученная для разработанного алгоритма посредством усреднения результатов экспериментов по всей тестовой выборке.

Для оценки влияния генетического алгоритма на качество встраивания на этом же графике представлена аналогичная зависимость для упрощенной версии разработанного алгоритма. В этой версии выбор пар схожих ДКП-блоков осуществлялся так же, как и в основном алгоритме, но решение об изменении того или

иного коэффициента в каждой паре принималось случайным образом.

Можно увидеть, что применение генетического алгоритма позволяет существенно повысить качество встраивания, улучшение составляет более 1 дБ.

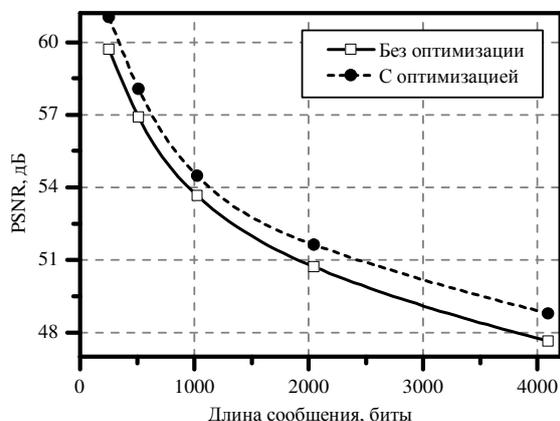


Рис. 3. Зависимость качества стегоконтейнера от объема встраиваемого сообщения.

На рис. 4 представлены примеры стегоизображений, полученных с помощью предложенного алгоритма. Объем встраиваемой информации в каждом случае составлял 4096 битов. Можно увидеть, что встраивание не приводит к появлению на стегоизображениях заметных артефактов.



Рис. 4. Примеры стегоизображений

В качестве недостатка нашего алгоритма в сравнении с известными аналогами можно отметить меньшую максимальную емкость. Однако это характерно для всех стеганографических алгоритмов, которые используют для встраивания отдельных битов сообщения группы ДКП-коэффициентов, либо осуществляют встраивание битов сообщения в одиночные ДКП-коэффициенты, выбирая их по некоторому критерию из множества имеющихся ДКП-коэффициентов.

Примеры подобных алгоритмов представлены в работах [7, 13]. Результаты вычислительных экспериментов с данными алгоритмами показаны в табл. 1. В данных

экспериментах использовались те же тестовые изображения, что и при исследовании алгоритма, полученного в настоящей работе.

Длина сообщения, биты	PSNR, дБ		
	Алгоритм [7]	Алгоритм [13]	Полученный алгоритм
2 712	40,71	—	50,65
4 096	—	34,41	48,79

Табл. 1. Сравнение полученного алгоритма с аналогами

Можно увидеть, что предложенный алгоритм обеспечивает значительное преимущество по сравнению с алгоритмами-аналогами. Данное преимущество объясняется тем, что операция встраивания, на которой основывается предложенный алгоритм, вносит меньшее количество искажений в изображение-контейнер по сравнению с операциями, используемыми в [7, 13]. Дополнительное преимущество достигается за счет применения генетического алгоритма для оптимизации встраивания.

6. Заключение

В данной работе представлен новый алгоритм встраивания информации в сжатые JPEG-изображения, отличающийся использованием пар схожих блоков ДКП-коэффициентов для формирования пространства сокрытия и применением генетического алгоритма для повышения качества встраивания.

Предлагаемый подход к встраиванию информации в сжатые JPEG-изображения, реализованный в представленном алгоритме, дает потенциал для новых исследований, направленных на получение новых эффективных стеганографических алгоритмов. В частности, можно предложить новые решения по выбору пар схожих ДКП-блоков и ввести более сложные критерии схожести. Развитию данного подхода будут посвящены новые исследования авторов работы.

Отдельно необходимо отметить, что сформулированная в разделе 3.2 задача оптимизации может быть перенесена на иные подходы к встраиванию сообщения в наборы изображений, представленные в известных исследованиях.

7. Благодарности

Данная работа выполнена при поддержке РФФИ (проект № 16-47-700350 p_a).

8. Литература

- [1] Евсютин О.О. Алгоритм встраивания информации в сжатые цифровые изображения на основе операции замены с применением оптимизации / О.О. Евсютин, А.А. Шелупанов, Р.В. Мещеряков, Д.О. Бондаренко // Компьютерная оптика. – 2017. – Т. 41, № 3. – С. 412–421.
- [2] Chang C.C. A steganographic method based upon JPEG and quantization table modification / C.C. Chang, T.S. Chen, L.Z. Chung // Information Sciences. – 2002. – Vol. 141(1–2). – P. 123–138.
- [3] Chen B. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding / B. Chen, G.W. Wornell // IEEE Transactions on Information Theory. – 2001. – Vol. 47(4). – P. 1423–1443.
- [4] Denmark T. Steganography with Multiple JPEG Images of the Same Scene / T. Denmark, J. Fridrich // IEEE Transactions on Information Forensics and Security. – 2017. – Vol. 12(10). – P. 2308–2319.

- [5] Denmark T. Steganography with Two JPEGs of the Same Scene / T. Denmark, J. Fridrich // Proceedings of the 42nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017). — USA, New Orleans, LA. — 2017.
- [6] Fridrich J. Steganography in Digital Media: Principles, Algorithms, and Applications. — Cambridge: Cambridge University Press, 2010. — 437 p.
- [7] Fujimura M. New data hiding scheme using method of embedding two bits data into two DCT coefficients / M. Fujimura, T. Takano, S. Baba, H. Kuroda // Proceedings of the International Conferences on Signal Processing, Image Processing and Pattern Recognition (SIP 2010) and Multimedia, Computer Graphics and Broadcasting (MulGraB 2010). — Korea Jeju, Island. — 2010. — P. 156–164.
- [8] Huang H.C. Survey of Bio-inspired Computing for Information Hiding // Journal of Information Hiding and Multimedia Signal Processing. — 2015. — Vol. 6(3). — P. 430–443.
- [9] Li X. A steganographic method based upon JPEG and particle swarm optimization algorithm / X. Li, J. Wang // Information Sciences. — 2007. — Vol. 177(15). — P. 3099–3109.
- [10] Noda H. High-performance JPEG steganography using quantization index modulation in DCT domain / H. Noda, M. Niimi, E. Kawaguchi // Pattern Recognition Letters. — 2006. — Vol. 27(5). — P. 455–461.
- [11] Yang W. Reversible DCT-based data hiding in stereo images / W. Yang, L. Chen // Multimedia Tools and Applications. — 2015. — Vol. 74(17). — P. 7181–7193.
- [12] Yu L. PM1 steganography in JPEG images using genetic algorithm / L. Yu, Y. Zhao, R. Ni, Z. Zhu // Soft Computing. — 2009. — Vol. 13(4). — P. 393–400.
- [13] Zhao J. Embedding robust labels into images for copyright protection / J. Zhao, E. Koch // Proceedings of the International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies (KnowRight'95). — Austria, Vienna. — 1995. — P. 242–251.
- [14] SIPI Image Database [Электронный ресурс]. — URL: <http://sipi.usc.edu/database/> (дата обращения 01.12.2016).

Об авторах

Евсютин Олег Олегович, канд. техн. наук, доцент кафедры безопасности информационных систем факультета безопасности Томского государственного университета систем управления и радиоэлектроники. Его e-mail eo@keva.tusur.ru.

Мещеряков Роман Валерьевич, д-р техн. наук, заведующий кафедрой безопасности информационных систем факультета безопасности Томского государственного университета систем управления и радиоэлектроники. Его e-mail mrv@keva.tusur.ru.

Ращупкина Анжелика Владимировна, инженер кафедры безопасности информационных систем факультета безопасности Томского государственного университета систем управления и радиоэлектроники. Ее e-mail angelinara@mail.ru.

Ассоциация сцен в эндоскопических видео

Д.А. Коваленко¹, В.С. Гнатюк²

dmitri.a.kovalenko@gmail.com|alfredfid@gmail.com

¹ Computer Science Center, Санкт-Петербург, Россия;

² Москва, Россия

Эндоскопическое наблюдение является распространенным методом раннего выявления ряда заболеваний пищевода. Видеопоток с эндоскопа подвергается ручному анализу, в ходе которого эксперт соотносит изображения участков ткани с их более ранними снимками для установления динамики протекания болезни. Предложен алгоритм ассоциации сцен для автоматизации этого процесса. Подход основан на глобальном анализе изображения и построении системы соседства визуально схожих кадров. Исследование применимости метода проведено на видеоданных реальных эндоскопических операций. Менее специфическая проблема обнаружения границ сцен в видеопотоке рассмотрена; определена ее взаимосвязь с выбранной задачей.

Ключевые слова: ассоциация сцен, обнаружение границ сцен, анализ эндоскопических видео.

Scene association in endoscopic videos

D.A. Kovalenko¹, V.S. Gnatyuk²

¹ Computer Science Center, Saint-Petersburg, Russia;

² Moscow, Russia

Endoscopic surveillance is a common method of monitoring several diseases of the esophagus. Such a surveillance includes the manual matching of images of tissue regions with their earlier observations from previous surveillance operations to control disease dynamics. A scene association algorithm is proposed to automate this process. The algorithm is based on a global approach to image analysis and utilizes construction of neighborhood system for visually resemblant frames. The experimental evaluation of method has been conducted on a set of in vivo endoscopic videos. The article explores the relationship between the current problem and the problem of scene boundary detection.

Keywords: scene association, scene boundary detection, surgical vision.

1. Введение

Обнаружение границ сцен в видеопотоке является темой активных исследований в течении последних двух десятилетий. Технология важна для ряда практических приложений: индексации и навигации по базам видеоданных, оптимизации методов сжатия в продвинутых видеокодеках, реализации шага предобработки для высокоуровневого семантического анализа видеопотока. Решения, основанные на обнаружении границ сцен в видеопотоке, применяются хостингами (пример: Youtube) для построения информативной раскадровки – визуальной подсказки при перематке. Кроме того, реализации технологии встраиваются в популярные средства редактирования видео*. Национальный институт стандартов и технологий США в рамках конференции TREC проводит соревнование TRECVID [5], в рамках которого с 2001 по 2007 год одной из задач являлось обнаружение границ сцен в видеопотоке. Настоящая статья рассматривает ее обобщение до задачи ассоциации сцен и ее применение в сфере обработки эндоскопических изображений.

Гастроэнтерологическая эндоскопия является одним из методов предупреждения развития аномальных образований в пищеводе. Пример такого заболевания – синдром Барретта, проявляющийся в метаплазии слизистой и считающийся фактором,

повышающим риск аденокарциномы пищевода. Эндоскопический мониторинг состояния аномальных образований включает в себя задачу обработки и сопоставления видеотректов участков ткани, снятых в разное время. Задача трудоемка и подвержена ошибкам разметки, вызванным человеческим фактором, поэтому частичная автоматизация процесса является мотивацией исследования.

Определение 1. Обнаружение границ сцен в видеопотоке заключается в разделении видеопотока на связанные последовательности кадров, объединенных по их семантическим или морфологическим признакам. Для потока кадров $F_0 \dots F_i$, где индексы являются их порядковыми номерами, нужно дать ответ, принадлежит ли кадр F_i той же сцене, что и кадр F_{i-1} .

Определение 2. Ассоциация сцен в видеопотоке заключается в разделении видеопотока на подмножества кадров (не обязательно последовательных), сгруппированных по их семантическим или морфологическим признакам.

В протоколе тестирования TRECVID, в соответствии с определением 1, идеальная разметка представляет собой последовательность индексов кадров, расположенных на границах сцен. Метрики качества вычисляются при сравнении индексов «пограничных» кадров, предсказанных алгоритмом, с индексами, содержащимися в идеальной разметке.

* FFmpeg Issue Tracker – <https://trac.ffmpeg.org/ticket/442>



Рис. 1. Пример видеоданных. Первый ряд – металлизированная ткань; Второй ряд – здоровая ткань.

Данный протокол не подходит для проблемы ассоциации сцен, так как принадлежность кадра определенной сцене не следует из множества границ, обнаруженных в видеопотоке. Например, сцена часто состоит из нескольких непоследовательных фрагментов видеопотока, что отражено на рис. 2. Такое разбиение на сцены следует из того, что специалист при проведении исследования перемещает эндоскоп по сложной траектории, многократно возвращаясь к уже знакомым участкам ткани.

Вкладом статьи являются:

- метод ассоциации сцен, основанный на глобальном анализе изображения и построении системы соседства визуально схожих кадров
- протокол тестирования решений для ассоциации сцен

2. Обзор предшествующих результатов

Методы, решающие проблему обнаружения границ сцен в видеопотоке сводятся к отысканию критерия, который обеспечит устойчивую двухклассовую классификацию (см. опр. 1). Обрабатываемые кадры видеопотока подвергаются операции снижения размерности (часто – построение RGB-гистограммы). Эти компактные представления информации о кадре служат входом для критерия, в роли которого может выступать метод Отцу [3] или иной метод вычисления адаптивного порога [6].

В [2] решение принимается с помощью ансамбля таких критериев, оперирующих сегментами изображения, используя инфраструктуру и модель обработки h264 кодировщика. Результаты, полученные в [4] с использованием более сложного признакового вектора и SVM в качестве критерия демонстрируют высокие результаты на TRECVID2006. Однако авторы отмечают, что подход плохо масштабируется на данные, в которых мгновенная граница сцен отсутствует, уступая место плавному переходу между сценами в течении нескольких кадров (alpha-blending, transition, cross-fade, zoom in/out).

В эндоскопических видео переход наделен теми же сложностями, которые можно встретить в задачах общего обнаружения границ сцен:

- изменение яркости и контрастности
- размытие, вызванное движением

- появление бестекстурных участков на большей части изображения

Кроме того, добавляются проблемы, специфические для приложения:

- малая вариативность и слабая уникальность сцен (см. рис. 1)
- нарушение предположения о твердотельности мира: ткани неизменно изменяются в моменты перистальтики
- перекрытие объектива потоками жидкости, инструментами хирурга

В [1] авторы обращаются к каждой из этих проблем, и предлагают решение, основанное на синергии трекера по методу оптического потока, каскадного визуального классификатора по типу Виолы-Джонса и графической вероятностной модели, использующей геометрические свойства ключевых точек для удаления статистических выбросов, пропущенных классификатором. В данной модели создание новой сцены возможно только после срыва трекера.

3. Ассоциация сцен

Настоящая работа предлагает альтернативный подход к проблеме, основанный на глобальном анализе кадра. Отказ от обработки ключевых точек и их отслеживания вызван характером данных: многие участки видео лишены сильно текстурированных и контрастных элементов, оставаясь в то же время узнаваемыми. Подход, успешно применяемый в смежных областях (пример: биометрия [7]) основан именно на построении статистической модели, описывающей общий облик кадра.

Постановка задачи Дадим формальное определение задачи ассоциации сцен в статическом сценарии использования: анализ видеопотока производится после завершения операции. Входом алгоритма является последовательность RGB кадров F_1, \dots, F_n эндоскопического видеопотока. Целью выступает объединение в группы кадров, являющихся изображениями одного и того же физического участка ткани пациента. Примеры кадров видеопотока на рис. 1, примеры сцен на фрагменте видеопотока – на рис. 2.

Определение 3. Выход алгоритма Последовательность 3-кортежей длины n^2 вида $\{(i, j, c_{ij}^j)\}_{i,j=1}^{n,n}$, где c_{ij}^j – мера уверенности в том, что кадры i и j принадлежат одной сцене, $c_{ij}^j \in [0 \dots 1]$, n – количество кадров в видеопотоке.

Опишем основные шаги работы предложенного алгоритма:

1. Обучение промежуточных представлений
2. Индексация входного видеопотока
3. Парное сопоставление кадров входного видеопотока полным перебором (см. опр. 3)

Обучение выполняется единожды и не относится к процедурам статистического вывода, выполняемым

алгоритмом в штатном режиме работы. Оно вынесено в отдельный пункт в силу того, что ряд операций, введенных для обучения, будет переиспользоваться при выводе. Таким образом, предложенный алгоритм является двупроходным оффлайн-решением для ассоциации сцен (пункты 2 и 3).

4. Обучение промежуточных представлений

В предложенном методе в качестве средства снижения размерности входных данных выступает распространенный метод - сумка слов (BoW), где визуальным «словом» является дескриптор ключевой точки на изображении. Тогда промежуточным представлением, требующем обучения (без учителя) является словарь сумки слов – совокупность самых характерных дескрипторов (центроидов), встречающихся на ключевых точках обучающей выборки. Используемый в решении словарь обучен на постороннем наборе данных общего назначения Oxford 102 flowers dataset [8].

Алгоритм 1. Построение словаря сумки слов

Вход: $F_{1\dots m}$

Выход: v (словарь)

- 1: $D := \{\emptyset\}$
 - 2: для $i \in \{1 \dots m\}$
 - 3: плотное вычисление ключевых точек на F_i
 - 4: вычисление дескрипторов полученных точек d_i
 - 5: сохранение полученных дескрипторов:
 $D := D \cup d_i$
 - 6: обучение словаря методом k -средних
 $v := \text{kmeans_train}(D, k)$
-

Процесс обучения описан в алгоритме 1.. Данная процедура осуществляют плотную и равномерную выборку ключевых точек на изображении и вычисление дескрипторов их окрестности [9]. Далее, множество дескрипторов изображений обучающей выборки использовано для их кластеризации по методу k -средних.

Тогда, множество дескрипторов нового входного изображения и словарь из k центроидов можно использовать для построения гистограммы на носителе $[1 \dots k]$, которая описывает изображение в терминах частоты встречаемости характерных ключевых точек k классов. Переход от изображения к гистограмме использован в шаге 2 алгоритма 2. и в шаге 3. алгоритма 3.. Вычисление этого перехода идентично шагам 2 - 3 алгоритма 1..

5. Индексация входного видеопотока

Первый этап обработки входного видеофайла заключается в его индексации, а именно построении системы соседства визуально схожих кадров. Мерой визуальной схожести выступает расстояние между гистограммами. Последовательно вычислив гистограммы всех изображений входного видеофайла, можно выстроить систему соседства по методу k -

ближайших соседей (в его постановке без учителя). Эта процедура описана в алгоритме 2..

Алгоритм 2. Индексация видеофайла

Вход: $F_{1\dots m}, v$

Выход: n (система соседства)

- 1: $H := \{\emptyset\}$
 - 2: для $i \in \{1 \dots m\}$
 - 3: вычисление гистограммы:
 $h_i := \text{hist}(v, F_i)$
 - 4: сохранение гистограммы:
 $H := H \cup h_i$
 - 5: обучение knn классификатора
 $n := \text{knn_train}(H, k)$
-

6. Парное сопоставление кадров видеопотока

Этап работы алгоритма, следующий за индексацией, заключается в полном парном сравнении кадров и построении списка этих пар с соответствующими мерами (см. опр. 1). Алгоритм 3. описывает вычисления, проводимые для отдельно взятой пары (F_i, F_j) , в результате которых получается c_i^j . Процедуру сопоставления можно разделить на два шага:

- определение, является ли пара (F_i, F_j) кадрами-кандидатами
- вычисление c_i^j , меры уверенности в принадлежности кадров i, j общей сцене

Первым шагом вычисляются k -ближайших соседей кадра F_i , используя словарь сумки слов и построенную прежде систему соседства. Процедура фильтрации полученных соседей удаляет тех, для которых расстояние на гистограмме до F_i больше порога. Если F_j находится в отфильтрованном множестве соседей F_i , считаем их кадрами-кандидатами и переходим к вычислению меры уверенности (см. строку 4 алгоритма 3.).

Сопоставление разделено на два шага с целью оптимизации времени работы и повышения точности. Поэтому вычислительноемкая процедура получения c_i^j вызывается только для кадров-кандидатов, а не для любой пары. Этот факт важен ввиду того, что ручная разметка видеоданных показала, что большинство кадров не представляют интереса (см. рис. 2). Фильтрация кадров и многократное сокращение числа обращений к процедуре score понижает количество ложноположительных срабатываний.

Критерии вычисления мер уверенности В работе предложен нейросетевой критерий вычисления мер уверенности в принадлежности пары кадров эндоскопического видео одному участку ткани. Для сравнения, в качестве базового алгоритма, применен алгоритм на основе проективного преобразования плоскости [11]. Для базового алгоритма, далее именуемого гомографическим, мерой является доля ключевых точек, согласующихся с проектив-

Алгоритм 3. Построение меры уверенности в принадлежности кадров сцене

Вход: F_i, F_j, v, n

Выход: c ;

- 1: вычисление гистограммы
 $h_i := \text{hist}(v, F_i)$
- 2: вычисление k -ближайших соседей для кадра F_i
 $\text{neighborhood}_i := \text{knn}(n, h_i)$
- 3: для $m \in \text{neighborhood}_i$: $\text{dst}(h_m, h_i) > t_h$
 $\text{neighborhood}_i := \text{neighborhood}_i \setminus \{m\}$
- 4: если $j \in \text{neighborhood}_i$ то
 выход $c := \text{score}(F_i, F_j)$
- 5: иначе
 выход $c := 0$

ным преобразованием, вычисленным в ходе процедуры RANSAC.

Нейросетевой критерий использует архитектуру сети Крыжевского [14]. Входом сети является разница между кадрами F_i и F_j . Выходной слой которой производит двухклассовую классификацию, где $0 : (F_i, F_j) \notin S$, $1 : (F_i, F_j) \in S$. Тогда 1-ая компонента вектора весов может быть использована как мера уверенности c_i^j . По причине ограниченности доступного набора данных, обучение сети с нуля для решения задачи не представлялось возможным. Был использован распространенный подход дообучения сети, описанный в[†]. Процесс обучения сошелся менее чем за 20000 эпох на обучающей выборке из 15000 изображений. Качественный и количественный анализ вышеописанных критериев приведен в последующих разделах.

7. Эксперименты и результаты

Данные для проведения экспериментов предоставлены Имперским колледжем Лондона [1] и являются записями реальных эндоскопических операций. Набор данных содержит 6 видео, от 500 до 1400 кадров в каждом. Разметка истинной принадлежности кадров сценам была выполнена не экспертом. Следующие параграфы описывают проведенные эксперименты и их отличия от протокола TRACVID.

Протокол тестирования Структура, являющаяся выходом алгоритма (см. опр. 3), выступает в роли входа для процедуры тестирования. Каждая задача информационного поиска зависит от компромисса между требованиями к точности результата и к полноте выдачи. Введем эти определения.

Определение 4. Точность – доля верно ассоциированных кадров среди тех, которым алгоритм присвоил сцену. Формально: $P = \frac{TP}{TP+FP}$, где TP – число верно ассоциированных кадров, FP – число кадров, отнесенных к сцене, которой они не принадлежат.

[†]Fine-tuning CaffeNet for Style Recognition on Flickr Style Data – http://caffe.berkeleyvision.org/gathered/examples/finetune_flickr_style.html

Определение 5. Полнота – доля верно ассоциированных кадров среди тех, которым присвоена сцена на истинной разметке. Формально: $R = \frac{TP}{TP+FN}$, где TP – число верно ассоциированных кадров, FN – число кадров, имеющих присвоенную им сцену в истинной разметке, но отброшенных алгоритмом.

Для выбранной проблемы пользователи заинтересованы в максимальных значениях полноты при фиксированной точности, близкой к 100% [1]. Однако, было отмечено, что разметка проведена не экспертами, поэтому результату со 100% показателем точности, вычисленному на неверной разметке, доверять нельзя. В силу этого оценка алгоритма производится путем построения кривых точности/полноты.

Суть предложенного протокола тестирования заключается в том, что ни в ручной разметке, ни в выходных данных алгоритма сцены не определены явно. При этом, тест позволяет сказать для каждого кадра видеопотока был ли он отнесен алгоритмом к той же сцене, что и в ручной разметке. Такая постановка желательна в силу:

- упрощения работы по разметке видеопотока
- возможности строить более простые модели

Процедура тестирования обрабатывает две выходные последовательности (см. опр. 3) для получения кривых точности/полноты. Первая выдана алгоритмом, вторая – идеальная ручная разметка. При этом, ручная разметка будет содержать только меры уверенности $c_i^j = 1$. Далее описан метод вывода множества сцен из выходной последовательности. Ей в соответствие можно поставить следующий объект: ненаправленный граф с индексами кадров в вершинах и мерами уверенности c_i^j на ребрах. В графе удаляются все ребра, меньшие текущего порога на кривой. Тогда любая пара кадров, имеющая путь в графе, считается принадлежащей одной сцене. Множество сцен видеопотока – список непесекающихся множеств индексов кадров, получаемый из графа по [12].

Так восстановлены два множества сцен, идеальных и выведенных алгоритмом. Счетчик положительных срабатываний инкрементируется покадрово, если пересечение между сценами, к которым кадр отнесен в алгоритмической и ручной разметке, больше порога. Подсчет ложноположительных и ложноотрицательных срабатываний производится аналогично.

Исследователи, решающие задачи ассоциации, нуждаются в средствах качественной оценки результата на данном видеофайле. Предложено средство для сверки алгоритмически восстановленных и истинных сцен, изображающее на графике последовательные ассоциированные кадры в виде столбцов соответствующего цвета. Данная визуализация приведена на рис. 2. Подход вдохновлен решением [13] другой проблемы сопоставления визуальных данных.

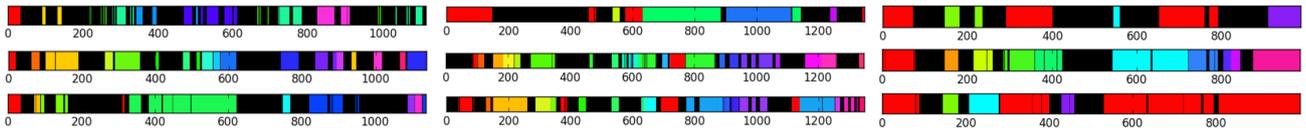


Рис. 2. Сравнение автоматической ассоциации сцен с ручной разметкой. Столбцы: результаты для 2, 4, 10 видеофайлов из набора данных; Строки: результаты разметки: идеальной, гомографическим методом, нейросетевым методом. Цветные блоки – обнаруженные сцены, черные блоки – кадры, не отнесенные ни к одной из сцен.

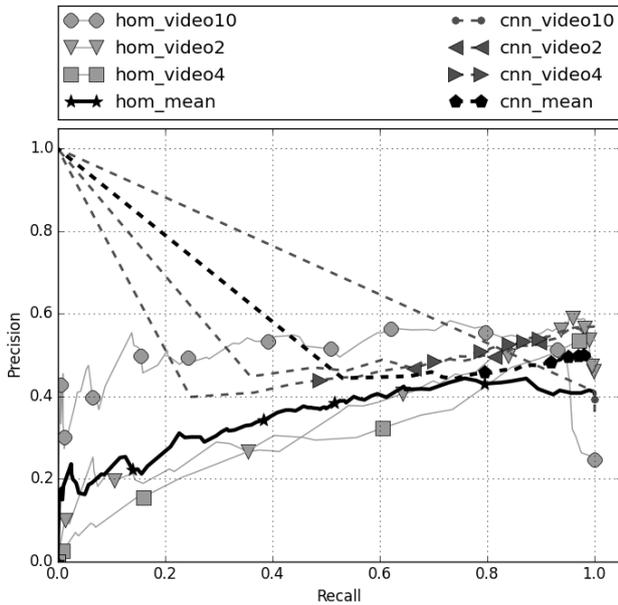


Рис. 3. Кривые точности и полноты. Кривые с префиксами *hom* и *cnn* – результаты для гомографического и нейросетевого критериев на соответствующих видеофайлах. Кривые с суффиксом *mean* – усредненный по набору тестируемых данных результат.

Эксперименты Ввиду недоступности авторских реализаций алгоритма [1], предложенный метод протестирован в сравнении с решением, обеспечивающим базовый результат. Их показатели производительности, усредненные по набору данных, и точечные для трех видеофайлов приведены на рис. 3.

Траектории изображают изменение параметров точности и полноты в зависимости от выбранного значения порога, применяемого к c_i^j . На рис. 3 значение порога изменяется слева направо от 1 до 0. Закономерно, в правой части графика прямые сходятся к примерно одним и тем же значениям для разных критериев, так как классификация перестает работать (порог стремится к нулю) и решения об ассоциации начинают в большей степени зависеть от предыдущей фазы алгоритма – выбора кадров-кандидатов. Как было упомянуто ранее, с практической точки зрения более интересны результаты из области графика, соответствующей $[0 \dots 0.5]$ по оси абсцисс, предпочитающие большую точность при

меньшей полноте. Можно заметить, что в этой области преимущество нейросетевого критерия над гомографическим видно явно, хотя первый и превосходит второй на любом значении порога. Результаты замеров точности и полноты коррелируют с качественной оценкой, приведенной на рис. 2. На рис. 2 можно пронаблюдать, что нейросетевым критерий чаще чем базовый метод отнесит кадр к верной сцене. Это особенно характерно для случаев, когда эндоскоп вернулся к исходному участку ткани спустя некоторое время (кадры отнесены к той же сцене, которую пронаблюдали несколькими сотнями кадров ранее). Недостатком базового метода является недостаточно выразительный критерий, который допускает больше ложноположительных срабатываний при больших значениях порога. Это приводит к падению точности одновременно с падением полноты на интервале $[0 \dots 0.4]$. Средняя точность ассоциации при средней полноте 90% равна 57% для предложенного алгоритма. В [1] докладывают большую точность, однако, важно отметить некорректность прямого сравнения этих результатов в силу использования разных протоколов тестирования. Решение [1] использует концепцию времени и априорные предположения о длительности и характере смены сцен, тем самым снижая размерность информационного поиска и делая разбиения на сцены, подобные изображенным на рис. 2 невозможными.

8. Выводы

Предложенный метод ассоциации сцен для эндоскопических видео позволяет выделить ключевые сцены с помощью построения системы соседства похожих кадров и глобального анализа похожих кадров путем применения нейросетевого статистического критерия. Работа формализует и описывает протокол тестирования алгоритма. Приведены результаты сравнения производительности базового и предложенного алгоритмов на наборе данных реальных эндоскопических операций.

Программная реализация алгоритмов ассоциации сцен и средств тестирования и визуализации открыта для использования исследователями[‡].

[‡]https://github.com/wf34/scene_ass

9. Благодарности

Работа выполнена при финансовой поддержке Фонда Содействия Инновациям, грант 10092ГУ2/2015.

10. Литература

- [1] *Ye M., Johns E., Giannarou S., Yang G.* Online Scene Association for Endoscopic Navigation //International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI, 2014
- [2] *Dimou A., Nemethova O., Rupp M.* Scene change detection for H. 264 using dynamic threshold techniques //Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Service 2005
- [3] *Sahoo P. K., Soltani S., Wong A. K.C., Chen Y.C.* A survey of thresholding techniques //Computer vision, graphics, and image processing. – 1988. – Т. 41. – №. 2. – С. 233-260.
- [4] *Chávez G. C., Cord M., Philipp-Foliguet S., Precioso F., Araujo A.* Robust scene cut detection by supervised learning //Signal Processing Conference, 2006 14th European. – IEEE, 2006. – С. 1-5.
- [5] *Smeaton A. F., Over P., Kraaij W.* Evaluation campaigns and TRECVID //MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval //Proceedings of the 8th ACM international workshop on Multimedia information retrieval. – ACM, 2006. – С. 321-330.
- [6] *Kovalenko D.A., Potapev I.A.* Increasing Performance of the Scene Boundary Detection System with use of OpenCL // XIX International Conference Modern Technique and Technologies, 2013 – Vol.2 – С.426-428. portal.tpu.ru/files/conferences/ctt/proceedings/ctt-2013-2-Tom.pdf.
- [7] *Bharadwaj S., Bhatt H., Vatsa M., Singh R.* Periocular biometrics: When iris recognition fails //Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. – IEEE, 2010. – С. 1-6.
- [8] *Nilsback M. E., Zisserman A.* Automated flower classification over a large number of classes //Computer Vision, Graphics Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on. – IEEE, 2008. – С. 722-729.
- [9] *Tola E., Lepetit V., Fua P.* Daisy: An efficient dense descriptor applied to wide-baseline stereo //IEEE transactions on pattern analysis and machine intelligence. – 2010. – Т. 32. – №. 5. – С. 815-830.
- [10] *Lowe D. G.* Object recognition from local scale-invariant features //Computer vision, 1999. The proceedings of the seventh IEEE international conference on. – Ieee, 1999. – Т. 2. – С. 1150-1157.
- [11] *Hartley R., Zisserman A.* Multiple view geometry in computer vision – Cambridge university press, 2003. – С. 325-340
- [12] *Tarjan R.* Depth-first search and linear graph algorithms //SIAM journal on computing. – 1972. – Т. 1. – №. 2. – С. 146-160.
- [13] *Zhu Y. et al.* Aligning books and movies: Towards story-like visual explanations by watching movies and reading books //Proceedings of the IEEE International Conference on Computer Vision. – 2015. – С. 19-27.
- [14] *Krizhevsky A., Sutskever I., Hinton G. E.* Imagenet classification with deep convolutional neural networks //Advances in neural information processing systems. – 2012. – С. 1097-1105.

Точный автоматический алгоритм детектирования цефалометрических точек на КТ-изображениях

Д.А. Лачинов¹, А.А. Белокаменская¹, В.Е. Турлапов¹
 dlachinov@gmail.com|alexandra.belokamenskaya@gmail.com|vadim.turlapov@gmail.com
¹ННГУ им. Н.И. Лобачевского, г. Нижний Новгород, Российская Федерация

Предложен алгоритм для автоматического детектирования цефалометрических точек (ЦТ) по 3D-данным томограммы пациента, достигающий существенно большей точности по сравнению с публикациями. Алгоритм ищет отображение множества точек черепа-шаблона в множество точек КТ-томограммы с использованием метода Coherent Point Drift (CPD). Череп-шаблон размечен на анатомические зоны, включающие базис геометрического дескриптора ЦТ. В результате совмещения мы получаем соответствие областей томограммы анатомическим зонам. Далее, в каждой найденной зоне происходит поиск цефалометрической точки по ее геометрическому дескриптору. Мерой качества предложенного алгоритма является Евклидово расстояние между размеченными вручную и найденными автоматически точками на нескольких томограммах. Алгоритм проверен количественно на ЦТ орбиталей для 5 томограмм с деформациями черепа. В результате, 87% найденных точек лежат в радиусе 2мм от искомым точек и 100% в радиусе 2.1мм. Ключевой особенностью и источником точности подхода является использование CPD для персонализации структуры шаблона и линейной регрессии с регуляризацией Тихонова для уточнения базиса дескриптора при позиционировании цефалометрической точки.

Ключевые слова: цефалометрия, ключевые точки, 3D, автоматическое детектирование, компьютерная томография.

Precise automatic cephalometric landmark detection algorithm for CT images

D. Lachinov¹, A. Belokamenskaya¹, V. Turlapov¹
 dlachinov@gmail.com|alexandra.belokamenskaya@gmail.com|vadim.turlapov@gmail.com
¹Lobachevsky State University (UNN), Nizhni Novgorod, Russia

New approach for automatic cephalometric landmark detection is proposed. Described algorithm projects labeled template onto examined CT Image using Coherent Point Drift (CPD), thereby Region of Interest (ROI) for each landmark is determined. In each ROI special target function is constructed according to the description of corresponding landmark and then optimized with restrictions defined by ROI-box. The key features of this approach are template personalization using CPD and local coordinate system refinement using ridge regression during landmark detection. The Euclidian distance between detected and manually identified landmarks is used as a measure of quality for series of images. The result is 87% of automatically detected landmarks lie within 2mm interval and 100% - within 2.1mm interval, what is pretty accurate result. This approach achieves great precision compared to previously proposed methods. Such performance is achieved by precise template personalization and local coordinate system refinement for each ROI.

Keywords: cephalometry, landmarks, automatic detection, computer tomography, point set registration.

1. Введение

Компьютерная томография (КТ, 1972) является сегодня неотъемлемым инструментом для оценки и планирования лечения пациентов с тяжелым челюстно-лицевыми заболеваниями и доступна практически во всем мире. Челюстно-лицевая хирургия с использованием КТ становится все более популярной. Но необходимое для этого точное позиционирование ключевых точек черепа (цефалометрических точек) на слоях томограммы выполняется до сих пор вручную, требует много времени, значительных усилий и опыта. Это, в свою очередь, ставит задачу разработки автоматизированной процедуры разметки.

Достижения в области компьютерных технологий позволили с достаточной точностью обнаруживать цефалометрические точки на рентгенограммах. Алгоритмы, разработанные для КТ изображений, не достигают такой же точности и, зачастую, обладают, как и 2D методы, рядом недостатков, затрудняющих их применение на практике.

Целью данной публикации является разработка метода автоматического детектирования ключевых точек черепа с высокой точностью.

2. Предметная область

Цефалометрический анализ представляет из себя измерение углов, расстояний и отношений между различными морфологическими структурами черепа и целиком базируется на поиске цефалометрических точек. Разметка цефалометрических точек до сих пор является самой трудоемкой и требовательной к квалификации задачей, как в ручном, так и в автоматическом варианте. Для автоматической регистрации цефалометрические точки определяются при помощи их формального признакового описания (дескриптора), исходя из применяемого метода анализа [3,6] (в нашем случае цефалометрические точки определены геометрическими условиями). Основные лицевые цефалометрические точки представлены в таблице 1.

Таблица 1. Описание основных точек [3,6].

Название	Сокр.	Описание (Дескриптор)
Orbitale left	O_{r_l}	самая нижняя точка на костном крае левой орбиты
Orbitale right	O_{r_r}	самая нижняя точка на костном крае правой орбиты
Nasion	N	соединение лобной и носовой кости
Porion left	P_l	самая верхняя точка левого наружного слухового прохода
Porion right	P_r	самая верхняя точка правого наружного

		слухового прохода
Menton	Me	самая нижняя точка на симфизе нижней челюсти
Pronasale	Pn	наиболее выдающаяся часть носа

3. Предложенный метод

На примере точек орбиталей O_{r_l} и O_{r_r} , предлагается подход к автоматическому определению основных цефалометрических точек, позволяющий заметно повысить точность их локализации по сравнению с данными публикаций. Подход может быть применен для любых других цефалометрических точек. Его ключевой особенностью является применение нелинейной и нежесткой процедуры для задачи совмещения точек шаблона, принятого за эталон и содержащего базис геометрического дескриптора, с особыми точками, экстрагированными из томограммы пациента. После совмещения мы можем считать базис дескриптора позиционированным, в первом приближении, на множестве особых точек пациента. Это совмещающее преобразование ищется при помощи метода Coherent Point Drift (CPD) [4]. Точки на томограмме пациента, как и точки исходного шаблона, могут быть получены из данных томограммы при помощи любого детектора, в частности, в работе использован детектор Determinant of Hessian (DoH) [8]. После отображения шаблона на томограмму пациента становятся таким образом определенными базисы всех дескрипторов (в виде систем координат или точечных базисов). Шаблон, с особыми точками, с размеченными базисами и дескрипторами цефалометрических точек, готовится для алгоритма заранее с участием специалиста-морфолога. Далее алгоритм включает следующие шаги:

- 1) детектирование множества особых точек томограммы пациента;
- 2) совмещение множества точек размеченного шаблона с множеством особых точек томограммы пациента, определяющее первое приближение базиса дескриптора;
- 3) уточнение базиса дескриптора на множестве особых точек области его определения (опционально);
- 4) поиск цефалометрической точки внутри каждого базиса.

3.1 Детектирование особых точек

Для детектирования особых точек в работе использовался метод Determinant of Hessian (DoH) [8].

Метод DoH относится к методам нахождения областей изображения, которые отличаются от других областей в рамках некоторого показателя, например, интенсивности. Причем этот показатель внутри самого региона (блоба) практически неизменен.

Исходное изображение $f(x, y, z)$ подвергается сглаживанию при помощи фильтра Гаусса $g(x, y, z, t)$ при разных параметрах t фильтра – масштабах.

$$L(x, y, z, t) = g(x, y, z, t) * f(x, y, z),$$

где * - операция свертки.

При каждом масштабе t вычисляется определитель матрицы Гессе $\det[H_{norm}(L)] = t^2(L_{xx}L_{yy} - L_{xy}^2)$. Точка изображения, имеющая координаты $(\hat{x}, \hat{y}, \hat{z})$, считается принадлежащей блобу, если величина определителя матрицы Гессе в этой точке является локальным максимумом:

$$(\hat{x}, \hat{y}, \hat{z}) = \operatorname{argmax}_{(x,y,z)} \det[H_{norm}(L(x, y, z, t))].$$

Для отсеечения заведомо непригодных точек используется заранее заданный порог (в работе использовался порог равный 200). Найденные точки, в

основном, располагаются на углах, изломах, неоднородностях. Недостатком данного детектора является чувствительность к шуму. Этот алгоритм без существенных изменений обобщается на трехмерный случай. Пример применения алгоритма предоставлен на рисунке (Рис. 1).

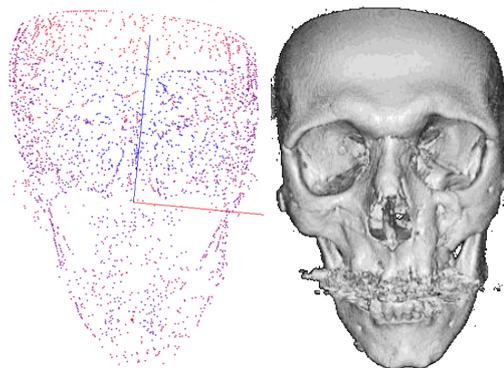


Рис. 1. Пример особых точек (слева) для костей черепа (справа), реконструированных по данным КТ пациента. Цвет точек задает глубину и изменяется от красного (ближние) до синего (дальние).

Точки шаблона извлекаются из томограммы, принятой за шаблон, таким же методом. Некоторым точкам шаблона вручную присваиваются метки (таблица 2).

Таблица 2. Описание меток шаблона.

Метка	Описание
0	Точки, не представляющие интереса, метка назначается автоматически
1	Точки, соответствующие правой скуловой кости
2	Точки, соответствующие левой скуловой кости
3	Цефалометрическая точка O_{r_l} шаблона
4	Точка на левом краю левой орбиты, задает точечный базис
5	Точка на правом краю левой орбиты, задает точечный базис
6	Цефалометрическая точка O_{r_r} шаблона
7	Точка на правом краю правой орбиты, задает точечный базис
8	Точка на левом краю правой орбиты, задает точечный базис
9	Точки, лежащие на верхнем краю левой орбиты, задают точечный базис
10	Точки, лежащие на верхнем краю правой орбиты, задают точечный базис
11	Точки, лежащие на поверхности левой орбиты, используются для регрессии
12	Точки, лежащие на поверхности правой орбиты, используются для регрессии

3.2 Совмещение шаблона и точек пациента

Преобразование, совмещающее множество точек размеченного шаблона с множеством особых точек КТ пациента осуществляется, как упоминалось, при помощи метода Coherent Point Drift (CPD), описанного в работе [4].

CPD основан на алгоритме EM (Expectation Maximization) и является вероятностным итерационным алгоритмом. Множество точек шаблона представляется как множество центроидов распределенных по нормальному закону $N(\alpha_i, \sigma^2)$ точек, другое множество – особых точек томограммы пациента. Центры Гауссиан подстраиваются под данные пациента, тем самым максимизируя функцию правдоподобия. При достижении оптимума множества точек выравниваются, соответствия между множествами вычисляется при помощи апостериорных вероятностей.

Для этого минимизируется функция

$$E(W) = - \sum_{n=1}^N \log \sum_{m=1}^M e^{-\frac{1}{2\sigma^2} \|x_n - y_{0m} - \sum_{k=1}^M w_k G(y_{0k} - y_{0m})\|^2} + \frac{\lambda}{2} \text{tr}(W^T G W),$$

где $Y = G W + Y_0$ - обновляется на каждом шаге,

G - матрица Грамма с элементами $g_{ij} = e^{-\frac{1}{2\beta^2} \|y_{0i} - y_{0j}\|^2}$,

$W = (w_1, w_2, \dots, w_m)^T$ - матрица весов.

Верхняя оценка функции $E(W)$ определяется как

$$Q = \sum_{n=1}^N \sum_{m=1}^M p_{old}(m|x_n) \frac{\|x_n - y_{0m} - G(m, \cdot) W\|^2}{2\sigma^2} + \frac{\lambda}{2} \text{tr}(W^T G W),$$

где p_{old} - апостериорные вероятности.

Минимизируя верхнюю оценку Q получаем систему уравнений относительно W :

$$(\text{diag}(P \cdot 1)G + \lambda \sigma^2 I)W = P X - \text{diag}(P \cdot 1)Y_0, \quad (1)$$

где P - матрица апостериорных вероятностей

$$p_{mn} = \frac{e^{-\frac{1}{2\sigma^2} \|y_{0m}^{old} - x_n\|^2}}{\sum_{m=1}^M e^{-\frac{1}{2\sigma^2} \|y_{0m}^{old} - x_n\|^2}}, \quad (2)$$

1 - вектор из единиц длины N ,

y^{old} - результаты предыдущей итерации.

На каждой итерации происходит вычисление матрицы апостериорных вероятностей P по формуле (2) и решается система уравнений (1), относительно W . На каждой итерации Y , матрица точек, обновляется как $Y = Y_0 + G W$. Пример итеративного процесса из двух режимов, жесткого (rigid) и нежесткого (non-rigid), как в нашем случае, только двумерный, показан на рисунке (Рис. 2).

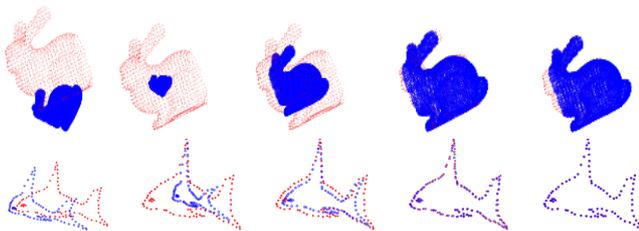


Рис. 2. Пример из [4] rigid (верхний) и non-rigid (нижний) совмещения множества Y (синим цветом) и множества X (красным цветом). Показаны итерации 0, 10, 20, 30, 40.

На этом основании можно считать, что метод CPD [4] является адаптивным, и к аффинным преобразованиям, и к нелинейным деформациям. Пример применения алгоритма показан на рисунке (Рис. 3). При выполнении расчетов использовались значения параметров $\beta = 2, \lambda = 3$.

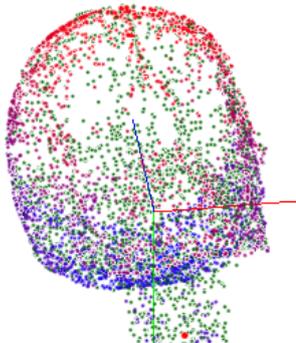


Рис. 3. Пример совмещения (вид черепа сзади) точек шаблона (красно-синие) с точками КТ-томограммы (зеленые).

3.3 Перенос разметки шаблона

Соответствие между размеченными точками шаблона и точками на КТ пациента находится при помощи матрицы P

апостериорных вероятностей. Каждой точке из множества X (точки томограммы пациента) ставится в соответствие точка из Y (точки шаблона) $X[i] \sim Y[\text{argmax}_j P(i, j)]$, где P задана формулой (2). Интересующие нас точки, такие как: выступающие точки левой и правой скуловых костей, положения других цефалометрических точек, вручную включены в шаблон. С использованием матрицы апостериорных вероятностей указанные цефалометрические точки переносятся на множество точек КТ пациента (Рис. 4).

После того, как соответствие между точками шаблона и КТ-изображения найдены, определяются области поиска и в каждой области вводятся локальные системы координат для уточнения каждой цефалометрической точки по дескриптору.

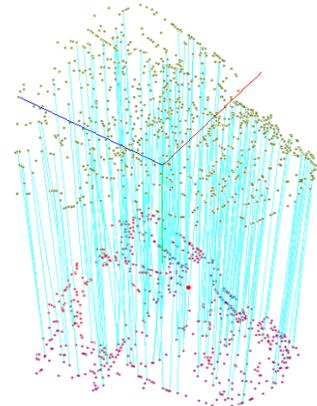


Рис.4. Пример найденных соответствий между точками шаблона (сверху) и точками томограммы (снизу); изображена лицевая часть черепа. Соответствия обозначены линиями, показано каждое десятое соответствие.

3.4 Базис дескриптора и его уточнение

Рассмотрим далее орбитальные цефалометрические точки под названием Or_l и Or_r . Опишем два подхода к применению базиса дескриптора (системы локальных координат или точечного базиса) в области определения дескриптора. Эта область определяется на шаблоне вместе с дескриптором как 3D-бюкс, который наполнен особыми точками, сопутствующими цефалометрической точке.

Для орбитальных цефалометрических точек в качестве горизонтальной оси локальной системы используется вектор \vec{h} между наиболее выступающими точками левой и правой скуловых костей, которые выделены на предыдущем шаге.

При первом подходе, без применения процедуры уточнения, для построения локальной системы координат дескриптора используется базис, непосредственно построенный на совмещенных точках шаблона. Плоскость орбиты задается при помощи трех точек - точки-центра региона поиска, левой точки орбиты, правой точки орбиты. Ось OX задается как проекция вектора \vec{h} на эту плоскость орбиты. Ось OY - нормаль к плоскости. Ось OZ - дополнение до правой тройки координат.

При втором подходе выполняется уточнение положения вертикальной плоскости глазницы. Она формируется при помощи линейной регрессии с регуляризацией Тихонова [1] (ridge regression). Используются множества точек томограммы пациента, попавшие после совмещения шаблона в области определения соответственно левой и правой орбит $\{x_l^{(i)}\}, \{x_r^{(j)}\}$. Решая систему

$$(X^T X + \lambda I)\beta = X^T y \quad (3)$$

находим коэффициенты регрессии β , где X - матрица $N \times 3$, в которой записаны по строкам две компоненты каждой точки одной из орбит, а последний столбец заполнен

единицами; λ - коэффициент регуляризации (при расчетах полагался равным 0.01); y - вектор третьих компонент точек соответствующей орбиты.

Решив систему линейных уравнений можно составить вектор нормали к искомой плоскости как $\vec{n}^T = (\beta_0; \beta_1; -1)$. Проецируя вектор \vec{h} на найденную плоскость находим вектор \vec{X} . Вектор \vec{Y} будет равен вектору \vec{n} . Вектор \vec{Z} является дополнением до правой тройки координат.

3.5 Поиск точек Or_l и Or_r

После построения локальной системы координат, внутри каждой области определения строится поверхность

$$S(x, y) = I(\operatorname{argmin}_z |I(x, y, z) - 200|),$$

где 200 - минимальная плотность кости по шкале Хаунсфилда. Для более точного построения поверхности используется линейная интерполяция сетки данных.

Далее задается кривая $C(x) = \max_y S(x, y)$. Искомая точка соответствует седловой точке поверхности и находится при достижении $\min_x C(x)$. Результат работы алгоритма - искомая цефалометрическая точка в анатомически выровненной системе координат (Рис. 5).

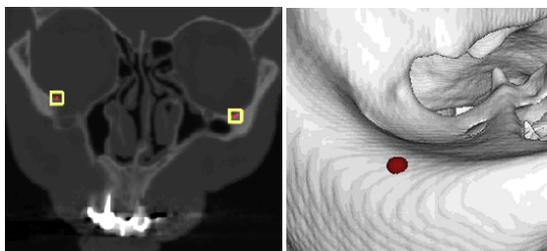


Рис. 5. Найденные точки, спроецированные на коронарный срез деформированного черепа пациента (слева). Найденная точка Or_r на реальной поверхности орбитали (справа).

4. Результаты

Для оценки результата на пяти КТ изображениях вручную были отмечены цефалометрические точки Or_l и Or_r с использованием визуализации в проекция XY, XZ, YZ и трехмерного отображения. Одно из размеченных изображений использовано как шаблон. Ключевой особенностью данных является то, что на всех КТ томограммах были представлены пациенты с теми или иными повреждениями костей черепа.

В таблице ниже (Табл. 3) представлено сравнение между двумя предложенными вариантами использования точечного базиса системы координат в области определения дескриптора, приведены величины математического ожидания и дисперсии ошибки. В результате, 87,5% точек орбиталей для всех 5 томограмм лежало внутри радиуса 2мм и 100% внутри радиуса 2.1мм от точек, размеченных вручную. По критериям введенным в [9], еще для 2D случая, алгоритм считается достаточно точным, если ошибка не превышает 2мм, и приемлемым, если ошибка не превышает 4мм.

Таблица 3. Результаты экспериментов.

	Регрессия	Точечный базис	[5] 2011	[7] 2014	[2] 2015
Point	mean/std	mean/std	mean/max	mean/max	mean/std
Or_l	1.19/0.71	2.53/0.60	-	-	1.78/1.36
Or_r	1.14/0.40	1.13/0.61	-	-	2.37/2.23
mean	1.17/0.53	1.83/0.60	3.15/6.27	3.4/7.10	2.01/1.23

Отличительной чертой и источником повышения точности предложенного алгоритма являются то, что разметка и локальная система координат шаблона переносится на особые точки КТ-томограммы пациента

путем нежесткого совмещения этих двух множеств методом CPD. Тогда как, в работе [2] использован шаблон, заданный исходя из некоторых средних показателей расположения цефалометрических точек друг относительно друга, что вводит систематическую ошибку и ограничивает применимость метода в медицинской практике. В работе [5] использовались подходы машинного обучения, но используемый метод и обучающая выборка не обеспечивали инвариантность к повороту. В работе [7] при трансформациях не учитывались возможные деформации черепа.

5. Заключение

Предложенный подход способен определять две рассмотренные цефалометрические точки с существенно более высокой точностью, чем в [2], [5], [7]. Он применим ко всем цефалометрическим точкам, так как их дескрипторы, так или иначе, соответствуют экстремумам анатомических поверхностей, реконструируемых по особым точкам томограммы пациента. Согласно результатам эксперимента, алгоритм может быть использован на практике для пациентов с повреждениями костей черепа.

В настоящей работе использован шаблон с меньшим числом цефалометрических точек, чем в работе [2], но однозначно можно утверждать, что предложенный подход будет точнее, благодаря использованию CPD для персонализации шаблона и линейной регрессии с регуляризацией Тихонова, которая оказалось более точной, чем использование исходного варианта точечного базиса при позиционировании дескриптора.

6. Литература

- [1] А.Н. Тихонов "О некорректных задачах линейной алгебры и устойчивом методе их решения", ДАН СССР, 1965 Том 163, № 3.
- [2] Abhishek Gupta, Om Prakash Kharbanda, Viren Sardana, Rajiv Balachandran, Harish Kumar Sardana "A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images" Int J CARS 2015 Nov;10(11):1737-52
- [3] Alexander Jacobson, Richard L. Jacobson "Radiographic Cephalometry: From Basics to Videoimaging" Quintessence Pub., 1995г., page 53-63.
- [4] A. Myronenko, & X. Song (2010). "Point set registration: Coherent point drift". IEEE transactions on pattern analysis and machine intelligence, 32(12), 2262-2275.
- [5] Chu, Chengwen, et al. "Fully automatic cephalometric x-ray landmark detection using random forest regression and sparse shape composition." submitted to Automatic Cephalometric X-ray Landmark Detection Challenge (2014).
- [6] Gwen R.J. Swennen, Filip Schutyser, Jarg-Erich Hausamen "Three-Dimensional Cephalometry. A Color Atlas and Manual" Springer-Verlag Berlin Heidelberg 2006. p 116-185
- [7] Shahidi S, Bahrapour E, Soltanimehr E, Zamani A, Oshagh M, Moattari M, Mehdizadeh A "The accuracy of a designed software for automated localization of craniofacial landmarks on CBCT images". BMC Med Imaging 2014 14(1):1471-2342
- [8] Tony Lindeberg "Scale Selection Properties of Generalized Scale-Space Interest Point Detectors" J Math Imaging Vis 2013, Volume 46, Issue 2, p 177-210.
- [9] Yue W, Yin D, Li C, Wang G, Xu T. Automated 2-D cephalometric analysis on X-ray images by a model-based approach (2006). IEEE Trans Biomed Eng 53(8): 1615-1623.

Метод оценки ядра размытия движением изображений на основе анализа хребтовых структур

А.А. Мельникова, А.С. Крылов

alexandra.melnikova.cmc.msu@gmail.com|kryl@cs.msu.ru

Факультет вычислительной математики и кибернетики

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

Предложен алгоритм оценки ядра размытия для изображений с равномерным линейным размытием. Алгоритм использует метод выделения хребтовых структур и анализ гистограмм распределения направлений. Полученные точности оценки направления размытия и степени размытия позволяют использовать данный метод для практических задач.

Ключевые слова: размытие изображений, размытие движением, оценка ядра размытия, выделение хребтовых структур, гистограммный метод.

Image motion blur kernel estimation using ridge analysis

A.A. Melnikova, A.S. Krylov

alexandra.melnikova.cmc.msu@gmail.com|kryl@cs.msu.ru

Faculty of Computational Mathematics and Cybernetics

Lomonosov Moscow State University, Moscow, Russia

An algorithm for image blur kernel estimation for the case of motion blur has been proposed. The algorithm uses ridge detection method and direction distribution histogram analysis. The obtained accuracy of the blur direction and blur lengths estimation allow to use the proposed method in practice.

Keywords: image blur, motion blur, blur kernel estimation, ridge detection, histogram method.

1. Введение

Размытие движением на изображении возникает при направленном относительно движении сцены и/или сенсора во время экспозиции. Такой эффект затрудняет распознавание и восприятие изображений. Для применения алгоритмов по восстановлению размытых изображений требуется оценить ядро размытия. Таким образом, успех восстановления размытого изображения существенно зависит от точной оценки параметров размытия.

Для оценки параметров размытия используются различные инструменты. Так, в [1] предложен метод, основанный на спектральном анализе, а в [2] оценка ядра размытия осуществляется исходя из пространственных свойств размытия, таких как форма, однородность и гладкость. В [3] предлагается анализировать корреляцию между преобразованием Фурье изображения и периодической функцией детектирования, здесь же проводится сравнение с методами [1] и [2]. В [4] осуществляется спектральный анализ градиентов изображения для определения параметров размытия, а в [5] предлагается анализировать преобразование Хафа изображения. В [6] авторы предложили модификации для преобразования Радона, с помощью которых можно осуществить анализ спектра изображения. В [7] для определения параметров был разработан новый способ анализа спектра изображения.

Оценка параметров размытия движением довольно востребованная задача, решение которой может использоваться в различных областях. Так, в [8] был

предложен метод для определения скорости движущихся автомобилей, основанный на анализе изображений, содержащих размытие движением, и определении параметров размытия. Также, определение параметров размытия применяется для улучшения работы и точности систем наблюдения [9] и восстановления фотографий текстовых документов, подвергнутых размытию [10].

1.1. Постановка задачи.

В рамках данной работы рассматривается равномерное линейное размытие на всем исходном изображении. Размытое изображение может быть представлено в виде:

$$g(x, y) = f(x, y) * h(x, y),$$

где $h(x, y)$ - функция размытия или PSF, $f(x, y)$ - неразмытое изображение, $g(x, y)$ - входное (размытое) изображение.

Функцию точечного источника PSF (Point Spread Function), или ядро размытия, в случае равномерного линейного размытия можно определить с помощью следующей формулы [11]:

$$h(x, y) = \begin{cases} \frac{1}{L}, & \sqrt{x^2 + y^2} \leq \frac{L}{2}, y = x \tan(\theta) \\ 0, & otherwise \end{cases},$$

где L, θ — длина и угол, задающий направление, соответственно. Таким образом, определение параметров размытия в рамках данной работы есть определение

длины и направления ядра размытия. В качестве основных инструментов анализа в данной работе используются метод выделения хребтовых структур (Ridge Detection) [12] и анализ гистограмм распределения направлений.

2. Алгоритм определения направления вектора размытия движением

Для анализа изображения необходимо выделить границы объектов. При использовании детектора границ Канни, в результирующей картинке для размытой одиночной точки появятся двойные линии границ. Также, плохо выделяются именно размытые объекты изображения за счет того, что при применении такого алгоритма учитывается лишь яркость пикселей. Поэтому, предлагается использовать вместо детектора границ Канни выделение хребтовых структур.

Хребтовой структурой функции двух переменных называются точки локального максимума (минимума) функции, достигающиеся, по крайней мере, по одной переменной. Хребтовые структуры, в которых достигается максимум, называются горами, минимум - долинами. Отметим, что для анализа размытия в данной работе используются горы.

Будем рассматривать изображение как функцию двух переменных $g(x, y)$. Для каждого пикселя составим матрицу:

$$\begin{bmatrix} \frac{\partial^2 g}{\partial x^2} & \frac{\partial^2 g}{\partial x \partial y} \\ \frac{\partial^2 g}{\partial x \partial y} & \frac{\partial^2 g}{\partial y^2} \end{bmatrix}.$$

Пусть λ_1 и λ_2 являются собственными числами этой матрицы (не ограничивая общности, будем считать что λ_1 - наибольшее по модулю из них).

Тогда определено 3 случая [12]:

- $|\lambda_1| \sim 0, |\lambda_2| \sim 0$ - в точке (x, y) на изображении нет особенностей.

- $|\lambda_1| \sim |\lambda_2|, \lambda_1 \gg 0, \lambda_2 \gg 0$ - в точке (x, y) на изображении находится угол или текстура.

- $|\lambda_1| \gg |\lambda_2|$ - в точке (x, y) на изображении находится хребтовая структура, при этом знак λ_1 определяет тип хребтовой структуры - гора или долина.

Пусть I - входное изображение, тогда если обозначить:

$$I_{xx} = \frac{\partial^2 g}{\partial x^2}, \quad I_{yy} = \frac{\partial^2 g}{\partial y^2}, \quad I_{xy} = \frac{\partial^2 g}{\partial x \partial y},$$

то хребтовые структуры определяются следующей величиной:

$$\lambda_{max} = \text{MAX}(\lambda_1, \lambda_2), \text{ где}$$

$$\lambda_{1,2} = \frac{I_{xx} + I_{yy} \pm \sqrt{(I_{xx} + I_{yy})^2 - 4(I_{xx}I_{yy} - I_{xy}I_{xy})}}{2}.$$

На практике вторые производные функции $g(x, y)$ вычисляются путем свертки изображения с соответствующими вторыми производными функции Гаусса,

что позволяет отфильтровать шумы и менять масштаб хребтовых структур при помощи изменения параметра σ .

Так как фильтр Гаусса

$$G_g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

является сепарабельным, для ускорения вычислений производных свертка производится независимо по осям x и y , что позволяет уменьшить число операций на порядок. Для того, чтобы выделить середины хребтовых структур применяется алгоритм "подавления немаксимумов" (Non-Maximum Suppression).

Для каждой точки выделенной хребтовой структуры можно найти угол наклона касательной к кривизне хребтовой структуры в этой точке по следующей формуле:

$$\theta = \text{tg}^{-1} \left(\frac{I_{xy}}{\lambda_{max} - I_{xx}} \right). \quad (1)$$

При размытии движением контуры изображения вытягиваются в соответствующем направлении, которое принимаем за искомым параметр. Построим гистограмму распределения направления, т.е. горизонтальная ось в гистограмме соответствует величине угла от 0° до 180° , а вертикальная - количеству пикселей с соответствующим значением угла, вычисленным по формуле (1). Заметим, что в формировании гистограммы участвуют пиксели, выделенные с помощью алгоритма подавления немаксимумов после выделения хребтовых структур. За искомое направление размытия примем значение угла в гистограмме (горизонтальная ось гистограммы), соответствующее максимальному количеству пикселей с таким углом (вертикальная ось гистограммы). Таким образом, становится возможным вычислить угол, задающий направление размытия движением как максимальное значение в гистограмме распределения направлений, которое соответствует "холму" в гистограмме. На Рис.1 представлен пример гистограммы направлений для изображения, размытого с параметрами $\theta = 30^\circ, \rho = 30$. На данном рисунке видно, что максимум соответствует углу $\theta = 30$.

3. Алгоритм определения длины вектора размытия движением

Для определения длины вектора размытия движением предлагается постепенно размывать изображение в направлении, которое является перпендикулярным к существующему размытию, постепенно увеличивая длину размытия. Начальная длина перпендикулярного размытия и шаг внутри цикла могут быть выбраны в зависимости от задачи. С каждой итерацией изображение становится все более расфокусированным, а гистограмма распределения направлений изменяется таким образом, что появляется второй "холм". Заметим, что второй "холм" появляется зеркально к первому относительно точки, соответствующей 90° .

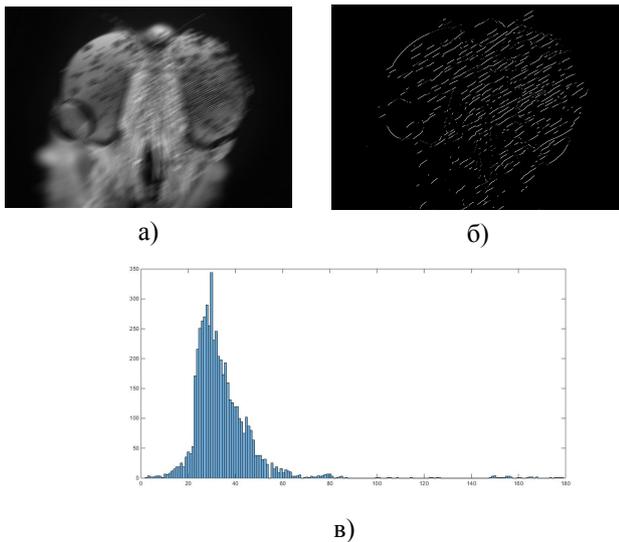


Рис. 1. Изображение head: (а) размытое с параметрами $\theta = 30^\circ$, $\rho = 30$; (б) хребтовые структуры; (в) гистограмма, вертикальная ось – количество пикселей, горизонтальная ось – угол в диапазоне от 0° до 180° .

Назовем *левой суммой* (*правой суммой*) *гистограммы распределения направлений* величину:

$$S_L = \sum_{i=0}^{90} hist(i) \quad (S_R = \sum_{i=90}^{180} hist(i)).$$

Таким образом, с каждой итерацией гистограмма стремится к моменту, когда сумма левой части гистограммы примерно равна сумме правой части гистограммы. Предлагается выходить из цикла в момент, когда появившийся "холм" становится выше, чем первоначальный, или когда соответствующая сумма гистограммы становится больше противоположной. За параметр длины принимаем значение итерационного параметра на предыдущем шаге. Для уменьшения времени работы алгоритма использован метод золотого сечения.

Условие выхода из цикла можно записать:

$$S_L > S_R \quad (S_R > S_L) \quad (*)$$

т.е. правая(левая) сумма больше левой(правой) в зависимости от части гистограммы, в которой находился максимум изначально.

Заметим, что размытие в перпендикулярном направлении в уменьшает число хребтов на изображении (этот эффект можно заметить на Рис.3 и Рис.4). Однако это не влияет на условие (*), поскольку для выхода из цикла важен момент изменения направления хребтовых структур, а не их количество. Таким образом, в условии (*) не важны значения правой и левой сумм, только знак их разности.

4. Общая схема метода

Таким образом, предлагаемый алгоритм состоит из следующих шагов (Рис. 2):

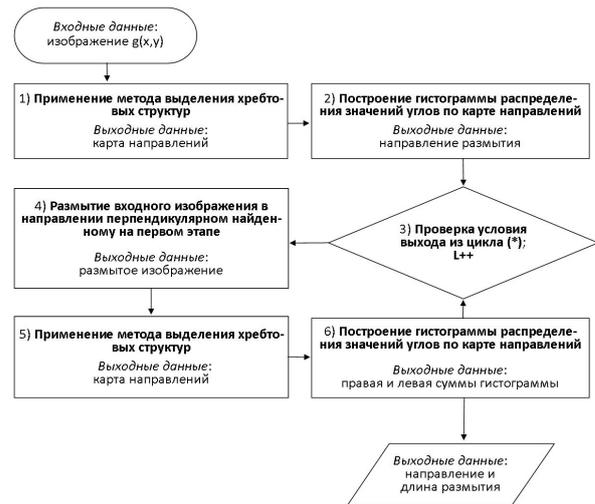


Рис. 2. Блок-схема предлагаемого алгоритма.

Определение направления вектора размытия движением.

1. Применяем метод выделения хребтовых структур. Выходными данными на этом шаге является карта направлений, построенная с помощью алгоритма выделения хребтовых структур и использования формулы (1).

2. На данном шаге строится гистограмма распределения направлений по карте, полученной на предыдущем шаге. За искомым параметр направления размытия принимаем значение угла, соответствующее максимальному значению в гистограмме.

Определение длины вектора размытия движением.

3. Осуществляем проверку условия (*) выхода из цикла. Пусть первоначально в гистограмме левая сумма больше, тогда выходим из цикла как только правая сумма начинает превосходить левую.

4. Если условие (*) ложно, применяем к изображению операцию размытия в направлении, перпендикулярном найденному на предыдущем этапе и увеличиваем длину вектора размытия.

5. Применение метода выделения хребтовых структур.

6. Построение гистограммы распределения направлений. Переходим к шагу 3).

7. Величина полученного параметра длины является вторым искомым параметром PSF.

В качестве иллюстрации рассмотрим метод в применении к тестовому изображению baboon, размытому с параметрами $\theta = 45^\circ$, $\rho = 10$ (Рис.3). На Рис.5 представлены гистограммы направлений, полученные в ходе итерационного процесса для входного изображения, размытого в направлении, перпендикулярном найденному (в данном случае 43°) с параметром $L = 0; 6; 8; 10$. На гистограммах заметно появление второго "холма". Рис.4 соответствует последней гистограмме (Рис.5) и показывает, как изменились направления хребтовых структур.

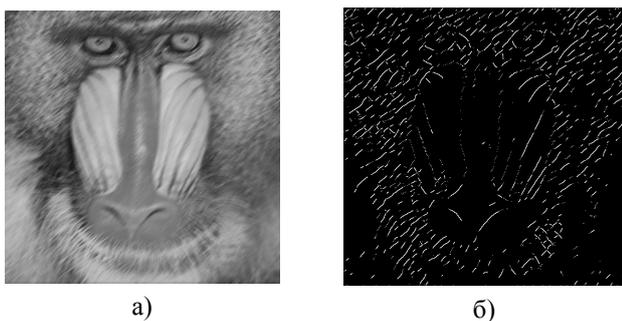


Рис. 3. Изображение baboon: (а) размытое с параметрами $\theta = 45^\circ, \rho = 10$; (б) хребтовые структуры.

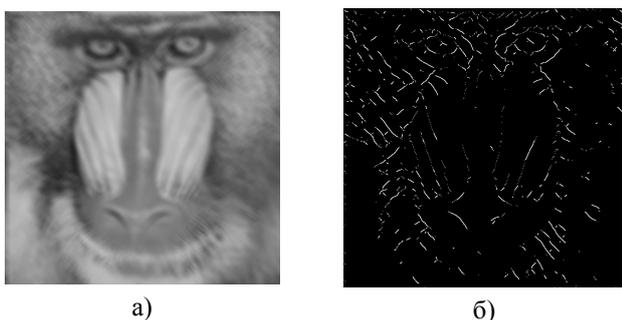


Рис. 4. (а) Изображение, последовательно размытое с параметрами $\theta = 45^\circ, \rho = 10$ и $\theta = 135^\circ, L = 10$; (б) хребтовые структуры изображения (а).

5. Тестирование и результаты

Результаты работы алгоритма примененного к изображениям без шума представлены в Таблице 1. В Таблице 2 представлены результаты работы алгоритма для трех тестовых изображений baboon ($\phi = 30, L = 5; 10; 15$), lena ($\phi = 60, L = 5; 10; 15$), sailboat ($\phi = 45, L = 5; 10; 15$) с вариацией Гауссовского шума 0, 0.0001, 0.001, 0.005, 0.01 соответственно. При сравнении полученных результатов с результатами работы [3] было обнаружено, что предложенный нами метод при точности определения длины размытия движением L дает существенно более устойчивые результаты в случае добавления Гауссовского шума.

Как видно из представленных результатов, алгоритм достаточно точно находит направление размытия даже в условиях шума. Отметим, что алгоритм показывает достаточно точные результаты по определению направления при удачном выборе σ в методе выделения хребтовых структур. В дальнейшем предполагается внедрение автоматического выбора σ . Также, отметим, что алгоритм будет показывать низкую точность оценки в случае наличия на изображении множества текстур с определенным направлением (примером может служить изображение с множеством однотонных полос), которые будут вносить помехи на этапе определения направления размытия.

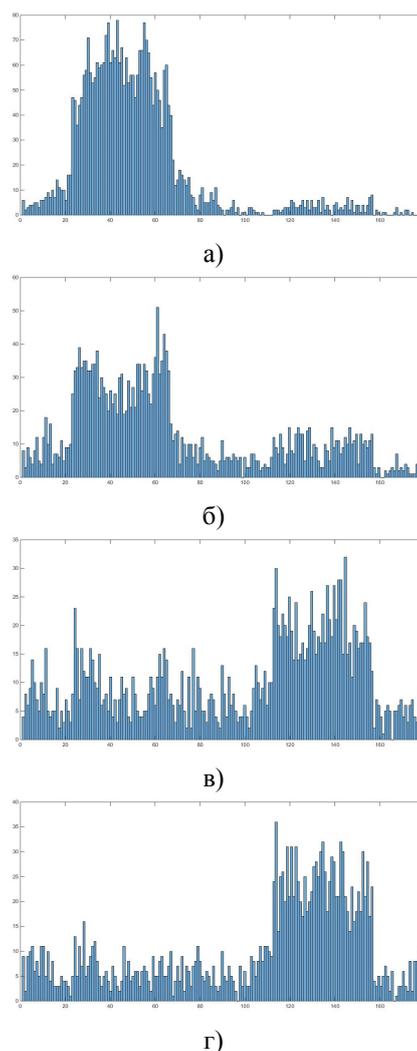


Рис. 5. Результаты применения алгоритма для изображения baboon($\phi = 45, L = 10$): гистограммы для размытия в направлении, перпендикулярном найденному, с $L=0$ (а); 6(б); 8(в); 10(г).

изображение	ϕ	L	ϕ найд.	L найд.
lighthouse2	135°	10	135°	9
head	30°	30	30°	30
head	30°	10	30°	9
lena	60°	5	58°	6
lena	60°	15	58°	16
baboon	45°	5	43°	5
baboon	45°	15	43°	13
baboon	30°	5	32°	4
baboon	30°	15	32°	14
sailboat	45°	5	42°	4
sailboat	45°	15	44°	12

Таблица 1. Результаты тестирования изображений без шума; (ϕ, L) - параметры размытия; (ϕ найд., L найд.) - параметры, найденные с помощью предложенного метода.

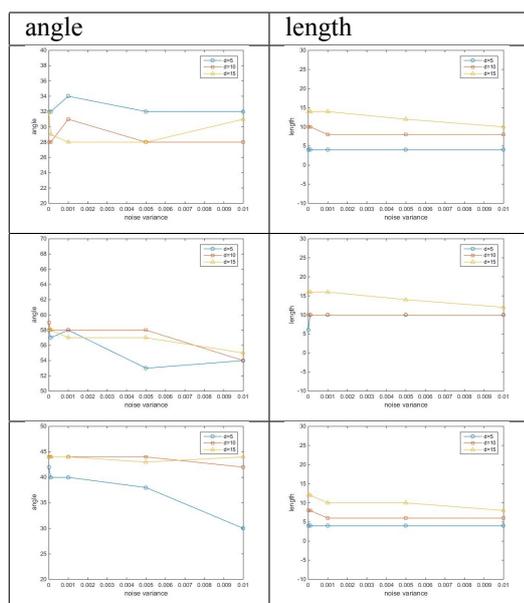


Таблица 2. Результаты для изображений baboon($\phi = 30, L = 5; 10; 15$), lena($\phi = 60, L = 5; 10; 15$), sailboat ($\phi = 45, L = 5; 10; 15$) с вариацией Гауссовского шума 0, 0.0001, 0.001, 0.005, 0.01 соответственно.

6. Заключение

Предложен метод оценки функции точечного источника размытия движением для автоматического восстановления изображений. Метод основан на анализе хребтовых структур. Рассматриваемый метод предполагает направленное пространственно-постоянное размытие движением. Метод показал достаточно хорошие результаты определения параметров ядра размытия движением, достаточно устойчивые к наличию Гауссовского шума. В дальнейшем предполагается использование в предложенном алгоритме многомасштабного аналога нахождения хребтовых структур.

7. Благодарности

Работа выполнена при финансовой поддержке РФФИ, грант 16-07-01207.

8. Литература

- [1] Cannon M. Blind deconvolution of spatially invariant image blurs with phase //IEEE Transactions on Acoustics, Speech, and Signal Processing. – 1976. – Т. 24. – №. 1. – С. 58-63.
- [2] Yitzhaky Y., Kopeika N. S. Identification of blur parameters from motion blurred images //Graphical models and image processing. – 1997. – Т. 59. – №. 5. – С. 310-320.
- [3] Tanaka M., Yoneji K., Okutomi M. Motion blur parameter identification from a linearly blurred image // Consumer Electronics, 2007. ICCE

2007. Digest of Technical Papers. International Conference on. – IEEE, 2007. – С. 1-2.

- [4] Ji H., Liu C. Motion blur identification from image gradients //Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. – IEEE, 2008. – С. 1-8.
- [5] Phansalkar N. Determination of linear motion point spread function using hough transform for image restoration // Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on. – IEEE, 2010. – С. 1-4.
- [6] Oliveira J. P., Figueiredo M. A. T., Bioucas-Dias J. M. Parametric blur estimation for blind restoration of natural images: Linear motion and out-of-focus // IEEE Transactions on Image Processing. – 2014. – Т. 23. – №. 1. – С. 466-477.
- [7] Wang Z., Yao Z., Wang Q. Improved scheme of estimating motion blur parameters for image restoration //Digital Signal Processing. – 2017. – Т. 65. – С. 11-18.
- [8] Lin H. Y., Li K. J. Motion blur removal and its application to vehicle speed detection // Image Processing, 2004. ICI'04. 2004 International Conference on. IEEE, 2004. – Т. 5. – С. 3407-3410.
- [9] Kang S. K., Min J. H., Paik J. K. Segmentation-based spatially adaptive motion blur removal and its application to surveillance systems // Image Processing, 2001. Proceedings. 2001 International Conference on. IEEE, 2001. – Т. 1. – С. 245-248.
- [10] Qi X. Y., Zhang L., Tan C. L Motion deblurring for optical character recognition // Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on. IEEE, 2005. – С. 389-393.
- [11] Aizenberg I. et al Blur identification by multilayer neural network based on multivalued neurons // IEEE Transactions on Neural Networks, 2008. – Т. 19. 5. – С. 883-898.
- [12] Serrat J., López A., Lloret D. On ridges and valleys [image registration] // Pattern Recognition, 2000. Proceedings. 15th International Conference on. – IEEE, 2000. – Т. 4. – С. 59-66.

Совмещение изображений флуоресцентной микроскопии для компенсации движения живых клеток: обзор

Д.В. Сорокин^{1,2}
dsorokin@cs.msu.ru

¹Лаборатория математических методов обработки изображений, факультет вычислительной математики и кибернетики, Московский Государственный Университет имени М.В.Ломоносова, Москва, Россия;

²Centre for Biomedical Image Analysis, Faculty of Informatics, Masaryk University, Brno, Czech Republic

Совмещение изображений является базовой задачей в области обработки и анализа микроскопических изображений живых клеток. С помощью совмещения изображений решается задача компенсации движения живых клеток, необходимая для изучения локального движения клеточных субструктур. В последнее время было предложено большое количество разнообразных методов совмещения микроскопических изображений. Однако, все методы имеют свои особенности и границы применимости, а выбор метода для реальных приложений обусловлен особенностями биологического эксперимента, и как следствие, полученных данных. В данной работе проведен обзор методов совмещения микроскопических изображений, обсуждены их достоинства и недостатки, а также границы применимости в реальных приложениях.

Ключевые слова: анализ микроскопических изображений, совмещение изображений, микроскопия живых клеток.

1. Введение

В последние десятилетия, одним из наиболее бурно развивающихся направлений биомедицины является клеточная биология. Для проведения экспериментов в этой области используются все более сложные приборы, такие, как флуоресцентные микроскопы. С появлением технологии флуоресцентных белков (Нобелевская премия по химии в 2008 году) расширились возможности проведения прижизненных наблюдений, что кардинально расширило возможности световой микроскопии. Бурно развивалась и микроскопическая техника. Особенное значение имело появление конфокального микроскопа и его вариантов (лазерный сканирующий, спиннинговый и т.п.). Недавно начала бурно развиваться микроскопия сверх-высокого разрешения (Нобелевская премия по химии в 2014 году). На данный момент мало какое исследование в области клеточной биологии обходится без экспериментов, связанных с микроскопией.

Среди различных методов микроскопии отдельно можно выделить микроскопию живых клеток, или прижизненные наблюдения за клеточными структурами. Данный метод исследования позволяет изучать протекание молекулярных процессов в живых клетках, что стимулирует развитие разных областей клеточной биологии, в том числе связанных с изучением механизмов онкогенеза и поиска новых лекарств.

Одной из наиболее актуальных задач в микроскопии живых клеток является анализ движения клеточных субструктур. Данная задача возникает при изучении механизмов восстановления молекулы ДНК [5], механизмов формирования и разбора ядрышек [13], вирусной защиты [9], и прочих процессов, необходимых для понимания природы функционирования клеток. Однако, анализ движения клеточных субструктур является сложной проблемой, т.к. в процессе получения данных живая клетка движется и деформируется.

Таким образом, наблюдаемое движение субструктур представляет собой суперпозицию локального движения субструктур и глобального движения клетки, включающего движение и деформацию клетки. Для того, чтобы определить чистое локальное движение субструктур, необходимо компенсировать глобальное движение клетки. Это делается с помощью алгоритмов совмещения изображений. А именно, все изображения из последовательности совмещаются с изображением, соответствующим начальному моменту времени.

В литературе представлено большое количество методов совмещения микроскопических изображений предназначенных для компенсации движения и деформации живых клеток. Данные методы можно разделить на две основные группы: методы, напрямую *использующие интенсивность изображения*, и методы, *не использующие напрямую интенсивность изображения* (такие методы используют результат сегментации изображения - бинарные маски). Также, в каждой из данных групп можно выделить алгоритмы, осуществляющие *жесткое совмещение* (клетка представляется, как жесткое тело, в результате чего компенсируется только движение — смещение, поворот, масштабирование), и алгоритмы осуществляющие *нежесткое совмещение* (клетка представляется, как нежесткое тело, в результате чего, помимо движения, компенсируется и деформация клетки). К сожалению, на данный момент не существует универсального метода, подходящего для любых данных и любой постановки задачи. Каждый из представленных подходов имеет свои достоинства и недостатки, а выбор подходящего метода обусловлен конкретной биологической проблемой, требующей решения.

В данной работе представлен обзор современных методов совмещения микроскопических изображений клеточных структур, используемых для компенсации движения и деформации клеток в задачах анализа движения клеточных субструктур. Рассмотрены достоин-

ства и недостатки данных методов. Проведен сравнительный анализ их эффективности и применимости в различных типовых постановках задачи.

2. Методы, использующие интенсивность изображения

Большинство методов совмещения изображений, представленных в литературе, используют значения интенсивности пикселей изображений напрямую для их совмещения.

2.1. Жесткие методы

В работе [24] был предложен метод совмещения изображений, основанный на фазовой корреляции. Другой корреляционный метод, позволяющий компенсировать сдвиги изображений был предложен в статье [7]. Baheerathan и соавторы [1] впервые предложили использовать фазовую корреляцию совместно с ключевыми точками, что позволило компенсировать движения соответствующие аффинному преобразованию. В подходе, предложенном в работе [16] был использован тензор инерции, который вычислялся с помощью окрашенных белков. Данный метод позволяет компенсировать поворот и смещение клетки. В работе [23] авторы использовали модель фотообесцвечивания для компенсации жесткого движения клеток при проведении FLIP (Fluorescence loss in photobleaching) экспериментов. Другой подход, использующий разбиение изображения на блоки с последующим поблочным совмещением был предложен в работе [15].

2.2. Нежесткие методы

Помимо методов жесткого совмещения, также был предложен ряд методов нежесткого совмещения. В работе [14] авторами была представлена параметрическая модель движения клетки, которая была использована для компенсации этого движения. Полуавтоматический метод, основанный на использовании ключевых точек и сплайнов тонкой пластинки, был предложен в работе [10]. Основанный на использовании ключевых точек метод был описан в работе [12], однако он был использован для коррекции движения клетки, соответствующего аффинному преобразованию изображения.

Отдельно стоит отметить семейство методов, основанных на модификации классического алгоритма оптического тока Lucas-Kanade. В работе [8] предложена базовая, двухкадровая версия данного подхода, где поле деформаций $\vec{u}(I_k, I_{k-1})$ между изображениями I_k и I_{k-1} последовательности, определяется с помощью итерационного алгоритма следующим образом:

$$\vec{u}_i(I_k, I_{k-1}) = G_D * (\vec{u}_{i-1}(I_k, I_{k-1}) + G_U * d\vec{u}_i(I_k^*, I_{k-1})), \quad (1)$$

где $\vec{u}_{i-1}(I_k, I_{k-1})$ — поле деформаций с предыдущей итерации, $d\vec{u}_i(I_k^*, I_{k-1})$ — изменение поля деформаций, а $I_k^* = I_k(\vec{x} + \vec{u}_{i-1}(I_k, I_{k-1}))$. Вектор \vec{x} обозначает пространственные координаты изображения, а G_D и

G_U — ядра свертки с Гауссианами ($\sigma_D = \sigma_U = 2$), с помощью которых проводится сглаживание полученного поля деформаций. Начальное поле деформаций $\vec{u}_{i-1}(I_k, I_{k-1}) \equiv 0$. Итерационный процесс останавливается, когда $\|I_{k-1} - I_k^*\| < \varepsilon$, где ε задается в качестве параметра алгоритма. При этом, изменение поля деформации определяется как:

$$d\vec{u}_i(I_k^*, I_{k-1}) = \min_{\vec{U} \in \mathbb{R}^2} \sum_{\vec{x} \in \Omega} [I_k(\vec{x} + \vec{U}) - I_{k-1}(\vec{x})]^2, \quad (2)$$

где Ω — некоторая окрестность точки \vec{x}_c изображения, \vec{U} — вектор поля деформации $\vec{u}_i(I_k, I_{k-1})$ в точке \vec{x}_c . Также, в работе [8] предложены более совершенные варианты данного базового двухкадрового алгоритма, полученные с помощью внесения изменений в минимизируемый функционал в (2), таких как симметричная модификация функционала, и симметричная взвешенная модификация функционала [8].

В работе [21] авторы развили данный подход, представив его многокадровую модификацию. Основная идея заключается в использовании нескольких кадров до текущего для построения поля деформаций между последовательными изображениями. Основным отличием данного подхода является иное определение изменения поля деформаций $d\vec{u}$ в (1), где вместо изображений I_k и предыдущего I_{k-1} , используется $N_l + 1$ изображений I_k, \dots, I_{k-N_l} :

$$d\vec{u}_i(I_k^*, I_{k-1}) = \min_{\vec{U} \in \mathbb{R}^2} \sum_{\vec{x} \in \Omega} \frac{1}{N_l} \sum_l [I_k(\vec{x} + \vec{U}) - I_l^*(\vec{x})]^2,$$

где $I_l^* = I_l(\vec{x} + \vec{u}(I_l, I_{k-1}))$ — изображение I_l , $l = k - N_l, \dots, k - 1$, преобразованное в систему координат изображения I_{k-1} , а N_l обозначает количество предыдущих кадров, используемых для совмещения.

Метод получил дальнейшее развитие в работе [22], в которой была представлена версия многокадрового алгоритма, где итоговое преобразование между соседними кадрами является диффеоморфизмом.

2.3. Достоинства и недостатки использования методов, основанных на интенсивности

Методы совмещения изображений, использующие интенсивность, являются достаточно эффективными, т.к. в данных подходах используется вся информация, содержащаяся в изображениях. Однако, данные методы демонстрируют меньшую эффективность (а иногда и вовсе неприменимость) в случаях, когда интенсивность изображений последовательности подвержена значительным изменениям с течением времени, например, в случае высокого уровня шума или резкого движения клетки между кадрами, что является достаточно характерным для большого числа практических случаев. Например, методы, основанные на принципе вычисления оптического потока [8, 21] или сопоставления блоков изображений [17], допускают лишь незначительные изменения интенсивности объектов между кадрами. Нарушение этого условия может привести к

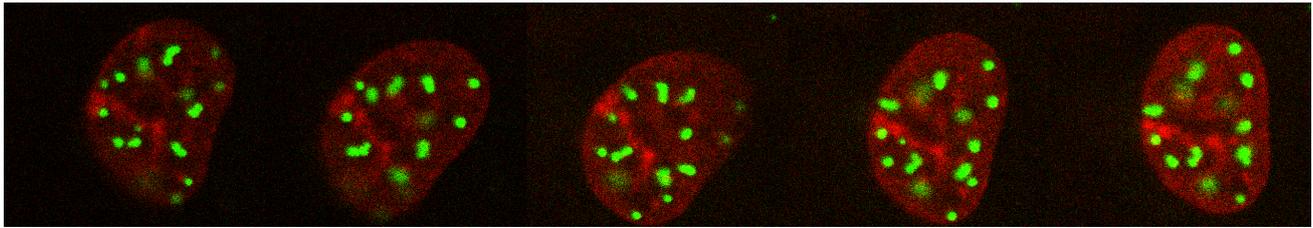


Рис. 1. Пример последовательности изображений флуоресцентной микроскопии, полученной в двух каналах, где в красном канале представлено клеточное ядро, а в зеленом канале представлены изучаемые клеточные субструктуры.

нестабильности метода. Стоит также отметить, что в ряде биологических приложений, данное условие заведомо невыполнимо, т.к. полученные последовательности изображений содержат резкие скачки интенсивности. В работе [14] был предложен подход для преодоления данной проблемы резкого изменения интенсивности, однако, данный метод использует аффинную модель движения и не может компенсировать сильные локальные деформации клетки. Алгоритмы основанные на ключевых точках [3, 10, 12] требуют наличия стабильных ключевых точек у движущихся клеток, что также является невыполнимым требованием для многих биологических приложений, где в изображениях движущейся клетки отсутствуют (либо не видны) заведомо стабильные структуры. Также, в случае изучения локального движения клеточных субструктур, методы, основанные на интенсивности, помимо компенсации глобального движения и деформации клетки, могут внести ошибку и в локальное движение изучаемых субструктур. Такое может произойти, т.к. основанные на интенсивности методы стремятся полностью компенсировать движение ярких объектов на изображениях, не делая различий между субструктурами и самой клеткой. Данная проблема может быть решена с помощью получения изображений в двух каналах [8, 21], где изучаемые субструктуры представлены в одном канале, а некоторые стабильные структуры, соответствующие движению клетки (например, хроматин), представлены в другом канале (см. Рис. 1). Однако, во многих биологических экспериментах данные возможно получить лишь в одном канале, что делает основанные на интенсивности подходы неприменимыми для таких данных. Это приводит к необходимости использования в таких случаях методов, не использующих интенсивность изображения.

3. Методы, не использующие интенсивность изображения

В литературе представлен ряд методов совмещения микроскопических изображений живых клеток, не использующих интенсивность изображения напрямую.

3.1. Жесткие методы

Среди этих методов можно выделить небольшое количество подходов, использующих жесткую модель. В работе [18] авторами был предложен подход, где клетка

дрожжей приближалась эллипсом. Далее было произведено совмещение полученных эллипсов для компенсации смещений и поворотов. В работе [11] был предложен оригинальный метод совмещения изображений клеток, основанный на жестком совмещении облаков точек. Облако точек определялось как положение исследуемых субструктур в каждом кадре. При этом, чтобы избежать "перерегистрации" (внесение ошибок в локальное движение самих субструктур, описанное в предыдущем разделе), была выбрана жесткая модель. Также в данной работе было показано, что применение некоторых ограничений по оси Z существенно увеличивает скорость работы метода, не оказывая большого влияния на качество результата. Еще одним жестким методом, не использующим интенсивность напрямую, является подход, предложенный в работе [4]. Данный метод основан на использовании предварительной сегментации клетки в изображениях последовательности, с последующим представлением формы клетки в виде специальной функции, определенной с помощью дискретного евклидова преобразования. Строго говоря, данный метод является нежестким, т.к. для совмещения используется минимизация функционала, где невязка параметризует сдвиг и поворот клетки, а регуляризатор позволяет параметризовать небольшие деформации. Однако, вклад нежесткой компоненты в данном методе незначительный, и его принято относить к жестким подходам.

3.2. Нежесткие методы

Помимо жестких методов, в литературе представлено несколько нежестких подходов. Все они используют в качестве входных данных сегментацию изображения, т.е. последовательность бинарных изображений, представляющих собой маски движущейся клетки. В работе [25] авторами был предложен подход, основанный на улучшенном алгоритме совмещения "демонов" (demons-based), который был применен непосредственно к бинарным маскам. Данный алгоритм позволил компенсировать незначительные движения и деформации клетки. Другой подход, предложенный в [2], направлен на компенсацию значительных перемещений клетки (более 30 пикселей) и значительных деформаций. Этот подход основан на поиске соответствия между контурами клетки в последовательных кадрах с помощью использования дескрипторов контуров, представляющих собой одномерные функции, и алгоритма

digital time warping для их совмещения. После нахождения соответствия между точками контуров, на контурах клетки задаются вектора смещения, которые затем интерполируются в остальных областях изображения с помощью метода интерполяции разреженных данных. В работе [20] авторами был предложен метод, который также использует контуры клетки. Однако в данном методе, для интерполяции разреженного поля деформации на контурах клетки, была использована физически обоснованная модель, основанная на теории упругости. А именно, клетка представляется как упругое тело, где деформация задана на границе объекта (с помощью совмещения контуров), а деформация внутри клетки моделируется с помощью уравнения Навье-Коши:

$$\begin{aligned} \mu \Delta \mathbf{u}(\mathbf{x}) + (\mu + \lambda) \nabla (\nabla \mathbf{u}(\mathbf{x})) &= 0, & \mathbf{x} \in Q, \\ \mathbf{u}(\mathbf{x}) &= \mathbf{u}_0, & \mathbf{x} \in \partial Q, \end{aligned} \quad (3)$$

где \mathbf{u} — это поле деформаций, \mathbf{x} — пространственная координата, μ и λ — константы Ламе, Q — область, соответствующая клетке, а ∂Q — граница клетки.

Данный подход был развит в работе [19]. В этой работе была использована более сложная нелинейная динамическая упругая модель, которая может быть описана вторым законом Ньютона для упругого тела:

$$\begin{aligned} M \vec{a} &= \vec{f}_{int} + \vec{f}, \\ \vec{f}_{int} &= \mu \Delta \vec{u} + (\mu + \lambda) \nabla (\nabla \cdot \vec{u}), \end{aligned} \quad (4)$$

где M — масса клетки, \vec{a} — ускорение, \vec{u} — поле деформаций, а \vec{f} представляет собой силы, действующие на границу клетки, которые приводят к совмещению изображений. В статическом методе [20] уравнение (3) по сути соответствует $\vec{f}_{int} = 0$ в уравнении (4), и задача решается в квази-статическом сценарии, когда для каждой пары последовательных изображений с помощью определения соответствия точек контуров клетки на этих изображениях, задается деформация на границе клетки, а затем решается уравнение Навье-Коши. В динамической модели (4) не требуется отдельного шага сопоставления контуров клетки в последовательных кадрах. Совмещение контуров происходит за счет определения внешних сил f специальным образом в ходе итерационного решения (4). Динамическая модель, использованная в [19], позволила улучшить результаты алгоритма совмещения [20], а также сделать метод значительно более устойчивым к параметрам. Стоит отметить, что оба метода [20] и [19] являются достаточно эффективными с вычислительной точки зрения, т.к. они основаны на методе конечных элементов.

3.3. Достоинства и недостатки использования методов, не использующих интенсивность изображений

Подходы, не использующие интенсивность изображений, как описано выше, часто используют в качестве входных данных результат сегментации клетки в последовательности изображений, а именно, после-

довательность бинарных изображений, представляющих собой маски клетки в различные моменты времени. Эффективность данных подходов была продемонстрирована как на реальных микроскопических данных, так и на синтетических данных. Данные методы являются применимыми как к изображениям, полученным в одном цветовом канале, так и к изображениям, полученным в двух и более цветовых каналах. Однако, у описанных методов существует ряд недостатков. Так, в работах [4, 11, 18] алгоритмы могут компенсировать только "жесткое" движение, а именно поворот и смещение клетки. В работе [25] авторы описывают нежесткий алгоритм совмещения, однако он основан на алгоритме "демонов" (demons algorithm), который не использует никакую физическую модель движения клетки. Метод, предложенный в [2] был специально разработан для компенсации сильных деформаций, однако использованная модель сплайнов тонкой пластинки является достаточно грубой аппроксимацией деформации клетки. Помимо этого, в данной работе используется равномерная сетка для дискретизации контуров клетки, что приводит к появлению и накоплению ошибки при поиске поля деформации для некоторых наборов данных. В работе [6] авторами предложен алгоритм преобразования клетки к некоторому сферическому объекту, что пригодно для части биологических приложений (где клетка зафиксирована, а изображение статично), однако неприменимо в случае изучения локального движения клеточных субструктур. Наиболее эффективными на данный момент являются методы, основанные на теории упругости [20] и [19], показывающие сравнимые результаты с методами, основанными на интенсивности, которые используют значительно больше информации. Однако данные подходы также не лишены недостатков. Во-первых, в случае невозможности точного определения границ клеток, точность данных подходов значительно падает. Также, в случае совмещения двухканальных изображений, где в одном из каналов представлены стабильные структуры, использование методов основанных на интенсивности является более надежным.

4. Заключение

В данной работе представлен обзор существующих на сегодняшний день методов совмещения микроскопических изображений, применяемых для компенсации движения и деформации клеток в последовательности двумерных или трехмерных изображений. Рассмотрены жесткие и нежесткие методы совмещения двух семейств: методы, использующие интенсивность изображения напрямую, и методы, не использующие интенсивность изображений напрямую. Обсуждены достоинства и недостатки описанных в литературе методов, а также границы их применимости.

5. Благодарности

Работа выполнена при финансовой поддержке РФФИ, грант 17-11-01279.

6. Литература

- [1] Baheerathan, S., Albrechtsen, F., and Danielsen, H. Registration of serial sections of mouse liver cell nuclei. *J. of Microscopy* 192, 1 (1998), 37–53.
- [2] De Vylder, J., De Vos, W. H., Manders, E. M., and Philips, W. 2d mapping of strongly deformable cell nuclei-based on contour matching. *Cytometry Part A* 79, 7 (2011), 580–588.
- [3] Du, T., and Wasser, M. 3d image stack reconstruction in live cell microscopy of drosophila muscles and its validation. *Cytometry Part A* 75, 4 (2009), 329–343.
- [4] Dzyubachyk, O., Essers, J., van Cappellen, W. a., Baldeyron, C., Inagaki, A., Niessen, W. J., and Meijering, E. Automated analysis of time-lapse fluorescence microscopy images: from live cell images to intracellular foci. *Bioinformatics* 26, 19 (2010), 2424–30.
- [5] Foltánková, V., Matula, P., Sorokin, D., Kozubek, S., and Bártoňová, E. Hybrid detectors improved time-lapse confocal microscopy of pml and 53bp1 nuclear body colocalization in dna lesions. *Microscopy and Microanalysis* 19, 02 (2013), 360–369.
- [6] Gladilin, E., Goetze, S., Mateos-Langerak, J., Van Driel, R., Eils, R., and Rohr, K. Shape normalization of 3d cell nuclei using elastic spherical mapping. *J. of Microsc.* 231, 1 (2008), 105–114.
- [7] Goobic, A. P., Tang, J., and Acton, S. T. Image stabilization and registration for tracking cells in the microvasculature. *IEEE Trans. Biomedical Engineering* 52, 2 (2005), 287–299.
- [8] Kim, I.-H., Chen, Y.-C. M., Spector, D. L., Eils, R., and Rohr, K. Nonrigid registration of 2-d and 3-d dynamic cell nuclei images for improved classification of subcellular particle motion. *IEEE Trans. Im. Proc.* 20, 4 (2011), 1011–1022.
- [9] Liu, S.-L., Zhang, Z.-L., Tian, Z.-Q., Zhao, H.-S., Liu, H., Sun, E.-Z., Xiao, G. F., Zhang, W., Wang, H.-Z., and Pang, D.-W. Effectively and efficiently dissecting the infection of influenza virus by quantum-dot-based single-particle tracking. *ACS Nano* 6, 1 (2012), 141–150.
- [10] Mattes, J., Nawroth, J., Boukamp, P., Eils, R., and Greulich-Bode, K. M. Analyzing motion and deformation of the cell nucleus for studying colocalizations of nuclear structures. In *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI'06)* (2006), pp. 1044–1047.
- [11] Matula, P., Matula, P., Kozubek, M., and Dvořák, V. Fast point-based 3-d alignment of live cells. *IEEE Trans. Im. Proc.* 15, 8 (2006), 2388–2396.
- [12] Medioni, C., Besse, F., Descombes, X., and Malandarne, G. Motion compensation in two-photon microscopy temporal series. In *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI'15)* (2015), pp. 1584–1587.
- [13] Musinova, Y. R., Lisitsyna, O. M., Sorokin, D. V., Arifulin, E. A., Smirnova, T. A., Zinovkin, R. A., Potashnikova, D. M., Vassetzky, Y. S., and Sheval, E. V. Rna-dependent disassembly of nuclear bodies. *J. of Cell Sci.* 129, 24 (2016), 4509–4520.
- [14] Ozere, S., Bouthemy, P., Spindler, F., Paul-Gilloteaux, P., and Kervrann, C. Robust parametric stabilization of moving cells with intensity correction in light microscopy image sequences. In *Proc. IEEE Int. Symp. Biomed. Imag. 2013 (ISBI'13)* (2013), pp. 468–471.
- [15] Raza, S.-E.-A., Humayun, A., Abouna, S., Nattkemper, T. W., Epstein, D. B. A., Khan, M., and Rajpoot, N. M. Ramtab: Robust alignment of multi-tag bioimages. *PLoS ONE* 7, 2 (2012), e30894.
- [16] Rieger, B., Molenaar, C., Dirks, R., and Van Vliet, L. Alignment of the cell nucleus from labeled proteins only for 4d in vivo imaging. *Microscopy Research and Technique* 64, 2 (2004), 142–150.
- [17] Saalfeld, S., Fetter, R., Cardona, A., and Tomancak, P. Elastic volume reconstruction from series of ultra-thin microscopy sections. *Nature Methods* 9, 7 (2012), 717–720.
- [18] Sage, D., Neumann, F. R., Hediger, F., Gasser, S. M., and Unser, M. Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. *IEEE Trans. Im. Proc.* 14, 9 (2005), 1372–83.
- [19] Sorokin, D. V., Peterlik, I., Tektonidis, M., Rohr, K., and Matula, P. Non-rigid Contour-Based Registration of Cell Nuclei in 2D Live Cell Microscopy Images Using a Dynamic Elasticity Model. *IEEE Transactions on Medical Imaging* (2017), accepted for publication 31 July 2017, DOI: 10.1109/TMI.2017.2734169.
- [20] Sorokin, D. V., Tektonidis, M., Rohr, K., and Matula, P. Non-rigid contour-based temporal registration of 2D cell nuclei images using the Navier equation. In *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI'14)* (2014), pp. 746–749.
- [21] Tektonidis, M., Kim, I.-H., Chen, Y.-C. M., Eils, R., Spector, D. L., and Rohr, K. Non-rigid multi-frame registration of cell nuclei in live cell fluorescence microscopy image data. *Med. Im. Anal.* 19, 1 (2015), 1–14.
- [22] Tektonidis, M., and Rohr, K. Diffeomorphic multi-frame non-rigid registration of cell nuclei in 2d and 3d live cell images. *IEEE Transactions on Image Processing* 26, 3 (March 2017), 1405–1417.
- [23] van de Giessen, M., van der Laan, A., Hendriks, E. A., Vidorreta, M., Reiber, J. H., Jost, C. R., Tanke, H. J., and Lelieveldt, B. P. Fully automated attenuation measurement and motion correction in flip image sequences. *IEEE Trans. Med. Imag.* 31, 2 (2012), 461–473.
- [24] Wilson, C. A., and Theriot, J. A. A correlation-based approach to calculate rotation and translation of moving cells. *IEEE Trans. Im. Proc.* 15, 7 (2006), 1939–1951.
- [25] Yang, S., Kohler, D., Teller, K., Cremer, T., Le Baccon, P., Heard, E., Eils, R., and Rohr, K. Nonrigid registration of 3-d multichannel microscopy images of cell nuclei. *IEEE Trans. Im. Proc.* 17, 4 (2008), 493–9.



Цифровая Земля и Большие Данные

Digital Earth and Big Data

GraphiCon 2017

24–28 сентября 2017
Пермь, Россия

Digital Earth как метод визуализации

А.Я. Аноприенко¹, Е.Н. Ерёмченко², С.В. Клименко³
 anoprien@gmail.com|eugene.eremchenko@gmail.com|stanislav.klimenko@gmail.com

¹ДонНТУ, Донецк, ДНР;

²МГУ, Москва, Россия;

³ИФТИ, Протвино, Россия

В работе даётся краткий обзор научных результатов, достигнутых при разработке концепции «Цифровой Земли» (Digital Earth) как нового метода геопространственной визуализации. Рассматриваются проблемы создания типологии визуализации и влияние этой концепции на семиотику. Описывается ход встраивания концепции в системы управления национального и континентального уровней в качестве среды поддержки принятия решений.

Ключевые слова: Цифровая Земля, визуализация, неогеография, типология, семиотика.

Digital Earth as Visualization Approach

A. Ya. Anoprienko¹, E. N. Eremchenko², S. V. Klimenko³
 anoprien@gmail.com|eugene.eremchenko@gmail.com|stanislav.klimenko@gmail.com

¹DonNTU, Donetsk, DPR;

²Lomonosov MSU, Moscow, Russia;

³ICPT, Protvino, Russia

In the article the brief review of scientific results achieved at the developing of Digital Earth concept as new approach for geospatial visualization is provided. Impacts of this concept on semiotics and the problem of visualization's typology are assessed. Process of embedding the concept in the management systems of the national and continental levels as an environment for decision support, is considered.

Keywords: Digital Earth, visualization, neogeography, typology, semiotics.

1. Введение

Практическая реализация концепции Цифровой Земли в виде Google Earth в 2005 году вызвала кризис в картографии, поскольку новый геосервис очевидным образом не подпадал под определение географической карты. В ходе возникшей полемики внимание было сконцентрировано на проблеме сравнительного анализа всего многообразия различных геопространственных продуктов за всю их историю. Одной из особенностей этого анализа являлась его беспрецедентная ретроспективная база — первые «настоящие» карты возникли, вероятно, не позднее чем за 8600 лет до нас и претендуют быть старейшими знаковыми системами.

При этом возникла естественная потребность систематизировать результаты анализа в виде классификации различных режимов визуализации в географии, а также попытаться выделить существенные классификационные признаки, отражающие внутреннюю логику эволюции картографического метода — т.е. выработать типологию геопространственных визуализаций. При этом обнаружились парадоксальные семиотические свойства новых геопродуктов. Масштабы возникших проблем и их отчётливо междисциплинарный характер, а также рост прикладной значимости концепции Цифровой Земли, ставший явным уже в 2017 году, позволяют говорить о целесообразности рассмотрения возникшего комплекса проблем с точки зрения научной визуализации как самостоятельной междисциплинарной исследовательской области.

2. Дискуссия

Концепция «Цифровой Земли» была выдвинута в 1998 году тогдашним вице-президентом США Элом (Альбертом)

Гором [17] в виде госзаказа на разработку качественно новой среды для визуализации геопространственной информации. При этом концепция опиралась на широко распространённые представления о том, каким должен быть «идеальный» географический продукт с точки зрения обычных пользователей¹. Спустя 7 лет после выступления А. Гора, в 2005 году, эта концепция нашла своё воплощение в геосервисе Google Earth [16].

Новизна, неожиданность и необычность Google Earth породили многочисленные вопросы и вызвали настоящий кризис в картографии. Новый продукт не только не соответствовал классическому определению географической карты во всех традиционных вариантах её реализации (обычные и электронные карты, ГИС и т.д.), но и явно противоречил ему в каждом из элементов без исключения.

Так, официальный российский ГОСТ 21667-76 определяет карту как [элементы определения выделены авторами] «1) построенное в картографической проекции, 2) уменьшенное, 3) обобщенное 4) изображение поверхности Земли, поверхности другого небесного тела или внеземного пространства, 5) показывающее расположенные на них объекты в определенной системе условных знаков» [3]. Вместе с тем очевидно, что реализованная в Google Earth концепция не предполагает обязательного использования ни одного из пяти элементов определения карты и, следовательно, не может быть отнесена к классу географических карт (рис. 1).

1 См., например, М. Булгаков, Мастер и Маргарита: «Маргарита наклонилась к глобусу и увидела, что квадратик земли расширился, многокрасочно расписался и превратился как бы в рельефную карту.» [2]



Рис. 1. Изображение одного и того же участка местности на карте (слева) и в Google Earth (справа). В последнем случае удаётся воссоздавать ракурсы просмотра обстановки, невозможные с точки зрения классической картографии.

При этом столь же очевидно, что Google Earth относится к предметной области, которую исторически принято идентифицировать как «картография». Следовательно, речь должна идти о существовании как минимум двух различных классов геопространственных продуктов — а, следовательно, о необходимости классификации всего их многообразия и выделения факторов, определяющих внутреннюю логику научной эволюции — т. е. типологизации.

Противоречие между очевидным отнесением Google Earth к картографии и столь же очевидным несоответствием его определению карты интенсивно обсуждалось в первые годы после появления Google Earth. В ходе этих дискуссий обозначились три основные группы позиций: 1) никакой качественной новизны в Google Earth нет и быть не может — это та же карта, улучшенная благодаря использованию неких инновационных технологий; 2) Google Earth является характерным представителем нового класса геопространственных продуктов, получивших условное обозначение «неогеография», полностью обособленных от классической «картографии», находящихся в состоянии противоборства с последней и фактически уже окончательно победивших её; 3) Google Earth стал итогом качественного развития картографического метода, и тем самым — новым научным принципом, преодолевшим ограничения, имманентно присущие картографическому методу в его классическом варианте.

Первая точка зрения постепенно сошла на нет, поскольку было очевидно отсутствие в Google Earth каких-либо специфических технологий и/или социальных практик, которыми можно было бы объяснить достигаемый с его помощью качественно новый уровень геопространственной визуализации. В Google Earth и в Цифровой Земле используются, конечно, наряду со старыми и новые технологии, однако каждая из них давно применяется и в классических картографических продуктах, и в ГИС, не приводя при этом к появлению какого-либо нового качества.

Вторую точку зрения отстаивал автор термина «неогеография» Эндрю Тёрнер [26], первым попытавшийся дать определение этому термину. Ему удалось проиллюстрировать методологическую изолированность неогеографии от классических карт и ГИС, однако его утверждение о наличии между ними непримиримой борьбы, уже увенчавшейся победой одной из сторон², следует, по всей видимости, признать преувеличением.

Третья точка зрения нашла выражение в первом эмпирическом определении неогеографии, сформулированном на рубеже 2007-2008 гг. Оно фактически обособило как отдельный класс Google Earth и его аналоги и тем самым отождествило Цифровую Землю и неогеографию.

² См., напр., презентацию Э. Тёрнера «Как неогеография уничтожила ГИС» (How Neogeography killed GIS) [27].

В этом определении [5] были выделены три признака, отличающие неогеографию от карт:

- 1) отказ от обязательного использования картографических проекций, использование непосредственно географических координат;
- 2) представление геопространственного контекста не опосредующими его картографическими условностями, но непосредственно растровыми изображениями — данными дистанционного зондирования;
- 3) использование гипертекстовых механизмов, в первую очередь для доступа к семантике.

Каждый из трёх вышеперечисленных признаков не является достаточным сам по себе, однако сочетание всех трёх позволяет получить принципиально новый режим геопространственной визуализации, отличный от карт и картографических продуктов. Режим этот обеспечивает одновременно всемасштабность и всеракурсность и достигается за счёт использования не редуцированных массивов данных. Карта, наоборот, является итогом необратимой трёхэтапной редукции исходных данных в процессе 1) проецирования, 2) генерализации, и 3) отнесения исходных данных к конечному набору объектных слоёв. Digital Earth, наоборот, позволяет обеспечить интерактивное и бесшовное изменение одновременно и ракурса просмотра сцены, и эффективной дистанции до неё как раз за счёт отсутствия обязательной редукции исходных данных.

Указанные признаки прямо противоречат базовым принципам классической картографии³, в силу чего можно говорить о возникновении нового принципа визуализации геоданных, отличного от картографического. При этом новые информационные продукты допускают включение в них любых картографических продуктов — мир карт и ГИС оказывается «погруженным» в гетерогенный мир Digital Earth, становясь одной из его возможных компонент [24]. Любая карта может быть визуализирована в среде Digital Earth — в то время как обратное невозможно.

Отнесение карт и Цифровой Земли к одной и той же предметной области, качественное различие обоих типов продуктов и этапность их возникновения позволяют разработать типологию картографической визуализации, т. е. классифицировать режимы геопространственной визуализации по их существенным признакам, отражающим логику их внутреннего развития. Следует отметить, что классификация и «цифровых земель», и в особенности карт — хорошо разработанная тема, которой посвящены многочисленные работы и целые научные направления (см. например, работы [1, 19]). В то же время создание единой классификации всех геопространственных продуктов от карт до Цифровой Земли является одновременно и новой, и актуальной темой.

В качестве классификационного признака для создания типологии логично выбрать степени свободы в выборе ракурса просмотра сцены и детальности её отображения в силу их фундаментальности. Обеспечение свободы, т. е. возможности произвольно, бесшовно и интерактивно изменять ракурс просмотра и удалённость от сцены, нуждаются в реализации в продуктах свойств полимасштабности и полиракурсности, а также, в случае

³ «Why, then, do we still flatten the Earth and still teach the frustrating mathematical complexity of map projections? Why did Google adopt the Web Mercator projection for Google Maps, instead of its own virtual globe? Now that most of the original reasons for projecting have disappeared.» [15]

геопространственной визуализации, достоверного отображения геопространственного контекста. При таком выборе классификационных признаков мы можем разделить всё множество геопространственных продуктов на два основных класса — класс одномасштабных и однокурсных продуктов (обычные и цифровые карты, ГИС), класс внесмасштабных продуктов (Цифровая Земля) и два промежуточных, паллиативных класса — глобусы и геопорталы [13] (рис. 2).

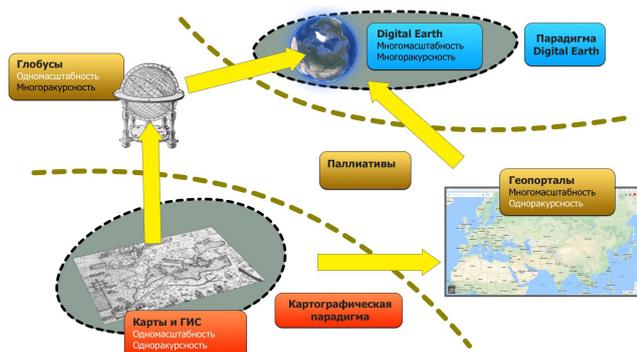


Рис. 2. Типология геопространственных визуализаций

Интересна ретроспектива развития геопространственных методов. Самой древней из известных науке полноценных географических карт, при создании которой осознанно были использованы все три фундаментальных принципа классической картографии, является, по-видимому, карта неолитического населённого пункта Чатал-Гююк в Малой Азии на территории нынешней Турции, с достаточной степенью надёжности датируемая 6600 г. до н.э. [25]⁴. Тем самым карта Чатал-Гююка оказывается как минимум в два раза старше самых древних из египетских пирамид и сопоставима по возрасту с древнейшими образцами знаковых письма.⁵ При этом на протяжении всей последующей истории человечества метод классической картографии принципиально не менялся — вплоть до наших дней, когда появилась Цифровая Земля. Но ни одно из уже обозначившихся направлений при этом своего развития не прекратило — например, появились цифровые глобусы, хотя они и имеют заведомо ограниченную функциональность и носят преимущественно имиджевый характер.

Можно ли распространить предложенную типологию картографических визуализаций на всю область научной визуализации? Данную постановку вопроса следует считать оправданной и актуальной [6], ведь типология визуализаций отсутствует, отсутствуют зачастую даже базовые определения отдельных методов. Так, по всей видимости, отсутствует дефиниция голографической визуализации как таковой, т.е. именно как специфического метода визуализации, в то время как методики создания голограмм и восстановления облика предметов и сцен с их помощью изучены весьма подробно. Парадоксальность такой ситуации подчёркивает востребованность голографии именно как специфического метода визуализации; она была даже предвосхищена и детально описана в литературе ещё до изобретения не только голографии, но даже лазера [8].

Идея голографической визуализации интуитивно понятна — это всеракурсность, позволяющая произвольно, интерактивно и бесшовно выбирать ракурс просмотра сцены. Востребованность такого режима визуализации очевидна. Однако столь же очевидно, что всеракурсность в визуализации должна дополняться ещё и всемасштабностью, позволяющей при необходимости визуализировать объект или сцену не только под любым ракурсом, но и в максимально широком диапазоне масштабов и без отрыва от контекста, в который погружена сцена. Особенно это необходимо в том случае, если отображаемой сценой выступает Земля или произвольный район Земли.

Классическая голография как метод получения и последующего восстановления частично-объёмных изображений обеспечить всемасштабность не может. В то же время в Цифровой Земле сочетаются и всемасштабность, и всеракурсность. Такой режим было предложено назвать «сверхголографией» [7]. С точки зрения предложенной типологии этот режим является предельным. Вопрос о необходимости включения в типологию визуализации других классификационных признаков (например, спектральных характеристик, наличия стереоскопии, и т.д.) нуждается в дальнейшем рассмотрении.

Всемасштабности удалось достичь благодаря использованию растровых изображений, не нуждающихся в генерализации и позволяющих обеспечить исключительно большой динамический диапазон. Так, в Google Earth мозаики, построенные из космических снимков высокого (субметрового) и низкого разрешения (LandSat), позволяют обеспечить уже сейчас, с доступными на рынке данными дистанционного зондирования разрешением около 0,5 м, корректное моделирование обстановки при изменении масштаба на 7-8 порядков, и нет никаких технических ограничений для ещё большего расширения этого диапазона за счёт всё более и более детальных данных. Такая гибкость разительно контрастирует с классическими картами, применимыми лишь в строго определённом масштабе и в строго определённой проекции. Цифровая Земля обеспечивает высокую эффективность визуализации, резкое снижение затрат за счёт отказа от необходимости формирования и обновления карт в широком диапазоне масштабов, и одновременно является существенной проблемой для семиотики — ведь новое качество достигается парадоксальным образом за счёт то ли возврата к простым и «примитивным» знакам, то ли отказа от использования знаков вообще.

С точки зрения семиотики, носителями информации могут являться знаки и только знаки, а сама информация определяется как мера знаков, или объём данных. В картографии используются и классические картографические знаки, и изображения, которые также считаются специфическими знаками — так называемыми знаками иконическими, или наглядными. Отождествление изображений со знаками⁶ следует признать оправданным в том случае, когда изображение создавалось человеком и субъективно в силу этого. Но правомерно ли отнесение изображений, полученных аппаратным способом — например, данных дистанционного зондирования — к иконическим знакам и к знакам вообще? Современная

4 Необходимо отметить, что в работе [21] взгляд на находку в Чатал-Гююке как на географическую карту оспаривается.

5 Тёртерийские таблицы – древнейшие образцы знаков в Европе – примерно на тысячу лет моложе [20]

6 “Изобразительный, или иконический знак подразумевает, что значение имеет единственное, естественно ему присущее выражение. Самый распространенный случай - рисунок.” Ю. Лотман [9].

аппаратура позволяет получать изображения чрезвычайно высокого качества, которые всё труднее отличать от прямого восприятия обстановки посредством органов чувств. Более того, из них поэтапно исключается даже такой минимальный источник субъективизма, как выбор композиции (глобальные мозаики космоснимков, сферические панорамы, сферическое видео). Постепенно стирается грань между восприятием обстановки с помощью геовизуализации и прямым, не опосредованным условностями восприятием. Оправданно ли современную геовизуализацию рассматривать в качестве знаковой реальности, или же речь должна идти о принципиально ином механизме передачи информации, не нуждающемся в знаках вообще?

Данный вопрос следует отнести к существенным проблемам современной семиотики. Она описывает знаковую, искусственную реальность, что изначально предполагает генезис знаков из незнаковых существей. Это подразумевает наличие восприятий, знаками не являющихся. Более того, описание знаковой реальности обязано сопровождаться ясным выделением знаков из «не-знаков», поскольку в противном случае мы приходим к ситуации порочного круга — любое восприятие по необходимости трактуется как проявление знаковой реальности, что обесмысливает саму идею знака. В то же время, семиотика и тем более картосемиотика не исследуют пороговую область между знаками и не-знаками, нулевыми знаками, неявно утверждая отсутствие иных средств восприятия обстановки, нежели знаки.

Для преодоления данного парадокса было предложено ввести в семиотику понятие о «нулевом знаке» [4] по аналогии с нулём в математике для обозначения существей, которые, не являясь знаками, способны передавать информацию и воссоздавать метрически достоверный образ обстановки. К таковым следует отнести данные дистанционного зондирования. В настоящее время предложение о введении в семиотику понятия о «нулевом знаке» и распространение его на данные дистанционного зондирования является предметом дискуссий [12]⁷.

3. Адаптация концепции Цифровой Земли

Появление концепции Цифровой Земли привело к образованию Международного Общества Цифровой Земли (International Society for Digital Earth, ISDE). Общество было зарегистрировано в Китае в 2006 году как неправительственная и некоммерческая международная организация, ставящая своей целью поддержку академических контактов, научных и технологических инноваций, образования, а также международного сотрудничества в создании Цифровой Земли.

Существенный импульс развитие Цифровой Земли получило десятилетие спустя после запуска Google Earth, когда доминировавшее на начальном этапе представление о Цифровой Земле как о технологической и социальной инновации сменилось растущим пониманием её как нового научного принципа в картографии и, вероятно, новой научной революции междисциплинарного и

общецивилизационного значения [14] в терминах теории научных революций Т. Куна [23].

Важным фактором развития Цифровой Земли стала континентальная инфраструктурная евразийская инициатива, известная как «Шёлковый пояс и путь». В апреле 2017 года под эгидой ISDE был основан альянс «Цифровой Шёлковый Путь» (Digital Silk Road Alliance, DSR), целью которого является формирование системы поддержки принятия решений континентального масштаба в рамках сообществ «Шёлковый Пояс» и «Шёлковый Путь» на базе трёх основных подходов: Цифровой Земли, Больших Данных, и космических данных дистанционного зондирования [18]. Менее чем месяц спустя первый проект континентального масштаба на базе Цифровой Земли анонсировала Австралия, выделив на эти задачи \$15,3 млн. Проект «Цифровая Земля Австралия» ставит своей целью создание обновляемого каждые пять суток покрытия континента и прилегающих акваторий космоснимками разрешением 10 м [28]. Наконец ещё через месяц, 22 мая 2017 года, совещание по вопросам политики в области дистанционного зондирования Земли прошло у Президента РФ; одновременно с этим Роскосмос анонсировал концепцию «Цифровой Земли» в качестве основы новой долгосрочной национальной политики России в области мониторинга Земли из космоса. Поставлена задача создания и периодического обновления покрытия всей планеты космическими снимками как глобальной геопространственной основы, необходимой для создания Цифровой Земли [10].

4. Заключение

Концепция Цифровой Земли была выработана как средство преодоления ограничений, имманентно присущих классическому картографическому методу, и привела к качественному скачку в области картографии. Новый картографический метод возник на базе новых технологий и социальных практик, но ни в коей мере не сводится только лишь к ним. Его основой является новый научный принцип, исключающий необходимость необратимого редуцирования данных и сводящийся к переходу от использования математического аппарата проекций к математическому аппарату подобия. Сравнительный анализ особенностей двух геопространственных методов позволил предложить типологию методов визуализации в картографии на базе двух существенных классификационных признаков — многоаккурности и многомасштабности.

Достичь такого результата удалось вследствие отказа от использования опосредующих реальность картографических знаков в качестве главного средства представления общегеографического контекста — в этом качестве их заменили прямые изображения местности, получаемые средствами дистанционного зондирования. С точки зрения семиотики эта ситуация парадоксальна, поскольку, на первый взгляд, противоречит доминирующей тенденции поэтапного перехода от простейших наглядных знаков-изображений ко всё более сложным и более абстрактным знакам и означает возврат то ли регрессию к иконическим знакам, то ли замещение знаков не-знаковыми носителями информации — нулевыми знаками». [11].

С 2017 года началось активное включение концепции «Цифровой Земли» в системы управления в качестве среды для поддержки выработки решений в Австралии (проект «Цифровая Земля Австралия»), в Евразии (альянс «Цифровой Шёлковый Путь»), а также в России (проект «Цифровая

⁷ Понимание логической неполноты семиотики без представления о нулевой исходной сущности присутствовало и раньше, однако, по всей видимости, лишь в качестве метафоры. См., например: «Peirce's basic triad: 'One, Two, Three'. 'One' is preceded by 'Zero', that 'emptiness' from whence the sign emerged» [22].

Земля»). Активное вовлечение Цифровой Земли в процессы управления свидетельствует о её высокой актуальности и о необходимости её дальнейшего изучения как специфического метода визуализации.

В целом можно отметить, что результаты, полученные в ходе разработки концепции Цифровой Земли, очевидным образом не просто выходят за рамки очередной сугубо технологической инновации, но и имеют отчётливые признаки смены научной парадигмы и в этом качестве заслуживают критического анализа с междисциплинарных позиций.

5. Благодарности

Работа выполнена при частичной поддержке грантов РФФИ №16-07-00955 А и 17-55-53109.

6. Литература

- [1] Багров, Лев. История картографии. - М.: ЗАО Центрполиграф. 2004. - 319 с. ISBN 5-9524-1078-2.
- [2] Булгаков, М.А. Мастер и Маргарита. - СПб.: «Азбука-Классика», 2006. - с. 416.
- [3] ГОСТ 21667-76. Картография. Термины и определения (с Изменением 1, 2). Статья 10 — Карта. Дата введения с и з м е н е н и я м и № 2 — 2 4 . 0 5 . 2 0 0 1 . <http://docs.cntd.ru/document/gost-21667-76> Проверено: 18 июня 2017 г.
- [4] Ерёмченко Е.Н. Концепция знака в контексте неогеографии // Информационные и математические технологии в науке и управлении. 2016. №1. С. 49-54.
- [5] Ерёмченко Е. Н. Неогеография: особенности и возможности // Материалы конференции «Неогеография XXI-2009» IX Международного Форума «Высокие технологии XXI века». Москва. 22–25 апреля 2008 г. 2008. - 170 с.
- [6] Ерёмченко Е.Н., Клименко С.В. К вопросу о типологии методов научной визуализации. Материалы конференции SCVRT2015-16. 2016. Протвино. С. 12-14.
- [7] Ерёмченко Е. Н., Тикунов В. С. Голографические возможности визуализации в географии. Вестник Московского университета. Серия 5. География. 2016. № 2. С. 22–29.
- [8] Ефремов И. Тень минувшего. Москва. Книга по требованию. 1945 (издание 2011) - 40 с.
- [9] Лотман Ю.М. Семиотика кино и проблемы киноэстетики. Т а л л и н : Ээсти Р а а м а т , 1 9 7 3 . http://lib.ru/CINEMA/kinolit/LOTMAN/kinoestetika.txt_with-big-pictures.html Проверено: 24 июня 2017 г.
- [10] Роскосмос. Проект «Цифровая Земля» стимулирует рост российской экономики. <https://www.roscosmos.ru/23571/> Accessed June 24, 2017.
- [11] Соломоник, А.. Наглядность в знаках // Геоконтекст. 2014. №1. сс. 31-38.
- [12] Соломоник, А. Неогеография и картография: семиотическая оценка // Геоконтекст. 2013. №1. сс. 14-25.
- [13] Eremchenko E., Tikunov V., Ivanov R., Massel L., Strobl J. Digital Earth and evolution of cartography. “4th International Young Scientist Conference on Computational Science, 2015” Procedia Computer Science. 2015. pp. 235-238.
- [14] Geospatial Revolution Project. The Pennsylvania State University. 2010. <http://geospatialrevolution.psu.edu/> Accessed: September 8, 2015.
- [15] Goodchild, M. Perspectives on the new cartography (Commentary). Environment and Planning A. 2015, volume 47, pages 1341–1345 doi:10.1177/0308518X15594911
- [16] Google Earth. <http://earth.google.com> Accessed June 29 2017.
- [17] Gore, A. The Digital Earth: Understanding our planet in the 21st Century. Al Gore speech at California Science Center, Los Angeles, California, on January 31, 1998.
- [18] ISDE10 held in Australia. <http://www.digitalearth-isde.org/news/804> Accessed June 24, 2017.
- [19] Kyeong P., Eunmi C.. The Enhanced Classification of Earth Globes. ICAMG-8 Conference Paper. 2015. https://www.researchgate.net/publication/283007830_The_Enhanced_Classification_of_Earth_Globes Accessed June 24, 2017.
- [20] Lazarovici G., Merlini M. New Archeological data refering to Tartaria tablets. Documenta Praehistorica XXXII. Univerza v Ljubljani. 2005. pp. 205-219.
- [21] Meece, S. A bird’s eye view - of a leopard’s spots. The Çatalhöyük ‘map’ and the development of cartographic representation in prehistory. Anatolian Studies 56:1-16. 2006. http://www.isde5.org/al_gore_speech.htm Accessed June 03, 2017
- [22] Merrell, F. Charles Sanders Pierce's Concept of the Sign. The Routledge Companion to Semiotics and Linguistics. Edited by Cobley P. N. Y.: Routledge, 2001. pp. 28-39.
- [23] Kuhn, T. The Structure of Scientific Revolutions. The University of Chicago, Chicago, 1970 (2nd edition, enlarged), 210 p.
- [24] Leonov A., Serebrov A., Anikushkin M., Belosokhov D., Bobkov A., Eremchenko E., Frolov P., Kazanskiy I., Klimenko A., Klimenko S., Leonova V., Rashidov A., Urazmetov V., Aleynikov A., Droznin V., Dvigalo V., Leonov V., Samoylenko S., Shpilenok T.. Virtual story in cyberspace: Valley of Geysers, Kamchatka. UNESCO Natural Heritage in Cyberworld // in Proc. CW2010 International Conference on CyberWorlds, 2010, Singapore, October 20-22, ISBN: 978-0-7695-4215-7, pp.247-253. DOI 10.1109/CW2010.38
- [25] Schmitt A., Danisik M., Aydar. E, et al. Identifying the Volcanic Eruption Depicted in a Neolithic Painting at Catalhoyuk, Central Anatolia, Turkey. PLoS ONE. V.9. I. 1. January 2014. pp. 1-9.
- [26] Turner, A. Introduction to neogeography. Sebastopol, CA: O’Reilly. 2006. P. 54.
- [27] Turner A. How neogeography killed GIS.. Presented at AGI UK GeoCommunity. 2009. <http://docslide.us/technology/how-neogeography-killed-gis.html> Accessed Sep 08, 2015.
- [28] ‘World first’ Digital Earth Australia announced in Federal Budget. <http://www.spatialsource.com.au/company-industry/world-first-digital-earth-australia-announced-federal-budget> Accessed June 24, 2017.

Об авторах

Аноприенко Александр Яковлевич, кандидат технических наук, профессор Донецкого национального технического университета. Его e-mail anoprien@gmail.com.

Ерёмченко Евгений Николаевич, научный сотрудник географического факультета Московского государственного университета имени Ломоносова, руководитель группы «Неогеография», научный сотрудник ГНЦ ПМБ, член совета и сопредседатель Outreach Committee ISDE. Его e-mail eugene.ermchenko@gmail.com.

Клименко Станислав Владимирович, доктор физико-математических наук, профессор Московского физико-технического института, научный руководитель кафедры МФТИ, директор Института физико-технической информатики (Протвино). Его e-mail stanislav.klimenko@gmail.com.

Создание цифровых моделей рельефа в трёхмерном пространстве методами фрактальной геометрии

Брылкин Ю.В.

maderator@yandex.ru

МФ ФГБОУ ВПО МГТУ им. Н.Э. Баумана, Мытищи, Россия

Статья посвящена созданию объёмной модели рельефа с помощью фрактальной геометрии для постановки первоначальных граничных условий при проведении землеустроительных работ и мониторинге местности. Для достижения поставленной цели в статье описывается и визуализируется работа алгоритма на основе синтеза шумов. Показаны положительные и отрицательные стороны данного алгоритма и возможности его практического применения.

Ключевые слова: цифровая модель, рельеф, фрактальная геометрия.

Digital elevation models creation in three-dimensional space fractal geometry methods

Brylkin Yu.V.

maderator@yandex.ru

Mytishinskiy branch FGBOU VPO «Bauman Moscow State Technical University», Mytishi, Russia

The article is devoted to creation of three-dimensional terrain model using fractal geometry for setting the initial boundary conditions when carrying out land management and area monitoring. To achieve this goal, the article describes and visualizes the algorithm is based on the noise synthesis. The positive and negative aspects of this algorithm and its practical application.

Keywords: digital model, terrain, fractal geometry.

1. Введение

Цифровые модели рельефа значительно уменьшают временные и трудовые затраты по отношению к традиционной технологии разметки топографических планов. Проектирование на основе моделей в трёхмерном пространстве предусматривает анализ объёмов работ, трудозатрат и факторов безопасности, для чего необходима полная информация о местности в широком диапазоне.

Наиболее используемыми методиками получения высотных отметок являются топографическая съёмка местности, радарная съёмка и использование лидарных систем.

2. Геометрическая модель рельефа

Процесс моделирования рельефа ограниченной территории по данным спутниковой съёмки хорошо применим для постановки граничных условий и сбора первоначальных данных о перепадах высот в пределах рассматриваемой области.

Получение имитационной модели местности подобным образом – наиболее простой и надёжный способ визуального моделирования, находящий применение при проведении землеустроительных, сельскохозяйственных и иных работ (рис. 1).

Основным недостатком данного способа является сложность моделирования пересечённой местности. Этот недостаток компенсируется простотой редактирования трёхмерной модели при проведении первичных геодезических работ и уточнением, например, аэросъёмкой, границ и координат положения объектов, попадающих в район предполагаемых работ.

В то же время, для решения проблемы первичной детализации на всех уровнях существует концепция фрактальной геометрии [7].

Модели, создаваемые с использованием фрактальной геометрии, на протяжении последних десятилетий являются основой многих процедурных алгоритмов для построения инженерных поверхностей [3].



Рис. 1. Построенная на основе спутниковых данных модель территории МФ ФГБОУ ВПО МГТУ им. Н.Э. Баумана, Мытищи, Россия.

Одним из методов фрактальной геометрии, получившим наибольшее распространение при работе с компьютерной графикой является дробное броуновское движение – шаговое интегрирование чисто случайного процесса по времени, аналогичное подбрасыванию монетки. Однако природный рельеф не может быть определен приближением, созданным с помощью дробного броуновского движения. Поэтому лучшие результаты дают поверхности, имеющие лишь некоторые качества истинной броуновской поверхности.

Кроме того, приближения могут быть вычислены поразному, и наиболее интересным методом создания поверхности является синтез шумов. Его принцип заключается в следующем:

– изображение, полученное с помощью спутниковой съёмки, двумерно преобразовывается с помощью случайного гауссовского белого шума с распределением Пуассона (рис. 2);

– полученный результат преобразуется в пространственную область с помощью обратного преобразования (рис. 2).

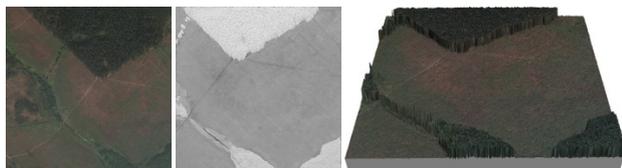


Рис. 2. Создание трёхмерной модели.

Спутниковая съёмка местности размером 1×1 км, изображение после применения фильтра гаусса с распределением Пуассона и объёмная визуализация.

Далее разрабатывается параметризованный алгоритм, который может использоваться многократно для создания более детальной визуализации того или иного участка поверхности. По сравнению с цифровыми моделями, построенными по данным, полученным вручную и использующимися только один раз, многократное использование синтеза шумов возможно [1] в виде:

$$H_{L_{min}}^{L_{max}}(x, y) = \sum_{l=L_{min}}^{L_{max}} R_{max} N(\lambda^l x, \lambda^l y) dS(x, y)$$

В этой формуле H – показатель Херста [6]; l – представляет собой уровень детализации; λ – шаг между каждым из последующих уровней шума. Тогда $\lambda^{L_{min}}$ и $\lambda^{L_{max}}$ являются соответственно мельчайшей и крупнейшей деталями, относительно наименьшего масштаба. Это означает, что $L_{max} = L_{min} + 1$ определяет количество суммируемых функций шума. Увеличение количества расчетных масштабных уровней добавляет дополнительные требования к вычислительным мощностям. Как результат, среднее число шумов удваивается примерно в два раза на каждом уровне. Из-за этого удвоения числа шумов, масштабные уровни называются октавами, что заимствовано из акустики. За R_{max} принята наибольшая высота профиля, согласно ГОСТ 2789-73, и в данном случае контролирует искусственную шероховатость. Соответственно в данном случае она может быть представлена в виде функции λ и экспоненты $\beta = 1 + 2H$, выведенной в работе [2]. Соотносятся между собой эти три параметра как $R_{max} = \lambda^{\beta/2}$.

После этого проводится расчёт фрактальной размерности D по формуле:

$$D = -\lim_{l \rightarrow 0} \frac{\ln S(\lambda^{L_{min}})}{\ln \lambda^{L_{min}}}$$

Здесь S – представляет собой количество объектов размера $\lambda^{L_{min}}$, требуемое для "покрытия" площади объекта. На основании этого определяется параметр самоподобия H (Херста). При этом фрактальная размерность находится в прямой зависимости от показателя Херста $D = 3 - H$. В случае, когда H находится в интервале $0 \div 1$, фрактальная размерность лежит между $2D$ объектом (плоской поверхностью) и $3D$ объектом (объёмной фигурой), при условии, что будет рассчитываться бесконечное число уровней. Это даёт возможность превратить выходной параметр H в следующий шаг обработки:

$$H'(x, y) = P(H(x, y))$$

Это лишь один пример использования процедурного алгоритма в качестве входных параметров. Следует отметить, что двухмерное изображение само по себе может быть процедурно созданным фракталом.

На основании данных спутниковой съёмки, эти простые преобразования делают прозрачными и понятными для пользователя механизмы создания имитационной математической модели рельефа, что даёт следующие преимущества:

- скорость создания и эффективность;
- не требуется видимость между соседними высотными отметками;

- возможность получения данных о рельефе в любой точке Земли;
- минимизация человеческих ресурсов на стадии сбора данных о рельефе;
- цифровая форма записи;
- меньший объём, требующийся для хранения и обработки данных о поверхности;

– применение стандартных форматов записи для последующего конвертирования в более удобную для ландшафтного дизайнера среду с привычным интерфейсом.

Надо отметить, что данный метод использовался при моделировании открытой местности в среде Autodesk 3Ds Max.

Использование имитационных цифровых моделей местности, созданных на основе спутниковых снимков, может найти применение при:

- определении границ участков;
- определении объёмов пространства, занимаемого лесами;
- выделении контуров вырубок, гарей, буреломов, участков, пораженных болезнями и вредителями, просек, дорог и т.д.;
- мониторинге вырубок и лесопатологий;
- моделировании процессов лесовосстановления;
- создании карт земель лесного фонда.

Повторные, регулярные или нерегулярные съёмки позволяют получать комплекс разнообразных характеристик и обеспечивать различные сервисы. Хорошим примером является оперативный сбор информации о последствиях природных катаклизмов. На их основе, с помощью фрактальных подходов делается перерасчёт объёмов пространства, занимаемого лесами.

3. Практическое применение

При наблюдении за территорией с координатами долготы и широты, соответствующими $61^{\circ}10'09''N$ и $55^{\circ}10'28''E$, на протяжении последних восьми лет, наглядно видно изменение лесного массива вследствие пожаров с 2010 года (рис. 3).



Рис. 3. Визуализация спутниковых снимков в 2009, 2010 и 2017 годах, пострадавшей от пожаров территории.

На основе данных моделей появляется возможность оценки площади лесного массива с помощью метода, описанного в [8]. Данный метод состоит в том, чтобы разделить площадь на небольшие квадраты с ребром δ . Площадь, занимаемая лесами, вычисляется по формуле:

$$A = N(\delta) \cdot \delta^2$$

Здесь $N(\delta)$ – число квадратов, необходимых для покрытия площади лесов, а δ^2 – площадь каждого из квадратов.

В результате было выяснено, что доля площади, занимаемая лесами на данном участке, сократилась на 31,88%.

В то же время, стоит отметить, что моделирование лесных пожаров и наблюдение разных стадий их развития должно производиться иными способами, например, с помощью авиационной гиперспектрометрии [5].

4. Заключение

С помощью моделей, воссоздающих рельеф местности в трёхмерном пространстве, становится возможным проведение инвентаризации лесных угодий; отслеживание

изменений, представляющих интерес; мониторинг общего состояния лесного покрова для государственных нужд или для частных компаний; контроль незаконных рубок. Кроме того, возможен оперативный сбор информации о последствиях природных или антропогенных (бури, лесные пожары) катаклизмов [4]. На их основе ставятся конкретные задачи по ликвидации последствий. Однако, в ряде случаев, для уточнения обстановки, могут потребоваться данные наземного наблюдения.

5. Литература

- [1] Giliam J.P. De Carpentier. Interactively synthesizing and editing virtual outdoor terrain. MA thesis. Delft University of Technology. – 2007.
- [2] Miller G.S. The definition and rendering of terrain maps // SIGGRAPH '86. – NY: ACM Press. – 1986, – p. 39-48.
- [3] Sakai S. et al. Fractal geometry of the ground surface and urban heat island // The seventh International Conference on Urban Climate. Yokohama, Japan, – 2009.
- [4] Евтушенко Н.В., Никитский А.Н., Митькиных Н.С., Черемисова А.М. Ежегодный спутниковый мониторинг весеннего половодья на реках России // Т-Comm № 2, – 2013, – с.46-48.
- [5] Козодеров В.В., Кондранин Т.В., Казанцев О.Ю., Бобылев В.И., Щербаков М.В., Борзяк В.В., Дмитриев Е.В., Егоров В.Д., Каменцев В.П., Беляков А.Ю., Логинов С.Б. Обработка и интерпретация данных гиперспектральных аэрокосмических измерений для дистанционной диагностики природно-техногенных объектов // Исслед. Земли из космоса. 2009. № 2. С. 36–54.
- [6] Мандельброт Б. «Фрактальная геометрия природы». М.-Ижевск: Ижевский институт компьютерных исследований, НИЦ «Регулярная и хаотическая динамика». – 2010 г., 656 стр.
- [7] Маркс А. Мониторинг лесов с помощью группировки спутников RapidEye // Геоматика, – 2011, – №3 (12), – с.58-66.
- [8] Федер Е. Фракталы. Пер. с англ. - М.: Мир, 1991. - 254с., ил., стр. 20-21.

Об авторах

Брылкин Юрий Владимирович, аспирант кафедры инженерной графики МФ ФГБОУ ВПО МГТУ им. Н.Э. Баумана. Его e-mail maderator@yandex.ru.

Пространственно-временная визуализация и перспективы прогнозирования геофизических полей

Т.П. Варшанина
vtp01@mail.ru

Адыгейский государственный университет, Научно-исследовательский институт комплексных проблем, Центр интеллектуальных геоинформационных технологий, Майкоп, Россия

Разработка методов визуализации пространственно-временной структуры геофизических полей рассматривается как необходимое основание для создания в рамках проекта «Цифровая Земля» модели глобального геоинформационного пространства, онтологически связывающего все земные сферы и геообъекты в единую саморазвивающуюся систему.

Ключевые слова Метод визуализации геофизических полей, прогнозирование геофизических процессов, глобальная онтологическая ГИС.

Spatial-temporal visualization and prospects of forecasting geophysical fields

T.P. Varshanina
vtp01@mail.ru

Adyge State University, the Research Institute of Complex Problems, the Center of Intellectual Geoinformational Technologies, Maykop, Russia

The development of methods for visualization of the space-time structure of geophysical fields is considered as a necessary basis for creating a model of a global geoinformation space within the "Digital Earth" project that ontologically links all terrestrial spheres and geobjects to a single self-developing system.

Keywords: Method of visualisation of geophysical fields, forecasting geophysical fields, ontological GIS of global space.

1. Введение

В возможности и стратегию развития проекта Цифровая Земля входит формирование глобального геоинформационного пространства, онтологически связывающего все земные сферы и геообъекты в единую систему нашей планеты. Осуществляемая в настоящее время формальная систематизация в ГИС пространственно-временных данных о геосферах и геообъектах производится на единой координатной основе. Методология визуализации многообразных и многомерных пространственно-временных географических данных на этой основе имеет многолетнюю историю становления и продолжает совершенствоваться [5, 6, 8]. Несмотря на это, современные виртуальные модели не обеспечивают отображения системных свойств географического пространства, вычисления интегральных параметров географических систем и входящих в них объектов, параметров их онтогенетической взаимообусловленности. Таким образом, в настоящее время создаётся систематизированный свод многообразных пространственно-временных координированных географических данных, непригодных для системного анализа и успешного их прогнозирования. Не приносят ощутимого успеха в прогнозировании и частные виртуальные модели, основанные на данных инструментального мониторинга параметров геосфер.

Для разработки онтологической классификации географических систем, основанной на параметрах - индикаторах их онтогенетической взаимообусловленности, необходимо опираться на фундаментальные законы современной квантово-полевой картины мира [4], в соответствие с которой все материальные объекты являются порождением взаимодействующих энергетических полей. Структура поля энергии у земной поверхности обусловлена интерференцией волновых, в первую очередь, гравитационных и электромагнитных

воздействий приходящих из космического пространства и из глубин Земли. Сложное взаимодействие воздействий приводит к тому, что в каждый отдельный интервал времени одни и те же составляющие воздействий могут в различной степени или усиливать друг друга или ослаблять. Это является одной из основных причин нелинейности и неопределённости природных процессов.

С этих позиций универсальной основой для выявления онтологической структуры всех геопространственных объектов, различающихся по вещественному составу и фазовому состоянию, является порождающая их структура поля энергии. В настоящее время использование этого подхода является единственно возможным путём системного представления глобального геоинформационного пространства Земли. Тот факт, что системообразующие геофизические поля когерентны полному спектру космических и планетарных воздействий, снимает проблему неопределённости при прогнозировании природных процессов.

2. Визуализация геофизических полей

Для перехода от формальной систематизации объектов географического пространства к их онтологической классификации и системному анализу применяются численные эксперименты визуализации геофизических полей на основании информации, заключённой в пространственно-временной структуре земных объектов. В географическом пространстве принято выделять три генеральных геофизических поля: эндогенное – результирующее поле сложной системы энергетических потоков в глубинах Земли; экзогенное, обусловленное, в первую очередь, приходящей солнечной радиацией; «реальное» геофизическое поле, возникающее в результате интерференции энергетических воздействий, приходящих из Космоса и глубин Земли.

Для параметрической визуализации геофизических полей разработан метод структурной маски энергетических географических полей (СМЭП) [1]. Метод обеспечивает визуализацию структурно подобного отображения геофизических полей через градиентные поля параметров, определяющих структуру геобъекта или геосферы, например: для рельефа – это параметры поля высот; для атмосферы – параметры поля температуры.

Так, напряжённость эндогенного геофизического поля в фокусе земной коры отвечает за формирование тектонического рельефа и является его параметром порядка, т.е. динамической переменной, определяющей структурную организацию системы тектонического рельефа и каждый уровень укрупнения и усложнения его структуры. Так как поле высот является структурообразующим параметром тектонического рельефа, то оно служит мерой его параметра порядка. Пространственно-временную динамику отметок высот на территории за предыдущие миллионы лет до настоящего времени характеризует положение потенциальных поверхностей выравнивания, проведённых через тальвеги однопорядковых водотоков. Мощность разностного слоя между смежными поверхностями выравнивания свидетельствует о скорости поднятия/опускания элементарных объёмов геологической среды за период между формированием смежных поверхностей выравнивания определённого геологического возраста. Наибольший градиент мощности разностного слоя характеризует относительную скорость и направление движения элементарных объёмов геологической среды в её реологических условиях.



Рис. 1. Генеральное поле тектонических напряжений, визуализированное в изогипсах (изолиниях мощности разностного слоя).

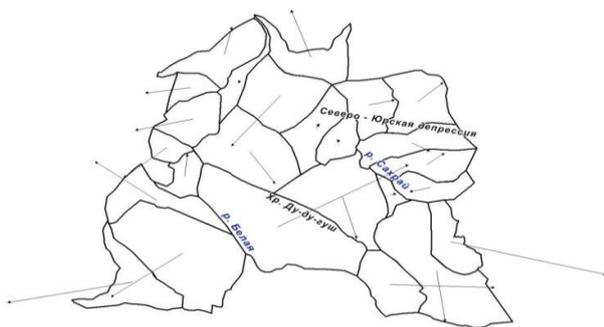


Рис. 2. Фоновое поле тектонических напряжений, визуализированное результирующими от центра тектонического блока наибольшими градиентами мощности разностного слоя.

Методом СМЭП вычисляется и визуализируется дискретно-континуальные, пространственно-временные иерархические геофизического поля тектонических

напряжений: континуальное генеральное (рис. 1); дискретное фоновое, движущее тектонические блоки (рис.2); локальное, возникающее при взаимодействии тектонических элементов (рис.3).

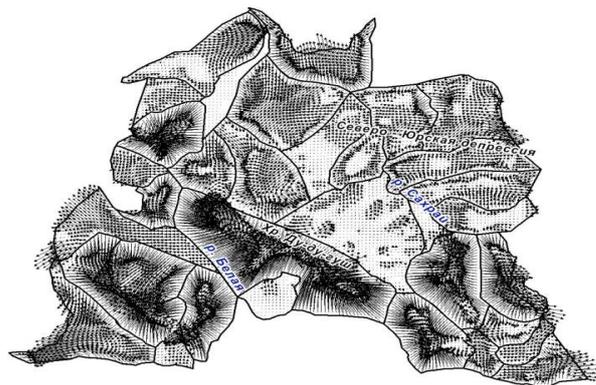


Рис. 3. Локальное поле тектонических напряжений, визуализированное результирующими наибольшими градиентами мощности разностного слоя относительно узлов равномерной решётки в границах тектонического блока.

Визуализации экзогенного энергетического поля в фокусе земной поверхности предшествует разработка модуля построения структурной 3D геометрической поверхности рельефа, так как геометрическая поверхность рельефа структурно подобная эндогенному полю энергии формирует структурно подобную этой поверхности поле потенциальной экзогенной энергии. Модуль информационно-математической модели рельефа классифицирует его гомоморфные поверхности в результате следующего порядка действий [7]:

- аппроксимация горизонталей NURBS (неоднородный рациональный сплайн) кривыми 4-го порядка;
- выделение авторским алгоритмом особых точек и структурных линий рельефа;
- идентификация с использованием продукционных правил экспертной системы элементарных поверхностей рельефа;
- построение 3D геометрической модели поверхности рельефа с помощью аппроксимации каждой элементарной поверхности поверхностью Безье (рис. 4).

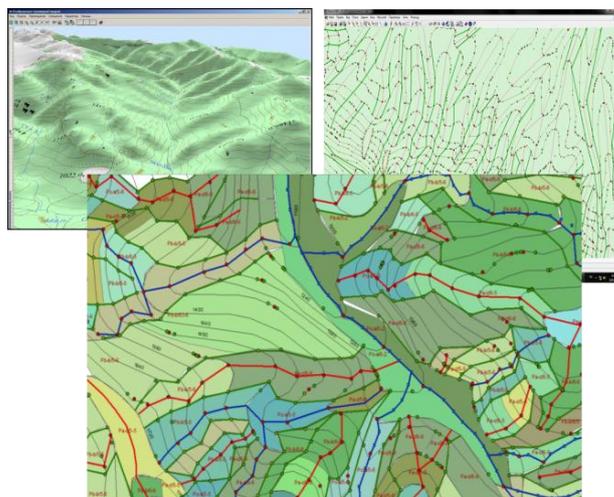


Рис. 4. Этапы построения структурной 3D геометрической поверхности рельефа.

Градиентное поле потенциальной экзогенной энергии рассчитывается относительно величины суммарной солнечной радиации, приходящей на каждую гомоморфную единицу структурной 3D геометрической поверхности рельефа, а также в границах их структурной иерархии.

В результате изложенного порядка действий визуализируется отображение онтогенетической связи страт эндогенного и потенциального экзогенного полей энергии, взаимообуславливающих формирование всех объектов геопространства.

При построении структурной маски «реального» геофизического поля в фокусе земной поверхности необходимо учитывать термодинамические особенности поверхности, характеризующие её энергетический потенциал – то, как эта поверхность принимает и расходует энергию. В данном случае индикатором может служить величина биопродуктивности. Таким образом, структура «реального» геофизического поля отображается путём построения градиентных полей интенсивности продуцирования биомассы по результатам дешифрирования мультиспектральных космоснимков по NDVI (нормализованный разностный индекс растительности).

Воссозданные страты геофизических полей:

- с одной стороны представляют собой структурную маску порождающих геообъекты энергетических полей, с другой – отображают структуру географических объектов или процессов;

- когерентны полному спектру космопланетарных энергетических воздействий, что снимает проблему неопределённости при прогнозировании;

- отображают инвариантное пространство состояний своих параметров, отвечающих за формирование геообъектов, в отличие от данных непродолжительных инструментальных наблюдений.

Пространство состояний трёх страт полей энергии, формирующих иерархию географических объектов в их онтогенетической взаимообусловленности и зависимости от воздействий космических и глубин планеты, формализует фундаментальный закон географической изменчивости и иерархической систематизации геообъектов [3].

Создаются условия для разработки глобальной онтологической ГИС, предназначенной для логического связывания в координированном поле наземной, подземной и надземной информации для междисциплинарных исследований саморазвивающейся системы планеты.

Онтологическую связь рассчитанных геофизических полей и географических процессов и поэтому возможность использования параметров геофизических полей в качестве предикторов доказывают разработанные впервые 1) модель точечного прогнозирования времени наступления и уровня паводков [2] и 2) модель точечного прогнозирования землетрясений.

1) Мерой параметра порядка синоптического процесса, отображающего структуру локального энергетического поля приземной атмосферы, служит ежесуточный градиент температуры, рассчитываемый в точке прогнозирования относительно территории равной площади барического образования. Ежесуточный ход градиента температуры приземной атмосферы является интегральным показателем, характеризующим:

- состояние энергетического поля приземной атмосферы в точке прогнозирования;

- ту часть барического образования, которая проходит через точку прогнозирования по параметрам ежесуточного ход градиента за 7-10 дней;

- скорости адвекции и трансформации воздушной массы, состояние подстилающей поверхности по параметрам суточного хода градиента.

Ход регионального градиента температуры позволяет классифицировать выборки, соответствующие типам синоптических ситуаций в точке прогнозирования.

Для уменьшения влияния свойства нелинейности природных процессов точечное прогнозирование паводка осуществляется с помощью математической модели нейронной сети. В модель включается ход параметров ежесуточного регионального градиента температуры за 10-20 лет и соответствующие ему уровни воды на гидрологическом посту.

Трехуровневая нейросетевая модель классифицирует синоптические ситуации в точке прогнозирования по их подобию и преобладанию и даёт точечный прогноз времени наступления и уровня паводка. Способ отработывался на примере четырех малых рек бассейна реки Кубань и показал высокую степень достоверности среднесрочного прогнозирования даты паводка и погрешность прогнозирования уровней от 3 до 14% (рис. 5).

Универсальность способа продемонстрирована на примере постпрогноза катастрофического паводка на р. Адагум в г. Крымск в 2012 (рис.6).

Разработанный способ обеспечивает точечное прогнозирование полного спектра гидрометеорологических явлений, имеющих ряды инструментальных наблюдений. Предполагается, что расчёт генерального геофизического поля приземной атмосферы и исследование его инвариантных состояний позволит идентифицировать вклад локальных антропогенных воздействий различного происхождения на изменение климата.

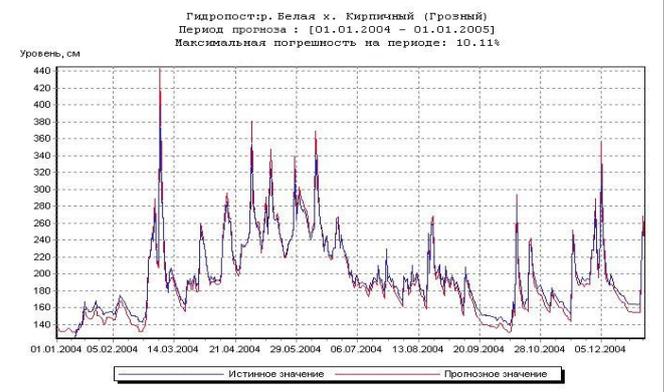


Рис. 5. График постпрогноза уровня подъема воды. Майкоп, р. Белая, 2004 год. Погрешность на периоде – 10,11%.

2) Открыты перспективы для точечного прогнозирования землетрясений. Наибольший градиент мощности разностного слоя также является интегральным показателем, характеризующим направление, относительную скорость движения тектонических блоков, их реологические свойства, тип тектонического взаимодействия (напряжения сжатия, растяжения, сдвига) [9].

Расчёт структуры локального энергетического поля тектонических напряжений произведён в макроблоке южной оконечности Скифской плиты. Фоновое значение наибольшего градиента мощности разностного слоя макроблока за четвертичный период выявляет невысокую относительную скорость движения на юго-восток с концентрацией тектонических напряжений в области

Скалистого хребта. В пределах макроблока за историческое время известно 21 сейсмическое событие магнитудой от 2,0 – 5,0 (рис. 7).

Относительно каждого эпицентра землетрясений вычислен результирующий наибольший градиент мощности разностного слоя – индикатор тектонического напряжения, характеризующий его среднюю величину за четвертичный период (за 2,5 млн. лет). Величина множественной корреляции между результирующими наибольшими градиентами мощности разностного слоя в макроблоке, вычисленными за четвертичный период, и магнитудой землетрясений зафиксированных в 20-ом веке составляе 0,43, в каждом блоке, входящем в макроблок она того же порядка – 0,45 [4, 7].

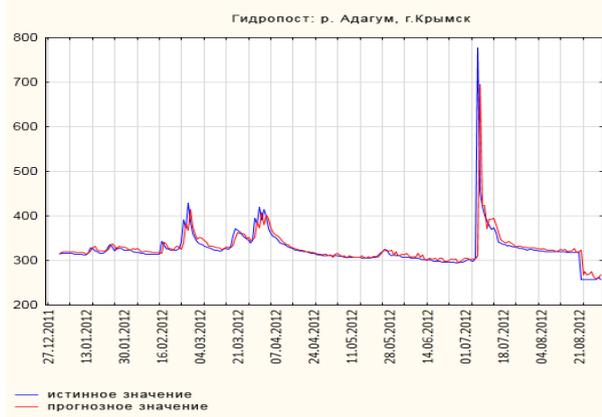


Рис. 6. График постпрогноза уровня подъема воды. Гидропост р. Адагум г. Крымск, 2012 год. Максимальная погрешность на периоде – 14,04%.

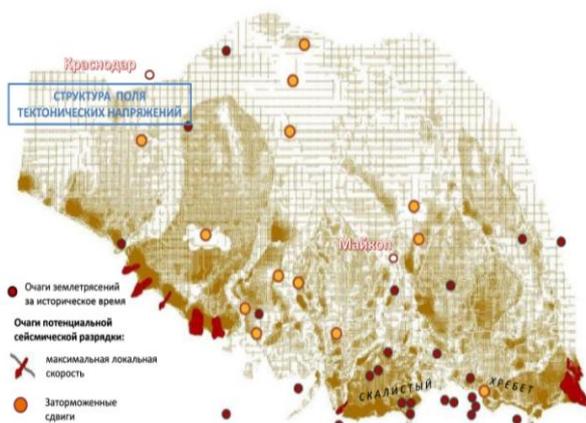


Рис. 7. Визуализация очагов потенциальной сейсмической разрядки в локальном поле тектонических напряжений.

Таким образом, мониторинг наибольшего градиента мощности разностного слоя, характеризующего напряжённость эндогенного геофизического поля в выявленных очагах потенциальной сейсмической разрядки, и магнитуды сейсмических проявлений позволит с помощью эвристических методов осуществлять точечное прогнозирование места, времени и магнитуды сейсмического события.

3. Заключение

Таким образом, визуализация результатов численного моделирования геофизических полей обеспечивает:

- формализацию фундаментального закона географической изменчивости и иерархической систематизации геобъектов, что открывает перспективы конструирования экологически сбалансированных культурных ландшафтов и прогнозирования развития каждого из них в соответствии с изменяющимися условиями среды;

- условия для разработки глобальной онтологической ГИС в целях междисциплинарных исследований законов саморазвития системы планеты;

- точечное прогнозирование полного спектра природных процессов, подтверждённых данными инструментальных наблюдений;

- наполняет новым содержанием проект «Цифровая Земля» для решения стратегически важных проблем в сфере наук о Земле.

4. Литература

- [1] Варшанина Т.П. Разработка хорошо структурированной модели геопространства на основе метода структурной маски энергетических геополей. Вестник Адыгейского государственного университета. Серия естественно-математических и технических наук. №4(110). 2012. С. 176-179.
- [2] Варшанина Т.П., Плисенко О.А., Коробков В.Н. Способ прогнозирования времени наступления и уровня паводков. Патент № 2480825, 2013г.
- [3] Варшанина Т.П., Хунагов Р.Д. ГИС научных исследований процессов изменчивости геосистем. Вестник Адыгейского государственного университета. Серия естественно-математических и технических наук. №4(147). 2014. С. 207-214.
- [4] Варшанина Т.П. Общенаучные предпосылки и перспективы прогнозирования геопространственных процессов в интересах устойчивого развития территорий. ИнтерКарто/ИнтерГИС 22. Геоинформационное обеспечение устойчивого развития территорий в условиях глобальных изменений климата. Т. 1. – М. Издательский дом «Научная библиотека». 2016. С. 3-13.
- [5] Е.Н. Еремченко, В.С. Тикунов. Голографические возможности визуализации в географии. Бюлл. Московского ун-та. Серия географии. 2016. N 2. С. 22-29.
- [6] Косиков А.Г., Ушакова Л.А. Виртуальное моделирование и многомерные географические модели. Успехи современного естествознания. Науки о Земле. №2, 2016. С. 164-169.
- [7] Плисенко О.А. Применение экспертных систем в ГИС для пространственного анализа рельефа. ИнтерКарто/ИнтерГИС 22. Геоинформационное обеспечение устойчивого развития территорий в условиях глобальных изменений климата. Т. 1. – М. Издательский дом «Научная библиотека». 2016. С. 157-167.
- [8] Тикунов В.С., Еремченко Е.Н. Цифровая земля и картография // Геодезия и картография. 2015. № 11. С. 14–23.
- [9] T. Varshanina, R. Khunagov, O. Plisenko, V. Korobkov, E. Shtelmakh. Modelling the Space-Time Field of Tectonic Stresses in the Area of Unlimited Space for Seismic Safety

Indian Journal of Science and Technology, Vol 8(29),
DOI: 10.17485/ijst/2015/v8i29/86864, November 2015.

Об авторах

Варшанина Татьяна Павловна, к. биол. наук, доцент кафедры географии Адыгейского государственного университета, заведующая Центром интеллектуальных геоинформационных технологий НИИ комплексных проблем АГУ, директор малого инновационного предприятия 'ГеоВертекс'.
e-mail vtp01@mail.ru

РАСПОЗНАВАНИЕ СИМВОЛОВ НА БОЛЬШЕФОРМАТНЫХ РАСТРОВЫХ ИЗОБРАЖЕНИЯХ ДОКУМЕНТОВ СО СЛАБО ФОРМАЛИЗОВАННЫМ ОПИСАНИЕМ ОБЪЕКТОВ

Д. Ю. Васин¹, М.А. Редкин¹
dm04@list.ru|maxop151@mail.ru

¹ ИИТММ Нижегородского Государственного Университета им. Н.И. Лобачевского, Нижний Новгород, Россия

В докладе рассмотрены алгоритмы формирования эффективных, устойчивых к искажениям признаков и решающих правил распознавания объектов, представленных на большеформатных растровых изображениях со слабо формализованным описанием объектов. Рассмотрен вопрос влияния метода бинаризации исходного растрового описания на качество последующего автоматического распознавания объектов.

Ключевые слова: растровые изображения, моментные характеристики изображения, низкоуровневая структурированная модель описания бинарного раstra, дискриминантные признаки, распознавание.

CHARACTER RECOGNITION ON LARGE-SCALE RASTER IMAGES OF DOCUMENTS WITH WEAKLY FORMALIZED DISCRIPTION OF OBJECTS

D. Y. Vasin¹, M. A. Redkin¹
dm04@list.ru|maxop151@mail.ru

¹ ITMM Nizhny Novgorod State University. N.I. Lobachevsky, Nizhny Novgorod, Russia

In the report the algorithms of building effective, resistant to distortion characteristics and decision rules of recognition of objects presented on large format raster images with weakly formalized description of objects. Considers influence of the method of binarization of the original raster descriptions on the quality of the subsequent automatic recognition of objects.

Keywords: bitmap, torque characteristics of the image, low-level structured model describing the binary image, discriminant features, recognition.

1. Введение

В настоящее время актуальной остается проблема создания электронных архивов большеформатных графических документов (топографических и морских навигационных карт, конструкторской документации, технических чертежей, электрических схем и др.). При этом источником, как правило, являются графические изображения на бумажных носителях. С целью автоматизации технологии ввода графических документов (ГД) в последние годы предложены различные информационные технологии на основе эвристических процедурных методов, а также на базе методов распознавания с учителем, которые эффективны для ограниченного набора объектов с жестким ограничением на размер и ориентацию. Однако, значительный класс большеформатных ГД выполняется ручным способом со слабо формализованным начертанием объектов, произвольной ориентацией, произвольными размерами символов, наличием большого количества межобъектных наложений, примыканий и пересечений (рис. 1).

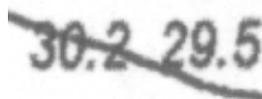


Рис. 1. Пример изображения ГД со слабо формализованным начертанием объектов.

Все это существенно снижает качество автоматического распознавания объектов, что, в конечном итоге, сказывается на эффективности всей технологической цепи создания электронных архивов большеформатных ГД.

Одной из основных проблем, возникающих при решении задачи распознавания образов и существенно влияющих как на эффективность распознавания, с одной стороны, так и на его качество, с другой, является выбор системы эффективных признаков классификации.

В работе [1] рассматривалось формирование решающих правил в задаче распознавания символов, находящихся в стандартной горизонтальной ориентации, но допускающих растяжение/сжатие по горизонтали/вертикали отдельных символов, на базе низкоуровневой структурированной модели описания бинарного растрового изображения (БРИ) [2]. Настоящая работа является продолжением исследований по созданию эффективных систем распознавания объектов по растровым данным. В частности, рассмотрен вопрос влияния различных методов бинаризации исходного полутонового растрового изображения на конечный результат распознавания объектов, расширен класс рассматриваемых документов.

Очевидно, что качество формируемых дискриминантных признаков, определяемых по низкоуровневой структурированной модели описания БРИ, напрямую зависит от качества выполнения операций предобработки и бинаризации исходного полутонового растрового изображения. Дело в том, что неудачный выбор порога бинаризации приводит к появлению различных шумов, которые вносят значительный вклад в снижение качества итогового результата распознавания. Среди основных помех можно выделить наличие на распознаваемых по БРИ объектах: неправомерных разрывов или слипаний, появление изолированных пустот внутри объектов, наличие «снега» – изолированных черных объектов малого размера, изрезанность краев объектов.

2. Моментные характеристики изображения

В работе [1] были введены моментные характеристики, нашедшие широкое применение во многих задачах обработки цифровых изображений [3].

Момент – характеристика контура изображения, объединённая (суммированная) со всеми пикселями контура. Момент $m_{p,q}$ для непрерывного изображения $f(x,y)$ определяется как:

$$m_{p,q} = \iint x^p y^q f(x,y) dx dy,$$

в случае дискретного растрового изображения:

$$m_{p,q} = \sum \sum x^p y^q f(x,y),$$

где p,q – мощность, на которой соответствующий компонент взят в сумме с другими отображенными. Если функция $f(x,y)$ рассматривается как бинарное изображение, тогда момент m_{00} вычисляет площадь чёрных пикселей этого изображения.

Для вычисления характеристик, инвариантных относительно сдвига, необходимо центрировать относительно средних значений $\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}}$ –

центр тяжести фигуры.

Центральный момент определяется как:

$$\mu_{p,q} = \sum \sum (x-\bar{x})^p (y-\bar{y})^q f(x,y).$$

Доказано, что все центральные моменты выражаются через обычные, по формуле:

$$\mu_{p,q} = \sum_{i=0}^p \sum_{j=0}^q C_p^i C_q^j (-\bar{x})^{p-i} (-\bar{y})^{q-j} m_{ij},$$

где $C_n^k = \frac{n!}{k!(n-k)!}$.

С помощью центральных моментов определяются характеристики, инвариантные к повороту изображения (моментные инварианты). Моментные инварианты – это признаки изображения, построенные на основе степенных моментов и описывающие силуэт некоторого объекта. В соответствии со своим названием данные признаки являются инвариантными к аффинным преобразованиям изображения. Они стали важнейшим инструментом для распознавания образов независимо от их особого положения, ориентации, угла обзора и других изменений, также они нечувствительны к различным геометрическим преобразованиям (сдвиг, поворот, масштабирование) [4]. Это определяет их эффективность при использовании в качестве признаков в задаче обнаружения и распознавания на изображении объектов неизвестной ориентации.

Имея набор характеристик, можно определить следующие аффинные преобразования моментных инвариантов, которые инвариантны относительно сдвигов, поворотов, а также растяжений и сжатий:

$$I^1 = \frac{\mu_{20} + \mu_{02}}{m_{00}^2}, \quad I^2 = \frac{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}{m_{00}^4},$$

$$I^3 = \frac{(\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{30})^2}{m_{00}^5},$$

$$I^4 = \frac{(\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2}{m_{00}^5}.$$

Значения $I^1 - I^4$ используются в качестве признаков классификации.

3. Низкоуровневая структурированная модель описания БРИ

Полное описание модели приведено в [2]. Пусть БРИ представлено в пиксельном формате, при этом функция яркости пикселей имеет вид:

$$f(x,y) = \begin{cases} 255, & \text{если пиксель фоновый;} \\ 0, & \text{если пиксель не фоновый.} \end{cases}$$

Непроизводным элементом модели является *штрих* – одномерный кластер связанных пикселей заданного цвета, представленный набором чисел $S=(t, N, K, C)$, где: t – номер строки раstra; N, K – координаты начала и конца связанной одноцветной пиксельной области вдоль линии раstra; C – код цвета связанных пикселей, образующих штрих. Для бинарных изображений штрих описывается тройкой типа $S=(t, N, K)$, где: t – номер строки раstra; N, K – координаты начала и конца штриха, определяемые соотношениями:

$$N = x \wedge f(x-1, y) \wedge \bar{f}(x, y), \quad (1)$$

$$K = x \wedge \bar{f}(x, y) \wedge f(x+1, y).$$

Совокупность штрихов для заданной строки раstra t определяет строку штрихов $SH=\{S_i | i=1,2,...,m\}$, где S_i – штрих. Описание раstra с помощью набора строк штрихов является однозначным. Отметим, что штриховой формат представления и хранения растровых данных является в 10–20 раз более компактным по сравнению с пиксельным.

Топологические свойства *штриха* $S=(t, N, K)$ определяются через его связность со штрихами смежных строк. Два штриха S_t и S_{t-1} смежных строк t и $t-1$ связаны, если $(N_{t-1} \geq N_t) \wedge (N_{t-1} \leq K_t) \vee (K_{t-1} \geq N_t) \wedge (K_{t-1} \leq K_t)$. Связный *штрих* – это набор $H = \{S, sw_p, sw_s\}$, sw_p , sw_s – количество связанных штрихов со штрихами смежных строк, а t – номер текущей строки штрихов раstra. Тогда $sw_p=0$, если в предыдущей строке $p=t-1$ нет ни одного штриха, для которого выполняется условие:

$$(N_p \geq N_t) \wedge (N_p \leq K_t) \vee (K_p \geq N_t) \wedge (K_p \leq K_t) \quad (2)$$

иначе, $sw_p = \mu_p$, μ_p – кратность выполнения условия (2), т.е. количество связанных штрихов предыдущей строки. Аналогично, $sw_s=0$, если в следующей строке $s=t+1$ нет ни одного штриха, для которого выполняется условие:

$$N_s \geq N_t \wedge N_s \leq K_t \vee K_s \geq N_t \wedge K_s \leq K_t \quad (3)$$

иначе, $sw_s = \mu_s$, μ_s – кратность выполнения условия (3), т.е. количество связанных штрихов последующей строки. Здесь N, K – координаты начала и конца исследуемого на связность штриха, N_t, K_t – координаты начала и конца текущего штриха предыдущей смежной строки штрихов, N_s, K_s – координаты начала и конца текущего штриха следующей смежной строки штрихов.

По суперпозиции значений sw_p и sw_s можно ввести следующую классификацию графических ситуаций. В текущей строке:

штрих изолированный (ШИ): $sw_p = 0 \wedge sw_s = 0$; штрих начала растрового объекта (ШН): $sw_p = 0 \wedge sw_s = 1$; штрих конца растрового объекта (ШК): $sw_p = 1 \wedge sw_s = 0$; штрих слияния растровых объектов (Шс): $sw_p > 1$; штрих расщепления растровых объектов (Шр): $sw_s > 1$; штрих слияния и расщепления растровых объектов (Шср): $sw_p > 1 \wedge sw_s > 1$.

Алгоритм нахождения указанных особых штрихов приведен в [2].

Растровый простой объект (РПО) – кластер связанных штрихов, не содержащий графических ситуаций типа слияния и расщепления, то есть для любого штриха кластера выполняется условие $sw_p < 2 \wedge sw_s < 2$.

Растровый составной объект (РСО) – кластер связанных штрихов, для каждого штриха которого выполняется условие: $sw_p > 0 \vee sw_s > 0$, и для любых двух элементов (штрихов или пикселей) найдется хотя бы одна

соединяющая их пиксельная траектория, состоящая из связанных смежных элементов.

4. Постановка задачи

Особенностью ГД со слабо формализованным описанием объектов является то, что, несмотря на отклонение от нормативных требований к изображению объектов, они обладают некоторой стилизованной формой представления. Это означает, что возможно создание набора эталонов объектов, основанных на работе с низкоуровневыми моделями графических изображений, максимально адаптированных для данного класса ГД, с учетом сильной зависимости эффективности признаков классификации от искажений входных данных. В дальнейшем, по мере поступления новых документов, этот набор может соответствующим образом дополняться. Вместе с тем, требует внимания выбор наиболее эффективного метода бинаризации исходного растрового изображения, с целью минимизации возникающих в процессе бинаризации искажений.

В качестве исходных документов были взяты 24-битные растровые изображения текстов на английском языке, а также планшеты гидрографической съемки (рис. 2).

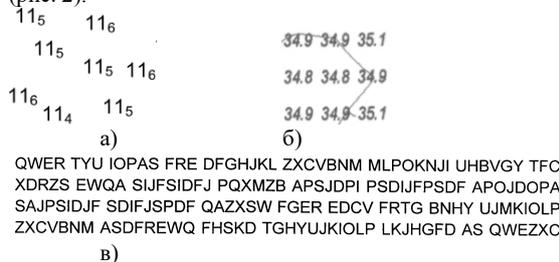


Рис. 2.Примеры исходных изображений: а) планшет гидрографической съемки хорошего качества; б) планшет гидрографической съемки среднего качества; в) отсканированный текст.

5. Признаки классификации

Для каждого объекта будем фиксировать следующие признаки:

1. Отношение сторон описанного вокруг символа прямоугольника со сторонами, параллельными осям координат (отношение аспекта).
2. Отношение площади символа к площади описанного вокруг символа прямоугольника со сторонами, параллельными осям координат (процент заполненности).
3. Значения моментных инвариантов $I^1 - I^4$.
4. Количество внутренних областей (ВО).
5. Количество особых штрихов: Шн, Шк, Шс, Шр.
6. Количество РПО.

Рассмотренный набор признаков не является окончательным и может быть расширен.

Число внутренних областей позволяет сразу разделить все объекты на 3 группы, т.е. объекты, у которых: ВО = 0 (К,Х,С, 1, 2 и др.), ВО = 1 (R,О,Р, 4,6,9 и др.) и ВО = 2 (В,8). Учтем, что «В» – единственный среди символов английского алфавита, а цифра 8 – единственная цифра среди арабских, имеющие ВО=2.



Рис. 3. РПО, выделенные на символах.

Количество РПО, а также их параметры, довольно хорошо характеризуют каждый символ (рис. 3) Символ “А” – 6 РПО, символ “К” – 5 РПО.

Дальнейшее распознавание объекта можно выполнять в пределах этих групп.

Число ВО – достаточно эффективный признак, отношение аспекта, и процент заполненности, более

устойчивы к различным искажениям распознаваемых данных.

6. Преобразование исходного РИГД в оттенки серого

Учитывая, что источником данных являются различные ГД, то основными цветами их представления будут белый (цвет фона) и чёрный (цвет объектов). Однако, изначальное сканирование исходного документа в режиме монохромного, может привести к существенным искажениям, т.к. довольно малое количество информации представлено чистыми черным и белым цветами. Для минимизации подобных искажений требуется постепенное уменьшение глубины цвета путём преобразования цвета всех пикселей изображения к требуемому диапазону. Изначально сканируем ГД с глубиной цвета 24 бит/пиксель. Данный режим сканирования является избыточным, но позволяет выполнить последующую программную бинаризацию РИГД с минимизацией возможных искажений. На первом этапе осуществляем конвертацию полноцветного РИГД в оттенки серого. Это осуществляется путем применения к каждому пикселю исходного РИГД преобразования вида:

$$Y = 0.299 R + 0.587 G + 0.144 B, \tag{4}$$

в котором R, G, B – красная, зелёная и синяя компонента цвета исходного РИГД соответственно, а Y – полученная яркость в оттенках серого.

6.1. Бинаризация РИГД

Нами было исследовано влияние различных методов бинаризации РИГД на конечный результат распознавания объектов.

Бинаризация РИГД методом Оцу. В методе Оцу для расчета порогового значения используется нормализованная гистограмма изображения [5]. Гистограмма строится по значениям $p_i = n_i / N$ где N – общее число пикселей изображения, n_i – число пикселей с уровнем яркости i , ($0 \leq i \leq L$). Гистограмма позволяет разделить все пиксели изображения на два класса: «полезные» и фоновые, с помощью порогового значения уровня яркости t (t – целое значение $0 \leq t \leq L$). Каждому классу соответствуют относительные частоты ω_0 и ω_1 :

$$\omega_0(t) = \sum_{i=1}^t p_i, \quad \omega_1(t) = \sum_{i=t+1}^L p_i = 1 - \omega_0(t).$$

Средние уровни для каждого из двух классов изображения: $\mu_0(t) = \frac{\sum_{i=1}^t ip_i}{\sum_{i=1}^t \omega_0}$, $\mu_1(t) = \frac{\sum_{i=t+1}^L ip_i}{\sum_{i=t+1}^L \omega_1}$.

Суть метода Оцу заключается в том, чтобы выставить порог между классами так, чтобы каждый из них был наиболее «плотным». Метод Оцу определяет порог, уменьшающий дисперсию внутри класса, которая определяется как взвешенная сумма дисперсий двух классов: $\delta_{\omega}^2(t) = \omega_1(t) \delta_1^2(t) + \omega_2(t) \delta_2^2(t)$.

Далее вычисляется максимальное значение оценки качества разделения изображения на две части:

$$\eta(t) = \max \left(\frac{\sigma_{кл}^2(t)}{\sigma_{\omega}^2(t)} \right), \quad \sigma_{кл}^2(t) = \omega_0(t) \omega_1(t) (\mu_1(t) - \mu_0(t))^2$$

– межклассовая дисперсия.

На рис. 4 приведен фрагмент удачно бинаризованного изображения планшета гидрографической съемки по методу Оцу.

Результат обработки имеет и явный недостаток: при обработке объектов, имеющих различные размеры, но близкие цвета, в результате бинаризации Оцу получаются темные пятна (рис. 5).

117 117 116
 117 116 116
 117 117

Рис. 4. Хороший результат бинаризации Оцу.

35.5 35.5 35.6
 35.4 35.6 35.6

Рис. 5. Неудачный выбор порога бинаризации Оцу.

Бинаризация с двойным ограничением. В этом случае пиксели, которые входят в диапазон $[t1, t2]$ – становятся черными, которые не входят – белыми [6].

$$F(x, y) = \begin{cases} 255, & f(x, y) < t1; \\ 0, & t1 \leq f(x, y) \leq t2; \\ 255, & f(x, y) > t2; \end{cases}$$

Пороговые значения яркости пикселей оценивались по гистограмме.

35.3 35.2 35.2
 35.3 35.4 35.6
 35.8 35.8 35.9

Рис. 6. Результат бинаризации с двойным ограничением.

На рис. 6 приведен фрагмент изображения планшета гидрографической съемки, обработанный по методу бинаризации с двойным ограничением.

Сравнивая результаты бинаризации (рис. 5 и рис. 6)

видно, что на данном классе документов метод бинаризации с двойным ограничением устраняет недостатки бинаризации по методу Оцу.

7. Этап подготовки изображения

На данном этапе изображения разбиваются на связные области чёрных пикселей (компоненты), каждая из которых является тем или иным объектом с присвоенной ему уникальной меткой и описанной вокруг нее прямоугольной областью. Отметим, что данная операция является наиболее затратной.

Во время выполнения данного этапа производится объектовая фильтрация БРИ, перевод БРИ в штриховой формат, а также выделение внутренних областей внутри компонент.

7.1. Алгоритм объектовой фильтрации исходного бинарного изображения

Заметим, что в целом, признаки, определяемые по низкоуровневой модели БРИ, не являются достаточно устойчивыми. Различного рода помехи на БРИ ведут к тем или иным изменениям штрихового описания, в результате чего могут быть выделены лишние штрихи расщепления, слияния, РПО а также внутренние области. Следовательно, эффективное использование штрихового описания для автоматического распознавания объектов требует выполнения предварительной фильтрации БРИ. Нами был разработан алгоритм объектовой фильтрации БРИ, суть которого в следующем. На вход алгоритма подается БРИ с распознаваемыми объектами. Входными параметрами являются код цвета удаляемой помехи и размеры описанного вокруг помехи прямоугольника со сторонами, параллельными осям координат, а также процент его заполненности пикселями цвета помехи. Работа алгоритма основана на выделении по исходному БРИ связных компонент, состоящих из пикселей с цветом, равным заданному, с последующим удалением (перекрашиванием в альтернативный цвет) тех найденных связных областей, которые по своим геометрическим характеристикам удовлетворяют параметрам помехи. Важным достоинством данного алгоритма является тот факт, что указанная фильтрация не вносит никаких искажений в те связные области, которые не распознаются алгоритмом как шумовые.

7.2. Алгоритм конвертации БРИ в штриховое представление

На вход алгоритма подается БРИ, представленное в пиксельном формате. На выходе алгоритма формируется БРИ, представленное в

штриховом формате. Конвертация осуществляется построчно, используя соотношения (1).

7.3. Алгоритм подсчета числа внутренних областей

Пусть имеется изображение, представленное в штриховом формате, с объектом, представленным единственной связной компонентой. В процессе работы на изображении находятся “особые” точки, хранящиеся в списке S .

1. Просматривая штриховой файл, находим штрихи расщепления.
2. Пусть штрих расщепления находится в столбце i . Просматриваем все связные с ним штрихи в столбце $(i+1)$, для каждого $H = (t, N, K)$ добавляем в список S особые точки $(i+1, N-1), (i+1, K+1)$.
3. Просматривая все особые точки (x, y) из списка:
 - 3.1. Если координаты точки выходят за рамки матрицы P , то удаляем её из списка как несуществующую.
 - 3.2. Запускаем поиск в ширину со стартом в (x, y) , при этом $цвет1$ – белый, $цвет2$ – любой дополнительный цвет (отличный от черного и белого).
 - 3.3. Проверяем все особые точки из S , следующие по списку за точкой (x, y) . Если они окрашены в $цвет2$, то удаляем их как лишние.

3.4. Просматриваем граничные точки всего изображения $[P_{ij}], [P_{Nj}], [P_{i1}], [P_{iM}]$, $i = 1, \dots, N; j = 1, \dots, M$. Если хотя бы одна из них будет закрашена в $цвет2$, значит точка (x, y) не принадлежит внутренней области объекта (замкнутая область белого цвета). Удаляем её из списка S . Количество элементов в списке S равно количеству внутренних областей в объекте.

7.4. Алгоритм построения классификатора

Просматривая все эталоны для каждого класса, фиксируем: для каждого признака принимаемые значения, а также математическое ожидание и доверительный интервал их значений.

7.5. Алгоритм автоматического распознавания символов

На вход алгоритма поступает вектор признаков связного объекта. На выходе распознанный символ алфавита базы эталонов или пустой символ. В алгоритме введено понятие штрафа – положительная величина, пропорциональная отношению расстояния до ближайшей границы интервала к длине этого интервала. Для признаков, сформированных по штриховому описанию: если значение признака меньше значения нижней границы интервала, то сразу начисляется штраф, больший порогового значения и класс убирается из дальнейшего рассмотрения.

Изначально формируем массив размером по количеству классов. В цикле по номеру признака:

1. Для каждого класса в соответствующую ему ячейку массива добавляется штраф, если значение признака не входит в допустимый интервал для данного класса.
2. Если штраф объекту для класса превосходит порог, то класс исключается из дальнейшего рассмотрения.
3. Если в рассмотрении остался один единственный класс, то соответствующий ему символ возвращается как результат распознавания.
4. Если были исключены все классы, то в качестве результата работы алгоритма возвращается пустой символ.

Если после завершения цикла осталось несколько классов, штраф которых не превзошел порог, то в качестве результата возвращается символ, соответствующий классу с минимальным штрафом.

7.6. Алгоритм автоматической сборки надписей

На вход алгоритма поступает набор описаний распознанных объектов. На выходе описание сформированной надписи в структуре выходной

тематической базы данных. Сборка осуществляется применением методов кластеризации. Из списка выбирается очередной непомеченный объект, который объявляется ядром нового кластера (надписи). Данный объект помечается уникальной меткой. Просматривается ϵ -окрестность относительно центра тяжести ядра кластера. Если в указанную ϵ -окрестность попадает центр тяжести очередного непомеченного объекта, то оценивается угол наклона вектора, соединяющего центр тяжести ядра кластера и объекта-кандидата на добавление в кластер. Если отклонение вектора от горизонтали не более, чем на заданную величину, то найденный объект присоединяется к кластеру, помечается меткой стартового объекта и для объектов, принадлежащих кластеру, высчитывается новый центр тяжести ядра, как среднее значение координат центров объектов, составляющих кластер. Поиск очередных объектов, принадлежащих формируемому кластеру, осуществляется относительно нового центра тяжести. Процесс повторяется до тех пор, пока на очередной итерации к кластеру не будет добавлено ни одного нового объекта. В результате, сформированный кластер может состоять либо из одиночного объекта, либо из нескольких объектов, расположенных на исходном документе вдоль строки. Процесс поиска новых кластеров завершается, если на входе нет непомеченных объектов.

8. Анализ сформированного признакового пространства

По обучающему тексту на английском языке, была сформирована выборка по 5 эталонов на символ. Анализируя значения мы пришли к выводу, что значений I^1 и I^2 достаточно чтобы выделить среди всех классов группы схожих, значения I^3 и I^4 необходимы для более тонкой классификации внутри группы. В целом, значения моментных инвариантов достаточно стабильны и потому распознавание начинается с них. Данные признаки с большой долей вероятности позволяют исключить множество классов, не относящихся к рассматриваемому объекту.

На практике, после использования моментных признаков остаётся 2 – 5 возможных классов, к которым может принадлежать распознаваемый объект. Для окончательного решения об отнесении объекта к определённому классу используются признаки, полученные из штрихового описания. У них есть одно замечательное свойство – чёткая нижняя граница возможных значений. Например, у символа “М” ни при каких условиях не может быть менее трех Шк, двух Шн и Шр, одного Шс и четырех РПО. Следовательно, по нижней границе штриховых признаков можно исключать классы, для которых объект нарушает указанные признаки. Данное утверждение справедливо для любого из распознаваемых нами объектов, при условии, что они представляются связными растровыми объектами.

9. Результаты распознавания

Шрифт написания обучающего и экзаменующего английских текстов – Consolas, размер шрифта – 14 пунктов. Обучающий текст содержал 1350, экзаменующий – 1850 символов без учета пробелов. Оба документа сканировались с глубиной цвета 24 бит/пиксель, с пространственным разрешением 300 dpi. Размеры полученной растровой матрицы 2500x3500 пикселей. На изображениях присутствовали только символы английского алфавита в стандартной ориентации. Изображения были последовательно переведены из полноцветных в изображения в оттенках серого, которые были бинаризованы по методу Оцу. По полученному обучающему БРИ сформирована выборка из 150 эталонов. Использование обучающего документа в качестве экзаменующего обнаружило следующие результаты распознавания: неправильно распознано – 8

объектов, непознанных объектов нет, качество распознавания более 99%. Высокий процент объясняется тем, что специально в качестве эталонных символов были выбраны наиболее различающиеся символы внутри своих классов. В результате были получены достаточно широкие интервалы допустимых значений, а сформированного набора признаков оказалось достаточно для распознавания. Эта же база эталонов была использована для проведения распознавания на экзаменующем документе. В результате: неправильно распознано 25 объектов, непознано – 15, качество распознавания $\approx 93 - 96\%$.

Для распознавания цифровых надписей на планшетах гидрографической съемки была сформирована база из 50 эталонов, по 5 эталонов на каждую цифру. При этом геометрический размер символов практически не играл роли в том смысле, что для разных размеров символов дискриминантные признаки принимают значения, хоть и слегка отличающиеся друг от друга, но гарантированно попадающие в границы рассчитанных доверительных интервалов. При этом качество распознавания было не ниже 85–90%. Основная нерешенная к настоящему времени проблема, снижающая качество распознавания – «рассыпанные» объекты–цифры и наложение на объекты–цифры тонких линейных объектов.

10. Благодарности

Работа выполнена при финансовой поддержке РФФИ – проекты №16-07-01214А и № 16-07-01198А.

11. Литература

- [1] Д.Ю. Васин, А.В. Аратский. Распознавание символов на основе инвариантных моментов графических изображений. // Труды юбилейной 25-ой Международной конференции GraphiCon 2015, 22-25 сентября 2015 г., Протвино, Парк Дракино, Московская область, стр.259 –264.
- [2] Д.Ю. Васин, Ю.Г. Васин, В.П. Громов. Структурное описание растровых данных. //Методы и средства обработки сложной графической информации: 6 Всероссийская конференция с участием стран СНГ, 25–27 сентября 2001г., Н.Новгород, стр. 21–23.
- [3] Dan S. Bloomberg, Gary E. Kopec and Lakshmi Dasari. Measuring document image skew and orientation. // Xerox Palo Alto Research Center. – Режим доступа: <http://www.leptonica.com/papers/skew-measurement.pdf>, свободный.
- [4] S. X. Liao, «Image Analysis by Moments,» *The Department of Electrical and Computer Engineering The University of Manitoba*, 1993.
- [5] N. Otsu. A thresholding selection method from gray-scale histogram // IEEE Transactions on System, Man, and Cybernetics, 1979. No. 9. P.62-66
- [6] M.I. Sezan, A peak detection algorithm and its application to histogram-based image data reduction //Computer Vision, Graphics, and Image Processing, 1990. Vol. 49 (1) P.36-51.

Об авторах

Васин Дмитрий Юрьевич – к.т.н., ведущий научный сотрудник Центра информатики и интеллектуальных информационных технологий Института Информационных технологий, математики и механики Нижегородского Государственного университета им. Н.И. Лобачевского (ИТММ ННГУ им. Н.И. Лобачевского).
E-mail: dm04@list.ru

Редькин Михаил Александрович – магистр 1 года обучения ИИТММ ННГУ им. Н.И. Лобачевского.
E-mail: maxop151@mail.ru

Разработка системы динамических тестов для начинающих и опытных водителей с применением технологий виртуальной и дополненной реальности

Н.Ю. Дудаков, И.Н. Мерзляков, А.Р. Ханнанов, Т.Н. Томчинская
dudakovnu@gmail.com|innermerzliakov@gmail.com|zionone@mail.ru|tomchinskaya@mail.ru
Нижегородский государственный технический университет

Исследуется создание приложения, предназначенного для визуализации дорожно-транспортных происшествий в реальном времени в режиме дополненной и виртуальной реальности на мобильной платформе с операционной системой Android. Приложение создано на базе Unity 3D, в нем реализованы сценарии развития различных моделей ДТП с привязкой к местности и взаимодействию с ситуацией на дороге. Приведены примеры, иллюстрирующие работу этого приложения.

Ключевые слова: дополненная реальность, виртуальная реальность, дорожно-транспортное происшествие, динамические тесты, виртуальная автошкола.

Development of a system of dynamic tests for beginners and experienced drivers using virtual and augmented reality technologies

N.Yu. Dudakov, Ig. N. Merzliakov, Ar.R. Khannanov, T.N. Tomchinskaya
dudakovnu@gmail.com|innermerzliakov@gmail.com|zionone@mail.ru|tomchinskaya@mail.ru
Alekseev State Technical University of Nizhny Novgorod, Russia

An application is considered for real-time visualization of augmented and virtual reality, enabling the consideration of traffic accidents on a mobile platform with the Android operating system. An example of an application developed by authors created on the basis of Unity 3D for visualizing various types of road accidents, linking them to terrain and interacting with integrated 3D scenes is given.

Keywords: augmented reality, virtual reality, road traffic accidents, dynamic tests, virtual driving school.

1. Введение

Проблема аварийности, связанная с автомобильным транспортом, в последнее десятилетие приобрела особую остроту в связи с несоответствием дорожно-транспортной инфраструктуры потребностям общества в безопасном дорожном движении, недостаточной эффективностью функционирования системы обеспечения безопасности дорожного движения и крайне низкой дисциплиной участников дорожного движения.

Согласно статистическим данным (stat.gibdd.ru), ежегодно на территории Российской Федерации происходит более 150 тысяч дорожно-транспортных происшествий. При этом свыше трех четвертей всех дорожно-транспортных происшествий связаны с нарушениями Правил дорожного движения Российской Федерации водителями транспортных средств и большой процент среди участников дорожно-транспортных происшествий составляют водители со стажем до трех лет.

Решение возникающих проблем не ограничивается развитием городской дорожной сети. Предлагаются различные интеллектуальные транспортные системы. Наиболее интересными, на наш взгляд, являются следующие работы [1-2]. Но, как правило, основной функцией этих систем является анализ аварийности, и они не поддерживают возможность реалистичной визуализации конкретного ДТП в живой обстановке. Одним из «новых трендов» в построении транспортных систем является технология отображения объектов в формате 3D с помощью дополненной реальности на лобовом стекле автомобиля. Приложения дополненной реальности в основном предназначены для навигации, помогающей водителям лучше ориентироваться на дороге [3-5]. Такое приложение может быть полезным для опытных водителей, т.к. закрывает частично лобовое стекло и начинающий

водитель, не умеющий достаточно быстро «сканировать» окружающую обстановку может не справиться с дополнительным потоком информации. Так, в ходе исследования было выявлено, что для начинающих водителей и курсантов автошкол наиболее острой проблемой является проблема корректного распознавания дорожной ситуации на дороге, в которой времени для принятия решения может почти и не быть. Действовать необходимо максимально быстро, при этом принять во внимание все аспекты текущего положения, такие как сигнал светофора, разметку, знаки, положение автомобилей и их скорости. К тому же, при попадании в экстремальную ситуацию, ученик может растеряться, что может привести к еще более плачевным результатам.

Решением данной проблемы может стать предлагаемое приложение для мобильной платформы предназначенное, прежде всего для начинающих водителей и курсантов автошкол.

Рассматриваются два пути обеспечения безопасности дорожного движения.

1. Предварительный просмотр незнакомого участка пути с использованием дополненной реальности. Наведя камеру смартфона на карту города, например, на Яндекс карту, водитель может увидеть реалистичную модель конкретного участка дорожной инфраструктуры на экране смартфона со всеми знаками и разметкой.

2. С помощью погружения в виртуальную реальность автомобилист может увидеть в стерео-режиме живую ситуацию произошедшего ДТП на конкретном участке реалистично смоделированной части конкретного города. Посмотреть данную ситуацию с разных точек зрения, как пострадавший, виновник или случайный свидетель и понять все причины данного происшествия. А также для курсантов автошкол предлагается система динамических тестов, выполненных на реально существующих участках с повышенной статистикой ДТП конкретных городов.

2. Блок дополненной реальности

Для реализации приложения на платформе Android использована интегрированная среда разработки Unity 3D. Сцены с дорожно-транспортными происшествиями на наиболее опасных участках города, предварительно разработаны в технологии низкополигонального моделирования в инструментальной среде Autodesk Maya и 3Ds Max, а затем сконвертированы под среду Unity 3D (рис.1).



Рис. 1. Низкополигональная трехмерная сцена с разметкой одной из площадей города.

В приложении создано несколько блоков, каждый блок отвечает за конкретный раздел. Используется интегрированный метод создания программы, заданы стандартные модули, а связи между ними прописаны с помощью языка программирования C#. На рисунке 2 представлен фрагмент меню данной подсистемы.

В блоке AR, дополненной реальности, с помощью плагина Vuforia запускается механизм поиска меток, заранее созданных на основе карт, и привязки к ней модели или всплывающей подсказки. На карту интегрируется сцена с наиболее опасными участками. С помощью плагина Vuforia распознается изображение одной из меток на карте, полученной с камеры смартфона, на основе информации о положении маркера в пространстве.

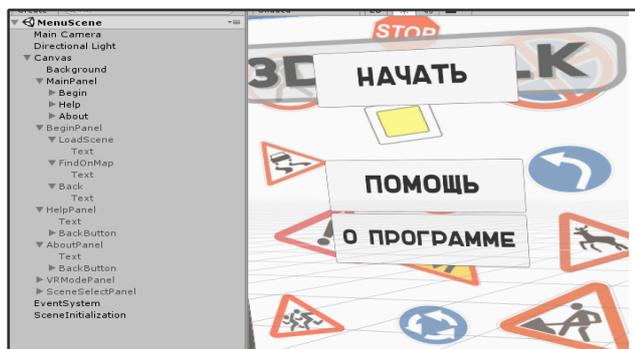


Рис. 2. Интерфейс приложения.

Затем происходит расчет опорных точек с одновременным трекингом, – тем самым трехмерный объект ориентируется в пространстве.

Приложение проецирует на метку соответствующую модель, и на экране загружается сцена случая ДТП. Сцена проигрывается в режиме просмотра от третьего лица (рис. 3).

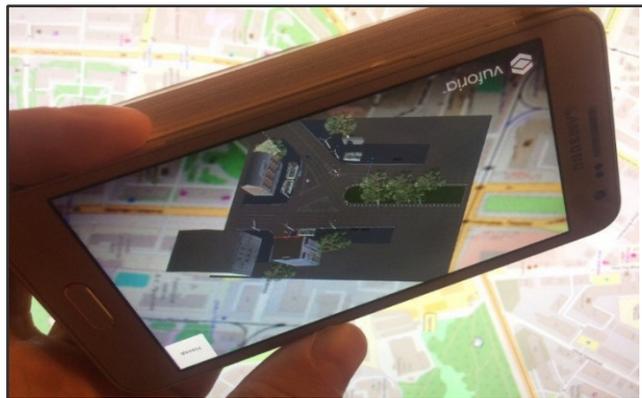


Рис. 3. Загрузка сцены на смартфон по метке карты.

3. Блок виртуальной реальности

Блок VR (виртуальная реальность) предназначен для просмотра модели в режиме виртуальной реальности без использования метки. Из списка предложенных сцен, пользователь выбирает интересующую его модель. На экране смартфона или планшета загружается сцена, с возможностью просмотра её с разных ракурсов и от разных участников дорожного движения с полным или частичным погружением.

Для полного погружения требуется шлем виртуальной реальности. На сцену добавляется виртуальная камера. Выбор сделан в пользу продукта Fibrum SDK, так у него имеются все необходимые настройки, и он не перегружает систему дополнительными плагинами. С помощью объекта Fibrum SDK установлено две камеры, для каждого глаза соответственно, расстояние между камерами, дисторсия изображения и требуемое цилиндрическое искажение. Для просмотра виртуальной реальности с полным погружением требуется экран с диагональю от 4 до 6 дюймов, так же необходимо иметь в телефоне датчики акселерометра и гироскопа. Это требуется для отслеживания положения смартфона в пространстве и как следствие – расположение виртуальной камеры во «внутриигровом» пространстве.

Для управления персонажем используется стандартный внутренний плагин Unity 3D - Characters Controller.

При загрузке сцены на смартфоне в режиме виртуальной реальности имеется возможность просмотра ситуации от различных лиц участников дорожного движения, будь то виновник, пострадавший или свободное перемещение по сцене. Когда пользователь находится в режиме свободного перемещения он может осмотреться вокруг, изучить местность и дорожную инфраструктуру окружающую его, или перемещаться по сцене с помощью стандартного джойстика управления (рис. 4). Кнопки выбора режима просмотра находятся сверху экрана. Кнопка назад – служит для выхода из сцены, в главное меню.

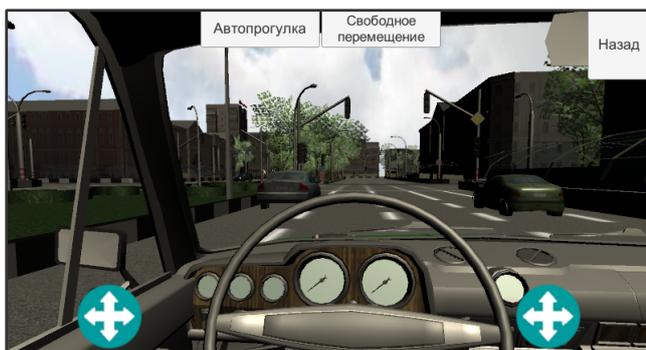


Рис. 4. Кнопки навигации и управления ориентацией в пространстве.

В режиме просмотра сцены виртуальной реальности с использованием очков виртуальной реальности (рис. 5) у пользователя нет возможности свободного перемещения по сцене. Для перехода к просмотру сцены от другого участника дорожного движения, пользователь должен «посмотреть» на кнопку до появления индикатора. После его заполнения, в течении нескольких секунд, программа перенаправит пользователя в выбранный пункт меню.

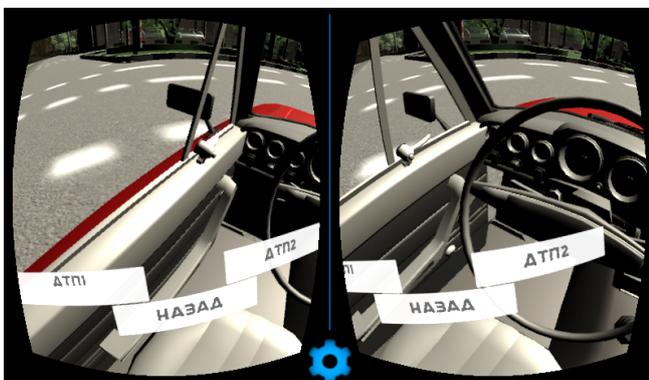


Рис. 5. Работа двух камер для получения стереоизображения.

Для перехода к просмотру сцены от другого участника дорожного движения, пользователю необходимо «посмотреть» на соответствующую кнопку на экране. Для более полного погружения, объекты внутри сцены отбрасывают тень от статических и динамических источников света. Для этого в родительский префаб RigidBody FPS Controller добавлен трехмерный объект цилиндр. Он отбрасывает тень от персонажа, но для того, чтобы в камеру не попадали полигоны объекта, он исключен из рендера. Чтобы тело имело свой объем и могло сталкиваться с другими объектами добавлен стандартный скрипт Mesh Collider, который делает все грани персонажа физическими и не дает ему проходить сквозь стены. Такой же плагин должен стоять на объектах, через которые пользователю запрещено проходить.

4. Блок динамического тестирования

В ходе обучения в автошколах курсантам предлагаются статичные иллюстрации в книжке с тестами, которые дают обучающемуся время на раздумье, против динамической ситуации на дороге, в которой времени для принятия решения крайне мало. Предлагается система динамических тестов, выполненных на реально существующих участках с повышенной статистикой ДТП, рассмотренных на примере конкретного города. Госавтоинспекция подготовила список

с наглядными схемами наиболее типовых нарушений ПДД. В ходе реализации проекта, за основу взяты наиболее опасные участки дорожно-транспортной развязки конкретного города.

В результате исследования этих участков и ДТП, составлен ряд тестовых заданий с ситуациями, в которых чаще всего происходят ДТП, а также ряд сложных участков дорог, с которыми возникают проблемы у водителей. На каждом из участков, рассматривается несколько вариантов заданий, исходя из реального расположения дорожных знаков, разметки и возможных препятствий (например, других автомобилей, нарушающих правила, в следствии чего водитель может не заметить дорожный знак).

Во время прохождения динамического теста, на экране монитора непосредственно перед обучающимся отображается вид из машины с водительского места (рис. 6). Тестируемый может вращать головой по сторонам, как это происходит в реальной жизни.



Рис. 6. Сравнение обычной системы тестов (на верхнем рисунке) и динамической.

Сама процедура прохождения происходит следующим образом. При приближении машины пользователя к этапу тестирования, например, перекрестку, тестирующемуся предлагается выбор из нескольких вариантов ответа, которые могут содержать то, или иное действие (рис. 7).

Если тестируемый выбрал правильный ответ, то испытуемому записывается балл, и машина продолжает движение в заранее заданную позицию. В случае выбора некорректного ответа, балл не записывается, о чем испытуемый будет проинформирован. В отдельных случаях, если некорректный выбор может привести к ДТП, воспроизводится сцена аварии.

Всего на пути тестируемого 20 тестовых этапов (заданий), по результатам которых оглашается окончательный балл и даются советы по неправильно пройденным этапам, если таковые имеются. На каждом этапе реализовано от 1 до 5 возможных исходов, где каждый исход – это задача, подгружаемая случайным образом в систему в зависимости от сделанного ранее выбора.

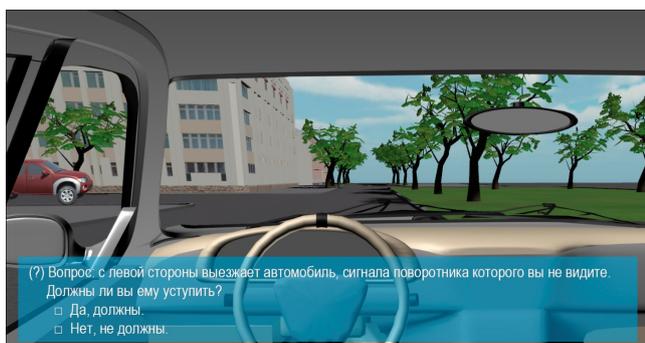


Рис. 7. Фрагмент теста.

Данное решение принято по причине того, что некоторые испытуемые могут запомнить положение ответа на вопрос, при этом не до конца понимая почему выбранный ими вариант корректен. Так же возможна перестановка вариантов ответов местами, чтобы еще больше усложнить задачу.

Можно сделать не более двух ошибок в одном блоке. Во время неправильного ответа программа добавляет пять случайных вопросов, в которых курсант, для успешной сдачи, не имеет права ошибиться.

Возможны три варианта загрузки тестов – в режиме моно- визуализации, в режиме стерео-визуализации и в режиме полного погружения с использованием шлема виртуальной реальности или его облегченного варианта - VR очков.

5. Заключение

Исследовано применение технологий виртуальной и дополненной реальности при тренировке водителя на реальных аварийно-опасных участках города. Для построения сценариев создания аварийно-опасных случаев использована статистика и ситуационное описание случаев ДТП. Предложен режим динамического тренинга и тестирования в режиме виртуальной и дополненной реальности с различными степенями погружения.

Предложена технология реализации приложения «Виртуальная автошкола» для мобильной платформы, способная реализовать сценарии обучения и тестирования на ДТП-опасных ситуациях в нескольких режимах использования виртуальной и дополненной реальности. Реализация выполнена в интегрированной среде разработки Unity 3D.

Предполагается использование приложения «Виртуальная автошкола» на практических занятиях в автошколах. Как известно в автошколах не имеется возможности проехать все сложные участки дорог и разобрать все нестандартные ситуации с курсантом. Поэтому применение данного приложения позволит заметно улучшить уровень подготовки курсантов автошкол, знакомя их со сложными ситуациями, возникающими в конкретных местах дорожной инфраструктуры, и тренируя их на динамических тестах в режиме виртуальной реальности.

Система динамических тестов позволит курсанту автошколы быстрее и удобнее «вжиться» во все сложности и аспекты передвижения по дороге. Такая система позволит как начинающим, так и опытным водителям избежать многих аварийно-опасных ситуаций и, соответственно, будет способствовать повышению транспортной культуры всех участников движения.

Технология виртуальной и дополненной реальности может также применяться при введении в эксплуатацию новых сложных развязок. Например, в Нижнем Новгороде,

при введении в эксплуатацию метромоста, многие водители не использовали его по причине сложных, по их мнению, подъездов к нему как с одной стороны реки, так и с другой. Своевременное "виртуальное информирование" водителей позволит избежать многих аварийно-опасных ситуаций и, соответственно, будет способствовать повышению безопасности дорожного движения.

6. Литература

- [1] Agarwal Shaurya, Kachroo Pushkin, Regentova Emma. A hybrid model using logistic regression and wavelet transformation to detect traffic incidents. // IATSS Research 40, 56–63, 2016.
- [2] Paz A., Veeramisti N., Khaddar R., de la Fuente-Mella H., and Modorcea L. Traffic and Driving Simulator Based on Architecture of Interactive Motion. // The Scientific World Journal, vol. 2015, Article ID 340576, 9 pages, 2015.
- [3] Системы современного автомобиля http://systemsauto.ru/another/augmented_reality.html [Электронный доступ] (последний доступ 30.06.17).
- [4] С. Карасёв. «Konica Minolta создала автомобильную HUD-систему формата 3D» // <https://3dnews.ru/953017/?feed> [Электронный доступ] (последний доступ 30.06.17).
- [5] С.Б. Ефремов. «Тип коммуникаций между водителем и автомобилем, основанный на дополненной реальности: «новый тренд» в построении интеллектуальных транспортных систем» // Эл. библ. М ГППУ <http://psychlib.ru/otherdocs/guide/208.html> [Электронный доступ] (последний доступ 30.06.17).

Об авторах

Дудаков Николай Юрьевич – магистрант кафедры «Графические информационные системы» НГТУ им. П.Е. Алексеева; e-mail: dudakovnu@gmail.com.

Мерзляков Игорь Николаевич – к.т.н., доцент, заведующий кафедрой «Графические информационные системы» НГТУ им. П.Е. Алексеева, email: inmerzliakov@gmail.com.

Ханнанов Артем Раисович – магистрант кафедры «Графические информационные системы» НГТУ им. П.Е. Алексеева; e-mail: zionone@mail.ru.

Томчинская Татьяна Николаевна – к.т.н., доцент кафедры «Графические информационные системы» НГТУ им. П.Е. Алексеева; e-mail: tomchinskaya@mail.ru.

Построение рельефа при реконструкции сцены городского пространства с помощью сервисов Google Maps

Кудряшов Алексей¹, Соловьёв Игорь²
KudryashovA@dvo.ru | igorek.solovyev@mail.ru

¹Институт автоматизации и процессов управления ДВО РАН, Владивосток, Россия;

²Институт прикладной математики ДВО РАН, Владивосток, Россия

В данной работе предложен метод построения рельефа для реконструкции трехмерной сцены городского окружения. Реконструкция производится на основе распознавания данных с топографического плана. Для построения рельефа используются геолокационные данные сервиса Google Maps Elevation.

Ключевые слова: Трехмерная реконструкция, рельеф, топографический план, Google Maps, GPS.

Relief generation for urban scene reconstruction using Google Maps services

A. Kudryashov¹, I. Soloviev²
KudryashovA@dvo.ru | igorek.solovyev@mail.ru

¹Institute of Automation and Control Processes, Far Eastern Branch of RAS, Vladivostok, Russia;

²Institute of Applied Mathematics, Far Eastern Branch of RAS, Vladivostok, Russia

A method of the relief generation for the reconstruction of a three-dimensional scene for urban environment is proposed. Reconstruction using data from the topographic plan. The geolocation data of the Google Maps Elevation service is used to build the relief.

Keywords: 3D reconstruction, relief, topographic plan, Google Maps, GPS.

1. Введение

Одной из важных практических задач компьютерного зрения является реконструкция реальных сцен и объектов. Одним из видов такой реконструкции является реконструкция городской обстановки [1,2,3]. Входными данными могут служить спутниковые или аэроснимки, измерения с помощью лазерных дальномеров, так и уже существующие топографические планы местности [3]. В зависимости от этого, будут меняться и подходы к реконструкции. Однако, актуальность построенной трехмерной сцены реальной городской обстановки, вне зависимости от выбранного подхода, зависит не только от точности восстановленных объектов (зданий и др.), а также и от рельефа местности.

Для реконструкции городского пространства авторами был выбран подход для реконструкции сцены по топографическому плану. Для этого на изображении топографического плана выделяются все возможные объекты (контуры здания, текст, деревья, уличные фонари, светофоры, клумбы и т.д.). Выделение информации на изображении происходит с использованием волнового алгоритма, применение и модификация которого описана в [4]. На основании всех полученных данных строится трехмерная модель сцены [5]. Однако, рельеф сцены остается плоским, что как правило, не соответствует окружающей действительности. Для решения этой проблемы был предложен следующий метод.

2. Описание алгоритма

2.1 Исходные данные

На топографическом плане (Рис. 1) в схематическом виде содержится информация для представления о структуре некоторой местности.

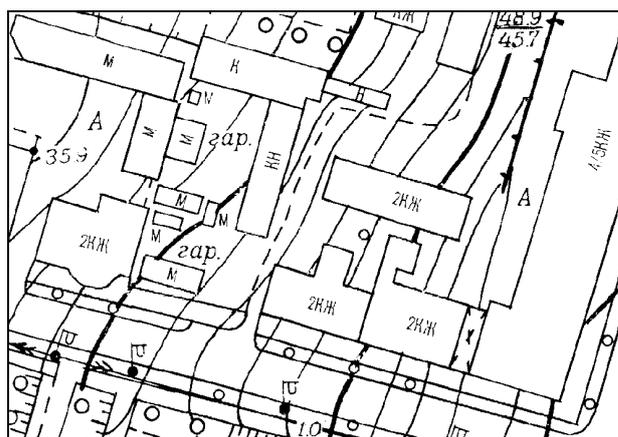


Рис. 1. Топографический план.

Также топографический план содержит информацию об изолиниях и уровнях высот в некоторых точках территории. Однако, несмотря на это, затруднительно точно выделить автоматически эти данные, т.к. они многократно перекрываются другими объектами. В связи с этим, было решено использовать уже существующую базу данных, содержащих значения высот над уровнем моря, для построения рельефа. Одной из них является сервис Google Maps Elevation API. На основании переданных данных о географической точке сервис возвращает высоту над уровнем моря в ней.

2.2 Возможности сервиса

Сервис Google Maps Elevation API позволяет получать информацию о высоте над уровнем моря для массива точек в одном запросе. Количество точек в массиве ограничено только длиной запроса к сервису, который не должен превышать 2000 знаков. Чтобы увеличить количество возможных точек в массиве в сервисе предусмотрена возможность «сжатия» записи точек в более компактный вид [6]. Запрос по получению высот для массива точек будет выглядеть следующим образом:

<https://maps.googleapis.com/maps/api/elevation/xml?locations=58.00720,56.18640|58.00439,56.18580>

После кодирования запрос будет следующим:

https://maps.googleapis.com/maps/api/elevation/xml?locations=enc:p`aJ_uIpPvB

Это позволяет за один запрос получить информацию сразу о десятках точек. На основании географических точек также возможно получить текстуру поверхности рельефа. Для этого используется сервис Google Maps API. Сервис позволяет получить изображение местности со спутника по географическим координатам точки. Например, следующий запрос:

<http://maps.googleapis.com/maps/api/staticmap?zoom=20&size=640x640¢er=58.00720,56.18640&maptype=satellite> возвращает текстуру рельефа (Рис. 2).



Рис. 2. Результат запроса получения текстуры рельефа.

За один запрос максимально возможное по размеру изображение можно получить с размерами 640x640 пикселей. На топографическом плане строится сетка из точек. Для каждой точки будет получаться изображение. Количество точек по горизонтали и вертикали равны согласно (1):

$$K_w = \left\lceil \frac{W}{640} \right\rceil, K_h = \left\lceil \frac{H}{640} \right\rceil, \quad (1)$$

где W , H – соответственно ширина и высота топографического плана, K_w , K_h – соответственно минимальное количество точек по горизонтали и вертикали для построения текстуры рельефа.

Для построения всей текстуры рельефа требуется выполнить несколько запросов, количество которых равно $K_w \times K_h$. Для определения ширины и высоты каждого изображения, которое будет получаться с сервиса, используется (2).

$$w = \left\lceil \frac{W}{K_w} \right\rceil, h = \left\lceil \frac{H}{K_h} \right\rceil. \quad (2)$$

Далее из всех полученных изображений собирается готовая текстура и масштабируется до размеров топографического плана.

Например, для топографического плана с размерами 1248x496. Требуется выполнить 2 запроса. При этом каждый раз запрос получает изображение с размерами 624x496. Результат получения текстуры рельефа продемонстрирован на Рис 3.



Рис. 3. Текстура рельефа.

2.3 Нахождение сетки высот

Построение рельефа местности требует наличия данных о высотах в каждой её точке. Рельеф для сцены будет

представлять из себя состоит карту высот размера $N \times M$. Промежуточные данные будут интерполироваться.

Горизонтальные и вертикальные пиксельные координаты любой точки этой сетки вычисляются по формуле (3):

$$\begin{aligned} x &= \left[\frac{W}{N-1} c \right], \\ y &= \left[\frac{H}{M-1} r \right], \end{aligned} \quad (3)$$

где W и H – соответственно ширина и высота топографического плана, c – номер колонки точки в сетке $[0..N-1]$, r – номер строки точки в сетке $[0..M-1]$.

Топографические планы строятся для территорий, площадь которых, как правило, не превышает нескольких квадратных километров. При этом построение осуществляется без учета кривизны поверхности Земли. Из этого следует, что пиксельные расстояния между точками топографического плана практически пропорциональны географическим расстояниям между точками местности.

Пусть известны две пары точек a , b и A , B , в которых географические координаты (A_{lon} , A_{lat} , B_{lon} , B_{lat}) соответствуют пиксельным (a_x , a_y , b_x , b_y).

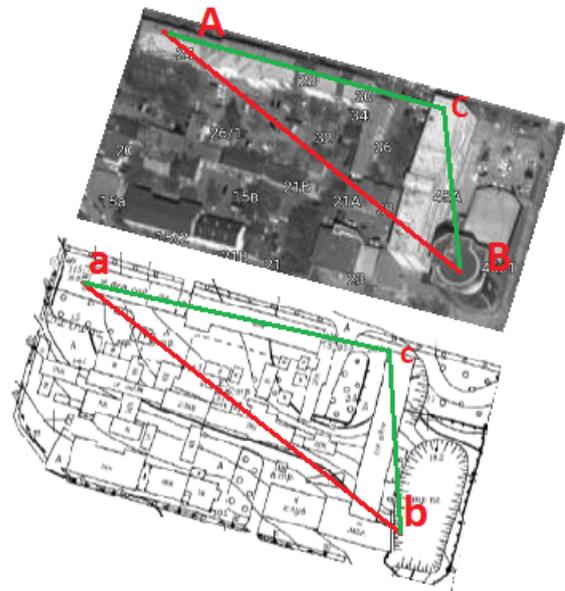


Рис. 4. Точки на топоплане (a , b , c) и соответствующие им географические точки (A , B , C).

Для третьей точки C (Рис. 4) с известными пиксельными координатами C_x и C_y для её географических координат C_{lon} и C_{lat} будет верно (4):

$$\begin{aligned} C_{lon} &= \frac{c_x - a_x}{b_x - a_x} (B_{lon} - A_{lon}) + A_{lon}, \\ C_{lat} &= \frac{c_y - a_y}{b_y - a_y} (B_{lat} - A_{lat}) + A_{lat}. \end{aligned} \quad (4)$$

Для определения высот в каждой точке сетки требуется рассчитать её географические координаты согласно (4). После этого формируются запросы к сервису *Google Maps Elevation API*. На основании данных о высоте строится триангуляционная поверхность.

2.4 Доработка сцены с учетом рельефа

После добавления рельефа в сцену все объекты, которые раньше располагались на плоскости должны быть смещены вертикально на соответствующий уровень высоты, чтобы не возникали зазоры между рельефом и объектами.

Так как рельеф является триангуляционной поверхностью, то определяется треугольник, к которому принадлежит точка пересечения перпендикуляра и рельефа. После этого определяется точка пересечения перпендикуляра и плоскости, образованной тремя точками треугольника. Длина отрезка, соединяющего эту точку с

точкой основания перпендикуляра, даст требуемую высоту смещения объекта.

Здания имеют несколько точек соприкосновения с основанием. В каждой точке высота рельефа может быть разной. Однако, уровень перепада высот по отношению к размерам здания, как правило, не велик. Поэтому для смещения здания по высоте определяется минимальное расстояние от каждой точки основания здания до рельефа аналогично определению высоты для других объектов. После этого здание поднимается на это расстояние.

3. Результаты

Произведена реконструкция некоторых реальных городских сцен для демонстрации работы описанного алгоритма. Текстура для зданий строится автоматически, в зависимости от типа здания [5].

Топографический план сцены 1 (Рис. 5, Рис. 6, Рис. 7) имеет размеры 2480x1356 пикселей, что в реальности составляет 310x170 метров. Время работы алгоритма реконструкции – 16 сек. Время обращения к сервису – 4 сек. Время построения рельефа – 1 сек.

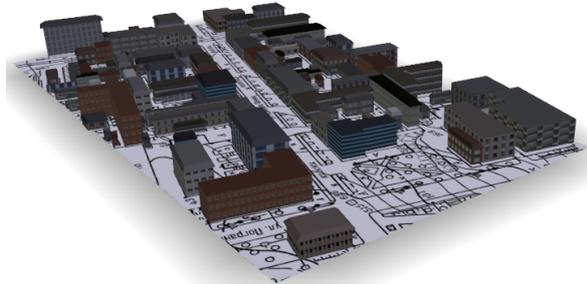


Рис. 5. Сцена 1. Реконструкция без рельефа.

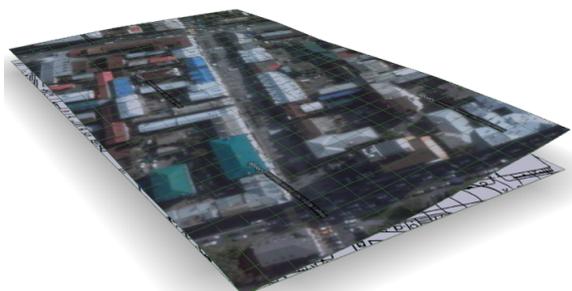


Рис. 6. Сцена 1. Реконструированный рельеф

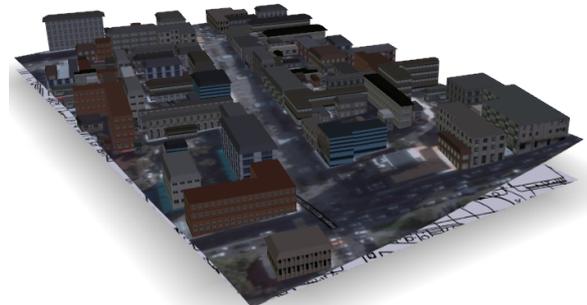


Рис. 7. Сцена 1. Реконструкция с рельефом

Топографический план сцены 2 (Рис. 8) имеет размеры 804x1020 пикселей, что в реальности составляет 100x128 метров. Время работы алгоритма реконструкции – 7 сек. Время обращения к сервису – 3 сек. Время построения рельефа – 1 сек.

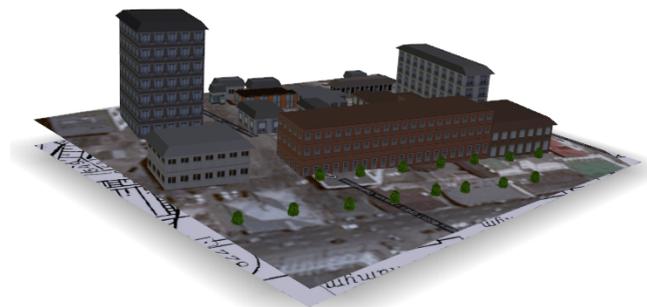


Рис. 8. Сцена 2. Реконструкция с рельефом.

4. Заключение

Предлагаемый метод построения рельефа позволяет значительно увеличить актуальность и наглядность реконструированной сцены городской обстановки. Продемонстрированные результаты показывают надежность и точность предлагаемого подхода реконструкции с использованием топографического плана и геолокационных данных. Процесс полностью автоматизирован после внесения исходных данных.

5. Литература

- [1] Knopp J., Prasad M., Gool L.V., “Scene cut: Class-specific object detection and segmentation in 3D scenes”, Proceedings of international conference on 3D imaging, modeling, processing, visualization and transmission. pp. 180–187, 2011.
- [2] Lin B. C., You R. J. “Tensor-based quality prediction for building model reconstruction from LIDAR data and topographic map”, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXIX-B7, pp. 437-441, 2012
- [3] Vezhnevets V., Konushin A., Ignatenko A., “Interactive image-based urban modeling”, PIA-2007, pp. 63-68. 2007
- [4] Кудряшов А.П., Соловьёв И.В. “Выделение контуров зданий и распознавание служебных символов для трехмерной реконструкции объектов городской обстановки по топографическому плану”. Международная конференция по компьютерной графике и зрению (GraphiCon 2013). 16-20 сентября 2013 г., Россия, Владивосток. с. 290-293.
- [5] Кудряшов А.П., Соловьёв И.В., “Распознавание контуров зданий на топографическом плане для реконструкций городских сцен”, “Вестник компьютерных и информационных технологий”. 2015. № 2. С. 3-8.
- [6] <https://developers.google.com/maps/documentation/utilities/polylinealgorithm?hl=ru> (дата обращения 11.12.2016)

Об авторах

Кудряшов Алексей Павлович, к.т.н., младший научный сотрудник лаборатории машинной графики института автоматики и процессов управления ДВО РАН.

Соловьёв Игорь Владимирович, аспирант, Институт прикладной математики ДВО РАН.

Автоматическая генерализация графических объектов с учетом иерархии классов

Н.С. Пономарев, А.С. Тарасенко, Т.Ш. Утешева
ponom.nikita@gmail.com|tarasenk93@gmail.com|uts13@yandex.ru

Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

Разработан алгоритм цензово-нормативного отбора дискретных картографических объектов, основанный на использовании двух критериев: близости и удельной плотности с учетом иерархии приоритетов классификационных кодов и значений атрибутов.

Ключевые слова: автоматическая генерализация, картографические объекты, иерархия классов.

Automatic generalization of graphic objects taking into account the hierarchy of classes

N.S. Ponomarev, A.S. Tarasenko, T.Sh. Utesheva
ponom.nikita@gmail.com|tarasenk93@gmail.com|uts13@yandex.ru

Lobachevsky State University of Nizhni Novgorod, Russia

The algorithm of censor-standard selection of discrete cartographical objects, based on the use of two criteria: proximity and specific density, taking into account the hierarchy of classification code priorities, is developed.

Keywords: automatic generalization, cartographical objects, hierarchy of classes.

1. Введение

Важнейшей характеристикой современных цифровых карт, определяющей их интеллектуализацию, является мультимасштабность [1]. В [2-4] дан обзор этапов возникновения и развития мультимасштабного картографирования как научного направления. Предпосылки для возникновения мультимасштабного картографирования были заложены в работах по автоматизированной генерализации, выполнявшихся в 60-е - 70-е годы прошлого столетия [7, 10]. Проблема формализации процесса генерализации была обозначена в 80-х годах, были сформулированы первые принципы данного процесса [5]. Идея множественных представлений (multiple representations) была заложена в работе [9] и получила развитие в [6, 8]. В 2000-е годы разрабатывается идея создания мультимасштабных карт на основе баз данных, большое внимание уделяется графическим аспектам. В это же время в широком доступе появились мультимасштабные карты OpenStreetMap (2004), Яндекс Карты (2004), Google Maps (2005).

В основе механизма мультимасштабности лежит автоматическая генерализация картографических объектов. На данный момент правила автоматической генерализации слабо формализованы. В работе [8] предложена структура классификации операторов преобразования набора данных из источника в целевую модель, основанная на общепринятой модели обобщения Грюнрайха. В работе [9] рассматривается проблема автоматизированной генерализации, основанной на качественном и количественном анализе. Следует отметить, что в данной статье особое внимание уделяется наборам топографических данных, наборы кадастровых данных, ортофотоснимки и цифровые модели рельефа не рассматриваются. В статье [12] представлены исследования, в которых реализован полностью автоматизированный процесс генерализации топографической карты. Целью исследований было показать, что решение задачи полностью автоматической генерализации технологически возможно. Были разработаны и реализованы операции обобщения,

содержащие соответствующие алгоритмы и значения параметров.

Процесс формализации правил автоматической генерализации осложнен тем, что для большого количества классификационных кодов картографических объектов генерализация выполняется по своим особым правилам (алгоритмам).

В работе рассматривается процесс генерализации дискретных картографических объектов (знаков) планшетов гидрографической съемки. Под дискретным понимается объект цифровой карты, местоположение которого описывается координатами одной точки и имеющий различные пространственно – временные и качественные атрибуты.

2. Постановка задачи

Количество дискретных объектов, сохраняемых при генерализации, определяется в соответствии с цензами и нормами отбора (норма представительства) соответствующего масштаба. Цензы отбора определяют объекты, удаляемые с карты, или объекты, обязательно сохраняемые на карте. Нормы отбора определяют количество объектов, сохраняемых на карте (в процентном отношении) в соответствии с густотой объектов исходного материала.

Предложена формальная математическая модель, а также алгоритм отбора дискретных объектов с учетом иерархии приоритетов классификационных кодов (например, опорные пункты имеют более высокий приоритет, чем отметки высот и в иерархии классов занимают более высокий уровень).

Алгоритм осуществляет отбор по графической плотности и близости. Возможно использование дополнительных параметров отбора, настраиваемых в зависимости от типа и классов дискретных объектов (например, для отметок глубин – величина разницы значений глубин).

3. Модель операции отбора дискретных объектов с учетом иерархии классов

Пусть $D = \{d_i\}$ – совокупность дискретных объектов и на $\{d_i\}$ задано отношение порядка P , называемое иерархией дискретных объектов со свойствами линейности и транзитивности.

Линейность порядка P означает, что для любой пары объектов d_1 и d_2 из D всегда выполняется либо $d_1 \leq_P d_2$, либо $d_2 \leq_P d_1$.

Транзитивность отношений означает, что если $d_1 \leq_P d_2$ и $d_2 \leq_P d_3$, то $d_1 \leq_P d_3$.

Отношение порядка P на $\{d_i\}$ индуцирует отношение эквивалентности R . Объекты d_1 и d_2 считаются эквивалентными (или находятся в отношении эквивалентности), если одновременно выполняется $d_1 \leq_P d_2$ и $d_2 \leq_P d_1$. Отношение эквивалентности определяет разбиение D на классы эквивалентности D_j , такое, что:

$$\cup_j D_j = D \text{ и } D_j \cap D_k = \emptyset, \text{ если } j \neq k.$$

Два класса D_1 и D_2 из $\{D_j\}$ находятся в отношении π , индуцированном отношением P , если в этих классах можно выбрать по представителю $\tilde{d} \in D_1$ и $\tilde{d} \in D_2$ так, что выполнено отношение

$$\tilde{d} \leq_P \tilde{d}.$$

Отношение на множестве классов эквивалентности $\{D_j\}$, индуцированное отношением иерархии обладает следующими свойствами:

1) является строгим линейным порядком, то есть для любой пары представителей классов разбиения $\tilde{d} \in D_1$ и $\tilde{d} \in D_2$ при $D_1 = D_2$ выполняется либо $\tilde{d} \leq_P \tilde{d}$, либо $\tilde{d} \leq_P \tilde{d}$;

2) во множестве упорядоченных классов нет классов совпадающих ($D_i \neq D_j$).

Обозначим через номер класса $D = D_{k(i)}$, для которого $d_i \in D_{k(i)}$. Величины расстояний между объектами d_i и d_j образуют матрицу расстояний ρ_{ij} .

Каждой паре объектов d_{i_0} и d_i и соответствующим $D_{k(i_0)}$ и $D_{k(i)}$ сопоставляется число $R_{(k(i), k(i_0))}$ и ε – окрестность ε_{i_0} такая, что

$$\varepsilon_{i_0} = \{t | \rho(t, d_{i_0}) < R_{(k(i), k(i_0))}\},$$

где t – точки изображения.

Каждому d_{i_0} и соответствующему $D_{k(i_0)}$ сопоставляется число $R_{k(i_0)}$ и ε – окрестность ε_{i_0} такая, что

$$\varepsilon_{i_0}^* = \{t | \rho(t, d_{i_0}) < R_{k(i_0)}^*\}.$$

Для заданной совокупности дискретных объектов предлагается два формальных критерия отбора с учетом их иерархии: критерий близости и критерий удельной плотности.

4. Критерий близости

Критерий близости определяет дискретные объекты, подлежащие исключению в окрестности заданного знака.

Пусть i_0 – текущий номер дискретного объекта, в окрестности которого ведется отбор, i – текущий номер отбираемых дискретных знаков, $i \neq i_0$. ε_{i_0} – ε – окрестность объекта d_{i_0} .

Тогда критерий близости формулируется следующим образом. В ε_{i_0} объекта d_{i_0} исключить такие $d_i \in D_k$, для которых

$$\rho(i_0, i) < R_{(k(i), k(i_0))}.$$

Упрощением данного критерия является следующий. В $\varepsilon_{i_0}^*$ исключить все $d_i \in D_k$, для которых $\rho(i_0, i) < R_{k(i_0)}^*$, где $R_{k(i_0)}$ не зависит от $k(i)$.

5. Критерий удельной плотности

Критерий удельной плотности определяет количество дискретных объектов, отбираемых на элементе площади.

Пусть i_0 – текущий номер дискретного объекта, в окрестности которого ведется отбор, i – текущий номер отбираемых знаков, $i \neq i_0$; $S_{k(i_0)}$ – элемент площади в окрестности $d_{i_0} \in D_k$; T_k – число объектов k -го класса на элементе площади S ; $S_{k(i)}$ – элемент площади в окрестности $d_i \in D_k$. И дана зависимость удельной плотности от класса отбираемых дискретных объектов: $T = T(k)$.

Тогда критерий удельной плотности формулируется следующим образом.

Число объектов k -го класса на $S_{k(i)}$ не должно превышать $T(k)$:

$$|\{i | i \in S_{k(i)} \cap D_k\}| \leq T(k).$$

Если положить, что для каждого класса объектов D_k задается значение элемента площади $S_{k(i)}$, то критерий удельной площади формулируется следующим образом. Число объектов k -го класса на элементе площади $S_{k(i)}$ не должно превышать T^* :

$$|\{i | i \in S_{k(i)} \cap D_k\}| \leq T^* \text{ при } T(k) \equiv T^*.$$

6. Реализация

На основе изложенного подхода разработан и реализован алгоритм генерализации точечных объектов планшетов гидрографической съемки (ПГС). Данный картографический документ является исходным для составления морских карт (навигационных, рельефа морского дна и др.) и, как правило, имеет очень высокую графическую плотность.

Высший приоритет в иерархии классов был присвоен объектам категории "навигационная опасность" (скала, риф, банка, затонувшее судно и др.). Картографические объекты "отметки глубин", как правило, имеют наибольшую графическую плотность на ПГС и подлежат основному отбору. При реализации алгоритма был введен параметр, определяющий дополнительную иерархию объектов внутри данного классификационного кода и выбираемый в зависимости от тематической направленности, формируемой в процессе генерализации карты.

При составлении морских навигационных карт обязательным является требование сохранения минимальных глубин (глубин, меньших и близких к предельно безопасным).

Для карт, отражающих структуру и рельеф морского дна, в целях достоверности градиентов важным является сохранение экстремальных значений глубин.

При составлении карт, предназначенных для реалистической 3D визуализации рельефа морского дна, логично использовать средние значения совокупностей (кластеров) отметок глубин, заменяемых центроидами совокупностей.

Таким образом, в зависимости от параметра, определяющего цель генерализации, объектам одного классификационного кода присваивается либо равный приоритет, либо различные приоритеты, в зависимости от значений атрибута. При наличии иерархии объектов внутри одного классификационного кода, осуществляется предварительная сортировка объектов по значению определяющего приоритет атрибута.

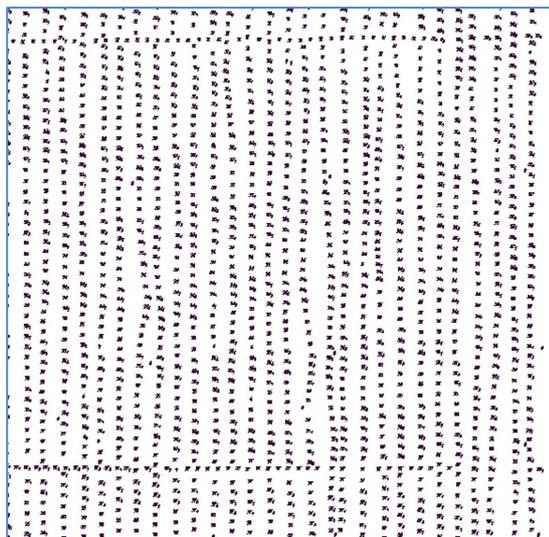


Рис. 1. Фрагмент исходного ПГС.

Процесс отбора объектов, основанный на критериях близости и удельной плотности, осуществляется последовательно от высших классов к низшим с учетом внутренней иерархии каждого классификационного кода.

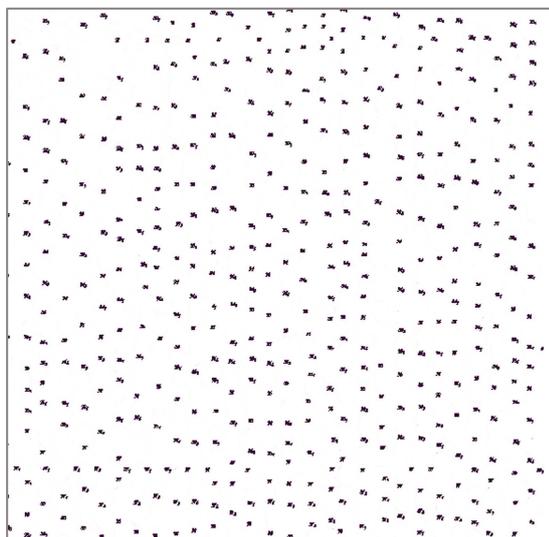


Рис. 2. Результат генерализации с сохранением минимальных глубин.

На рис.1. приведен фрагмент планшета гидрографической съемки. На рис. 2. представлен тот же фрагмент результата генерализации с приоритетом меньших глубин и заданным порогом минимальной разницы значений глубин.

7. Заключение

Введение гибкой системы приоритетов – как классификационных кодов, так и объектов внутри классов – позволяет повысить качество генерализации с учетом тематики результирующей карты. Однако приемлемость получаемого результата определяется пока экспертом. Дальнейшие исследования планируется направить на разработку критериев и методов автоматической оценки качества генерализации.

8. Благодарности

Работа выполнена при финансовой поддержке РФФИ (проекты №16-07-01198А, №16-07-1214А).

9. Литература

- [1] Берлянт А.М. Теория геоизображений. – М.: ГЕОС, 2006. – 262 с.
- [2] Гук, А. П. Проблемы автоматической генерализации при тематическом картографировании в среде ГИС / А. П. Гук, С. С. Дышлок, А. Б. Женибекова // Известия вузов. Геодезия и аэрофотосъемка. – 2014. – № 4/С. – С. 97–100.
- [3] Лурье И.К., Самсонов Т.Е. Структура и содержание базы пространственных данных для мультимасштабного картографирования // Геодезия и картография — 2010, №11. – с. 17–23.
- [4] Самсонов Т.Е. Мультимасштабное картографирование — новое направление картографии / Современная географическая картография / Под ред. И.К. Лурье и В.И. Кравцовой. М.: Дата+, 2012. с. 21-35.
- [5] Свентэк, Ю. В. Теоретические и прикладные аспекты современной картографии / Ю.В. Свентэк; МГУ им. М.В.Ломоносова. - М. : ЭДИТОРИАЛ УРСС, 1999. - 80с.
- [6] Buttenfield В. Research Initiative 3: Multiple Representations. Closing Report, National Center for Geographic Information and Analysis, Buffalo, 1993.
- [7] Douglas D. H., Peucker T. K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. // The Canadian Cartographer, 10 (2), 1973, pp. 112–122.
- [8] Foerster, T, J Stoter, and B Köbben Towards a Formal Classification of Generalization Operators. In: Proceedings of the 23th International Cartographic Conference. Moscow, Russia, 2007.
- [9] Foerster Theodor, Jantien Stoter and Menno-Jan Kraak Challenges for Automated Generalisation at European Mapping Agencies: A Qualitative and Quantitative Analysis / The Cartographic Journal Vol. 47 No. 1 pp. 41–54 February 2010
- [10] McMaster R.B., Shea K.S. Generalization in Digital Cartography. — Association of American Cartographers, Washington, DC, 1992.
- [11] Marble D. Geographic Information Systems: an overview // Proceedings of Pecora 9, Sioux Falls, SD, 1984, pp 18–24.
- [12] Stotera Jantien, Marc Postb, Vincent van Altenab, Ron Nijhuisb and Ben Brunsb Fully automated generalization of a 1:50k map from 1:10k data / Cartography and Geographic Information Science, 2014, Vol. 41, No. 1, 1–13.
- [13] Tobler W. R. Numerical map generalization. Michigan Inter-University Community of mathematical Geographers Discussion Paper 8. — Dept. of Geography, University of Michigan, An Arbor, MI, USA, 1966.

Об авторах

Никита Сергеевич Пономарев – студент 2 курса магистратуры ИИТММ ННГУ им. Н.И. Лобачевского
Алексей Сергеевич Тарасенко – студент 2 курса магистратуры ИИТММ ННГУ им. Н.И. Лобачевского.

Тамара Шатовна Утешева – ведущий научный сотрудник ЦИИИТ ИИТММ ННГУ им. Н.И. Лобачевского.



Геометрическое моделирование.
Компьютерная графика в образовании

Geometric Modeling.
Computer Graphics in Education

GraphiCon 2017

24–28 сентября 2017
Пермь, Россия

Graph-based feature recognition and suppression on the solid models

A. Malyshev¹, S. Slyadnev¹, V. Turlapov¹

al.s.malyshev@rambler.ru|sergey.slyadnev@gmail.com|vadim.turlapov@itmm.unn.ru

¹Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

We consider one of the main problems in the field of CAD / CAM, which is feature recognition for CAD models represented in boundary representation form. We take into account volumetric features only. We use the recognition results for CAD model simplification. Our recognition procedures are based on detection of specific patterns in adjacency graphs constructed for the input CAD parts. Recognition procedures for bosses, holes, pockets, and for arbitrary cavities are proposed. Simplification is conducted by suppression of the detected features using a dedicated topological operator.

Keywords: feature recognition, feature suppression, boundary representation, CAD model simplification, graph-based methods.

1. Introduction

The problem of feature recognition and suppression from the so-called "dumb" model is one of the most frequently encountered problems in the CAD field. "Dumb" means that digital input model does not contain a history of construction: only pure geometric and topological data are present.

The vast majority of CAD-systems implement history-based modeling. However, a way of storing this information is vendor-specific and is not readily available for integration with third-party software. A "dumb" model may emerge in various scenarios like a lossy translation from one format to another or reverse engineering. Although the international STEP standard (ISO 10303) supports features as a part of AP224 protocol, usually, CAD-systems do not fully use this possibility.

Features are widely covered by academic literature. However, a "feature" term may have a different meaning in different engineering domains. In our paper, the "feature" term is used in a way proposed by Shah and Mantyla [9]. The description, properties, and semantics of the most commonly used features can be found in their fundamental book.

Feature-based modeling is close in nature to the *constructive solid geometry (CSG)*. The main difference between the two techniques is how a building tree is organized. In the feature-based modeling the building tree consists of various features, while in the CSG, it consists of Boolean operators and primitives. The concept of the building tree can be found in the article [1], where the problem of reconstructing three-dimensional models from two-dimensional drawings is considered.

Feature-based modeling is a more advanced design technique compared to other design methods. In feature-based modeling paradigm, an engineer works in terms of meaningful groups of geometric elements (features), and not in terms of *vertices, edges, and faces*. Feature recognition is opposed to design-by-features paradigm in sense that it does not rely on a construction tree. Feature recognition allows performing the following conversions:

- Defeaturing (feature suppression) on a model.
- Creation of feature model associated with a "dumb" model.

The defeated model can be used in the following contexts:

- Protection of the intellectual property when exchanging data.
- Preparation of a model to engineering analysis.
- Creating levels of details for efficient visualization of massive data (e.g. large mechanical assemblies).
- Preparation of a model for visualization on mobile devices.

2. Related work

A general overview of recognition methods is given by Han, Pratt, and Regli in [3]. Joshi and Chang proposed *graph-based methods of recognition* in [5]. The main idea besides the graph-based methods is to analyze connectivity between the boundary elements of a model, these methods convert the input model to one or several graphs, assign some attributes to the vertices and edges of these graphs and after that perform recognition procedure.

Joshi and Chang propose to use *attributed adjacency graph (AAG)*. Its structure is presented below:

- **Vertices** - represent faces of the input model.
- **Edges** - represent adjacency relationship between faces.
- **Attributes** - represent convexity of dihedral angles.

Gao and Shah [2] introduce new graphs into recognition process. They use manufacturing face adjacency graph, minimal condition sub-graph, concave adjacency graph, and other medium structures. The main goal of their work is to recognize interacting features which is the key problem in feature recognition field.

Verma and Rajotia in [11] propose to use a matrix representation of a graph instead of isomorphism in the recognition process. The so-called feature vector is calculated for each feature of interest. The feature vector is then used in recognition heuristic. This transformation decreases recognition complexity from $o(n^k)$ to $o(n^3)$.

Lockett and Guenov in [6] propose to build AAG on the *mid-surface representation* of a geometric model. The main problem of their approach is to adapt AAG concept to the mid-surface representation. This technique significantly reduces a number of nodes and vertices in the graph and improves recognition speed.

Hayasi and Asiabanpour [4] consider the platform-dependent approach of conversion design features into manufacturing features.

Nasr and others [7] unite different recognition procedures into the sequential workflow to construct building tree for a model.

The major drawback of graph-based methods is the impossibility of guaranteed separation of interacting features. This problem is partially solved by Gao and Shah, but it remains actual.

Usually, the result of recognition is presented as a set of faces, which is an advantage, since it allows using this method for recognition of *volumetric and surface* features. The concept of volumetric and surface features can be found in [10].

The method, which is proposed in this paper, is based on the original attributed adjacency graph. Therefore it can be implemented in any CAD-system since no specific functionality

is required. The recognition procedures presented below are based on the *rule-based* approach, initially invented in [10]. This approach reduces complexity of the problem and allows flexible tuning of the *recognition rules*.

3. Recognition procedures

The development of efficient feature recognition algorithm is hardly possible without constraining the input. In this work, the following assumptions are made:

- **Input model consists of one or more valid solids.** This restriction enables convexity analysis for dihedral angles. Such analysis relies on proper orientations of faces in a solid.
- **Face maximization was applied to the input model.** The definition of the maximization procedure can be found in [8].
- **All features are placed in interior of faces.** It is a principal constraint which is actively used in the recognition and suppression procedures.
- **The host geometry of faces is canonical.** This limitation is used in the first three procedures and does not apply to the last method.

The common stage which is shared by all recognition procedures proposed in our paper is detection of a *base face*. In general, a base face is a face where features can exist. In our study, the base faces are the faces with inner wires (Figure 1).

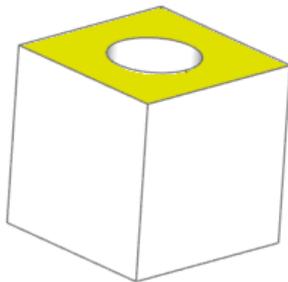


Figure 1: Cube model and base face (yellow color).

Here and later, it is assumed that attributed adjacency graph is built. Particular features can be recognized using additional constraints, which are proposed for each feature type.

3.1 Holes recognition

Hole is cavity feature, so the dihedral angle between base face and feature faces should be convex. Feature faces which are adjacent to the base face should have cylindrical host geometry. After that, the list of feature faces can be extended by adjacent faces with non-cylindrical host geometry. For instance, it is possible to allow transition from cylindrical geometry to the planar to handle nested holes case. For clarity, the simplest case is considered – the stopping criterion for the recognition procedure is another base face. The main parameter determining the hole is radius. It seems reasonable to filter out all holes with radius bigger than some given value. Modified recognition procedures can be proposed using the requirements of a particular subject area.

The figure below shows the detected holes on the ANC101 test model with a radius not exceeding a given threshold (Figure 2):

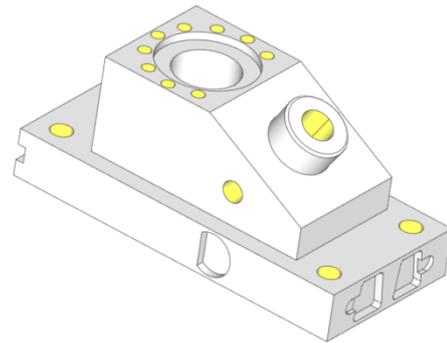


Figure 2: ANC101 model and recognized holes (yellow color).

3.2 Pockets recognition

Pockets are another machining features, the main difference between holes and pockets is that the pocket is always *blind*. The dihedral angles on the edges of the inner contour must also be convex. The underlying geometry on the feature faces which are adjacent to the base face can have any type. The feature is constructed by adding adjacent faces. The recognition process stops when a plane face without internal contours is detected, parallel to the base face. The pocket can be defined by two parameters – the depth and the contour that forms it. For example, the filtering can be carried out using the area of the figure bounded by the contour of the pocket. The following illustration shows the result of this recognition rule (Figure 3):

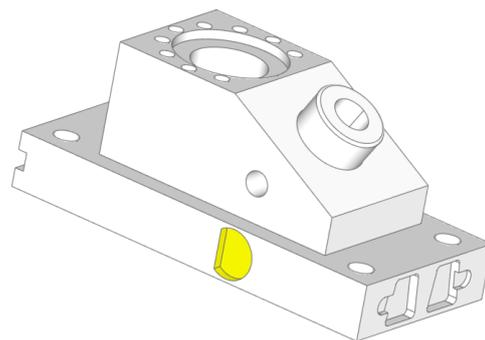


Figure 3: ANC101 model and recognized pockets (yellow color).

3.3 Bosses recognition

A boss is a protrusion feature in a work piece. Therefore, the dihedral angles between the boss faces and the base face are concave. The faces constituting a boss feature yield an isolated connected component in the adjacency graph with eliminated base face (as in the case of a pocket). New faces are added to the feature similarly to the case with pockets. Filtering bosses by size can be conducted using the diagonals of their corresponding bounding boxes.

Recognition speed can be improved by limiting the search depth in the adjacency graph. For instance, the search depth equal to two can be used to find the simplest bosses (the search begins from the base face). Figure 4 shows the boss feature found in the MBB Gehause Rohteil model.

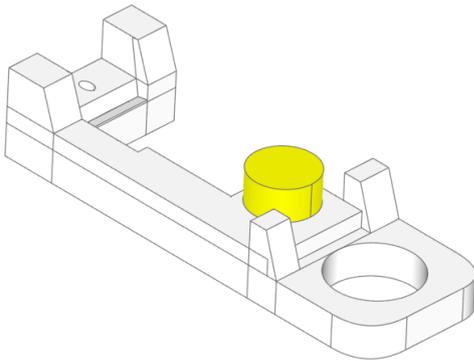


Figure 4: Modified MBB Gehäuse rohteil model and recognized boss (yellow color).

3.4 General holes recognition

All the procedures considered earlier are designed to operate on canonical geometry. In this section, a more general method is proposed. The features detected by the following procedure do not correspond to simple machining operations such as drilling or milling. We call this new feature type a “general hole” since it represents an arbitrary cavity. Like in the case with a pocket or a hole feature, the dihedral angles for general holes are convex. More faces are added to the feature according to the procedure already described. There are two stopping criteria:

- For through general holes, the feature faces are located between two base faces.
- For blind general holes, the feature faces yield an isolated connectivity component in the adjacency graph.

Filtering by size can be based on diagonal of a bounding box calculated for the detected feature faces.

The described recognition procedure covers both the procedure for holes and the procedure for pockets. Figure 5 and Figure 6 show the recognition result for ANC101 model.

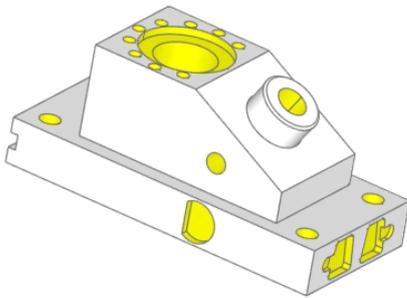


Figure 5: ANC101 model and recognized general holes.

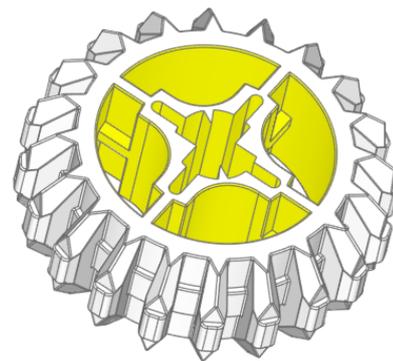


Figure 6: Gear model and recognized general holes (yellow color).

4. Feature suppression

When all the features that should be suppressed are recognized, they can be suppressed to simplify a model. This process involves two main steps:

1. Removing of the feature faces from B-Rep representation.
2. Updating all affected faces to get valid B-Rep model after suppression.

In general, the update procedure is very difficult since it requires both topology and geometry modification. Restrictions introduced earlier significantly reduce the complexity of this step. It is necessary to delete only inner wires on the base faces, and no additional actions are needed to bring geometry into valid state. It is the reason to detect isolated non-interacting features which represented as isolated connectivity component on the graph.

5. Conclusion and further work

Feature recognition and suppression procedures were proposed for holes, pockets, bosses, and general holes. These methods are implemented using OpenCascade geometric modeling kernel.

The algorithms were tested on a wide set of cases, including the well-known benchmarking models (MBB Gehäuse Rohteil, ANC101) and real industrial parts.

Figure 7 and Figure 8 demonstrate the defeaturing process based on the proposed recognition procedures.

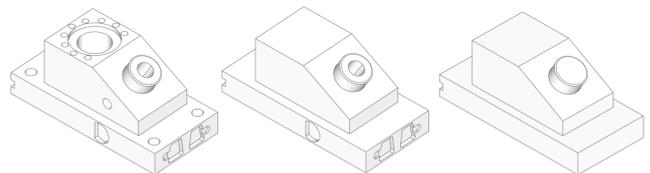


Figure 7: Feature suppression on the ANC101 model.

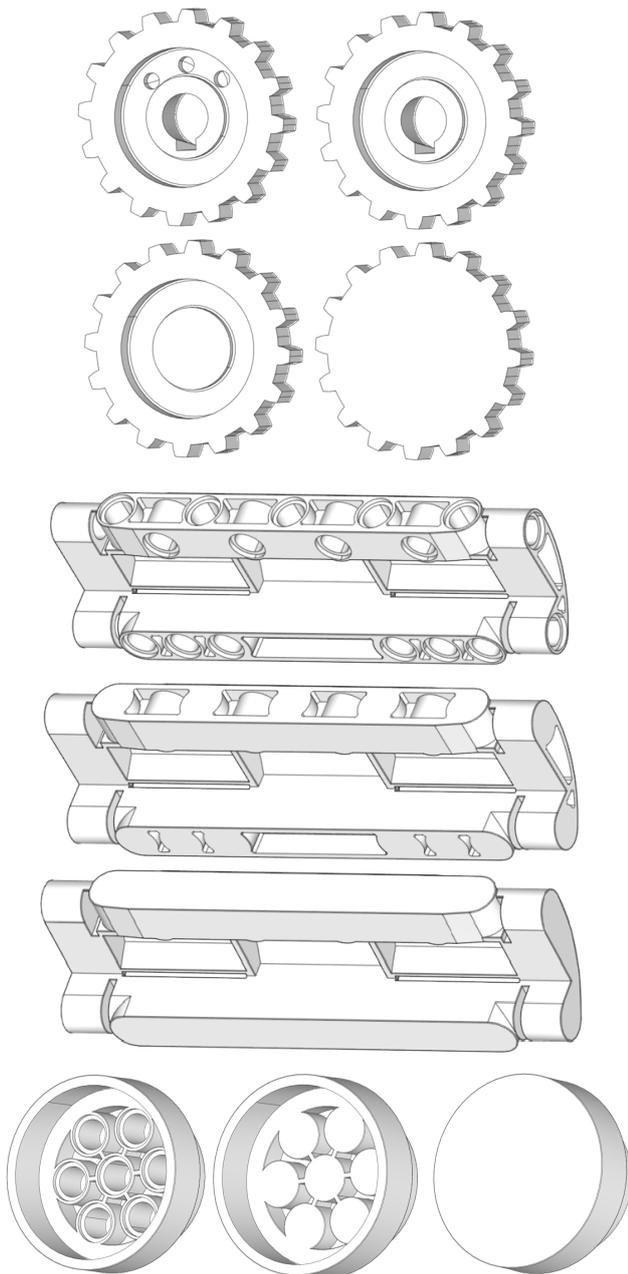


Figure 8: Feature suppression demonstration, from left to right, from top to bottom.

In future, we plan to improve the recognition methods in the following aspects:

- Extension of the supported feature types.
- Recognition of non-isolated features.
- Migration from explicit-based rules to the parameterized rules.
- Automatic generation of the building tree.

6. References

- [1] Жилина, Н., Лагунова, М., Мошкова, Т., Ротков, С. и Тюрина, В. 2014. Автоматизация процесса чтения чертежа с использованием системы Компас -3D. ГрафиКон 2014: труды 24 междунар. конф. по компьютерной графике и зрению, 17–19.
- [2] Gao, S. and Shah, J.J. 1998. Automatic recognition of interacting machining features based on minimal condition subgraph. *Computer-Aided Design* 30, 9, 727–739.
- [3] Han JungHyun, Pratt, M., and Regli, W.C. 2000. Manufacturing feature recognition from solid models: a status report. *IEEE Transactions on Robotics and Automation* 16, 6, 782–796. АвторД. Источник.
- [4] Hayasi, M.T. and Asiabanpour, B. 2009. Extraction of manufacturing information from Design-by-feature solid model through feature recognition. *International Journal of Advanced Manufacturing Technology* 44, 11–12, 1191–1203.
- [5] Joshi, S. and Chang, T.C. 1988. Graph-based heuristics for recognition of machined features from a 3D solid model. *Computer-Aided Design* 20, 2, 58–66.
- [6] Lockett, H.L. and Guenov, M.D. 2005. Graph-based feature recognition for injection moulding based on a mid-surface approach. *CAD Computer Aided Design* 37, 2, 251–262.
- [7] Nasr, E.S.A., Khan, A.A., Alahmari, A.M., and Hussein, H.M.A. 2014. A Feature Recognition System Using Geometric Reasoning. *Procedia CIRP* 18, 238–243.
- [8] Requicha, A.A.G. and Voelcker, H.B. 1985. Boolean operations in solid modeling: Boundary evaluation and merging algorithms. *Proceedings of the IEEE* 73, 1, 30–44.
- [9] Shah, J.J. and Mantyla, M. 1995. *Parametric and Feature Based CAD/Cam: Concepts, Techniques, and Applications*.
- [10] Vandenbrande, J.H. and Requicha, A.A.G. 1993. Spatial Reasoning for the Automatic Recognition of Machinable Features in Solid Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 12, 1269–1285.
- [11] Verma, A. K. and Rajotia, S. 2004. Feature vector: a graph-based feature recognition methodology. *International Journal of Production Research* 42, 16, 3219–3234.

About authors

Vadim Turlapov (vadim.turlapov@itmm.unn.ru): professor of Computer Science, Lobachevsky Nizhny Novgorod University.

Alexander Malyshev (al.s.malyshev@rambler.ru): PhD student, Lobachevsky Nizhny Novgorod University.

Sergey Slyadnev (sergey.slyadnev@gmail.com): PhD student, Lobachevsky Nizhny Novgorod University.

CAD model inspection utility and prototyping framework based on OpenCascade

S. Slyadnev, A. Malyshev, V. Turlapov
sergey.slyadnev@gmail.com|al.s.malyshev@rambler.ru|vadim.turlapov@itmm.unn.ru
Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

We present an open-source, extensible software which assists in the development of CAD geometric algorithms based on Boundary Representation data structures. The software is coupled with OpenCascade library which is an open-source geometric modeling kernel offering a broad range of modeling operations for the development of multi-purpose CAD systems and utilities. The new software serves as a framework for constructing geometric modeling algorithms. We provide several examples of geometric utilities developed with the help of the present framework. Furthermore, our software offers some advanced facilities for inspection of geometric and topological structures of existing CAD parts which can be imported from any popular CAD system.

Keywords: CAD, geometric modeling, boundary representation, OpenCascade, VTK, software framework, cognitive visualization.

1. Introduction

Research and development in geometric modeling field for Computer-Aided Design is hardly possible without specialized tools for visualization and analysis of geometry defined in Boundary Representation (B-Rep) form. In this paper, we present an open-source software which assists in handling a certain amount of common geometric issues. The software is based on the OpenCascade library as it is the only currently available open-source geometric modeling kernel. Another reason to use OpenCascade is that the data structures of the library are compatible with ISO 10303-42 (STEP). The latter fact brings a certain level of generality to the present approach. The similar kind of software is known to exist for other geometric kernels. However, other kernels are closed-source and are not readily available for an unrestricted academic and commercial use.

The remainder of this paper is organized as follows. Section 2 describes the motivations for starting yet another open-source project and lists the problems the project should be able to solve. Section 3 provides an overview of similar existing tools. Section 4 describes B-Rep inspection facilities of the new software. Section 5 examines how to use the software as a framework for prototyping new CAD-oriented algorithms. Section 5 concludes with several examples illustrating some real use cases where the software has been successfully used as a prototyping framework. Section 6 outlines our plans for future work.

2. Motivations

There are two primary motivations behind the idea to start yet another free and open-source (FOSS) software project dealing with CAD geometry algorithms. The first reason is the lack of freely available inspection utilities for Boundary Representation (B-Rep) geometry. It is widely recognized that a better visualization of a problem leads to a better understanding of the underlying science (see [Jern 1989]). The traditional visualization engines for Boundary Representation models are CAD systems. However, such systems do not allow for an easy inspection of boundary elements and their geometries. The existing visualization engines are primarily designed for end users and not for developers, although advanced CAD systems may offer comprehensive development kits in addition to mainstream software. The second motivation for starting this FOSS project is to allow for the so-called *reproducibility of research results* which is similar to the motivation behind the popular open-source MeshLab project presented by [Cignoni et al. 2008]. Our software simplifies prototyping new CAD

modeling algorithms and renders the results of the research group's work readily available for academic and engineering communities. Availability of the research results in the form of software prototypes is a key to solving the *technology transfer problem* raised by [Brown 1982].

Inspection of geometric and topological structures is of crucial importance when developing or profiling geometric modeling algorithms, such as Boolean Operations, fillets, offsets, etc. During such tedious working sessions, a lot of common questions may arise. Some of them are as follows: a) what is the host geometry (curve or surface type) of a particular edge or a face? b) what are the topological and geometric orientations of the geometric primitives? c) what is the topological structure of the model? d) how to find a particular boundary element in the model? e) what is the quality of a curve or a surface parameterization? f) is it necessary to repair the working geometry and how to do that? Answering these and many other similar geometric questions often requires laborious manual work. In this paper, we introduce a software which assists in geometric analysis and substantially reduces the efforts needed to "debug CAD model".

In addition to its inspection functions, the software can also serve as a framework for the development of geometric modeling algorithms. The software provides STEP-compliant geometric and topological primitives. These primitives are rendered using a set of readily available VTK pipelines. An algorithm which is developed within the framework of our system may benefit from the convenience interfaces which facilitate error reporting and graphical dumps (similar in nature to the commonly used textual diagnostic outputs). Therefore, the software establishes a framework for implementing multipurpose geometric tools for Computer-Aided Design applications. The software permitted to develop several CAD algorithms including some commercial ones.

Furthermore, it was discovered that the developed inspection and prototyping facilities are helpful in teaching the basic CAD concepts. According to [Rossignac 2004], CAD learners suffer from the unavailability of simplistic and intuitive models of the concepts they have to deal with during the education process. For example, students and even experienced engineers often confront such common problem as a misleading visual appearance of a CAD model. The geometry may look visually perfect, but contain a variety of internal quality phenomena, including wild or irregular parametrization of curves and surfaces, presence of redundant topological primitives, large tolerance gaps, and overlaps, etc. The first stage is to explain that what is displayed on the screen is not the real definition of a CAD part (What You See is NOT What You Get). This is exactly where a boundary representation

inspection utility can be advantageous. In this regard, the present software contributes to the *education-driven research* in CAD as formulated by [Rossignac 2004].

Finally, the efficient development of geometric modeling algorithms involves some mathematical abstractions which cannot be visualized directly in the modeling space. Such abstractions include adjacency graphs for boundary elements, parametric domains of faces, level sets of objective functions being optimized, etc. The visualization of such abstract data often gives a useful insight in a problematic situation where a human imagination may fail to recognize the cause of a problem. As mentioned by [Zenkin 1991], a system providing such *cognitive visualization* facilities is the so called “brainware” because it assists humans in their creative process.

3. Related work

Among the available frameworks for geometry processing, it is worth mentioning MeshLab (see [Cignoni et al. 2008]) and OpenFlipper (see [Mobius and Kobbelt 2012]) software tools. Both systems were designed with the similar motivation to make the development of geometric algorithms more efficient, and to bridge the gap between academic researchers and industry. Our framework pursues the same goals, but focuses on precise CAD instead of polyhedral geometry. Although OpenFlipper provides a generic architecture which applies to curved geometry, it does not use a full-featured embedded CAD kernel. Therefore, the development of precise CAD algorithms will be problematic in such systems.

[Averbukh et al. 2015] presented a framework for deriving specialized systems for cognitive visualization. The cognitive visualization tools facilitate the investigation of subtle algorithmic problems, and even the creation of new knowledge. According to [Brooks 1996], such systems serve as “intelligence amplification” tools. The ability to visualize information which does not correspond to any real object but represents some abstract or multi-dimensional data is of great importance for the efficient and thoughtful development of geometric algorithms.

In practice, CAD systems and specialized CAD tools are often used for the development of geometric algorithms. These systems provide extensible frameworks which can be exploited by software developers and researchers. Data communication between an algorithm and a CAD system can be either direct (an algorithm calls the corresponding API of a system) or indirect with the use of exchange file formats (e.g. STEP). In the latter case, a geometric algorithm maintains a certain level of independence, as it can be used in a CAD-neutral environment. CAD software which integrates a newly developed geometric algorithm should offer sufficient functionality which is sometimes rather advanced. For example, [Lee et al. 2005] base their experiments on a commercial platform which provides enough functionality to prepare CAD models for the engineering analysis. Likewise, [Seo et al. 2005] base their algorithm on Parasolid kernel which offers an advanced face removal operator. These examples illustrate that a particular technology can be essential to conduct research in a specific field.

The alternative approach consists in developing technology-neutral algorithms which communicate with an underlying geometric kernel using an abstract interface. Researchers have made several attempts to create such interface. Probably the most remarkable one is Djinn interface proposed by [Armstrong et al. 2002]. The Djinn interface eliminates dependency on the underlying geometric kernel. Moreover, it unifies representation of a geometric model with the use of the concept of *cellular*

topology. According to [Stroud and Nagy 2011], Djinn’s authors aimed to create a standard interface which could be used to share applications between researchers and to demonstrate research applications within a commercial environment. The existence of such interface could lead to a closer collaboration between researchers and industry. Therefore, Djinn contributes to solving the technology transfer problem raised by [Brown 1982]. Unfortunately, the concept of Djinn does not take into account commercial considerations. Indeed, software vendors prefer to have a code that is closely tied to their system. According to [Stroud and Nagy 2011], having Djinn interface in a CAD system would allow researchers to add new functions to a CAD system, such as feature recognition, analysis or manufacturing code. However, to develop a software system which is fully compliant with Djinn is a challenge.

We have chosen another approach which is essentially technology-dependent. At the same time, the employed technology (OpenCascade) is open-sourced and readily available for anyone from academia or industry. Moreover, this modeling kernel is known to provide only a small subset of functionality which is available in commercial libraries such as ACIS or Parasolid. Therefore, any algorithm based on OpenCascade can be transmitted to another library, at least in principle. From this point of view, OpenCascade-based algorithms can be considered as prototypes for exchange between academia and industry.

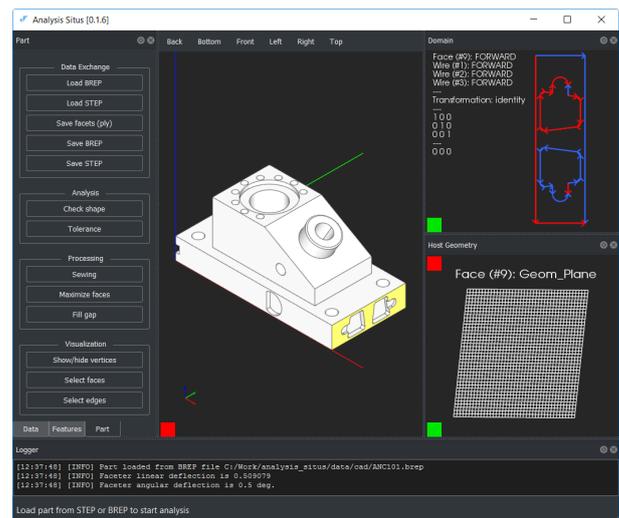


Fig. 1. The general view of the software. The desktop is composed of three viewers: part viewer, face domain viewer and host geometry viewer.

OpenCascade library provides its own prototyping framework called Draw Test Harness (or simply Draw). Although our software is not as multi-purpose as Draw, there are several advantages which make it more practical than Draw in some cases. First of all, it offers a combined layout of three principal geometric viewers, as shown in Figure 1. These viewers are interactive and do not require any console inputs like Draw. Among other features are explicit indexation of topological elements, availability of graph models, curvature maps and combs for geometric analysis, convexity analysis for edges, developer-oriented visualization and many other utilities which are not readily available in Draw.

4. Inspection facilities

For the three-dimensional visualization, VTK library by [Schroeder et al. 2006] was employed. The architecture of visualization pipelines offered by VTK allows for a clear

distinction between the data transformation *before* rendering and the rendering itself. Another remarkable advantage of VTK is a broad set of available visualization features which are commonly used by many scientists and software engineers all over the world.

4.1 B-Rep inspection

According to [Corney and Lim 2001], although different geometric kernels have their subtleties in B-Rep realization, there is one consensus that emerged earlier on. The well-turned idea was to separate geometric definitions of boundary elements from the information about how these elements are connected. This idea which divorces *topology* from *geometry* is presented in any commercial multi-purpose CAD geometric kernel.

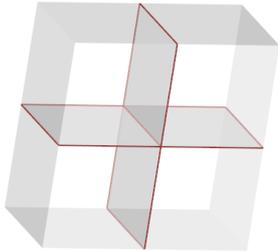


Fig. 2. Non-manifold edges (dark-red color) in a model composed of four boxes sharing their faces.

The visualization facilities for B-Rep structures introduced in our software are not unique. Similar techniques for insightful rendering of boundary representation structures are described by [Corney and Lim 2001]. When designing a visualization system, it is first necessary to define how boundary representation structures of a particular modeling kernel are implemented. Modern libraries have to deal with both *manifold* and *non-manifold* objects. OpenCascade, as well as ACIS, utilize OBJECT/CO-OBJECT (e.g. EDGE/CO-EDGE) data structure which is capable of representing non-manifold models (Figure 2).

4.2 Adjacency graphs

According to [Corney and Lim 2001], the great value of graphs in geometric modeling comes from their ability to represent the connectivity between geometric elements. The Scheme AIDE utility of ACIS allows for a visualization of different graphs including vertex-edge and face adjacency graphs.

In our software, two graphs can be visualized for a CAD part: the topology graph (Figure 3), which provides full information on how topological elements are nested into each other, and the attributed adjacency graph (Figure 4) representing adjacency relations between individual faces as explained by [Joshi and Chang 1988]. Both graphs give insight into the internal organization of the topological structure of a particular part (which is not visible without specialized inspection tools).

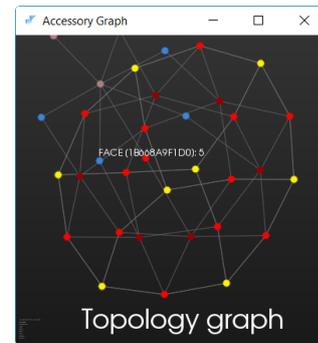


Fig. 3. Topology graph of a simple box model. Different colors are used to distinguish between different types of topological elements (compounds, solids, shells, faces, etc.).

Both graphs are synchronized with the part viewer which allows for easy searching of the face of interest in the topological structure of the model.

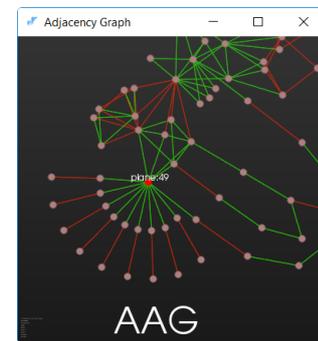


Fig. 4. Attributed adjacency graph of ANC101 model. The nodes correspond to faces; the arcs represent adjacency relations.

Graph representations are especially useful in algorithms which involve analysis of the connection between boundary elements of a model. Such information is fully used in *feature recognition* scenarios to extract the engineering semantics of a model. Other applications are *model simplification* for CAE, *local operations* (see [Fahlbusch and Roser 1995] for introduction), etc.

4.3 Geometry analysis

As mentioned by [Greiner et al. 1995], visualization is the indispensable tool for controlling the quality of geometric objects including their differential and aesthetic properties. A shaded display is usually insufficient to draw a conclusion on the quality of a modeling result. Therefore, additional techniques such as curvature maps (Figure 5), or isoparametric wireframe views are employed to detect surface irregularities.

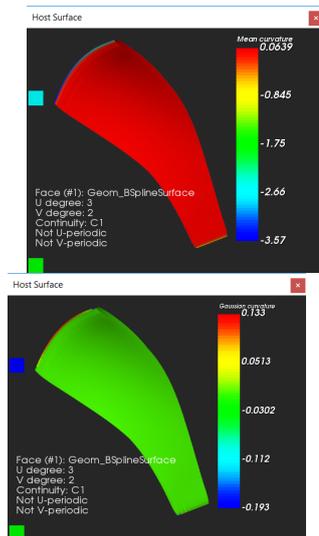


Fig. 5. Mean (on the left) and Gaussian (on the right) curvature maps for a B-spline blade surface.

5. Prototyping framework

According to [Mobius et al. 2013], researchers tend to confine their development efforts to implementing the algorithmic core and to *spend less time on software engineering*. Therefore, the authors of the OpenFlipper framework have implemented a set of common operations for data processing, visualization, I/O and many other components, including data structures, which are constantly and vainly re-invented by researchers and software engineers. Having all such components readily available allows developers to implement their innovative ideas in a rapid and goal-oriented way. Our main driving force was the idea to create new software, which will provide a similar framework for CAD-oriented research.

5.1 General architecture

Our software is designed to operate with a single CAD part. This restriction simplifies the general architecture of the software, thus allowing it to operate as a test bench for a single part geometry. The software is also capable of working with meshes and point clouds at the basic level. The framework provides such commonly used features as hierarchical organization of data objects, open/save in binary format, undo/redo, multi-purpose 3D and 2D viewers, etc.

Compared to MeshLab or OpenFlipper utilities which offer a plugin mechanism for software extension, we use a more straightforward prototyping approach. Our software provides a collection of components, including data structures, basic algorithms, visualization pipelines and typical GUI controls. A minimalistic application is provided to put these components together and to enable interactive B-Rep inspection. Every new algorithm being implemented within the framework should be supplied with its own independent executable. A prototype developer is free to choose particular GUI controls and to configure their communication. Such flexibility comes at a price. The developer is forced to write some sort of non-scientific code to bring everything together.

5.2 Use cases

In the following paragraphs, we present several use cases where our system helped to facilitate coding and debugging. Some of the mentioned algorithms were integrated into commercial software.

5.2.1 Feature recognition

Feature recognition allows to recover the semantics of a geometric model when the information about its features is unavailable. Feature recognition is a primary application for model inspection utilities, including edge and face indexation, edge convexity analysis (Figure 6), traversing adjacency graphs, etc. This sort of functionality has found a broad use in manufacturing planning where such features as holes, pockets, and bosses are detected in an input 3D model.

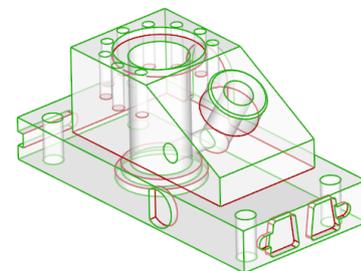


Fig. 6. Visualization of convex (green) and concave (red) edges for ANC101 model.

Figure 7 illustrates a sheet metal part (which was imported from a STEP file without the information on features) before and after recognition.

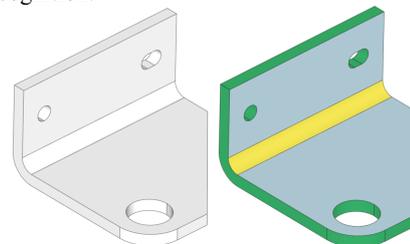


Fig. 7. The sheet metal part imported from STEP file (on the left) and feature recognition result (on the right) shown in different colors depending on the feature type.

In sheet metal manufacturing, the information about features can be used for fabrication cost estimation, validation of design by unfolding, preparation of 2D drawings, etc.

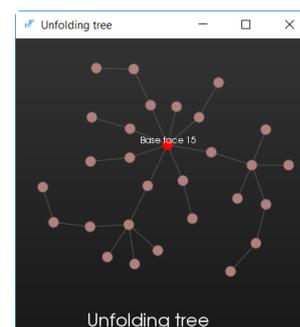


Fig. 8. Flattening graph for a sheet metal part. Each node corresponds to a flange being unfolded to a flat pattern.

The prototyping system has not only helped to perform an interactive inspection of sheet metal features, but has also assisted in the construction of explicit unfolding graphs (Figure 8). Such graphs provide information about the order of flattening and transformations associated with every flange.

5.2.2 As-built reconstruction

The present framework is used in research and development for automated reconstruction of CAD models from three-dimensional point clouds and meshes. We use VTK to visualize

massive point clouds and to let users interactively operate on their portions for segmentation and surface fitting. One example is the as-built reconstruction of industrial pipelines from laser-scan data (Figure 9).

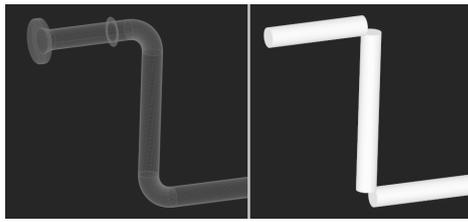


Fig. 9. The initial point cloud (on the left) and the extracted cylindrical primitives (on the right).

5.2.3 Contour capture

Contour capture algorithm is a variation of non-solid Boolean operation used to extract a certain piece of the two-dimensional (shell) model. The user interactively constructs a closed contour which follows the input geometry with a prescribed tolerance. The resulting polyline is then smoothed out using a dedicated reapproximation technique. Once the contour is smoothed, it can be then imprinted on the model by projection. At the last stage, the piece of geometry limited by the imprinted contour is extracted (Figure 10). Like any Boolean operation, this algorithm consists of several internal stages which can be implemented and tested separately. To facilitate the maintenance of this algorithm, we have used the novel framework which allowed to save a significant amount of time on debugging.

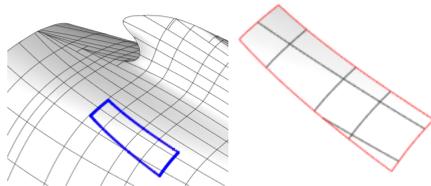


Fig. 10. Manually specified contour to capture (closed loop of line segments) and the result of capturing algorithm.

6. Conclusion and further work

The present software is available online at <http://www.analysisstus.org>. In future versions, we are planning to add new inspection facilities, i.e. to check fairness of curves and surfaces following the ideas presented in [Burchard et al. 1994], [Greiner et al. 1995], [Hagen et al. 1995], [Sapidis and Koras 1997] and other works.

7. References

- [1] Armstrong, C.G., Corney, J.R., Salmon, J.C., and Cameron, S. 2002. Djinn. A Geometric Interface for Solid Modelling.
- [2] Averbukh, V.L., Bakhterev, M.O., Vasev, P.A., Manakov, D. V. and Starodubtsev, I.S. 2015. Development of approaches to realization of specialized visualization systems. GraphiCon, 17–21.
- [3] Brooks, F.P. 1996. Toolsmith II. Communications of the ACM 39, 3, 61–68.
- [4] Brown, C.M. 1982. PADL-2: A Technical Summary.
- [5] Burchard, H.G., Ayers, J.A., Frey, W.H., and Sapidis, N.S. 1994. Approximation with Aesthetic Constraints. Designing Fair Curves and Surfaces, 3–28.
- [6] Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., and Ranzuglia, G. 2008. MeshLab: an Open-Source Mesh Processing Tool. Sixth Eurographics Italian Chapter Conference, 129–136.
- [7] Corney, J.R. and Lim, T. 2001. 3D Modelling with ACIS.
- [8] Fahlbusch, K.-P. and Roser, T.D. 1995. HP PE / SolidDesigner : Dynamic Modeling for Three-Dimensional Computer-Aided Design. Hewlett-Packard Journal, 6–13.
- [9] Greiner, G., Kolb, A., Pfeifle, R., Seidel, H., Encarna, M., and Klein, R. 1995. A Platform for Visualizing Curves and Surfaces. Computer-Aided Design 27, 7, 559–566.
- [10] Hagen, H., Hahmann, S., and Schreiber, T. 1995. Visualization and computation of curvature behaviour of freeform curves and surfaces. Computer-Aided Design 27, 7, 545–552.
- [11] Jern, M. 1989. Visualization of Scientific Data. Proceeding Advances in Computer Graphics V (Tutorials from Eurographics'89 Conf.), 1–17.
- [12] Joshi, S. and Chang, T.C. 1988. Graph-based heuristics for recognition of machined features from a 3D solid model. Computer-Aided Design 20, 2, 58–66.
- [13] Lee, K., Armstrong, C., and Price, M. 2005. A small feature suppression/unsuppression system for preparing B-rep models for analysis. Proceedings of the 2005 1, 212, 113–124.
- [14] Mobius, J. and Kobbelt, L. 2012. OpenFlipper: An open source geometry processing and rendering framework. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6920 LNCS, 488–500.
- [15] Mobius, J., Kremer, M., and Kobbelt, L. 2013. OpenFlipper - A highly modular framework for processing and visualization of complex geometric models. 2013 6th Workshop on Software Engineering and Architectures for Realtime Interactive Systems, SEARIS 2013; Co-located with the 2013 Virtual Reality Conference - Proceedings, 25–32.
- [16] Rossignac, J. 2004. Education-driven research in CAD. CAD Computer Aided Design 36, 14, 1461–1469.
- [17] Sapidis, N.S. and Koras, G.D. 1997. Visualization of curvature plots and evaluation of fairness: an analysis of the effect of 'scaling.' Computer Aided Geometric Design 14, 96, 299–311.
- [18] Seo, J., Song, Y., Kim, S., Lee, K., Choi, Y., and Chae, S. 2005. Wrap-around operation for multi-resolution CAD model. Computer-Aided Design and Applications 2, 1-4, 67–76.
- [19] Schroeder, W., Martin, K., Lorensen, B. 2006. Visualization Toolkit: An Object-Oriented Approach to 3D Graphics. Kitware, Colombia.
- [20] Stroud, I. and Nagy, H. 2011. Solid Modelling and CAD Systems: How to Survive a CAD System.
- [21] Zenkin, A. 1991. Cognitive Computer Graphics.

About authors

Vadim Turlapov (vadim.turlapov@itmm.unn.ru): professor of Computer Science, Lobachevsky Nizhny Novgorod University.

Alexander Malyshev (al.s.malyshev@rambler.ru): PhD student, Lobachevsky Nizhny Novgorod University.

Sergey Slyadnev (sergey.slyadnev@gmail.com): PhD student, Lobachevsky Nizhny Novgorod University.

T-сплайны: геометрическая гибкость и возможность локального изменения представления поверхности

А.С. Бакенов¹

bakenov_a@auca.kg

¹Американский университет в Центральной Азии, Бишкек, Кыргызская Республика

В этой статье представлено обобщение B-сплайн-поверхностей, выраженных в виде тензорных произведений, называемых T-сплайнами. Общепринятые B-сплайн-поверхности, выраженные в виде тензорных произведений, представляют собой прямоугольную сетку, а границы прямоугольника – это узловые линии. Представление T-сплайнами допускает существование T-узлов. Вместо использования векторов узлов информация об узле выражается путем назначения «узловых интервалов» для каждого ребра (для нечетного случая) или контрольной точки (для четного случая) контрольной сетки. T-сплайны обеспечивают локальное улучшение поверхности гораздо более естественным образом, чем B-сплайны: одна контрольная точка может быть вставлена вдоль любого края управляющей сетки или один сегмент линии может быть вставлен между любыми двумя краями управляющей сетки.

Ключевые слова: T-сплайны, неоднородные рациональные B-сплайны (NURBS).

T-splines: geometrical flexibility and local modification

A.S. Bakenov¹

bakenov_a@auca.kg

¹American University of Central Asia, Bishkek, Kyrgyz Republic

This article introduces a useful generalization of tensor-product B-spline surfaces called T-splines. In conventional tensor-product B-spline surfaces, the domain is a rectangular grid and rectangle boundaries are knot lines. A T-spline domain allows for the existence of T-junctions. Instead of using knot vectors, knot information is expressed by assigning a “knot interval” to each edge (for odd degree) or control point (for even degree) of the control grid. T-splines allow for local refinement in a much more natural manner than do B-splines: a single control point can be inserted along any edge of the control grid, or a single line segment can be inserted between any two edges of the control grid.

Keywords: T-splines, NURBS.

1. Введение

Почти 15 лет прошло с того времени, когда мне посчастливилось поработать у профессора Седерберга над исследованием [1] по решению проблемы соединения нескольких B-сплайн-поверхностей (далее патчей), выраженных в виде тензорного произведения, в одну единую поверхность. Математически строгие решения [2] существовали для выполнения такой операции слияния с использованием вставки узлов, но, к сожалению, это имело побочный эффект, приводивший к тому, что общее количество контрольных точек в объединенном B-сплайне часто было значительно больше, чем общее количество контрольных точек в исходных поверхностях. Данный недостаток показан на рис. 1, что приводит к необходимости обработки огромного количества данных.

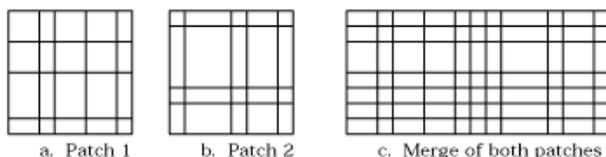


Рис. 1. Слияние двух B-сплайн-поверхностей.

В результате проведенного исследования было предложено обобщение B-сплайн-поверхностей, выраженных в виде тензорных произведений, называемых T-сплайнами.

Топологически контрольная сетка поверхности B-сплайна, выраженного тензорным произведением, должна быть прямоугольной решеткой (рис. 1). Контрольной сетке для T-сплайнов разрешено иметь T-узлы, как показано на

рисунке 2. Это свойство наделяет T-сплайны гораздо большей гибкостью, чем B-сплайны. Это позволяет T-сплайнам решать проблему «взрывной» генерации данных, присущей слиянию B-сплайновых поверхностей.

2. T-сплайны

Данный раздел содержит краткий обзор технологии на основе T-сплайнов, представленных в [3-4]. Контрольная сетка для T-сплайн-поверхности называется T-сеткой. Если T-сетка образует прямоугольную сетку, то T-сплайн вырождается в B-сплайн-поверхность.

Информация об узле для T-сплайнов выражается с помощью узлового интервала, неотрицательного числа, которое указывает разницу между двумя узлами. Каждому ребру в T-сетке присваивается интервал узла. На рисунке 2 показано предварительное изображение части пространства параметров T-сетки, выраженной в параметрическом пространстве $(s; t)$, где d_i и e_i обозначают интервалы узлов. Интервалы узлов должны удовлетворять условию: сумма всех узловых интервалов вдоль одной стороны любой грани должна быть равна сумме интервалов узлов на противоположной стороне. Например, на рис. 2 на грани F_1 $e_3 + e_4 = e_6 + e_7$ и на грани F_2 $d_6 + d_7 = d_9$.

Для определения системы координат сначала надо установить их начало. Например, на рис. 2 обозначим (s_0, t_0) началом координат. Далее можно найти координаты каждого узла сетки. Итак, основываясь на выборе начала координат, имеем $s_0 = t_0 = 0$, $s_1 = d_1$, $s_2 = d_1 + d_2$, $s_3 = d_1 + d_2 + d_3$, $t_1 = e_1$, $t_2 = e_1 + e_2$ и т.д.

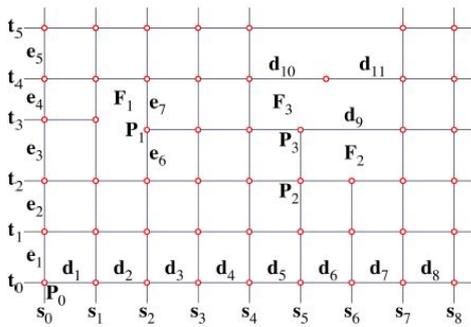


Рис. 2. Контрольная сетка для Т-сплайнов.

Кроме того, каждая контрольная точка имеет узловые координаты. Например, для $P_1 (s_2, t_2 + e_6)$.

Еще одно правило для Т-сетки из [3]: для возможности соединения Т-образного узла одной стороны грани с Т-образным точкой на противоположной стороне грани, нужно, чтобы сумма узловых векторов с обеих сторон была одинаковой. Например, для грани F_1 необходимо выполнение условия: $e_3 = e_6$ и $e_4 = e_7$.

Т-сплайн-поверхность определяется следующим образом:

$$P(s, t) = (x(s, t), y(s, t), z(s, t), w(s, t)) = \sum_{i=1}^n PiBi(s, t), \quad (1)$$

где $P_i(x_i, y_i, z_i)$ – контрольные точки, w_i – их веса соответственно, чьи декартовы координаты определяются $\frac{1}{w_i}(x_i, y_i, z_i)$.

Декартовы координаты точек на плоскости:

$$\frac{\sum_{i=1}^n (x_i, y_i, z_i) Bi(s, t)}{\sum_{i=1}^n w_i Bi(s, t)}. \quad (2)$$

$$B_i(s, t) = N[s_{i0}, s_{i1}, s_{i2}, s_{i3}, s_{i4}](s) N[t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}](t), \quad (3)$$

где $N[s_{i0}, s_{i1}, s_{i2}, s_{i3}, s_{i4}](s)$ – это кубическая базовая функция В-сплайна, которая связана с узловым вектором

$$s_i = [s_{i0}, s_{i1}, s_{i2}, s_{i3}, s_{i4}] \quad (4)$$

Аналогично, $N[t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}](t)$ связана с узловым вектором

$$t_i = [t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}] \quad (5)$$

Уравнение Т-сплайн-поверхности очень похоже на уравнение тензорного произведения для рациональной В-сплайн-поверхности. Разница между Т-сплайн-уравнением и В-сплайн-уравнением в том, как узловые векторы s_i, t_i определяются для каждой стыковочной (blending) функции $B_i(s, t)$.

NURBS-поверхности определяются с помощью набора контрольных точек, лежащих топологически в прямоугольной сетке, как показано на рис. 3 (а). Это означает, что большая часть точек в NURBS являются лишними, так как они не содержат большого количества геометрической информации, а необходимы лишь для удовлетворения топологических ограничений. С другой стороны, Т-сплайны могут иметь только часть точек в каждой строке сети, как показано на рис. 3 (b). Частичный ряд точек заканчивается Т-образным соединением – отсюда и название Т-сплайнов. Для процесса проектирования чем меньше количество точек, тем быстрее моделирование. Более того, для того чтобы сделать уточнение для Т-сплайнов, достаточно лишь сделать вставку одной контрольной точки. В случае с NURBS понадобится вставка точек вдоль всей строки. Следующим ограничением NURBS является то, что в большинстве случаев объект будет состоять из нескольких NURBS-поверхностей и после «склеивания» частей, возможно,

будут присутствовать разрывы. Т-сплайны позволяют конструировать гладкие без разрывов (или их еще называют «водонепроницаемые») поверхности.

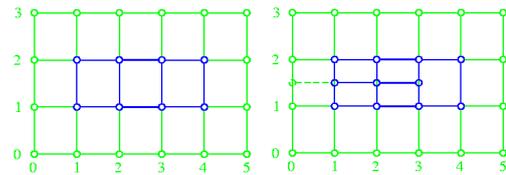


Рис. 3. Полигональная сеть для: а – NURBS, б – Т-сплайн.

Локальное уточнения кривых и поверхностей NURBS осуществляется через узел вставки, в процессе чего изменяется форма кривой или поверхности. Существует несколько алгоритмов для вставки узлов. Алгоритм вставки узла является фундаментальным, его можно использовать как математический инструмент для понимания и анализа В-сплайнов, так и в качестве практического инструмента для управления и построения В-сплайн кривых и поверхностей. Вставка узла хорошо отрабатывает для локального уточнения В-сплайн-кривой: вставка одного узла порождает обновления только нескольких точек в локальном участке. Тем не менее, для В-сплайн-поверхностей, выраженных в виде тензорного произведения, локальное улучшение невозможно, поскольку вставка одного узла в один из узловых векторов поверхности вызывает вставку новых точек в целую строку или столбец.

3. Алгоритм локального улучшения с помощью Т-сплайна

Процесс локального улучшения состоит в том, чтобы произвести вставку одной или нескольких точек в Т-сетку без изменения формы Т-сплайн-поверхности. Данная процедура также может называться локальной вставкой узла, так как контрольные точки для Т-сетки должны влиять на вставку дополнительных узлов при построении стыковочных функций.

Локальный алгоритм состоит из двух этапов: топологического и геометрического. На первом этапе определяется, какие контрольные точки надо будет вставить дополнительно. После того как все необходимые новые контрольные точки определены, вычисляются декартовы координаты и веса для искомой Т-сетки.

Рассмотрим теперь алгоритм топологического этапа. Важным ключом к пониманию решения проблемы является понимание того, как в Т-сплайне тесно связаны между собой стыковочные функции и Т-сетка: каждой контрольной точке соответствует стыковочная функция, и для каждой стыковочной функции по правилу 1 определяется узловой вектор.

Правило 1: определим узловые векторы s_i (4) и t_i (5) для точки P_i . Тогда узловой вектор s_i будет состоять из первых двух узлов, являющиеся общими для луча, выходящего из узла (s_i, t_i) вдоль оси s , и первыми двумя ребрами, с которыми луч пересекается. Аналогично определяются координаты узлового вектора t_i .

Например, на рис. 2 для точки P_1 узловые векторы имеют координаты $s_1 = [s_0, s_1, s_2, s_3, s_4]$ и $t_1 = [t_1, t_2, t_2 + e_6, t_4, t_5]$. Для P_2 : $s_2 = [s_3, s_4, s_5, s_6, s_7]$ и $t_2 = [t_0, t_1, t_2, t_2 + e_6, t_4]$. Для P_3 : $s_3 = [s_3, s_4, s_5, s_8]$ и $t_3 = [t_1, t_2, t_2 + e_6, t_4, t_5]$.

Предположим, что на некоторое время могут существовать стыковочные функции, которые будут нарушать правило 1. Рассмотрим три возможных нарушения, которые могут возникнуть в алгоритме.

- Нарушение 1: в стыковочной функции отсутствует узел, который удовлетворяет правилу 1 для текущей Т-сетки.

- Нарушение 2: стыковочная функция имеет узел, который не удовлетворяет правилу 1 для текущей Т-сетки.
- Нарушение 3: контрольная точка не имеет стыковочной функций, которая с ней связана.

Если нарушений не существует, Т-сплайн построен без нарушений. Если нарушения присутствуют, то алгоритм будет решать их по одному, пока никаких дальнейших нарушений не будет существовать. Топологический этап нашего алгоритма локального улучшения состоит из следующих действий:

1. Вставить все необходимые контрольные точки в Т-сетку.
2. Если стыковочные функции «виновны» в нарушении 1, выполнить необходимые вставки узла для таких стыковочных функций.
3. Если стыковочные функции «виновны» в нарушении 2, добавить соответствующие контрольные точки в Т-сетку.
4. Повторить шаги 2 и 3, пока не будут больше нарушений.

Решение всех случаев нарушения 1 и 2 автоматически решит все случаи нарушения 3.

Проиллюстрируем алгоритм на примере. На рис 4 (a) изображена начальная Т-сетка, в которую мы хотим вставить одну контрольную точку P_2 . Поскольку Т-сетка на рис. 4 (a) верна, то никаких нарушений нет. Но если мы просто вставим P_2 в Т-сетку (рис. 4, b) без изменения стыковочных функций, то сразу получим несколько нарушений. Поскольку P_2 имеет координаты узлов (s_3, t_2) , то четыре стыковочные функции становятся виновными в нарушении 1: с центром в точках (s_1, t_2) , (s_2, t_2) , (s_4, t_2) и (s_5, t_2) . Чтобы устранить эти нарушения, мы должны вставить узел s_3 в каждой из этих функций. Стыковочные функции с центром в точке (s_2, t_2) :

$$N[s_0, s_1, s_2, s_4, s_5](s) N[t_0, t_1, t_2, t_3, t_4](t)$$

Вставка узла s_3 в вектор s разбивает стыковочную функцию на две функции:

$$c_2 N[s_0, s_1, s_2, s_3, s_4](s) N[t_0, t_1, t_2, t_3, t_4](t) \text{ (рис. 4, c) и}$$

$$d_2 N[s_1, s_2, s_3, s_4, s_5](s) N[t_0, t_1, t_2, t_3, t_4](t) \text{ (рис. 4, d).}$$

Первая из функций удовлетворяет правилу 1. Кроме того, функции с центром в точках (s_1, t_2) , (s_4, t_2) и (s_5, t_2) удовлетворяют правилу 1. Тем не менее узловой вектор t стыковочной функции $d_2 N[s_1, s_2, s_3, s_4, s_5](s) N[t_0, t_1, t_2, t_3, t_4](t)$ имеет нарушение 2, поскольку для вектора $t = [t_0, t_1, t_2, t_3, t_4]$ правило 1 не выполняется для узла t_3 . Эта проблема не может быть устранена путем переработки стыковочной функции – следует добавить дополнительную контрольную точку в Т-сетку. Необходима контрольная точка P_3 на рис. 4 (e). Вставка контрольной точки устраняет нарушение 2, но возникает новый случай нарушения 1. Как показано на рис. 4(f) стыковочная функция с центром в (s_2, t_3) имеет узловой вектор s , который не включает в себя узел s_3 , что необходимо для удовлетворения правила 1. Вставка s_3 в узловой вектор решает проблему, и нет дальнейших нарушений правила 1.

Данный алгоритм всегда гарантированно заканчивает свое исполнение. В худшем случае алгоритм будет распространять везде ряды точек, чтобы пересечь всю поверхность. На практике, как правило, алгоритм требует нескольких дополнительных точек, кроме тех, которые пользователь хочет вставить. Таким образом, данный алгоритм улучшения имеет два существенных преимущества: алгоритм всегда обрабатывает и минимизирует количество дополнительных узлов для вставки.

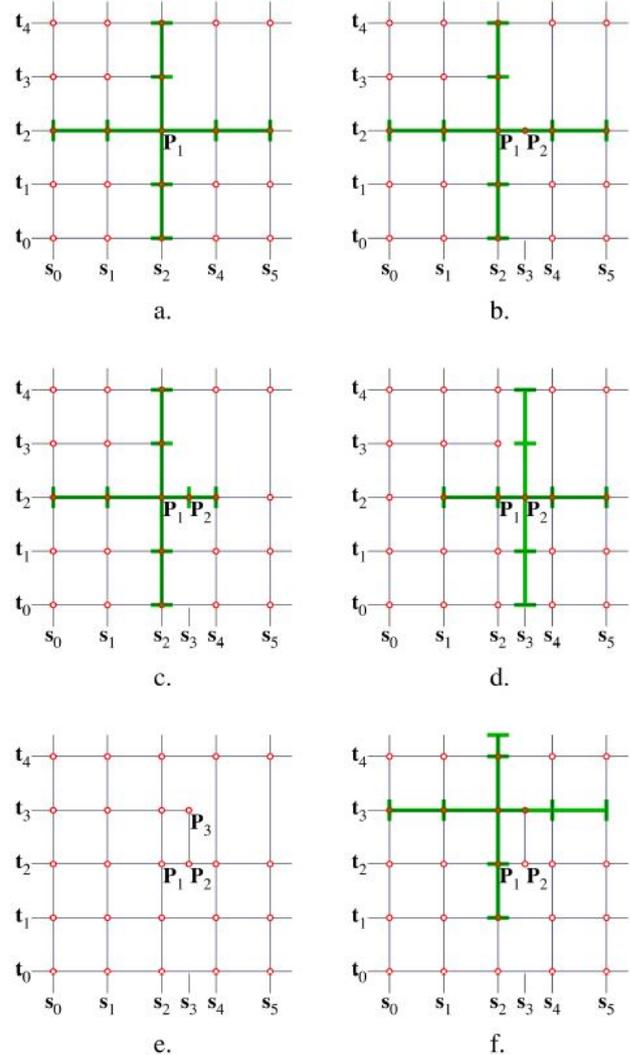


Рис. 4. Пример выполнения алгоритма локального улучшения.

4. Заключение

Т-сплайны представляют собой расширение NURBS-технологии, которое предоставляет возможность локального улучшения поверхности, бесшовного соединения нескольких патчей и решение проблемы обрезанной поверхности. Эти функции очень желательны в контексте дизайна, распространения 3D-печати и в последнее время стали доступны в системах проектирования компании Autodesk (Fusion360) [5].

5. Благодарности

Написание данной статьи после долгого перерыва с момента появления Т-сплайн-технологии стало возможным благодаря консультациям профессора МГУ Местецкого Л.М., участвовавшего по нашему приглашению в защите дипломных проектов студентов-бакалавров АУЦА в мае 2017 года.

6. Примеры фигур, созданных с помощью T-сплайн-технологии



Рис. 5. Заготовка для матрешки, созданная в программе Fusion360 в режиме «Sculpt» («Ваять»), функционирующей на основе T-сплайн-технологии.

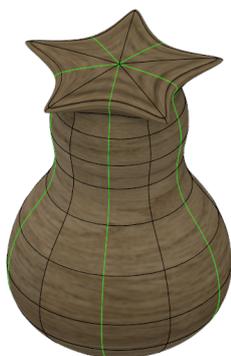


Рис. 6. Пример локального изменения с помощью T-сплайн-технологии.



Рис. 7. Узел, созданный по аналогии гравюры на дереве Маурица Корнелиса Эшера «Узел», 1965 г. [6].



Рис. 8. Тот же узел с показанными гранями T-сплайн-поверхности.

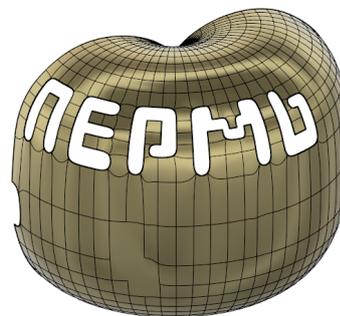


Рис. 9. На поверхности золотого пол-яблока путем удаления граней и вставок новых ребер продемонстрированы возможности T-сплайн-технологии. Получившаяся поверхность с надписью состоит из 1999 ребер и 962 граней.

7. Литература

- [1] Bakenov A. Tensor product B-spline surfaces with T-junctions. Master's thesis, Brigham Young University, Department of Computer Science, 2001.
- [2] Gerald Farin. Curves and surfaces for Computer Aided Geometric Design. Academic Press, Boston, 1988.
- [3] Sederberg T. W., Zheng J., Bakenov A., Nasri A. 2003. T-splines and T-nurcs. ACM Transactions on Graphics 22, 3 (July), 477–484.
- [4] Sederberg T. W., David L. Cardon, G. Thomas Finnigan, Nicholas S. North, Jianmin Zheng, and Tom Lyche. 2004. T-spline Simplification and Local Refinement. ACM Transactions on Graphics 23, 3 (July), 276–283.
- [5] <https://www.autodesk.com/products/fusion-360/>
- [6] <http://mcesher.ru/uspeh.html>

Об авторе

Бакенов Алмаз Сатыбалдиевич, директор Национального центра информационных технологий Кыргызской Республики и руководитель программы разработки программного обеспечения Американского Университета в Центральной Азии. Его e-mail bakenov_a@auca.kg.

Конструирование дуг обвода из кривых одного отношения

И.Г. Балуоба¹, Е.В. Конопацкий²
e.v.konopatskiy@donnasa.ru

¹Донецк, Донецкая Народная Республика;

²ГОУ ВПО «Донбасская национальная академия строительства и архитектуры», Макеевка,
Донецкая Народная Республика

В данной статье предлагается геометрический алгоритм и точечное описание дуги обвода, полученной на основе геометрической схемы конструирования кривых одного отношения. Приводится пример конструирования дуги обвода 1-го и 2-го порядка гладкости, после чего сделано обобщение способа конструирования плоской дуги обвода n-го порядка гладкости.

Ключевые слова: дуга обвода, кривая одного отношения, БН-исчисление, порядок гладкости, кривая Безье.

Constructing contour arcs from one relation curves

I.G. Baliuba¹, E.V. Konopatskiy²
e.v.konopatskiy@donnasa.ru

¹ Donetsk, Donetsk People's Republic;

²Donbas National Academy of Civil Engineering and Architecture, Makeyevka, Donetsk People's Republic

In this paper, we propose a geometric algorithm and a dot description of the arc contour obtained on the basis of a geometric scheme for constructing from the one relation curves. An example is given of constructing an arc contour of the smoothness by the first and second order, after which a generalization is made of the method for constructing a plane arc contour of the smoothness by n-th order.

Keywords: arc contour, the one relation curve, BN-calculation, smoothness by order, Bezier curve.

1. Введение

В инженерной практике часто встречаются задачи конструирования геометрических объектов нелинейной формы. Одним из основных способов решения такого класса задач является использование кусочных-гладких функций, к которым относятся различного рода обводы и сплайны. Здесь возникает отдельная задача – стыковка этих дуг между собой с наперед заданным порядком гладкости. Аналитически порядок гладкости в точке стыковки дуг определяется порядком производной функции так, чтобы значения производных в конечной точке исходной дуги и начальной точки последующей дуги совпадали. При этом большое значение имеет как вид дуги обвода, так и геометрическая схема стыковки дуг обвода между собой. Например, стыковка обводов 1-го порядка гладкости геометрически достигается за счёт совпадения касательных в точке стыковки. Обеспечить такое условие достаточно просто. Сложнее обеспечить 2-й порядок гладкости, который подразумевает совпадение кривизны в точке стыковки. И т.д., чем выше порядок гладкости, тем сложнее геометрический алгоритм формирования обвода, а соответственно и его аналитическое описание. Однако задачу построения обвода необходимого порядка гладкости можно значительно упростить, если все необходимые свойства заложить на стадии конструирования дуги обвода.

2. Из истории возникновения кривых одного отношения

Кривые одного отношения имеют достаточно широкое распространение в инженерной практике. Они обладают легкоуправляемой гибкостью и при этом описываются системой линейных уравнений. Например, знаменитая дуга кривой Безье (как квадратная, так и кубическая) является не чем иным, как кривой одного отношения, и может быть аналитически описана системой линейных точечных уравнений [1, 2].

Термин «кривые одного отношения» был впервые введен профессором Балуобой И.Г. в его диссертационной работе на соискание ученой степени доктора технических наук [3], который использовал кривые одного отношения для конструирования выпуклых замкнутых и незамкнутых обводов 1-го порядка гладкости. Дальнейшее обобщение и развитие кривые одного отношения получили в работах учеников Балуобы И.Г. [4-5].

3. Дуга обвода 1-го порядка гладкости

Рассмотрим более подробно геометрическую схему конструирования дуги квадратичной кривой Безье (рис. 1).

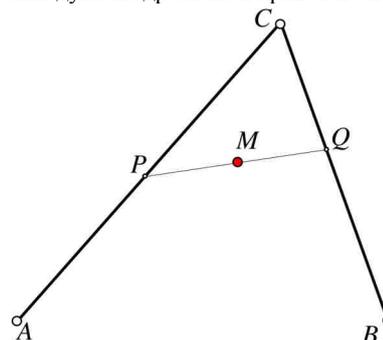


Рисунок 1: Геометрическая схема конструирования дуги квадратичной кривой Безье.

Дуга кривой Безье описывается текущей точкой M , движение которой определяется согласованным движением точек s с помощью одинакового отношения $\frac{AP}{AC} = \frac{CQ}{CB} = \frac{PM}{PQ} = t$, и имеет касательные AC и BC соответственно в точках A и B . Точечное уравнение такой дуги кривой будет иметь следующий вид:

$$M_1 = (A - C)t^2 + (B - C)t + C,$$

где A, B и C – исходные точки, определяющие симплекс двухмерного пространства;

t – текущий параметр, который изменяется от 0 до 1;

$\bar{t} = 1 - t$ – дополнение текущего параметра до 1.

Следует отметить замечательную особенность, которая заключается в том, что благодаря согласованному одинаковым параметром движению точек при значении параметра $t=0$ текущие точки P и M совпадают с исходной точкой A , а при $t=1$ текущие точки Q и M – с исходной точкой B . Выдвинем гипотезу, что именно это совпадение точек $A \equiv P \equiv M$ и $B \equiv Q \equiv M$ обеспечивает соответствующие касательные AC и BC , а следовательно и 1-й порядок гладкости в случае стыковки таких дуг. Проверим, подтвердится ли эта гипотеза для дуги обвода 2-го порядка гладкости.

4. Дуга обвода 2-го порядка гладкости

Модифицируем геометрическую схему конструирования дуги квадратичной кривой Безье (рис. 1), таким образом, чтобы в начальной и конечной точках дуги кривой выполнялись соответственно условия: $A \equiv P_1 \equiv P_2 \equiv M$ и $B \equiv Q_1 \equiv Q_2 \equiv M$. Для выполнения этого условия отношение соответствующих отрезков должно быть постоянным. Примем это отношение в качестве параметра: $\frac{AP_1}{AP_2} = \frac{P_1P_2}{P_1C} = \frac{CQ_1}{CQ_2} = \frac{Q_1Q_2}{Q_1B} = \frac{P_2M}{P_2Q_1} = t$ (рис. 2).

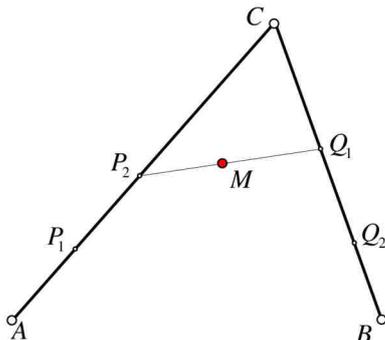


Рисунок 2: Геометрическая схема конструирования дуги кривой 2-го порядка гладкости.

Таким образом, получим систему линейных точечных уравнений, решая которую поочередно находим точки: P_1, P_2, Q_1, Q_2 и M .

$$P_1 = A \frac{\bar{t}}{1-t} + C \frac{t^2}{1-t}; \quad P_2 = A \frac{\bar{t}^2}{1-t} + C \frac{t}{1-t};$$

$$Q_1 = B \frac{t^2}{1-t} + C \frac{\bar{t}}{1-t}; \quad Q_2 = B \frac{t}{1-t} + C \frac{\bar{t}^2}{1-t};$$

$$M_2 = (A-C) \frac{\bar{t}^3}{1-t} + (B-C) \frac{t^3}{1-t} + C.$$

Перейдем от точечных уравнений, которые по своей сути являются символьной записью, к системе параметрических уравнений:

$$\begin{cases} x_{M_2} = (x_A - x_C) \frac{\bar{t}^3}{1-t} + (x_B - x_C) \frac{t^3}{1-t} + x_C; \\ y_{M_2} = (y_A - y_C) \frac{\bar{t}^3}{1-t} + (y_B - y_C) \frac{t^3}{1-t} + y_C; \\ z_{M_2} = (z_A - z_C) \frac{\bar{t}^3}{1-t} + (z_B - z_C) \frac{t^3}{1-t} + z_C. \end{cases}$$

Проверим, обладает ли полученная дуга кривой заявленными ранее свойствами. Для этого вычислим кривизну полученной дуги кривой в начальной и конечной

точках. В соответствии с [6], кривизна в декартовых координатах выражается следующей формулой:

$$k = \frac{\sqrt{(\dot{y}\ddot{z} - \dot{z}\ddot{y})^2 + (\dot{z}\ddot{x} - \dot{x}\ddot{z})^2 + (\dot{x}\ddot{y} - \dot{y}\ddot{x})^2}}{(x^2 + y^2 + z^2)^{\frac{3}{2}}}.$$

Воспользуемся программным пакетом *Maple* для расчёта кривизны в начальной и конечной точках. В результате, в обоих точках получаем одинаковое значение кривизны равно нулю. Причём это значение не зависит от координат точек симплекса ABC . Таким образом, полученная дуга обвода в начальной и конечной точках имеет точки спрямления, что может эффективно использоваться при моделировании обводов 2-го порядка гладкости инвариантных по отношению к геометрической схеме моделирования обвода. Т.е. при моделировании обвода появляется возможность использования уже известных ранее и хорошо изученных геометрических алгоритмов моделирования обвода 1-го порядка гладкости, модернизировав их предложенной дугой обвода.

5. Обобщение геометрической схемы конструирования дуг обвода n-го порядка гладкости

Предложенный геометрический алгоритм конструирования дуги обвода (рис. 1 и рис. 2) легко обобщить для построения дуг обвода высших порядков гладкости (более 2). Таким образом, были получены точечные уравнения дуг обвода 3-го, 4-го и 5-го порядка гладкости:

$$M_3 = (A-C) \frac{\bar{t}^4}{1-2t} + (B-C) \frac{t^4}{1-2t} + C,$$

$$M_4 = (A-C) \frac{\bar{t}^5}{1-3t+t^2} + (B-C) \frac{t^5}{1-3t+t^2} + C,$$

$$M_5 = (A-C) \frac{\bar{t}^6}{1-4t+3t^2} + (B-C) \frac{t^6}{1-4t+3t^2} + C.$$

В данном случае индекс при точке M соответствует порядку гладкости дуги кривой. Чтобы перейти от точечной формы представления уравнений к параметрической, достаточно выполнить по координатный расчёт, аналогично дуге обвода 2-го порядка гладкости.

Воспользуемся программным пакетом *Maple* для визуализации полученных дуг обвода. На рисунке 3 различным цветом показана эволюция дуг обвода с повышением порядка гладкости.

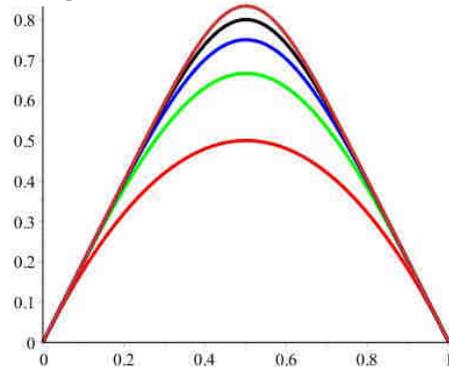


Рисунок 3: Визуализация дуг обвода различного порядка гладкости

Следует отметить, что при увеличении порядка гладкости наблюдается некоторое спрямление дуги обвода к начальной и конечной её точкам. Однако этот эффект был ожидаемым, поскольку это свойство заложено в геометрическую схему конструирования дуги обвода.

Как видно из полученных уравнений дуга обвода определяется только точками симплекса и текущим параметром t . Таким образом, для получения корректных результатов при построении выпуклого обвода необходимо наилучшим способом определить точку C , а точки A и B в любом случае обеспечат необходимые свойства стыковки дуг обвода.

6. Заключение

Представленный в работе способ конструирования дуг обвода, как кривых одного отношения, позволяет значительно упростить процесс стыковки дуг для моделирования кусочно-гладких кривых высоких порядков гладкости. К преимуществам предложенного способа конструирования дуг обвода можно отнести то, что необходимые свойства стыковки дуг между собой обеспечиваются геометрическим алгоритмом построения дуги и не зависят от геометрического алгоритма построения обвода.

7. Литература

- [1] Балюба, И. Г. Геометрическая сущность кривых Безье и их аналитическое представление [Текст] / И. Г. Балюба // Сучасні проблеми геометричного моделювання: Зб. праць Міжнародної науково-практичної конференції. Ч. 1. – Харків: ХПБ МВС України, 1998. – С. 178–182.
- [2] Балюба, И. Г. Конструювання плоских і просторових алгебраїчних кривих системою лінійних точкових рівнянь [Текст] / І.Г. Балюба, Є. В. Конопацький, Ж. В. Старченко // Праці / Таврійська державна агротехнічна академія. – Мелітополь : ТДАТА, 2002. – Вип. 4. Прикладна геометрія та інженерна графіка. Т. 17. – С. 66–67
- [3] Балюба, И. Г. Конструктивная геометрия многообразий в точечном исчислении [Текст]: дис. ... доктора техн. наук : 05.01.01 / И. Г. Балюба. – Макеевка, 1995. – 227 с.
- [4] Конопацький, Є. В. Геометричне моделювання алгебраїчних кривих та їх використання при конструюванні поверхонь у точковому численні Балюби-Найдиша [Текст]: дис. ... канд. техн. наук: 05.01.01 / Є. В. Конопацький. – Мелітополь, 2012. – 164 с.
- [5] Конопацький, Є. В. Криві третього порядку, як криві одного відношення [Текст] / Є.В. Конопацький, Ж.В. Старченко // Праці / Таврійський державний агротехнологічний університет. – Мелітополь: ТДАТУ, 2011. – Вип. 4. Прикладна геометрія та інженерна графіка. Т. 51. – С. 111–115.
- [6] Норден А.П. Краткий курс дифференциальной геометрии [Текст] / Норден А.П. // М.: ФИЗМАТГИЗ, 1958. – 244 с.

Об авторах

Балюба Иван Григорьевич, д.т.н., профессор, академик АНВО Украины, отличник образования Украины.

Конопацкий Евгений Викторович, к.т.н., доцент кафедры «Специализированные информационные технологии и системы» ГОУ ВПО «Донбасская национальная академия строительства и архитектуры». Его e-mail e.v.konopatskiy@donnasa.ru.

где f – функция, возвращающая значение целевой величины у для значений $x_{A,1}, \dots, x_{A,k}$ факторов, k, m – коэффициенты, позволяющие учесть разницу в единицах измерения значений величины y в эталоне и численном алгоритме. Если численные данные не требуют дополнительной настройки, значения их, соответственно, $k=m=1$. Параметр s позволяет единообразно рассматривать линейные и квадратичные отклонения.

Поскольку проверяемая полифигура задается набором точек, выбранным с некоторым шагом по каждому фактору (узлами k -мерной сетки), значение функции f известно только в узлах сетки, и для произвольного набора $x_{A,1}, \dots, x_{A,k}$ должно быть получено при помощи интерполяции [1, 6]. Чем меньше шаг сетки по каждому измерению (фактору), тем ближе к узлам сетки оказывается эталонный набор значений факторов $x_{A,1}, \dots, x_{A,k}$, тем выше достоверность результатов сравнения.

Процедура сравнения включает следующие шаги:

1. Формирование эталонной полифигуры a_s на основе контрольных данных.
2. Формирование проверяемой полифигуры a на основе результатов вычислений.
3. Сравнение полифигур в рамках общего способа наложения.

Поскольку проверяемая модель, как и эталон, состоит из одной фигуры, то проверку выполняет единственный вызов оператора определения меры различия двух полифигур [3].

3. Сравнение с учетом доверительного интервала

В случае сопоставления с набором экспериментальных данных интерес представляет сравнение в пределах некоторого диапазона для каждой точки. Организовать учет такого сравнения позволяет комплексная мера – многокомпонентная структура данных вида $\langle f_0, \dots, f_k \rangle$, каждый компонент которой формализует свой аспект различия сравниваемых объектов. При сложении мер компоненты суммируются независимо. Результат сравнения двух мер определяет старший компонент, в котором значения мер различны.

Пусть доверительный интервал для точки A задан значениями $[y_A - \delta_m; y_A + \delta_p]$, остальные координаты точки A и соответствующей точки проверяемой полифигуры равны $x_{A,1}, \dots, x_{A,k}$. Если значение $f(x_{A,1}, \dots, x_{A,k})$ не выходит из доверительного интервала, для точки A возвращается расстояние, равное $\langle \Delta_A; 0 \rangle$, если выходит – $\langle \Delta_A; 1 \rangle$. Тогда результирующая мера составит $\langle \Delta; N_f \rangle$, где Δ – суммарное или максимальное расстояние по формулам (1)-(2), а N_f – число точек, выходящих из доверительного интервала. Эталонная полифигура задается с указанием отклонений целевой величины в опорных точках $\langle x_{i,1}, \dots, x_{i,k}, y_i - \delta_{mi}, +\delta_{pi} \rangle$ или $\langle x_{i,1}, \dots, x_{i,k}, y_i, y_i - \delta_{mi}, y_i + \delta_{pi} \rangle$ либо с указанием доверительного интервала в абсолютных или относительных единицах $\langle x_{i,1}, \dots, x_{i,k}, y_i, \delta_i \rangle$.

4. Задание полифигур текстовым форматом

Для задания эталонной и проверяемой полифигур использовался простой и достаточно гибкий текстовый формат, в котором данные о координатах точек полифигуры располагаются в отдельных строках. Первые строки могут содержать дополнительные определения:

- $ids=$ структура строки данных. Во внутреннем представлении используются следующие названия основных осей координат – x, y, z, t . Ось y считается целевой; поэтому при сравнении полифигур, точки которых заданы перечислением значений, может понадобиться

поменять оси. Кроме названий осей в структуре строки данных используются y_{min} и y_{max} для явного указания границ доверительного интервала, $interval$ – для задания величины доверительного интервала координаты y для каждой точки. Примеры:

```
ids=x:y: _ymin:ymax
```

```
ids=y:x
```

- $interval=$ значение доверительного интервала в целом для полифигуры;

- $data=тип источника данных:параметры источника данных$. В качестве источника данных может выступать SQL-запрос к базе данных (тип *sql*), в этом случае параметры содержат строку запроса. В качестве источника данных для задач верификации рассматривается также аналитическое выражение (тип *law*), в этом случае параметры содержат текстовое представление выражения. Если источник данных не задан, далее следуют значения координат точек полифигуры:

```
ids=x:y:_:z
```

```
x1 y1 <лишнее значение1> z1
```

```
...
```

При описании данных допускаются счетчики и присваивания.

- $do=координата:начальное значение счетчика:шаг$.

Задаёт начальное значение указанной координаты и шаг изменения для следующей группы точек:

```
ids=y
```

```
do=x:0:0.01
```

```
y1
```

```
...
```

- $координата=общее значение для группы точек$. Задаёт общее значение координаты для группы точек:

```
ids=x:y
```

```
z=z1
```

```
x1 y1
```

```
...
```

Текстовое описание может быть легко получено средствами любого языка программирования или на основе данных, экспортированных из систем численного моделирования.

На рис. 1 показано сравнение расчетной полифигуры с эталонной на основе набора значений, снятых с осциллографа, из работы [7].

5. Проверка вычислительных подпрограмм

В случае сравнения расчетных данных с данными другой расчетной модели или аналитической зависимости (задача верификации) можно обеспечить совпадение значений факторов $x_{A,1}, \dots, x_{A,k}$ для проверяемой фигуры и эталонной, интерполяция в этом случае не требуется. Такое сравнение может быть полезно, например, для проверки вычислительных подпрограмм при их разработке, оптимизации и модификации.

Пусть требуется проверить корректность работы подпрограммы, реализующей вычисление некоторой функции в зависимости от значений k параметров – $y = f(x_1, \dots, x_k)$, и имеется таблица с достаточным числом контрольных значений величины y . Такая таблица может быть создана в другой версии подпрограммы, реализованной на другом языке программирования и вообще на другой вычислительной платформе.

Созданная один раз, далее она может применяться многократно. Преобразуем таблицу в эталонную полифигуру, как показано выше. Выполним вычисление y , используя проверяемую подпрограмму, и сформируем полифигуру для проверки. Осуществим сравнение полифигур, как показано выше.

Если подпрограмма реализует вычисление требуемой функции, результаты сравнения должны показать полную эквивалентность полифигур в пределах вычислительной погрешности. В строгом смысле такая эквивалентность не означает, что подпрограмма не реализует вычисление какой-то иной функции, но с увеличением числа контрольных точек вероятность ложной идентификации снижается.

Этот же подход может применяться для сравнения результатов работы различных конструктивно-геометрических алгоритмов. На рис. 2, а–в показана работа геометрических алгоритмов [2], реализующих так называемые ключевые (прогрессивный, лучевой и способ трапеции) способы образования поверхностей [4, 5] на одном и том же четырехзвенном контуре. На рис. 2, г–д показаны результаты сравнения поверхностей, полученных лучевым способом (λ) и способом трапеции (τ), с поверхностью, полученной прогрессивным способом (π), соответственно. Сравнение поверхностей λ и π показывает среднее расстояние между соответствующими точками – $3,9 \cdot 10^{-14}$ (максимальное – $4,6 \cdot 10^{-13}$), что с высокой вероятностью позволяет утверждать совпадение поверхностей (порядок 10^{-14} – 10^{-13} соответствует вычислительной погрешности).

6. Заключение

Было показано применение компонентов разработанной системы компьютерной проверки решений конструктивных задач на примере сравнения результатов численного моделирования с экспериментальными данными (задача валидации) и сравнения результатов работы двух алгоритмов формирования каркасов технических поверхностей (задача верификации).

Достоинство предложенного подхода заключается в автоматизации процедуры сравнения расчетных значений с контрольными при большом числе контрольных значений и в простоте представления эталонных и проверяемых данных, которые могут быть созданы в автоматическом, полуавтоматическом или ручном режиме средствами текстового редактора, любого языка программирования или получены из систем численного моделирования. Эталонные данные могут использоваться многократно.

Особенность предложенного алгоритма проверки способом наложения состоит в том, что используемый

в проверке эталон представляет собой, условно говоря, совокупность «рецепторов», сигнализирующих о тех или иных ошибках в ходе проверки, что позволяет наглядно представить ее результаты.

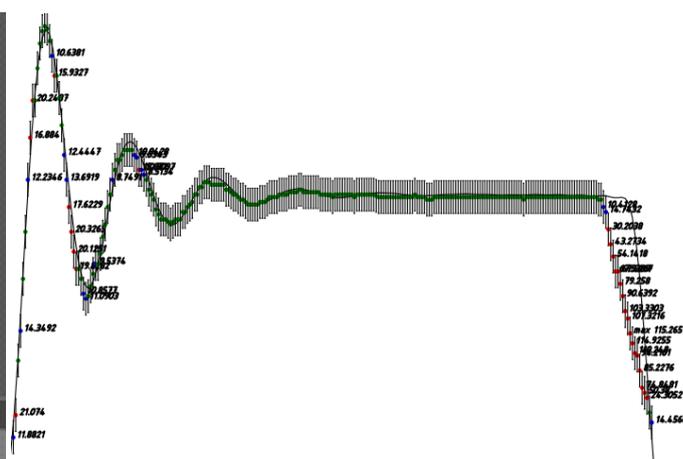
Таким образом предложенный алгоритм и компоненты разработанной системы могут применяться не только для автоматической проверки решений задач, связанных с представлением моделей геометрических объектов при помощи групп фигур, как это было показано в [3], но и в случае моделирования кривых и поверхностей на плоскости или в пространстве.

7. Литература

- [1] Бахвалов Н. С. Численные методы. – М.: Наука, 1985. – 631 с.
- [2] Бойков А. А. Элементы языка геометрических построений [электронный ресурс] // VII Междунар. научно-практ. интернет-конференция «Проблемы качества графической подготовки студентов в техническом вузе: традиции и инновации» (февраль – март 2017 г.). – Режим доступа: <http://dgng.pstu.ru/conf2017/papers/138/> (Дата обращения: 01.06.2017)
- [3] Бойков, А., Федотов, А. Автоматическая проверка решений задач инженерной геометрии // *GraphiCon 2016*. Труды Международной научной конференции. – Москва-Протвино, 2016. – С. 352–355.
- [4] Волошинов, Д. В. Конструктивное геометрическое моделирование / Д. В. Волошинов. – Saarbrücken: Lambert Academic Publishing, 2010. – 355 с..
- [5] Котов И. И. Геометрические основы ключевых способов построения поверхностей / И. И. Котов // Труды ВЗЭИ. Начертательная геометрия. – 1957. – Выпуск 10. – С. 15–36.
- [6] Роджерс Д., Адамс Дж. Математические основы машинной графики. – М.: Мир, 2001. – 604 с.
- [7] Шадриков, Т.Е. Особенности применения частотного анализа при расчете электрических цепей с транзисторными преобразователями напряжения / Т.Е. Шадриков, А.М. Соколов, А.В. Гусенков, В.Д. Лебедев, А.С. Страхов // *Электричество*. № 1. – 2016. – С. 4–12.



а)



б)

Рис. 1. Сравнение расчетной полифигуры с эталонной:

а – экспериментальная осциллограмма напряжения на нагрузке кабельной линии,

б – сравнение расчетной и экспериментальной кривой (фрагмент).

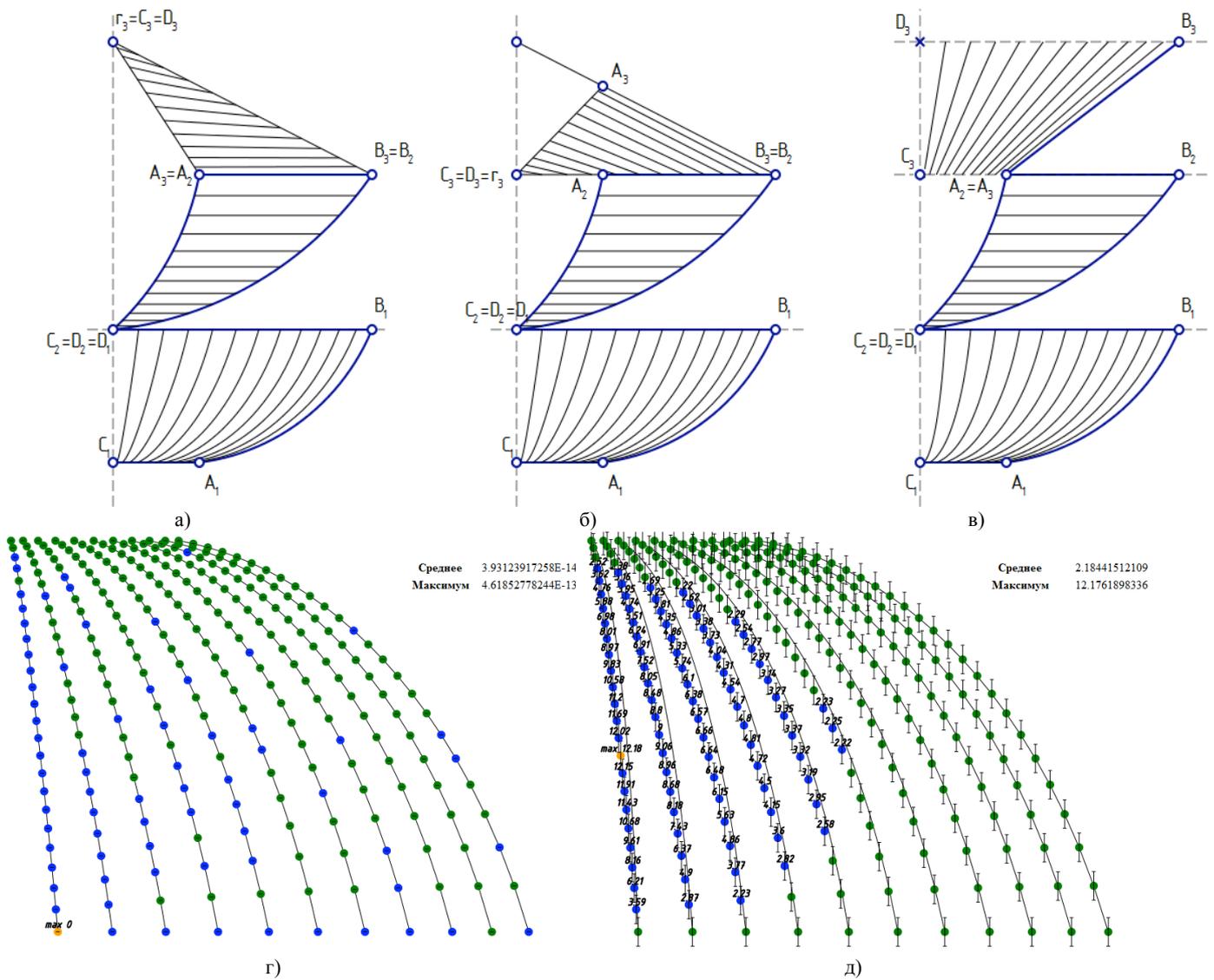


Рис. 2. Образование каркасов поверхностей ключевыми способами: прогрессивским π (а), лучевым λ (б) и способом трапеции τ (в), – и результаты поточечного сравнения каркасов поверхностей: λ с π (г), τ с π (д).

Визуализация различных модификаций алгоритма броуновского движения для построения фрактальных поверхностей

Брылкин Ю.В.

maderator@yandex.ru

МФ ФГБОУ ВПО МГТУ им. Н.Э. Баумана, Мытищи, Россия

Статья посвящена получению отсека поверхности, качественно аппроксимирующего геометрические параметры фрагмента реальной поверхности. Для достижения поставленной цели в статье описывается и визуализируется работа алгоритма на основе броуновского движения и нескольких его модификаций. Показано, что все алгоритмы приводят к построению поверхности с заданной фрактальной размерностью.

Ключевые слова: геометрическая модель, шероховатость, микроструктура, фрактал.

Brownian motion algorithm various modifications visualization to construct fractal surfaces

Brylkin Yu.V.

maderator@yandex.ru

Mytishinskiy branch FGBOU VPO «Bauman Moscow State Technical University», Mytishi, Russia

The article receiving surface compartment, efficiently approximating the real surface fragment geometrical parameters. To achieve this goal, the article describes and visualizes the algorithm is based on Brownian motion and a few modifications. It is shown that all algorithms lead to the surfaces construction with a given fractal dimension.

Keywords: geometric model, roughness, microstructure, fractal.

1. Введение

При испытаниях теплозащитных материалов изделий ракетно-космической техники особенно важно правильно интерпретировать получаемые данные. Сложность учёта геометрических характеристик шероховатости на микро- и наноуровне приводит к погрешностям при численном моделировании обтекания отсека поверхности потоком разреженного газа. В то же время, в работе [6] было показано, что реальная площадь поверхности медного сплава превышает площадь собственного отображения на плоскость в девять раз. Это означает, что скорость гетерогенных реакций на поверхности меди в девять раз больше, нежели на плоскости.

Оценка влияния микроструктуры поверхностного слоя возможна с помощью математического моделирования физико-химических процессов, протекающих на поверхности материала при тепловом воздействии. В ряде работ [2, 7, 10], при поиске универсального параметра, корректно описывающего геометрические характеристики шероховатости, было установлено, что фрактальная размерность [9], вычисленная для локального участка поверхности, заменяет целый комплекс амплитудных и шаговых параметров профиля поверхности. Опираясь на знание показателя фрактальной размерности, становится достаточно простым построение в программных комплексах негладких поверхностей, качественно аппроксимирующих геометрическую микроструктуру реальных материалов. При этом отпадает необходимость в получении данных о реальной геометрии поверхности исследуемого материала.

Созданию фрактальных моделей посвящено большое количество работ. Одним из самых интересных обобщений является работа [1], в которой описывается более десяти алгоритмов моделирования недифференцируемых форм для нужд компьютерной графики. Следует отметить, что наибольшее распространение получил итерационный метод с использованием броуновского движения. Количественной характеристикой данного алгоритма является число

итераций до достижения требуемого значения фрактальной размерности. Время на создание поверхности напрямую зависит от количества итераций.

2. Построение фрактальной поверхности

Работы [1, 11] описывают процесс создания фрактальной поверхности с помощью небольшого числа итераций. В то же время, в работе [8] показано, что данный способ моделирования требует значительного количества шагов. Здесь же предложено усовершенствование алгоритма на основе броуновского движения.

Рассмотрим более подробно процесс построения фрактальной поверхности с помощью алгоритма броуновского движения. За основу возьмём углеродосодержащее покрытие (рис. 1), нанесённое на керамику плазмохимическим способом [5]. Данные о геометрии микроструктуры получены с использованием сканирующей туннельной микроскопии [3].

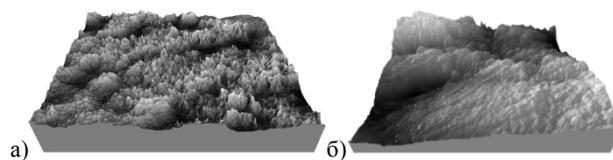


Рис. 1. Объёмная визуализация углеродосодержащего покрытия на керамике. а) отсек поверхности размером 150×150 мкм; б) отсек поверхности размером 2,05×2,05 мкм.

Вычисление фрактальной размерности D основано на использовании метода покрытия, т.е. построении эталонных элементов, покрывающих изучаемый отсек поверхности [8].

Для рассматриваемых структур вычисленная размерность составила $D = 2,46$. Это означает, что размерность лежит между 2D-объектом (проекцией отсека поверхности на координатную плоскость Oxy) и 3D-объектом (отсеком поверхности в пространстве).

Как было сказано выше, зная значение параметра D , появляется возможность построить эквивалентную геометрическую модель, с той же фрактальной размерностью.

Геометрический смысл алгоритма броуновского движения описанного в работе [1] заключается в проецировании сечений линейчатой поверхности на плоскость. На каждой итерации происходит изменение высотных отметок получаемых «осколков». Это наглядно продемонстрировано на рисунке 2.

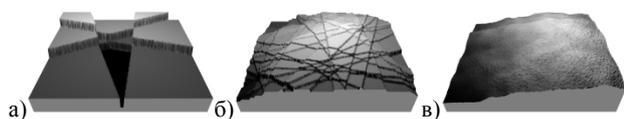


Рис. 2. Итерационное построение фрактальной поверхности проецированием сечений линейчатой поверхности.

Работа алгоритма останавливалась и визуализировалась после 4 итераций (рис. 2а), 32 итераций (рис. 2б) и по достижении фрактальной размерности $D \approx 2,5$ (рис. 2в). Фрактальная поверхность, полученная подобным образом, вполне корректно описывает параметры шероховатости реальной микроструктуры [8]. Такая модель становится пригодной для использования в численных расчётах физико-химических процессов, происходящих при взаимодействии разреженного теплового потока с поверхностью.

3. Модификация алгоритма построения

Тем не менее, как было показано в работах [4, 8] получение подобных результатов возможно сечением циклической поверхности (рис. 3а, рис. 3б, рис. 3в), сечением линейчато-циклической поверхности (рис. 3г, рис. 3д, рис. 3е) и проецированием пучка касательных к окружности (рис. 3ж, рис. 3з, рис. 3и). Однако, для этого требуется различное число итераций.

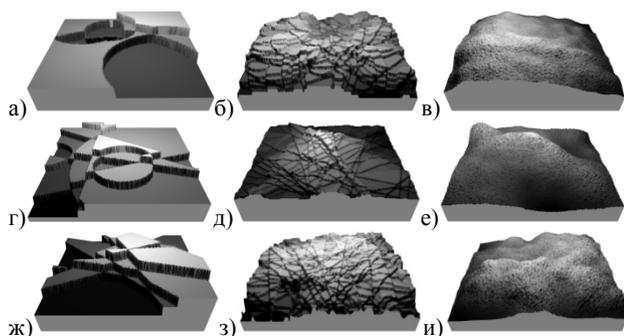


Рис. 3. Итерационное построение фрактальной поверхности.
а, г, ж) визуализация работы после 4 итераций;
б, д, з) визуализация работы после 32 итераций;
в, е, и) визуализация по достижении требуемого значения D .

Для дальнейшего снижения числа итераций была сделана попытка отойти от работы с двумерными объектами, что привело к модификации алгоритма с целью проецирования пучка касательных к сфере (рис. 4а), сечений тора (рис. 4б) и сечений торического узла (рис. 4в).

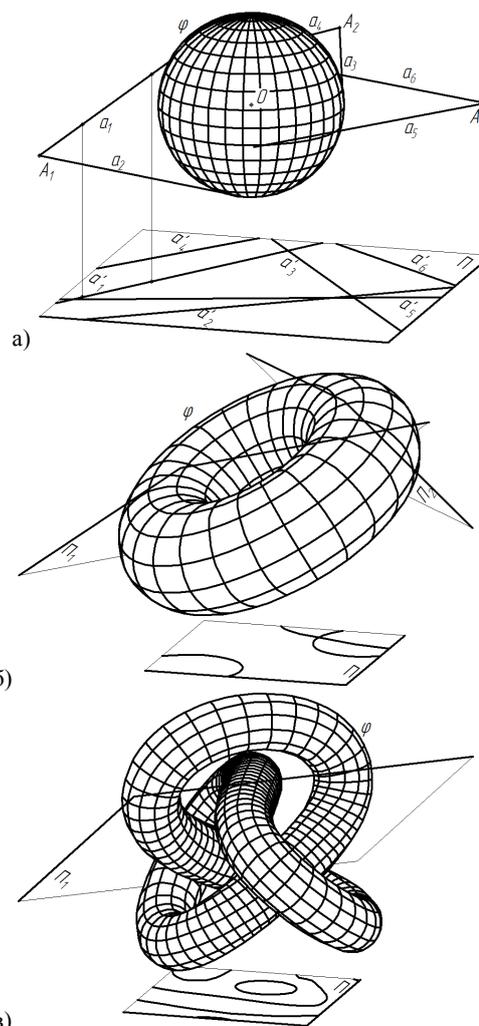


Рис. 4. Геометрическое представление работы алгоритмов.
а) проецирования пучка касательных к сфере;
б) проецирования сечений тора;
в) проецирования сечений торического узла.

Результаты работы алгоритмов проецирования пучка касательных к сфере (рис. 5а, рис. 5б, рис. 5в), сечений тора (рис. 5г, рис. 5д, рис. 5е) и сечений торического узла (рис. 5ж, рис. 5з, рис. 5и) так же приводят к построению фрактальной поверхности с размерностью $D \approx 2,5$.

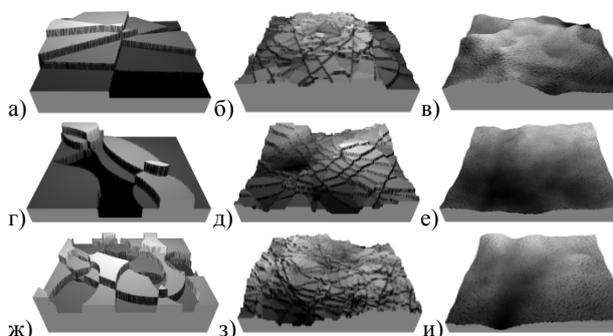


Рис. 5. Итерационное построение фрактальной поверхности.
а, г, ж) визуализация работы после 4 итераций;
б, д, з) визуализация работы после 32 итераций;
в, е, и) визуализация по достижении требуемого значения D .

Для оценки эффективности работы всех вышеперечисленных примеров построения фрактальной поверхности, рассмотрим рисунок 6.



Рис. 6. Количество итераций, характерное для каждого алгоритма.

Следует отметить, что на рисунке 6 показана диаграмма со средним количеством итераций n для каждого алгоритма. Данный параметр усреднялся по пяти результатам работы алгоритма. Разброс значений связан со случайным выбором точек при построении касательных к рассмотренным формам пространства, или их секущих плоскостей.

4. Заключение

Результаты работы всех алгоритмов, приводят к построению фрактальной поверхности, качественно аппроксимирующей микроструктуру реальной поверхности. Вследствие этого, при численном моделировании физико-химических процессов, происходящих на поверхности, отпадает необходимость использования экспериментальных данных, получаемых при исследовании материалов с помощью микроскопических методов.

Наименьшее число итераций до достижения требуемой фрактальной размерности обеспечивается алгоритмом использующим в построении сечения торического узла. Снижение количества итераций при использовании поверхностей в трёхмерном пространстве соответствует идее расслоения в решении многомерных геометрических задач и конструировании расслояемых кремниевых преобразований. Использование не только самоподобных и самоаффинных но и самопроективных и самобирациональных фракталов возможно позволит в дальнейшем достичь необходимых значений дробной размерности за ещё меньшее число шагов.

5. Благодарности

Автор выражает огромную благодарность своему научному руководителю, профессору Иванову Геннадию Сергеевичу за наставническую деятельность, а научному консультанту Кусову Андрею Леонидовичу за поддержку и помощь в освоении теории фракталов.

6. Литература

- [1] Giliam J.P. De Carpentier. Interactively synthesizing and editing virtual outdoor terrain. MA thesis. Delft University of Technology. – 2007.
- [2] Аксенова О.А. Фрактальное моделирование шероховатой поверхности при аэродинамическом расчете в разреженном газе // Аэродинамика. – СПб.: Изд-во С.-Петербург. Университета. – 2000. – С. 120-129.
- [3] Бахтизин Р.З. Сканирующая туннельная микроскопия - новый метод изучения поверхности твердых тел // Соросовский образовательный журнал. – 2000. – №11. – С. 83-89.

- [4] Брылкин Ю.В. Рационализация алгоритма моделирования поверхности методом броуновского движения по критерию минимизации количества итераций // Геометрия и графика. – 2017. – Т.5. – №1. – с.43-50.
- [5] Брылкин Ю.В., Залогин Г.Н., Красильников А.В., Рудин Н.Ф. Синтез углеродных наноструктур из метана плазмохимическим способом // Тверь: Твер. гос. ун-т. // Межвуз. сб. науч. тр. Физико-химические аспекты изучения кластеров, наноструктур и наноматериалов. – 2016. – Вып. 8. – с.76-81.
- [6] Брылкин Ю.В., Кусов А.Л. Исследование микро- и наноструктуры поверхности медного сплава с использованием теории фракталов // Космонавтика и ракетостроение. – 2016. – №5(90). – с.89-95.
- [7] Герасимова О.Е., Борисов С.Ф., Проценко С.П. Моделирование шероховатой поверхности // Математическое моделирование. – 2004. – Т. 16. – №6. – С. 40-43.
- [8] Иванов Г.С., Брылкин Ю.В. Фрактальная геометрическая модель микроповерхности // Геометрия и графика. – 2016. – Т. 4. – №1. – С. 4-11. DOI:10.12737/18053
- [9] Мандельброт Б. Фрактальная геометрия природы / Б. Мандельброт. – М.-Ижевск, Ижевский институт компьютерных исследований, НИЦ «Регулярная и хаотическая динамика», 2010. – 656 с.
- [10] Потапов А.А., Булавкин В.В., Герман В.А., Вячеслава О.Ф. Исследование микрорельефа обработанных поверхностей с помощью методов фрактальных сигнатур // Журнал технической физики. – 2005. – Т.75 – Вып.5.
- [11] Шишкин Е.И. Моделирование и анализ пространственных и временных фрактальных объектов / Е.И. Шишкин. – Екатеринбург: Урал. гос. ун-т. – 2004. – 88 с.

Об авторах

Брылкин Юрий Владимирович, аспирант кафедры инженерной графики МФ ФГБОУ ВПО МГТУ им. Н.Э. Баумана. Его e-mail maderator@yandex.ru.

Особенности обучения компьютерной графике в рамках курса «Информатика» для непрофильных специальностей и направлений

Н.Н. Василюк¹

nadia-vasiluk@yandex.ru

¹ПГНИУ, Пермь, Российская Федерация

В настоящей статье изложены особенности методического подхода к обучению компьютерной графике в рамках курса «Информатика» на непрофильных по отношению к этому предмету специальностях и направлениях. Сформулированы цель и задачи обучения в рамках указанной темы, описан отбор содержания обучения, прежде всего с точки зрения подбора программного обеспечения. Приводятся темы лабораторных работ и примерное описание заданий этих работ.

Ключевые слова: информатика, обучение, компьютерная графика, содержание обучения, лабораторные работы.

Aspects of computer graphics' learning within the course named «Informatics» for non-core specialties and directions

N.N. Vasiluk¹

nadia-vasiluk@yandex.ru

¹PSNIU, Perm, Russian Federation

In this article, the aspects of the methodical approach to teaching computer graphics within the course named «Informatics» on non-core specialties and directions in relation to this subject are described. The aim and objectives of the training are formulated within the framework of this topic, the selection of the content of the learning is described, primarily in terms of software selection. Themes of laboratory works and approximate tasks are given.

Keywords: informatics, learning, computer graphics, content of the learning, laboratory works.

Дисциплина «Информатика» в вузе входит в учебный план любой непрофильной по отношению к этой дисциплине специальности и направления. Ее целью является формирование знаний основ информатики и базовых умений и навыков в сфере информационных технологий, необходимых каждому образованному человеку и специалисту, с учетом компетенций, представленных в федеральном государственном образовательном стандарте. В свою очередь, формирование профессиональных, в узком смысле слова, компетенций, входит в задачи других дисциплин, изучаемых на старших курсах.

Стандартизация курса «Информатика» в ПГНИУ позволила унифицировать ее содержание, технологии и методики обучения, реализовать новые формы организации учебного процесса. В рамках уменьшающегося количества аудиторных часов особенно важно грамотно подбирать методы и средства обучения, с целью наиболее эффективного формирования необходимых компетенций. Прослеживание связей между школьным и вузовским курсом информатики помогает студентам более осознанно воспринимать как теоретическую информацию, так и овладевать практическими навыками.

Область информатики, связанная с компьютерной графикой, охватывает все виды и формы представления изображений, доступных для восприятия человеком либо на экране монитора, либо в виде копии на внешнем носителе. Она находит применение не только в компьютерном мире, но и в различных сферах человеческой деятельности: научных исследованиях (визуализация строения вещества, векторных полей и т. д.), медицине (компьютерная томография), опытно-конструкторских разработках и т. п. [2]

Компьютерная графика и анимация — необходимый инструмент в таких областях, как кино, реклама, искусство, архитектурные презентации, создание прототипов и имитации динамики, а также в создании компьютерных игр и обучающих программ [1]. Постоянно появляются новые области применения компьютерной графики, и

соответственно необходимы педагогические и методические подходы к подготовке как будущих специалистов в этой области, так и выпускников, не являющихся профессионалами в сфере ИТ.

Обучение компьютерной графике — одно из важнейших направлений использования персонального компьютера, на сегодняшний день рассматривается как важнейший компонент образования. Достижения в области ИКТ актуализируют вопросы подготовки специалиста в области представления информации в виде графических образов: чертежей, схем, рисунков, эскизов, презентаций, визуализаций, анимационных роликов, виртуальных миров и т. д. [4]

Безусловно, раздел «Компьютерная графика» является обязательным компонентом содержания обучения для дисциплины «Информатика». Поэтому вследствие неизбежного ограничения количеству часов, отводимых на изучение этой темы, нами выделена проблема исследования: каким образом выстроить методическое обеспечение темы «Компьютерная графика» для наилучшего достижения планируемых результатов обучения?

Согласно тематическому планированию курса «Информатика», изложенному в статье Е.К. Хеннера «Курс информатики в классическом университете» [3], тема «Графические редакторы» входит в модуль «Технологии обработки текстовой, числовой и графической информации, подготовки презентаций». На этот раздел отводится 9 часов, из них: 1 час — лекционный, 4 часа — лабораторные работы, 4 часа — самостоятельная работа студентов. Теперь подробнее рассмотрим цель, задачи, содержание, формы и методы, используемые при обучении студентов рассматриваемому разделу информатики.

Целью обучения любой дисциплины является усвоение обучающимися содержания на уровне, требуемом федеральными государственными образовательными стандартами высшего профессионального образования (ФГОС ВПО). В случае темы «Компьютерная графика» мы рассматриваем в качестве целевой следующую

компетенцию из ФГОС ВПО: «владение основными методами, способами и средствами получения, хранения, переработки информации, готовность использовать компьютер как средство работы с информацией».

Переосмыслив представленную выше компетенцию с точки зрения рассматриваемой темы, сформулируем следующие задачи, которые должны быть решены во время обучения. Итак, в ходе изучения темы «Компьютерная графика» студент должен:

- сформировать понятие о способах хранения графических изображений в памяти компьютера,
- овладеть методами создания и обработки графических изображений,
- выработать базовые навыки работы с различными видами графических редакторов.

Содержание темы «Компьютерная графика» отбирается с опорой на требования, предъявляемые ФГОС ВПО третьего поколения, а также с учетом ограниченного количества часов, отведенных на изучение темы.

Помимо этого, отбор содержания по теме «Компьютерная графика» производится исходя из следующих принципов:

- в содержании темы должно найти отражение научное содержание предметной области «Компьютерная графика»;
- должны быть реализованы требования каждого из ФГОС как по позиции «студент должен знать, уметь и владеть», так и по набору формируемых компетенций;
- должна сохраняться преемственность по отношению к школьному информатическому образованию;
- должны быть освоены графические редакторы общего назначения, на их основе сформированы общие умения и навыки создания и обработки изображений.

Представляется крайне важным, в современных стремительно развивающихся условиях и быстро устаревающим программном обеспечении подобрать теоретические сведения и составить лабораторные работы таким образом, чтобы после изучения темы студенты оказались в состоянии применять полученные умения и навыки для учебных или, возможно, каких-то личных целей. Теоретический материал, излагаемый во время лекционного занятия, должен отражать наиболее значимые сведения о принципах хранения изображений в ЭВМ, о различных типах графики, о всевозможных направлениях применения графических и мультимедийных редакторов.

Практическое же содержание темы с учетом выделенного количества аудиторных часов разбито на две лабораторных работы: «Растровая графика» и «Векторная и трехмерная графика». Подбор графических редакторов для лабораторных и самостоятельных работ осуществляется исходя из возможностей лаборатории, кафедры или учебного заведения (некоторые графические пакеты имеют довольно высокую стоимость). При этом не следует забывать, что в дальнейшей деятельности студенты, скорее всего, будут лишены возможности работать с дорогостоящим программным обеспечением, так как их будущая специальность может быть никак не связана с информационными технологиями. Оптимальным вариантом является выбор доступных для большинства непрофессионалов графических редакторов, которые либо являются свободно распространяемыми, либо относятся к бесплатным для некоммерческого использования с ограниченным функционалом.

В ПГНИУ для изучения темы «Компьютерная графика» используется следующее программное обеспечение. В рамках лабораторной работы «Растровая графика» студенты сначала работают с несложным встроенным графическим редактором Paint, с которым они, как правило,

уже познакомились в школьном курсе информатики. В работу включены два-три задания на работу с графическими примитивами, копированием и перемещением объектов, поворотом и отражением элементов изображения. Задания служат своеобразным введением в тему «Компьютерная графика», связующей нитью между школьным и вузовским курсом информатики.

Следующие задания лабораторной работы по этой теме нацелены на ознакомление с редакторами, обладающими более обширными функциональными возможностями. По соображениям, изложенным выше, для ознакомления был выбран бесплатно распространяемый редактор GIMP. В данной программной среде студентами отрабатываются навыки работы со сложными инструментами, а также работа со слоями и создание анимационных эффектов.

В рамках лабораторной работы «Векторная и трехмерная графика» студенты осваивают создание изображений при помощи свободно распространяемого векторного редактора Inkscape: выполняют задания, связанные с простейшими инструментами, манипулированием объектами, добавлением и форматированием текста. Вторым изучаемым редактором в рамках этой лабораторной работы является SketchUp – программа для моделирования относительно простых трехмерных объектов. В настоящее время нами используется версия, называемая SketchUp Make, которая бесплатна для некоммерческого использования, хотя и ограничена по функциональности. Студенты получают возможность в ходе выполнения простых заданий познакомиться с трехмерной графикой: создать трехмерный эскиз дома, посмотреть на него с различных сторон, заглянуть в окна и т.п.

По итогам каждой лабораторной работы студентам выдается индивидуальное задание для самостоятельной работы. Например, после лабораторной работы «Растровая графика», выдается следующее задание: создать коллаж с подписью из 4-5 фотографий на тему «Мой университет». Студенту необходимо подобрать фотографии (как вариант – самостоятельно запечатлеть то, что ему кажется самым интересным в университете, на цифровые снимки), выбрать редактор, в котором он будет создавать коллаж, обработать фотографии, создать надпись, поработав с текстом. Пример выполнения такого индивидуального задания представлен на рис. 1.



Рис. 1. Пример выполнения индивидуального задания по теме «Растровая графика».

Преобладающими методами обучения в рамках темы «Компьютерная графика» являются следующие: *репродуктивный* (при обработке практических навыков работы с графическими редакторами в компьютерном классе) и *частично-поисковый* (подбор инструментов для наилучшего выполнения индивидуального задания после каждой лабораторной работы). Сравним репродуктивные и

частично-поисковые задания, взяв для примера изучение графического редактора SketchUp.

В качестве *репродуктивного* задания представим создание трехмерной модели небольшого домика. Цель такого задания – познакомить обучаемых с инструментами редактора, продемонстрировать базовые возможности создания с их помощью трехмерного изображения. Для построения модели студентам выдается следующий алгоритм действий, сопровождаемый устным комментарием преподавателя:

1. Нарисуйте прямоугольник и сделайте из него коробку при помощи инструмента «Тяни\Толкай».

2. Вырежьте часть одной грани прямоугольника при помощи инструментов «Окружность» и «Тяни\Толкай».

3. Поверните модель и изучите ее со всех сторон, воспользовавшись инструментом «Орбита». Для масштабирования и перемещения примените «Панораму» и «Масштаб».

4. Нарисуйте на верхней грани коробки линию, следите за подсказками – линия должна идти от одной грани до другой из центральной точки («точка середины», подсвечивается синим цветом).

5. Выделите линию с помощью стрелки («инструмент «Выбрать») и, воспользовавшись инструментом «Переместить», потяните вверх, создавая крышу дома.

6. Создайте при помощи линий (обращая внимание на их цвет) прямоугольник на крыше, который послужит прообразом печной трубы.

7. Растяните этот прямоугольник при помощи инструмента «Тяни\Толкай», лишние линии удалите «Ластиком».

8. Сделайте для трубы заливку (инструмент «Заливка»), имитирующую кирпичи, для стен дома – имитирующую камень.

Таким образом, в ходе выполнения этого задания студенты применяют по образцу только что усвоенные знания. Подобные репродуктивные задания особенно эффективно содействуют отработке практических умений и навыков.

Примером *частично-поискового* задания может служить создание трехмерной модели дома мечты. Цель такого задания – закрепить навыки работы с инструментами графического редактора, а также научиться подбирать инструменты, цвета и заливку для построения задуманного изображения. Студентам предлагается придумать произвольное здание, в котором они хотели бы жить, и создать трехмерную модель такого дома в графическом редакторе SketchUp. На изображение накладывается ряд требований: модель не должна копировать те, что создавались по выданным ранее алгоритмам, а в процессе создания должны быть задействованы различные инструменты (не менее пяти), причем хотя бы один из тех, которые не использовались ранее.

Выполнение подобных частично-поисковых заданий способствует более осмысленному и самостоятельному овладению знаниями и развитию навыков творческой учебно-познавательной деятельности.

Основной формой обучения являются лабораторные работы. Выделим также индивидуальную форму обучения, которые применяются во время выполнения заданий в компьютерном классе, что обеспечивает комфортные условия работы и развитие самостоятельности.

Средствами контроля при изучении темы «Компьютерная графика» будут служить проверка выполнения практических заданий во время лабораторных работ в компьютерном классе, и индивидуальных заданий, которые студенты выполняют самостоятельно. Диагностика теоретических знаний осуществляется при

помощи компьютерного тестирования. Диагностика практических умений осуществляется в ходе проверки выполненных практических заданий, оценка за которые выставляется по двухбалльной шкале: «зачтено» – «не зачтено».

Результатом, диагностической целью изучения темы «Компьютерная графика» является готовность студента использовать графические редакторы для создания и обработки изображений в процессе обучения. Уровень сформированности подобной готовности у студентов мы можем оценить по итогам выполнения индивидуальных заданий.

Схематично изложенные выше особенности методики обучения компьютерной графике в рамках курса «Информатика» представлены на рис. 2.

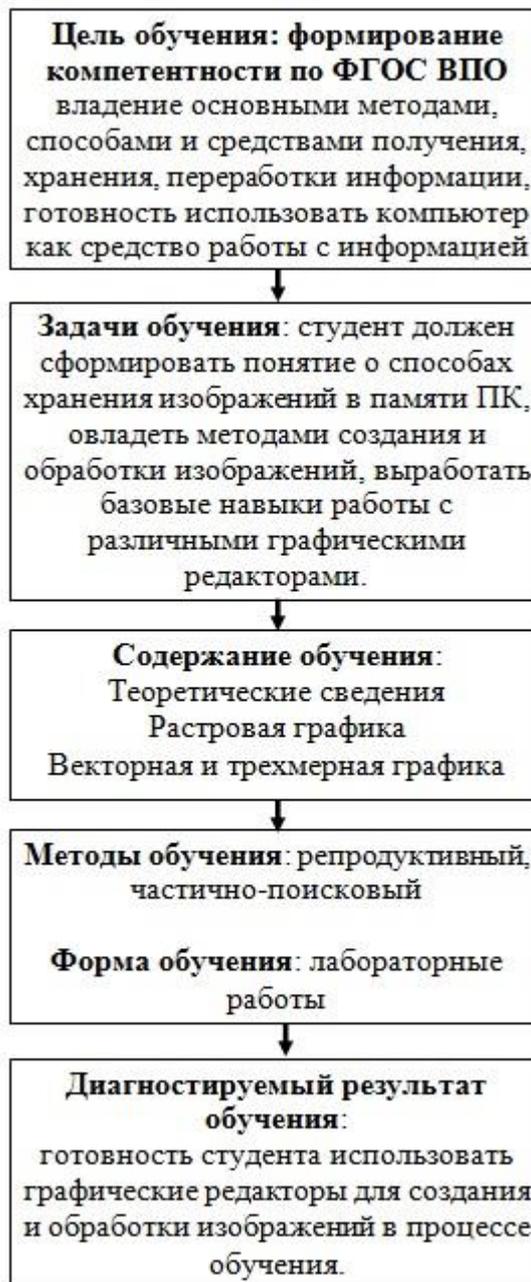


Рис. 2. Особенности методики обучения компьютерной графике в рамках курса «Информатика».

Итак, обучение компьютерной графике, безусловно, относится к одним из важнейших разделов обучения информатике в вузе, в том числе и на непрофильных

специальностях и направлениях. В условиях ограниченного количества часов, выделяемого на этот раздел, особое внимание следует уделить подбору содержимого и изучаемым в рамках темы графическим редакторам. Для достижения наилучшего результата следует переходить от простых программных средств к более сложным, обладающим большими функциональными возможностями. Рекомендуется плавный переход от репродуктивных заданий на воспроизведение к частично-поисковым заданиям, с подбором программных средств для их выполнения.

1. Литература

- [1] Коджаспирова Г. М., Петров К. В. Технические средства обучения и методика их использования: учеб. пособие для студентов высш. учеб. заведений. 2-е изд., перераб. и доп. М. : Изд. центр «Академия», 2005.
- [2] Сакулина, Ю. В. Компьютерная графика как средство формирования профессиональных компетенций / Ю. В. Сакулина, И. В. Рожина // Педагогическое образование в России. – 2012. – № 6. – С. 76 – 80.
- [3] Хеннер Е.К. Курс информатики в классическом университете / Е.К. Хеннер, Н.Н. Василук // Педагогическая информатика. – 2013. – №2. – с.3-15.
- [4] Чернякова Т.В. Методика обучения компьютерной графике студентов вуза: автореферат дис. ... канд. пед. наук / Чернякова Т. В. – Екатеринбург, 2010.

Об авторах

Василук Надежда Николаевна, к. пед. н., доцент кафедры информационных технологий механико-математического факультета Пермского государственного национального исследовательского университета. Ее e-mail nadia-vasiluk@yandex.ru.

Обобщенный конструктивный алгоритм преобразования инверсии относительно окружности

Д.В. Волошинов, Е.С. Казначеева, Е.С. Хайбаракманова
denis.voloshinov@yandex.ru

Санкт-Петербургский государственный университет телекоммуникаций, Санкт-Петербург, Россия

Статья посвящена разработке унифицированного конструктивного геометрического алгоритма преобразования инверсии, позволяющего на единой методологической основе осуществлять преобразование как действительных, так и мнимых точек плоскости относительно действительных и мнимых окружностей.

Ключевые слова: инверсия, конструктивная геометрия, геометрическое моделирование, Симплекс.

A generic constructive algorithm for inversion with respect to the circumference

D.V. Voloshinov, E.S. Kaznacheeva, E.S. Khaybarakhmanova
denis.voloshinov@yandex.ru

Saint-Petersburg state University of Telecommunications, Saint-Petersburg, Russia

The article is devoted to the development of a unified design geometric transformation algorithm of the inversion, allowing for a unified methodological basis to carry out the conversion of both real and imaginary points of the plane relative to the real and imaginary circles.

Keywords: inversion, constructive geometry, geometric modeling, Simplex.

1. Введение

Инверсия относится к достаточно хорошо изученным геометрическим преобразованиям [1]. Она находит применение при решении многих геометрических задач, в частности, задач на сопряжение прямых и окружностей. Инверсия является одним из центральных «геометрических стержней», связывающих через проективные преобразования объекты, на которых строится решение задачи Аполлония. Однако в силу ограниченности возможностей традиционных геометрических инструментов [2-6], область применения преобразования инверсии ограничивается лишь задачами, относящимися к действительным (вещественным) объектам плоскости. Однако с чисто математической точки зрения преобразование инверсии может быть легко перенесено и на мнимые образы плоскости, причем и сама окружность, относительно которой осуществляется преобразование инверсии, также может быть мнимой, в том числе иметь мнимые центр и радиус. Возможность реализации компьютерными средствами специализированных геометрических инструментов позволяет поставить вопрос о разработке единого конструктивного геометрического алгоритма, одинаково пригодного для реализации преобразования как в действительных, так и в мнимых образах проективной плоскости.

2. Общий алгоритм инверсии

Инверсией I точки A относительно окружности d называется преобразование, переводящее некоторую точку плоскости A , называемую прообразом, в точку-образ A' таким образом, что A и A' находятся на одной прямой линии с центром C окружности f , а произведение расстояний от центра C до точек A и A' равняется квадрату радиуса R_f окружности f : $|CA||CA'| = R_f^2$. Для построения образа точки A при

помощи чертежных инструментов используется следующий алгоритм:

Укажем точку A внутри круга, ограниченного окружностью f (рис. 1);

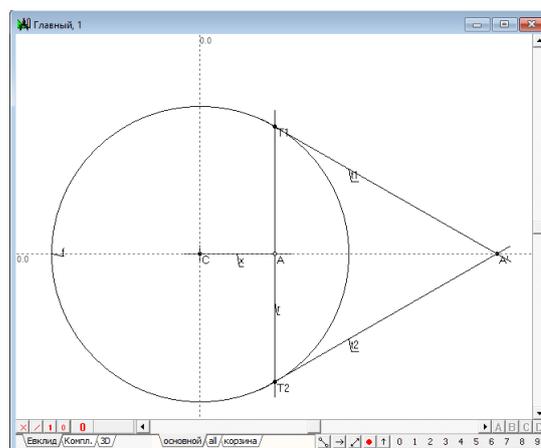


Рис. 1. Первая часть конструктивного алгоритма преобразования точки в инверсии.

1. Найдем центр окружности f – точку C ;
2. Соединим точку C с точкой A прямой линией x ;
3. Проведем через точку A прямую линию r , перпендикулярную x ;
4. Найдем точки T_1 и T_2 пересечения прямой r с окружностью f : $T_1, T_2 = r \times f$.
5. Проведем через точки T_1 и T_2 прямые t_1 и t_2 , касательные к окружности f ;
6. Найдем точку A' в пересечении прямых t_1 и t_2 : $A' = t_1 \times t_2$. Понятно, что полученная точка A' инцидентна с прямой x .

Пусть теперь точка A находится вне круга, ограниченного окружностью f (рис. 2). Выполним следующий алгоритм, устанавливающий связь между точками A и A' :

1. Найдем центр окружности f – точку C .
2. Соединим точку C с точкой A прямой линией x .
3. Опустим из точки A касательные прямые t_1 и t_2 на окружность f .
4. Обозначим точки касания прямых t_1 и t_2 с окружностью f через T_1 и T_2 .
5. Проведем через точки T_1 и T_2 прямую линию r .
6. Найдем точку A' пересечения прямых x и r : $A' = x \times r$.

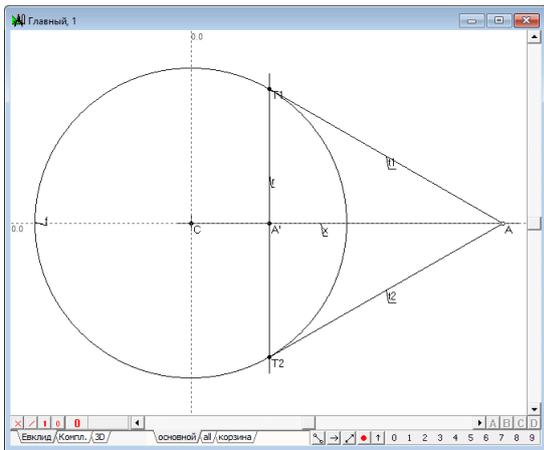


Рис. 2. Вторая часть конструктивного алгоритма преобразования точки в инверсии.

Из построений видно, что алгоритм несимметричен и его исполнение зависит от того, какое геометрическое место по отношению к окружности f занимает точка A .

Покажем теперь, что преобразование инверсии обладает следующим, отмеченным ранее, свойством: произведение расстояний от центра окружности инверсий до точек A и A' равно квадрату радиуса R_f окружности f (рис. 3).

Введем следующие обозначения: радиус окружности f равен R_f , расстояние от точки C до точки A есть r_1 , расстояние от точки C до точки A' есть r_2 . Обозначим расстояние между точками T_1 и A через $c = |T_1A|$, расстояние между точками A и A' через $a = |AA'|$, расстояние между точками T_1 и A' через $b = |T_1A'|$. Запишем равенства, следующие из чертежа: $r_1 + a = r_2$, $r_1^2 + c^2 = R_f^2$, $b^2 = c^2 + a^2$, $R_f^2 + b^2 = (r_1 + a)^2$. Вычтем из второго равенства третье. Получим: $r_1^2 - a^2 = R_f^2 - b^2$. Сложим полученный результат с четвертым равенством. В результате имеем: $2R_f^2 = r_1^2 + 2r_1a + a^2 + r_1^2 - a^2 = 2r_1(r_1 + a)$. Поскольку $r_1 + a = r_2$, то $2R_f^2 = 2r_1 \cdot r_2$, что и требовалось показать.

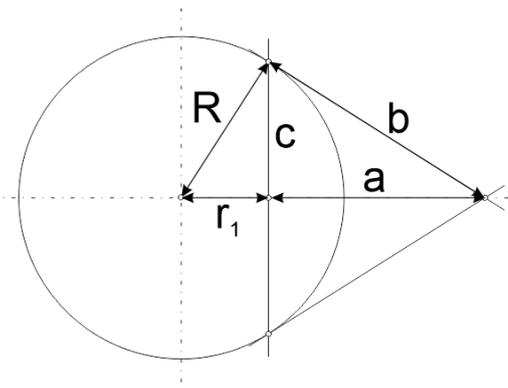


Рис. 3. К доказательству основного свойства инверсии.

В работах [2-5] показано, каким образом можно реализовать компьютеризированные геометрические инструменты, реализующие построения точек пересечения прямых линий и окружностей во всевозможных сочетаниях. Там же показано, что рассмотренные геометрические операции обладают общностью, заключающейся в том, что эти операции можно применять как к действительным, так и мнимым прямым и окружностям плоскости. В связи с этим возникает задача о распространении полученных в [2-5] результатов на преобразование инверсии, в которой можно было бы переводить как действительные, так и мнимые точки плоскости по отношению к окружностям, также имеющим действительную или мнимую геометрическую природу. Для решения этой задачи следует разработать конструктивный геометрический алгоритм преобразования инверсии, построенный исключительно на операциях пересечения прямых линий и окружностей (во всевозможных комбинациях), причем желательно, чтобы этот алгоритм был инвариантен по отношению к положению точки A в отношении окружности инверсии f .

Рассмотрим следующую геометрическую схему (рис. 4):

Пусть на плоскости задана некоторая окружность f . Укажем также точку A произвольно (но не совпадающую с центром окружности f). Выполним преобразование точки A в точку A' в инверсии I_f с использованием следующего алгоритма:

Найдем центр окружности f – точку C .

Соединим точку C с точкой A прямой линией x .

Построим окружность v с центром в точке A , проходящую через центр C окружности инверсии f .

Найдем точки P и Q пересечения окружности v с окружностью инверсии f .

Проведем через точки P и Q прямую линию $m = PQ$. Прямая m является радикальной осью окружностей f и v .

Найдем точку пересечения U прямых линий m и x : $U = m \times x$.

Проведем окружность w с центром в точке U и проходящую через центр C окружности инверсии f .

Точка A' находится, как одна из точек пересечения окружности w с прямой x , не совпадающая с точкой C .

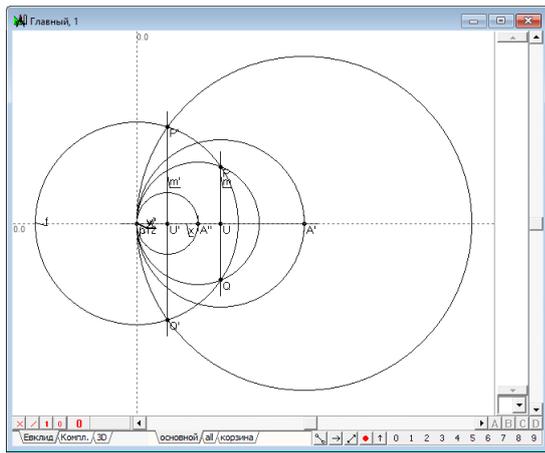


Рис. 4. Общий алгоритм инверсии.

Аналогичным образом алгоритм работает и в «обратном» направлении. Выберем теперь в качестве исходной точки точку A' и преобразуем ее в инверсии I_f в точку A'' относительно той же окружности f . Естественно предположить, что совпадение точек A и A'' будет обеспечено.

Проведем с центром в точке A' окружность v' , проходящую через центр C окружности инверсии f .

Найдем точки P' и Q' пересечения окружностей v' и f : $P', Q' = f \times v'$.

Соединим точки P' и Q' прямой $m' = P'Q'$.

Найдем точку U' пересечения прямых m' и x : $U' = m' \times x$.

Проведем с центром в точке U' окружность w' , проходящую через центр C окружности инверсии f .

Найдем две точки пересечения окружности w' с прямой x . Та из них, которая не совпадает с точкой C и является искомой точкой A'' – образом точки A' в преобразовании инверсии I_f относительно окружности f .

Докажем теперь, что данное геометрическое построение полностью соответствует преобразованию инверсии.

Обозначим расстояние от точки C до точки A через $r_1 = |CA|$, а расстояние от точки C до точки A' через $r_2 = |CA'|$. Пусть радиус окружности f равен R . Требуется доказать, что $r_1 \cdot r_2 = R^2$.

Обозначим расстояние между точками A и U через $a = |AU|$, а расстояние между точками P и U через $c = |PU|$. Тогда можно записать следующие равенства:

$$2(r_1 + a) = r_2, \quad a^2 + c^2 = r_1^2 \quad \text{и} \quad R^2 = (r_1 + a)^2 + c^2.$$

Поскольку $c^2 = r_1^2 - a^2$, то

$$R^2 = (r_1 + a)^2 + r_1^2 - a^2 = 2r_1(r_1 + a) = 2r_1 \frac{r_2}{2} = r_1 \cdot r_2, \quad \text{что и требовалось доказать.}$$

Мы видим, что $A' = I_f(A)$ и $A'' = I_f(A')$.

Преобразование инверсии инволюционно, то есть $I_f = I_f^{-1}$, поскольку оно переводит прообраз в образ и наоборот.

3. Сопряжения на мнимых образах

Рассмотрим типовую задачу на построение окружности, проходящей через заданную точку, касательную к двум заданным окружностям.

Пусть имеются две окружности a и b , а также точка p , через которую требуется провести окружности u и v , касательные к a и b (рис. 5). Пусть также известно, что a и b пересекаются явным образом, то есть образуют в пересечении действительные точки P и Q . Построим, например, окружность d_1 с центром в точке P . Преобразовав в инверсии относительно d_1 окружности a и b в прямые линии l и m , а точку p в точку p_2 , построим сопряжение l и m с точкой p_2 окружностями d_2 и d_3 . Такое сопряжение инструментально выполняется проще, чем исходное.

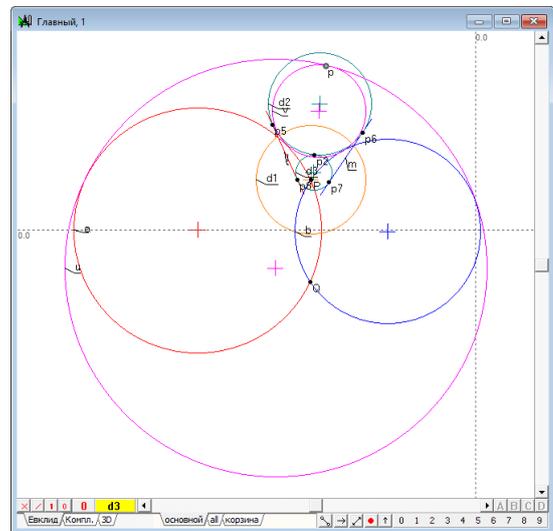


Рис. 5. Построение окружностей, касательных к двум заданным, проходящих через точку.

Остается полученные окружности d_2 и d_3 преобразовать в инверсии еще раз (поскольку инверсия инволюционна), чтобы получить окончательное решение – искомые окружности u и v .

Нетрудно заметить, однако, что данное построение выполнимо лишь в том случае, если окружности a и b пересекаются в действительных точках P и Q . Если же окружности a и b располагаются таким образом, что точки P и Q получают мнимые значения, то это будет означать, что центр окружности инверсии d_1 – точка P будет мнимым при, вообще говоря, вещественном радиусе. Заметим также, что невыполнимость обычными средствами операции инверсии не означает, что построение сопряжения окружностей a и b при заданных условиях невозможно (рис. 6).

Это, в свою очередь, означает, что «потерянная» конструктивная взаимосвязь объектов кроется в мнимости объектов, ее составляющих.

Исследование картины преобразования инверсии при решении задачи сопряжения с использованием мнимой окружности выполнено в среде системы конструктивного геометрического моделирования Симплекс [6], функциональный состав которой расширен действиями над мнимыми образами.

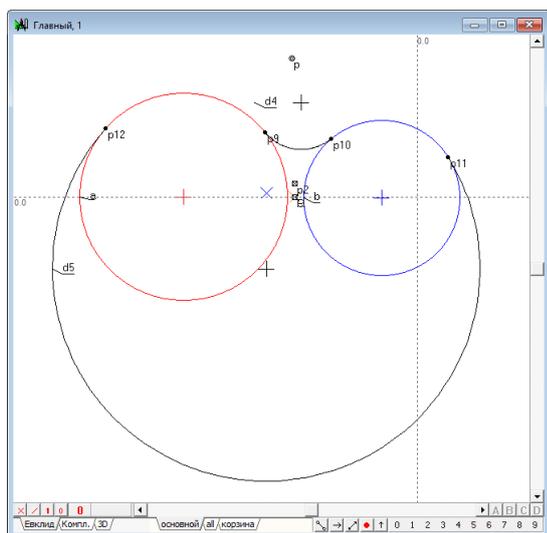


Рис. 6. Иллюстрация существования решения задачи на построение сопряжения при «вырождении» объектов операции инверсии.

На рис. 7, 8. Можно наблюдать картину сопряжения мнимых образов и их взаимосвязь с действительным решением при выполнении обобщенного алгоритма инверсии, рассмотренного ранее в данной статье.

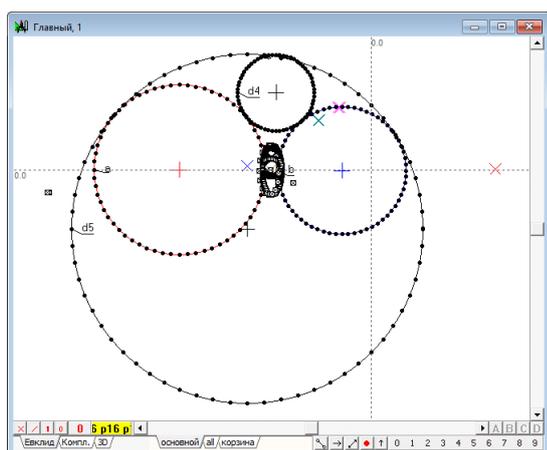


Рис. 7. Картина преобразования, иллюстрирующая образование мнимых точек на образах при отображении их действительных частей.

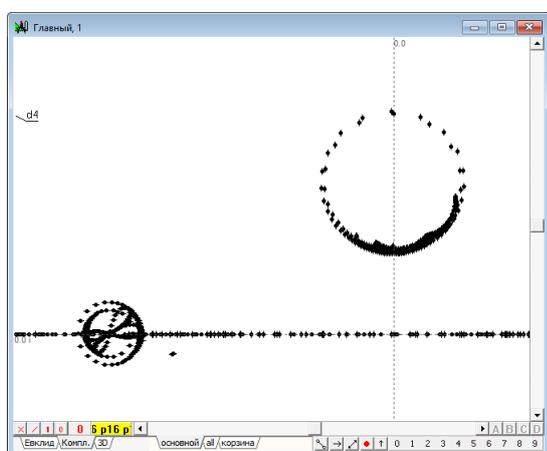


Рис. 8. Картина преобразования, иллюстрирующая образование мнимых точек на образах при отображении мнимыми точечными парами.

4. Заключение

Таким образом, в статье показана возможность реализации компьютеризированных геометрических инструментов, расширяющих возможности решения геометрических задач, не получавших ранее инструментальной поддержки из-за ограниченности возможностей средств предметного воплощения геометрических алгоритмов. Показана общность алгоритма преобразования точки в инверсии относительно окружности на множестве мнимых образов и возможность его применения для решения задач и синтеза геометрических форм.

5. Литература

- [1] Бакельман И.Я. Инверсия. М.: Наука. 1966, 79 с.
- [2] Волошинов Д.В. Геометрическая лаборатория. Закладываем основы // Качество графической подготовки: проблемы, традиции и инновации: Материалы VII международной Интернет-конференции. Февраль - март 2017 г. Пермь, 2017.
- [3] Волошинов Д.В. Геометрическая лаборатория. Инструменты ортогональности // Качество графической подготовки: проблемы, традиции и инновации: Материалы VII международной Интернет-конференции. Февраль - март 2017 г. Пермь, 2017.
- [4] Волошинов Д.В. Геометрическая лаборатория. Новый геометрический инструмент // Качество графической подготовки: проблемы, традиции и инновации: Материалы VII международной Интернет-конференции. Февраль - март 2017 г. Пермь, 2017.
- [5] Волошинов Д.В. Казначеева Е.С., Хайбахманова Е.С. Автоматизация проектирования поверхностей на основе конструктивных геометрических моделей // Качество графической подготовки: проблемы, традиции и инновации: Материалы VII международной Интернет-конференции. Февраль - март 2017 г. Пермь, 2017.
- [6] Волошинов Д.В. Конструктивное геометрическое моделирование. Теория, практика, автоматизация: монография / Д.В.Волошинов. – Saarbrücken: Lambert Academic Publishing, 2010. – 355 с.

Об авторах

Волошинов Денис Вячеславович, д.т.н., заведующий кафедрой информатики и компьютерного дизайна Санкт-Петербургского государственного университета телекоммуникаций им. М.А.Бонч-Бруевича. E-mail: denis.voloshinov@yandex.ru.

Казначеева Екатерина Сергеевна, магистрант кафедры информатики и компьютерного дизайна Санкт-Петербургского государственного университета телекоммуникаций им. М.А.Бонч-Бруевича.

Хайбахманова Екатерина Сергеевна, магистрант кафедры информатики и компьютерного дизайна Санкт-Петербургского государственного университета телекоммуникаций им. М.А.Бонч-Бруевича.

Синтез параметрической конструктивной модели 3D-объекта

Н.Д. Жилина, М.В. Лагунова, Т.В. Мошкова, С.А. Роменский, С.И. Ротков, М.М. Смычѣк, В.А. Тюрина
zhilina@nngasu.ru | mvlmn@mail.ru | ng.forever.mtv@gmail.com | romensky.serge@gmail.com | rotkovs@mail.ru | mariasmychek@gmail.com | 55555_73@mail.ru

Нижегородский государственный архитектурно-строительный университет, Нижний Новгород, Россия

В статье рассмотрен процесс синтеза конструктивной модели 3D-объекта на основе теории параметризации. Описано создание банка производных фигур (геометрических примитивов) с учётом формообразования поверхностей, входящих в их состав. Банк фигур предназначен для использования на этапах декомпозиции и последующей сборки с помощью теоретико-множественных операций при формировании конструктивной твердотельной модели.

Ключевые слова: электронная геометрическая модель изделия, параметризация, параметры формы, параметры положения, производная фигура, составная фигура, бинарное дерево (граф сборки).

Synthesis of a parametric constructive solid geometry model of a 3D-object

N.D. Zhilina, M.V. Lagunova, T.V. Moshkova, S.A. Romensky, S.I. Rotkov, M.M. Smychek, V.A. Tyurina
zhilina@nngasu.ru | mvlmn@mail.ru | ng.forever.mtv@gmail.com | romensky.serge@gmail.com | rotkovs@mail.ru | mariasmychek@gmail.com | 55555_73@mail.ru

Nizhny Novgorod State University of Architecture and Civil Engineering, Nizhny Novgorod, Russia

In this article considered process of synthesis of constructive solid model of 3D object based upon parametrization theory. Also it describes developing storage of non-derivative figures (geometry primitives) with considering their surface's law of formation. Storage of non-derivative figures designed for using on stages of decomposition and following construction with boolean operations of creation of constructive solid model.

Keywords: electronic geometric model of the product, parametrization, form parameters, position parameters, non-derivative figure, composite figure, binary tree (assembly graph).

1. Введение

Процесс проектирования объекта в современных условиях невозможно осуществить без использования компьютерных программ, ускоряющих и повышающих эффективность всего процесса. Общей тенденцией развития систем трехмерного геометрического моделирования является расширение возможностей параметрического подхода при создании электронных моделей объектов. Отличие параметрической модели от обычной состоит в том, что в ней предусмотрены топологические и геометрические взаимосвязи и ограничения между элементами, составляющими модель.

Параметрический режим описания объекта имеет особенности в различных системах геометрического моделирования, определяемые функциональной полнотой той или иной системы. Проведенный анализ показывает, что ограничения систем часто связаны с трудностями параметрического задания сложных поверхностей и недостаточным ассортиментом операций, позволяющих задавать поверхности в соответствии с их аналитическим описанием. Для преодоления подобных противоречий пользователям систем приходится прибегать к различным приемам и способам упрощения, аппроксимации или замены необходимых поверхностей, что значительно снижает качество и эффективность работы при создании электронных моделей.

Исследование законов формообразования поверхностей на основе классической теории параметризации, формализация этих законов и применение их при 3D-моделировании изделий в CAD-системах,

представляет научный и практический интерес в целях повышения эффективности процесса создания электронных конструктивных моделей трехмерных объектов.

2. Конструктивная модель 3D-объекта

В соответствии с ГОСТ 2.052-2015 Электронная геометрическая модель изделия (рис.1) представляет собой совокупность геометрических элементов и атрибутов модели, которые совместно определяют геометрию изделия и его свойства, зависящие от формы и размеров [1].

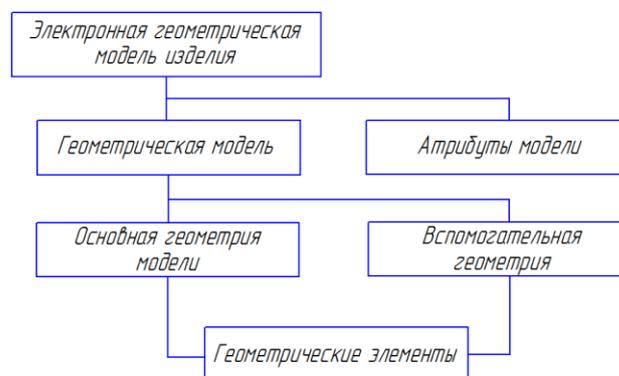


Рис. 1. Схематический состав электронной геометрической модели изделия

Конструирование геометрической модели является сложной и многовариантной задачей, связанной с

различными способами декомпозиции изделия на конструктивно-технологические примитивы и выбором способов получения той или иной поверхности.

Назначение конструктивных параметров геометрических элементов, определяющих основную и вспомогательную геометрию модели, может быть не связано с технологическими особенностями и часто носит субъективный характер. В этом случае возможно возникновение избыточной или, напротив, неполной геометрической информации о модели.

Теоретическим основам метода параметризации посвящены труды В.С. Полозова, И.И. Котова и других ученых, в которых сформулирована и описана теория параметризации [2], [3].

Параметром будем считать независимую величину, позволяющую из множества выделить элемент, либо подмножество элементов. Под параметризацией будем понимать процесс выбора и подсчета количества параметров, позволяющих однозначно выделить единственную геометрическую фигуру в пространстве. Параметрическое число фигуры [2, 4] определяет минимально необходимое, но достаточное, количество размеров на поле чертежа для однозначной интерпретации геометрии объекта как в двумерном, так и в трехмерном пространстве измерений.

Твердотельная геометрическая модель (твердотельная модель) определяется ГОСТ 2-052-2015 как «трехмерная геометрическая модель, представляющая форму изделия как результат композиции множества геометрических элементов с применением операций булевой алгебры к этим геометрическим элементам» [1]. Конкретизируем это определение.

Твердотельной (конструктивной) моделью будем называть совокупность $CM = \{G(R), B(M), A\}$,

где $G(R)$ - граф сборки конструкции, либо бинарная древовидная структура, определяющие порядок и характер взаимодействия составляющих объектов,

R - ребра графа, атрибутами которого могут быть параметры формы и положения i -ой модели относительно j -ой модели и знаки регуляризованных теоретико-множественных операций объединения, вычитания и пересечения.

Параметрами формы могут быть числовые значения масштабных коэффициентов по координатным осям.

Параметрами положения могут быть вектор переноса T , вектор поворота E , задающий углы поворота i -го объекта на заданные углы относительно координатных осей.

$B(M)$ - банк граничных моделей пространственных объектов, каждый из которых может быть как производной, так и составной фигурой. Модели объектов, составляющих банк, могут быть сформированы различными средствами и способами и заданы в своей собственной системе координат, не связанной с координатной системой проектируемого объекта. Каждой модели поставлено в соответствие символическое имя, по которому идет поиск информации, и имя файла, в котором хранится вся информация, относящаяся к модели.

A - атрибуты моделей (материал, плотность, цвет и т.д.) [4].

Модель материального тела или конструктивная (constructive solid geometry, CSG) модель дает возможность определять качественные характеристики объекта, его структуру, топологические и геометрические связи между составляющими его элементами.

Имея набор граничных моделей и информацию о связях между ними, можно получать конструктивные модели сложных объектов.

Конструируемую модель 3D-объекта будем рассматривать как составную фигуру (СФ), состоящую из геометрических примитивов, имеющих определенные параметры и способ образования. Такие геометрические примитивы будем называть «непроизводными фигурами» (НФ) и считать их неделимыми в данном классе рассматриваемых задач.

Для практического использования данного подхода необходим банк непроизводных фигур, параметры которых могут задаваться интерактивно.

Поскольку поверхности, ограничивающие одну и ту же непроизводную фигуру можно образовать различными формообразующими операциями, то в основу определения минимального, но достаточного количества параметров необходимо положить анализ определителя создаваемой поверхности. Поверхность считается заданной, если относительно любой точки пространства можно однозначно решить вопрос о ее принадлежности данной поверхности.

Определителем поверхности называют совокупность условий, задающих поверхность. Определитель фигуры состоит из двух частей – геометрической (ГЧО) и алгоритмической (АЧО) [3]. В геометрическую часть определителя входят геометрические образы и параметры постоянной формы, положения и величины. Алгоритмическая часть определителя поверхности представляет собой алгоритм построения точек и линий поверхности, занимающих на ней переменное положение.

Анализ закона образования поверхности позволяет выделить ее определитель. А алгоритмическая часть определителя дает возможность составить формальную модель конструирования непрерывного каркаса поверхности.

3. Синтез конструктивной модели 3D-объекта

Рассмотрим процесс синтеза конструктивной модели трехмерного объекта - составной фигуры, состоящей из стандартных геометрических примитивов, выбранных из некоторого банка непроизводных фигур. Объект может быть задан своими проекционными изображениями, например, стандартными ортогональными видами, либо аксонометрией (оба варианта задания показаны на рис.2).

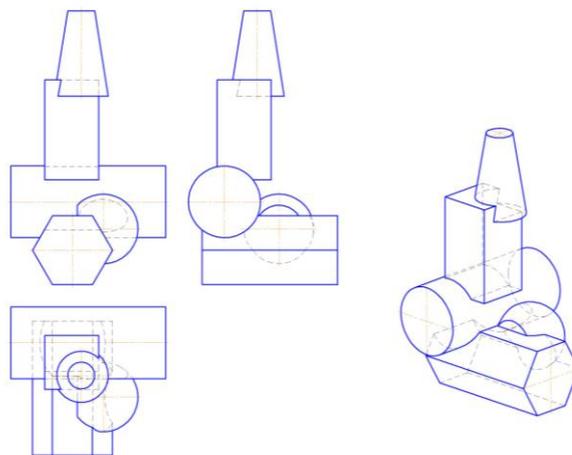


Рис. 2. Проекционные изображения составной фигуры

В процессе анализа изображений необходимо мысленно выделить из состава фигуры ряд производных фигур (рис.3), а также топологические и геометрические связи между ними.

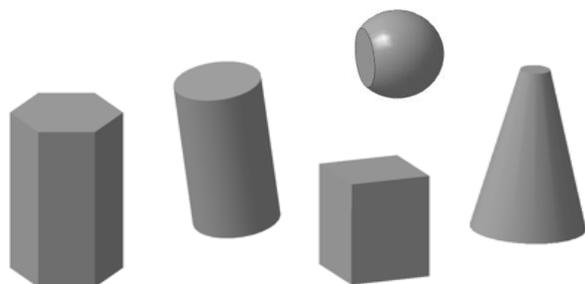


Рис. 3. Набор производных фигур, входящих в состав фигуры

Для уяснения сути процесса конструирования, а именно таких его этапов, как декомпозиция и последующая сборка составной фигуры, авторами статьи разработан и успешно применяется в различных аспектах научной и педагогической деятельности «Электронный конструктор графа сборки 3D-объекта» (рис.4) [5]. Конструктор предназначен для генерирования графа сборки составной фигуры (бинарного дерева) и наглядной демонстрации топологических связей между примитивами.

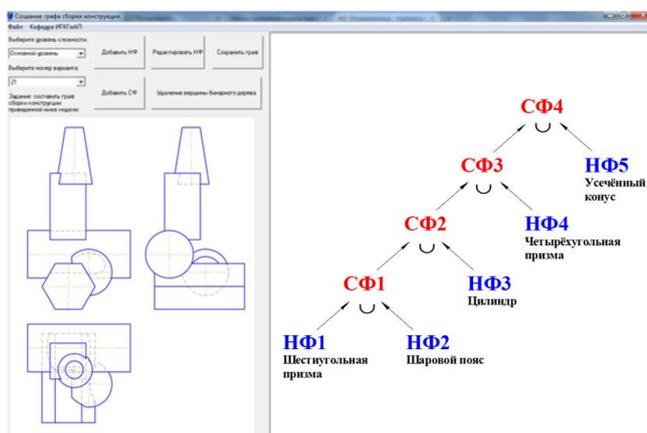


Рис. 4. Интерфейс электронного конструктора графа сборки 3D-объекта

Пользователь может выбирать производные фигуры (примитивы) из банка объектов, входящего в состав конструктора (рис.5).

При отсутствии необходимого примитива следует использовать функционал используемой системы геометрического моделирования для создания собственной параметрической производной фигуры (НФ).

После выбора производной фигуры, определяется способ образования ограничивающей ее поверхности. Например, примитив «конус» может иметь разные геометрическую и алгоритмическую части определителя и, следовательно, быть образован с использованием разных кинематических процедур (рис.6).

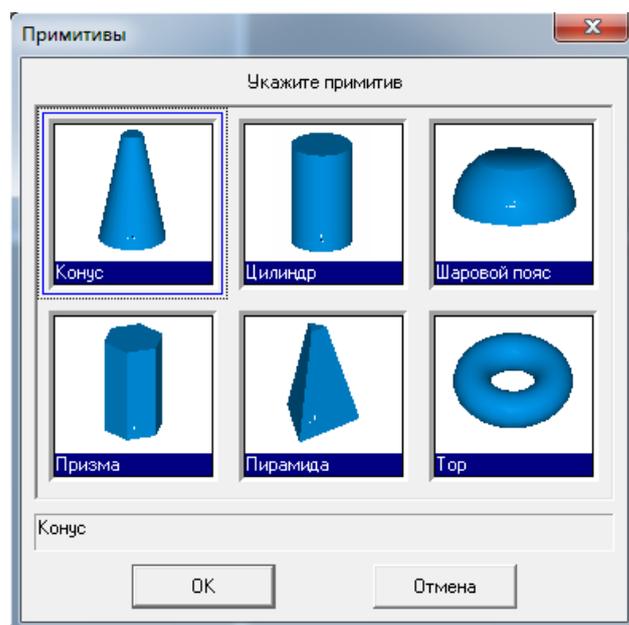


Рис. 5. Интерфейс диалогового окна для выбора прототипа производной фигуры

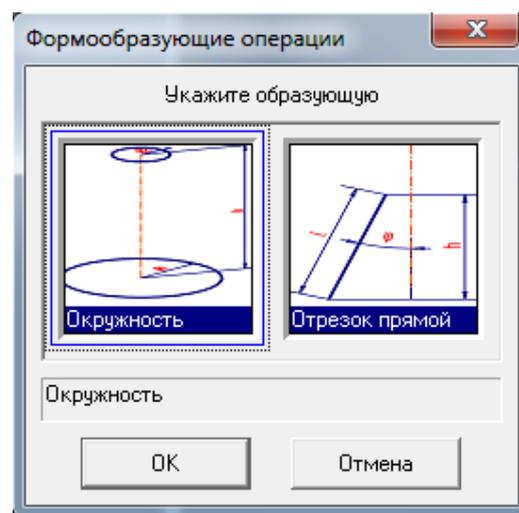


Рис. 6. Интерфейс диалоговых окон для выбора способа формообразования конической поверхности

Выбор вида образующей для кинематической формообразующей операции определяет набор параметров формы прототипа. Например, если выбран вид образующей «Окружность», то для задания элементов геометрической части определителя поверхности необходимо задать три параметра: радиусы образующих, определяющих основания конуса ($R1$, $R2$) и расстояние между ними (h). В этом случае используется формообразующая операция «По сечениям» (рис.7).

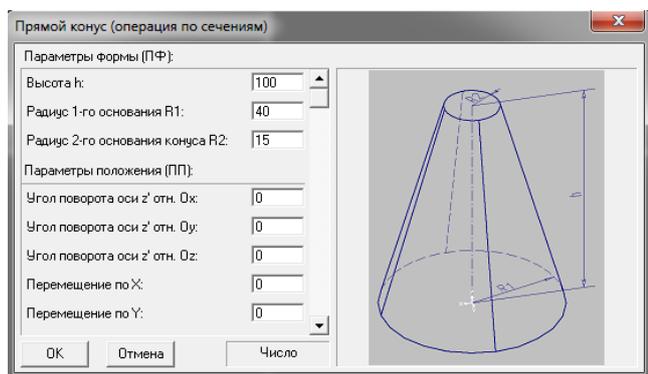


Рис. 7. Интерфейс диалогового окна для выбора параметров прямого конуса в случае вида образующей «Окружность» - операция «По сечениям»

Если выбран вид образующей «Отрезок прямой», то для задания элементов геометрической части определителя поверхности необходимо задать также три параметра, представленные разными наборами, например, длиной образующей (l), углом наклона этой образующей (φ) и высотой (h). В этом случае используется формообразующая операция «Вращение» (рис. 8).

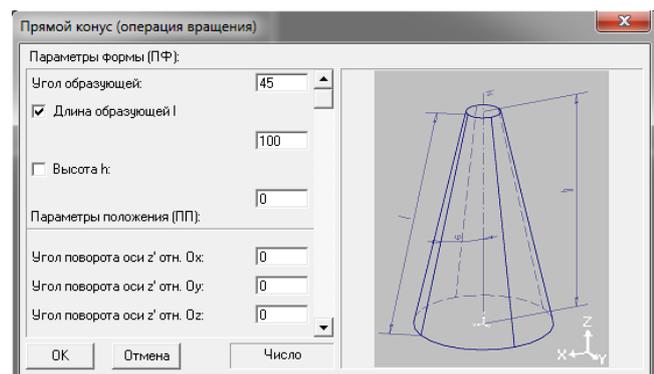


Рис. 8. Интерфейс диалогового окна для выбора параметров прямого конуса в случае вида образующей «Отрезок прямой» - операция «Вращение»

Параметры формы выделены во внутренней (локальной) системе параметризации (x,y,z), совмещенной с элементами геометрической части определителя поверхности. Например, в случае формообразующей операции «Вращение» ось z локальной системы координат совпадает с осью вращения конуса. Выбор параметров положения внутренней системы параметризации производной фигуры относительно внешней системы (XYZ) не зависит от параметров формы фигуры. Положение производной фигуры описывается в общем случае шестью параметрами положения внутренней системы относительно внешней системы параметризации.

Аналогично создаются другие производные фигуры, например, шаровой пояс, выбор способа формообразования которого показан на рис. 9, а процесс задания параметров формы и положения на рис. 10.

Примеры задания параметров формы и параметров положения внутренних (локальных) систем параметризации некоторых геометрических примитивов относительно внешней системы параметризации продемонстрированы на рис. 11.

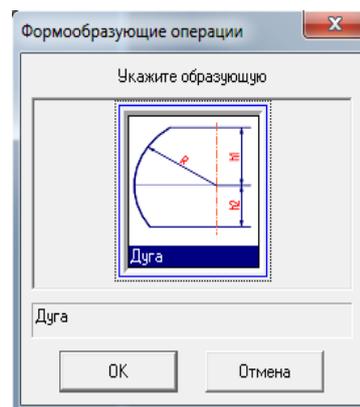


Рис. 9. Интерфейс диалогового окна для выбора способа формообразования сферической поверхности

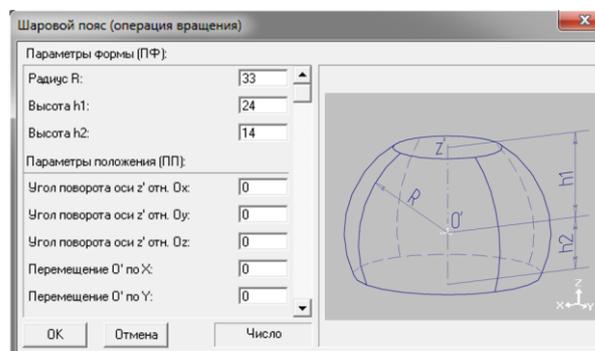


Рис. 10. Интерфейс диалогового окна для задания параметров формы и положения шарового пояса

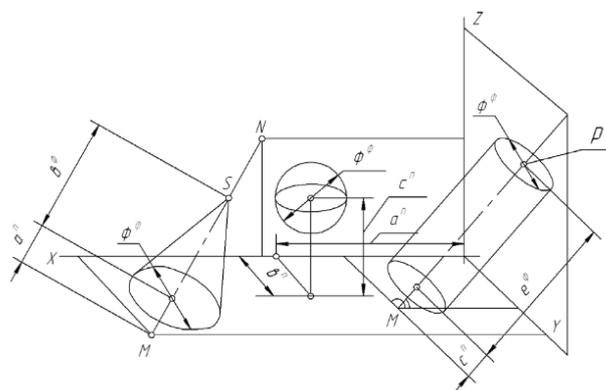


Рис. 11. Задание параметров производных фигур относительно внешней системы параметризации

Далее процесс сборки составной фигуры из геометрических примитивов (НФ) заключается в анализе связей между производными фигурами и применении соответствующих теоретико-множественных (булевых) операций [6].

Созданная таким образом электронная конструктивная модель составной фигуры позволяет получить различные изображения детали, например, многовидовый технический чертеж или аксонометрическое изображение. Такой подход позволяет соотнести и сопоставить геометрическую

информацию об объекте, представленную на разных геометрических моделях в наглядном и формализованном виде.

4. Заключение

Конструирование электронной геометрической модели изделия на параметрической основе способствует установлению точных геометрических и топологических связей в процессе создания 3D модели и создает условия для исследования процесса формообразования ограничивающих объект поверхностей за счет использования различных видов геометрических и алгоритмических частей определителей. Использование разработанных авторами приложений позволяет повысить эффективность процесса создания электронных конструктивных моделей трехмерных объектов с помощью стандартных систем геометрического моделирования.

5. Благодарности

Работа выполнена по грантам РФФИ №15-07-05110, №17-07-00543

6. Литература

- [1] ЕСКД. Единая система конструкторской документации. ГОСТ 2.052-2015 Электронная модель изделия. Общие положения / База нормативных документов [Электронный ресурс] – Режим доступа: <http://russgost.ru>.
- [2] Полозов, В.С. Начертательная геометрия (информационно-параметрический подход в инженерных графических задачах) : Учебное пособие. – Н.Новгород : Нижегород. гос. архит.-строит. ун-т, 2000, – 61 с.
- [3] Котов, И.И., Алгоритмы машинной графики / И.И.Котов, В.С. Полозов, Л.В. Широкова. – М., «Машиностроение», 1977, – 231 с.
- [4] Ротков С.И. Средства геометрического моделирования и компьютерной графики пространственных объектов для САЛС-технологий: Дис. докт. техн. наук: 05.01.01 / С.И.Ротков. – Н. Новгород, 1999. – 287 с.
- [5] Лагунова, М.В. Электронный конструктор графа сборки 3D-объекта / М.В. Лагунова, Т.В. Мошкова, С.А. Роменский, С.И. Ротков, М.М. Смычк, В.А. Тюрина // Междунар. науч. школа-семинар МГУ, Институт физико-технической информатики, – 21-24 нояб. 2016 г. – Москва-Протвино, 2016. – С. 11-13.
- [6] Жилина, Н.Д. Автоматизация процесса чтения чертежа с использованием системы Компас-3D / Н.Д.Жилина, М.В.Лагунова, Т.В. Мошкова, С.И. Ротков, В.А. Тюрина // ГрафиКон'2014 : 24-я междунар. конф. по компьютерной графике и зрению : Академия архитектуры и искусств ЮФУ. – Ростов-на-Дону, 2014. – С. 17-19.

Об авторах

Жилина Наталья Дмитриевна, к.п.н., доцент, профессор кафедры Инженерной геометрии, компьютерной графики и автоматизированного проектирования Нижегородского государственного архитектурно-строительного университета, zhilina@nngasu.ru

Лагунова Марина Викторовна, д.п.н., проф., профессор кафедры Инженерной геометрии, компьютерной графики и автоматизированного проектирования Нижегородского государственного архитектурно-строительного университета mvlmn@mail.ru

Мошкова Татьяна Владимировна, к.т.н., доцент, профессор кафедры Инженерной геометрии, компьютерной графики и автоматизированного проектирования Нижегородского государственного архитектурно-строительного университета, ng.forever.mtv@gmail.com

Роменский Сергей Александрович, аспирант кафедры Инженерной геометрии, компьютерной графики и автоматизированного проектирования Нижегородского государственного архитектурно-строительного университета, romensky.serge@gmail.com

Ротков Сергей Игоревич, д.т.н., проф., зав. кафедрой Инженерной геометрии, компьютерной графики и автоматизированного проектирования Нижегородского государственного архитектурно-строительного университета, rotkovs@mail.ru

Смычк Мария Михайловна, ассистент кафедры Инженерной геометрии, компьютерной графики и автоматизированного проектирования Нижегородского государственного архитектурно-строительного университета mariasmychek@gmail.com

Тюрина Валерия Александровна, к.т.н., доцент, профессор кафедры Инженерной геометрии, компьютерной графики и автоматизированного проектирования Нижегородского государственного архитектурно-строительного университета 55555_73@mail.ru

Визуальная компьютерная реконструкция квадратичных соответствий

В.А. Короткий
ospolina@mail.ru

Южно-Уральский государственный университет, кафедра инженерной и компьютерной графики
Челябинск, Россия

Рассматриваются две задачи реконструкции квадратичных соответствий, сформулированные еще в 19 веке, но так и не решенные до настоящего времени. Первая задача заключается в построении фундаментальных точек квадратичного преобразования плоскости, заданного семью парами соответственных точек. Вторая задача связана с поиском графического алгоритма реконструкции квадрики, заданной девятью точками. Интерес к этим задачам в настоящее время полностью утрачен, хотя именно сейчас, благодаря графическим средствам компьютерной визуализации, появилась возможность их точного конструктивного решения. Обе задачи, рассмотренные в статье, объединены как общим методом решения, так и общим инструментом, необходимым для конструктивной реализации метода. Для решения поставленных задач использован метод понижения размерности массива исходных данных. При поиске фундаментальных точек квадратичного преобразования, заданного 7 парами сопряженных точек, предлагается временно исключить из рассмотрения какую-либо пару данных сопряженных точек. При реконструкции квадрики, заданной 9 точками, также временно исключается из рассмотрения одна из данных точек, что позволяет построить пучок квадрик, инцидентных восьми оставшимся точкам. Для выделения из этого пучка искомой квадрики используется метод сечений. Для конструктивного решения обеих задач используется специальная компьютерная программа, вычерчивающая кривые второго порядка, без использования которой конструктивная реализация предложенных алгоритмов невозможна.

Ключевые слова: квадратичное бирациональное соответствие, преобразование Гирста, проективное преобразование, пучок конических сечений, пучок квадрик.

Visual computer reconstruction of quadratic mappings

V.A. Korotkij
ospolina@mail.ru

South Ural state University, Department of engineering and computer graphics
Chelyabinsk, Russia

We consider two tasks of reconstruction of quadratic correspondences, formulated in the 19th century, but has not been solved to date. The first task is to build fundamental points of a quadratic transformation of the plane defined by seven pairs of corresponding points. The second problem is connected with the search graphical algorithm of reconstruction of a quadric defined by the nine dots. Interest in these problems at present is completely lost, even though now, thanks to the graphical tools of computer visualization, there is a possibility of their accurate design solutions. Both tasks, considered in article, is united both by common solutions and common tool needed for the constructive implementation of the method. For the decision of tasks used the method of decreasing the size of the array the source data. When you search the fundamental points of a quadratic transformation given 7 pairs of conjugate points, is proposed to temporarily exclude from consideration any pair of conjugate points of data. During the reconstruction of quadric given 9 points, also temporarily eliminated from consideration one of the data points that allows you to build a bundle of quadrics incident eight remaining points. To highlight this beam the desired quadric method is used, sections. A constructive solution of both problems we use a special computer program for drawing conic curves, without the use of which a constructive implementation of the proposed algorithms is impossible.

Keywords: quadratic birational compliance, conversion Hirst, projective transformation, a beam of tapered cross-section, a bundle of quadrics.

1. Введение

Развитие средств компьютерной графики приводит не только к появлению принципиально новых методов геометрического моделирования, но и предоставляет возможность конструктивного решения классических задач теоретической и прикладной геометрии.

В статье рассматриваются две задачи, оставшиеся нам в наследство от великих геометров девятнадцатого века. Парадоксальным образом, интерес к этим задачам в настоящее время полностью утрачен. Парадокс состоит в том, что именно сейчас, благодаря средствам компьютерной визуализации, появилась возможность их точного конструктивного решения.

Первая из рассматриваемых задач заключается в построении фундаментальных точек квадратичного преобразования (F -точек), заданного семью парами сопряженных точек. Как известно (Стемона), сопряженностью семи пар точек устанавливается

соответствие любой точки плоскости Π определенной точке другой плоскости Π' , а следовательно – точек какого-либо прямолинейного ряда поля Π с точками некоторого конического сечения (гомалоида) поля Π' . Задача реконструкции (построения F -точек) может быть решена только на основе точной компьютерной визуализации. Более простая задача построения восьмой пары сопряженных точек рассматривалась К.А. Андреевым [1].

Вторая задача связана с поисками аналогии между свойствами поверхностей второго порядка и знаменитой теоремой Паскаля (М. Шаль). Безуспешные поиски такой аналогии приводят к задаче визуальной графической реконструкции поверхности второго порядка (ПВП) по девяти ее точкам.

Очевидно, задачи, связанные с исследованием квадратичных соответствий, могут быть конструктивно решены лишь при выполнении одного обязательного условия. Это условие состоит в том, что кривые второго порядка (КВП) должны быть такими же простыми и точно вычерчиваемыми линиями, как прямая и окружность.

Выполнение этого условия обеспечивается с помощью программного средства “Универсальный компьютерный коникограф” [2].

2. Визуализация F-точек

Квадратичное бирациональное соответствие Ω плоских точечных полей Π, Π' однозначно определено семью парами соответственных точек $A \sim A', B \sim B', C \sim C', D \sim D', E \sim E', R \sim R', T \sim T'$. Требуется найти F-точки соответствия Ω .

Выполним проективное преобразование Λ плоскости Π , переводящее четырехугольник $ABCD$ в четырехугольник, конгруэнтный четырехугольнику $A'B'C'D'$ поля Π' . Совмещая поля Π, Π' до совпадения конгруэнтных фигур $ABCD$ и $A'B'C'D'$, получаем задачу – на совмещенной плоскости $\Pi = \Pi'$ найти F-точки преобразования Ω , заданного семью парами сопряженных точек, из которых четыре пары точек совпадают.

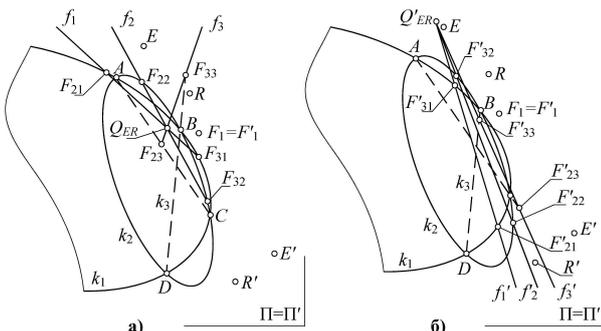


Рис. 1. Пучки хорд преобразования ω_{ER}

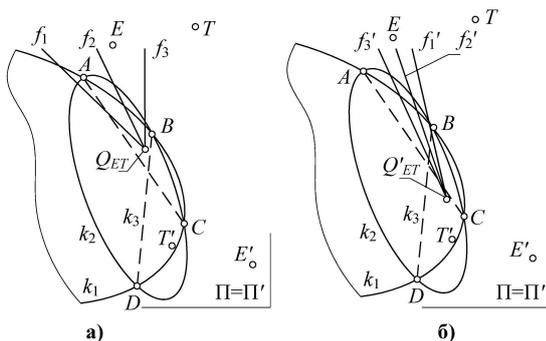


Рис. 2. Пучки хорд преобразования ω_{ET}

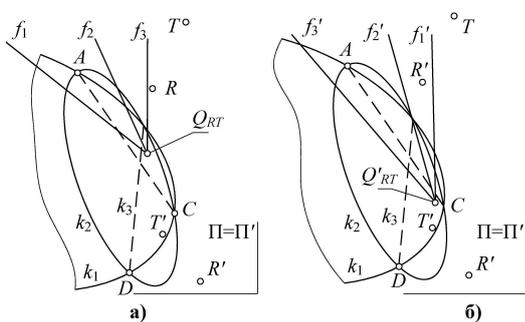


Рис. 3. Пучки хорд преобразования ω_{RT}

Действие 1. Исключаем из рассмотрения какую-либо пару не совпадающих сопряженных точек, например, $T \sim T'$. Рассмотрим однопараметрическое семейство квадратичных преобразований ω_{ER} , заданное четырьмя парами совпадающих сопряженных точек $A \sim A', B \sim B', C \sim C', D \sim D'$ и двумя парами не совпадающих сопряженных точек $E \sim E', R \sim R'$. Выделяя три произвольных конических сечения $k_1, k_2, k_3 = AC \cap BD$ пучка $\psi(ABCD)$, получаем три центральных квадратичных преобразования $(\omega_{ER})_1, (\omega_{ER})_2, (\omega_{ER})_3$ с общим

центром $F_1 = F'_1$ и инвариантными кониками k_1, k_2, k_3 (рис. 1). На каждой конике находим две пары ассоциированных точек $F_2 \sim F'_2, F_3 \sim F'_3$, из которых все сопряженные точки (за исключением точек $T \sim T'$) проецируются двумя парами проективных пучков. Каждой конике k_i ставится в соответствие хорда $f_i = F_{2i} \sim F_{3i}$ поля Π и хорда $f'_i = F'_{2i} \sim F'_{3i}$ поля Π' . Хорды f_i образуют пучок с вершиной Q_{ER} (рис. 1, а). Хорды f'_i образуют пучок с вершиной Q'_{ER} (рис. 1, б).

Лемма 1. Хорды f_i и f'_i образуют пучки прямых, проективные пучку конических сечений $\psi(ABCD)$. Доказательство леммы основано на свойствах проективных соответствий в образах второго порядка.

Действие 2. Повторяем действие 1, исключая из рассмотрения другую пару не совпадающих сопряженных точек $R \sim R'$. Получаем семейство квадратичных преобразований ω_{ET} , заданное шестью парами точек $A \sim A', B \sim B', C \sim C', D \sim D', E \sim E', T \sim T'$. Выделяя из ψ те же самые коники $k_1, k_2, k_3 = AC \cap BD$, находим на каждой из них две пары ассоциированных точек $F_2 \sim F'_2, F_3 \sim F'_3$, из которых все сопряженные точки (кроме $R \sim R'$) проецируются двумя парами соответственных пучков. Каждой конике $k_i (i=1,2,3)$ ставятся в соответствие хорды $f_i = F_{2i} \sim F_{3i}$ и $f'_i = F'_{2i} \sim F'_{3i}$ (рис. 2). Согласно лемме 1, получаем пучки хорд $Q_{ET}(f_1, f_2, f_3, \dots)$ и $Q'_{ET}(f'_1, f'_2, f'_3, \dots)$, проективные пучку коник ψ .

Действие 3. Повторяем действие 2, исключая из рассмотрения пару сопряженных точек $E \sim E'$. Рассматриваем семейство квадратичных преобразований ω_{RT} , заданное шестью парами точек $A \sim A', B \sim B', C \sim C', D \sim D', R \sim R', T \sim T'$. В каждом из преобразований $(\omega_{RT})_i$ инвариантной конике k_i ставятся в соответствие хорды $f_i = F_{2i} \sim F_{3i}$ и $f'_i = F'_{2i} \sim F'_{3i}$. Получаем пучки хорд $Q_{RT}(f_1, f_2, f_3, \dots)$ и $Q'_{RT}(f'_1, f'_2, f'_3, \dots)$, проективные пучку коник ψ (рис. 3).

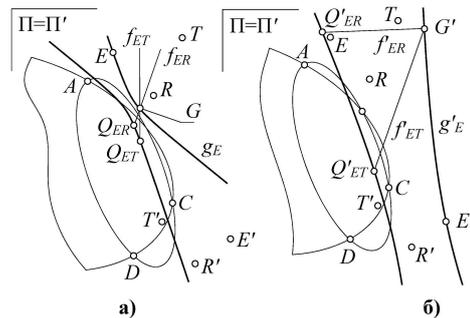


Рис. 4. Вспомогательные коники g_E и g'_E

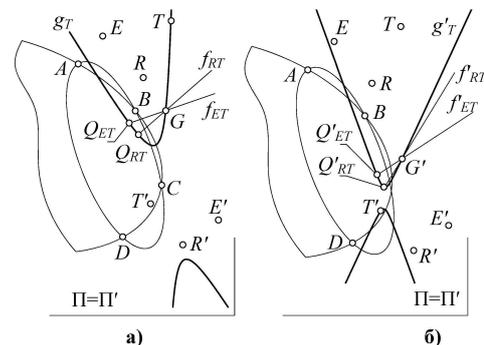


Рис. 5. Вспомогательные коники g_T и g'_T

Действие 4. Совместно рассмотрим преобразования ω_{ER} и ω_{ET} . Согласно лемме 1, пучки хорд, порожденные этой парой преобразований, проективные пучку коник ψ , следовательно, проективны между собой. Поэтому точки G пересечения проективно соответственных хорд f_{ER} и f_{ET} определяют конику g_E , проходящую через E и через вершины Q_{ER}, Q_{ET} (рис. 4, а). Аналогичным образом, точки $G' = f'_{ER} \cap f'_{ET}$ определяют конику g'_E , проходящую через E' и через вершины Q'_{ER}, Q'_{ET} (рис. 4, б).

Действие 5. Рассматривая совместно другую пару преобразований ω_{RT} и ω_{ET} , получаем проективные пучки хорд (Q_{RT}) и (Q_{ET}) . Точки пересечения проективно соответственных прямых этих пучков определяют коническое сечение g_T , проходящее через T и через вершины Q_{RT}, Q_{ET} (рис. 5, а). Точки пересечения соответственных прямых в проективных пучках (Q_{RT}') , (Q_{ET}') определяют конику g_T' , проходящую через T' и через вершины Q_{RT}', Q_{ET}' (рис. 5, б).

Лемма 2. В преобразовании Ω фундаментальные точки F_1, F_2, F_3 поля Π совпадают с точками пересечения вспомогательных конических сечений g_E и g_T , а точки F_1', F_2', F_3' поля Π совпадают с точками пересечения коник g_E' и g_T' .

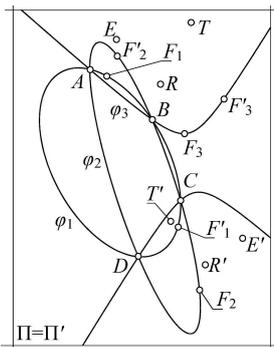


Рис. 6. Коники $\phi_1 \dots \phi_3$

Доказательство леммы 2 следует из того, что соответственные прямые в пучках F_i, F_i' определяют конику ϕ_i , инцидентную базисным точкам пучка ψ (рис. 6).

Из леммы 2 следует, что для визуализации F -точек преобразования Ω на совмещенной плоскости $\Pi=\Pi'$ достаточно отметить точки пересечения конических сечений $g_E \cap g_T$ и $g_E' \cap g_T'$ (рис. 7).

Чтобы вернуть найденные F -точки в исходное поле Π , следует выполнить обратное проективное преобразование Λ^{-1} .

Исследование. Вспомогательные коники g_E, g_T пересекаются в четырех точках (см. рис. 7, а). Чтобы выделить из них искомые точки F_1, F_2, F_3 , заметим, что эти коники проходят через вершину пучка (Q_{ET}) . Следовательно, точка Q_{ET} – одна из четырех точек пересечения коник g_E, g_T . Оставшиеся три точки указывают искомые F -точки преобразования Ω . Вспомогательные КВП g_E', g_T' также пересекаются в четырех точках, одна из которых совпадает с Q_{ET}' . Оставшиеся три точки указывают F' -точки преобразования Ω (см. рис. 7, б).

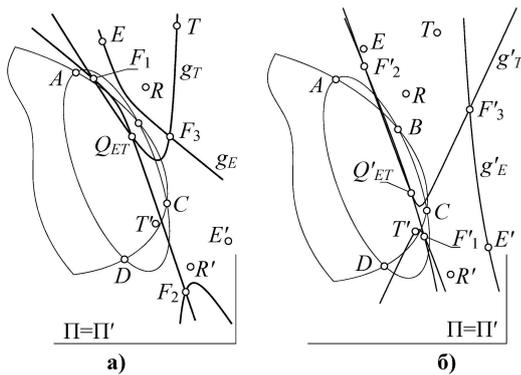


Рис. 7. Визуализация F -точек преобразования Ω

Заметим, что ассоциированные пары F -точек располагаются на одной и той же конике пучка ψ (см. рис. 6). Поэтому для выделения точки F_1' , ассоциированной с F_1 , следует начертить коническое сечение ϕ_1 , проходящее через F_1 и через базис пучка ψ . Коника ϕ_1 пройдет через искомую точку F_1' . Точно так же определяются две другие пары ассоциированных F -точек.

Две точки пересечения вспомогательных КВП могут быть мнимыми. В этом случае получаем квадратичное преобразование Ω с двумя парами мнимых сопряженных F -точек, которое может быть приведено к преобразованию Гирста с центром внутри инвариантной коники. В свою

очередь, преобразование Гирста с внутренней центральной точкой может быть сведено к инверсии относительно окружности.

Эксперимент. Для проверки решения найдем вспомогательную конику g_R , образованную проективными пучками (Q_{RT}) и (Q_{ER}) в преобразованиях ω_{RT} и ω_{ER} . Коника g_R проходит через R , через Q_{RT}, Q_{ER} и через ранее найденные F -точки преобразования Ω , что подтверждает достоверность разработанного алгоритма (рис. 8).

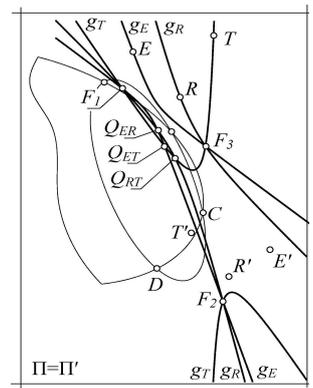


Рис. 8. Эксперимент

3. Реконструкция ПВП по 9 точкам

Девять произвольно указанных в пространстве точек A_j, B_j, C_j ($j=1,2,3$), не инцидентных по три одной прямой и не инцидентных по четыре одной плоскости, определяют единственную квадрику ϕ . Требуется составить алгоритм построения любого количества точек, инцидентных ϕ .

Схема решения

1. Выделяем три плоскости $\alpha(A_1, A_2, A_3), \beta(B_1, B_2, B_3), \gamma(C_1, C_2, C_3)$ и отмечаем линию m пересечения любой пары плоскостей, например, $m=\alpha\cap\beta$.
2. Как и в предыдущей задаче, одну из заданных точек (например, точку C_3) временно исключаем из рассмотрения.
3. Фиксируем на m произвольную точку M . Точка N_i , бегущая по m , индуцирует пучок квадрик $\Phi(\phi_i)$. Произвольная квадрика ϕ_i пучка Φ определена точкой C_1 и пересекающимися в точках M, N_i коническими сечениями $a_i(A_1, A_2, A_3, M, N_i), b_i(B_1, B_2, B_3, M, N_i)$ (рис. 9).
4. В сечении пучка квадрик $\Phi(\phi_i)$ плоскостью γ получаем пучок конических сечений $\gamma'(g_i)$. Находим в пучке

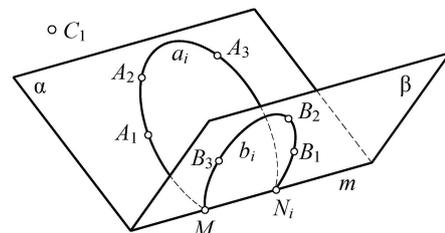


Рис. 9. Квадрика $\phi_i(a_i, b_i, C_1)$

$\gamma'(g_i)$ коническое сечение g_{C_2} , проходящее через точку C_2 . В пучке $\Phi(\phi_i)$ этому коническому сечению отвечает квадрика ϕ_{C_2} . Квадрика ϕ_{C_2} проходит через все заданные точки A_j, B_j, C_j , кроме временно исключенной из рассмотрения точки C_3 .

5. Отметим на m какую-либо точку M' , не совпадающую с M . Точка N_k , бегущая по m , индуцирует пучок квадрик $\Phi'(phi_k)$. Любая квадрика phi_k пучка Φ' определена точкой C_1 и пересекающимися в точках M', N_k коническими сечениями $a_k(A_1, A_2, A_3, M', N_k), b_k(B_1, B_2, B_3, M', N_k)$.

6. В сечении пучка квадрик $\Phi'(phi_k)$ плоскостью γ получаем пучок конических сечений $\gamma'(g_k)$. Находим в пучке $\gamma'(g_k)$ коническое сечение g'_{C_2} , проходящее через точку C_2 . В пучке $\Phi'(phi_k)$ этому коническому сечению отвечает квадрика phi'_{C_2} .

Таким образом, в результате выполнения действий 1...6 найдены квадрики ϕ_{C_2} и phi'_{C_2} , проходящие через все заданные точки A_j, B_j, C_j , кроме точки C_3 . Эти квадрики, пересекаясь, определяют пучок Θ поверхностей второго

порядка. Для окончательного решения задачи следует найти в пучке Θ квадрику, проходящую через точку C_3 .

На основании схемы составлен конструктивный алгоритм реконструкции квадрики, заданной 9 точками.

Алгоритм реконструкции ПВП

Действие 1. Выделив плоскости $\alpha(A_1, A_2, A_3)$, $\beta(B_1, B_2, B_3)$, $\gamma(C_1, C_2, C_3)$, отмечаем линии их пересечения $m=\alpha\cap\beta$, $l=\alpha\cap\gamma$, $n=\beta\cap\gamma$ (рис. 10).

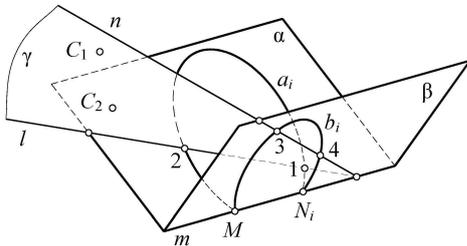


Рис. 10. Поиск квадрики Φ_{C2}

Действие 2 (построение пучка квадрик Φ). Фиксируем на m произвольную точку M . Точка N_i , бегущая по m , порождает в плоскостях α, β пучки коник $\alpha_M(a_i)$ и $\beta_M(b_i)$. Коника a_i пучка α_M определена точками A_1, A_2, A_3, M, N_i . Коника b_i пучка β_M определена точками B_1, B_2, B_3, M, N_i (на рис. 10 точки A_j, B_j условно не показаны). Точка C_1 совместно с любой парой коник a_i, b_i пучков α_M, β_M определяет пучок квадрик $\Phi(\varphi_i)$.

Действие 3 (построение пучка коник γ_M). Рассекая пучок квадрик Φ плоскостью γ , получаем пучок конических сечений $\gamma_M(g_i)$. Чтобы найти базисные точки, достаточно начертить две произвольные коники этого пучка. Для построения какой-либо коники g_1 пучка γ_M фиксируем на m произвольную точку N_1 , вычерчиваем коники $a_1(A_1, A_2, A_3, M, N_1)$, $b_1(B_1, B_2, B_3, M, N_1)$ и отмечаем точки $(1_1, 2_1)=a_1\cap l$, $(3_1, 4_1)=b_1\cap n$ пересечения коник a_1, b_1 с плоскостью γ . Через точки $1_1, 2_1, 3_1, 4_1$ и через точку C_1 проходит коника g_1 . Фиксируя на m другую произвольную точку N_2 и повторяя указанные построения, находим еще одну конику g_2 пучка γ_M . Найденные коники g_1, g_2 пересекаются в четырех точках (одна из которых – точка C_1), указывающих базисные точки пучка γ_M .

Действие 4 (поиск квадрики Φ_{C2}). В пучке γ_M находим конику g_{C2} , проходящую через точку C_2 . Отмечаем точки $(1_{C2}, 2_{C2})=g_{C2}\cap l$, $(3_{C2}, 4_{C2})=g_{C2}\cap n$ пересечения коники g_{C2} с плоскостями α и β . Через точки $A_1, A_2, A_3, 1_{C2}, 2_{C2}$, лежащие в α , проводим конику a_{C2} . Через точки $B_1, B_2, B_3, 3_{C2}, 4_{C2}$, лежащие в β , проводим конику b_{C2} . Конические сечения a_{C2}, b_{C2}, g_{C2} , попарно пересекаясь на прямых m, n, l , определяют квадрику Φ_{C2} , проходящую через произвольно указанную точку M и через все заданные точки A_j, B_j, C_j , кроме временно исключенной из рассмотрения точки C_3 .

Действие 5 (поиск квадрики Φ'_{C2}). Отмечая на m какую-либо точку M' , не совпадающую с M , и повторяя действия 2...4, находим квадрику Φ'_{C2} , проходящую через точку M' и через все заданные точки A_j, B_j, C_j , кроме временно исключенной из рассмотрения точки C_3 . Плоскость γ пересекает квадрику Φ'_{C2} по конике g'_{C2} .

Действие 6 (построение искомой квадрики Φ). Квадрики $\Phi_{C2}(M)$ и $\Phi'_{C2}(M')$, проходящие через все заданные точки, кроме точки C_3 , образуют пучок квадрик Θ . Чтобы найти в пучке Θ искомую квадрику φ , проходящую через C_3 , рассмотрим сечение пучка Θ плоскостью γ . Плоскость γ пересекает квадрики Φ_{C2} и Φ'_{C2} по ранее найденным коническим сечениям g_{C2} и g'_{C2} . Эти коники образуют пучок конических сечений $\omega(g_{C2}\cap g'_{C2})$ с четырьмя базисными точками, две из которых совпадают с точками C_1, C_2 . Между кониками пучка ω и квадриками пучка Θ установлено взаимнооднозначное соответствие, поэтому для построения квадрики φ пучка Θ , проходящей через C_3 ,

достаточно в пучке ω найти коническое сечение g , проходящее через точку C_3 . Коника g проходит через точки C_1, C_2, C_3 и при этом лежит на одной из квадрик φ пучка Θ . Эта квадрика является искомой. Для ее конструктивного выделения из пучка Θ достаточно отметить точки пересечения $(1_3, 2_3)$ и $(3_3, 4_3)$ коники g с плоскостями α, β и начертить конические сечения $a(B_1, B_2, B_3, 3_3, 4_3)$. Попарно пересекающиеся коники a, b, g , лежащие в плоскостях α, β, γ , полностью определяют реконструируемую поверхность второго порядка, заданную девятью точками.

Проверка.

Для проверки алгоритма использованы средства трехмерной компьютерной графики. Моделируем произвольную ПВП (например, эллипсоид вращения), отмечаем на его поверхности девять произвольных точек и “замораживаем” эллипсоид. Выполняя действия 1...6 алгоритма, находим конические сечения a, b, g , проходящие через заданные девять точек и пересекающиеся попарно на линиях m, n, l пересечения плоскостей α, β, γ (рис. 11). “Размораживая” эллипсоид, убеждаемся, что коники a, b, g инцидентны поверхности эллипсоида. Задача решена верно.

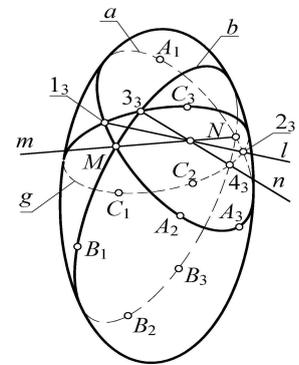


Рис. 11. Проверка

4. Заключение

Найдено графически точное визуальное решение двух известных геометрических задач, поставленных еще в 19 веке. Первая задача – построение F -точек квадратичного кременова преобразования, заданного 7 парами сопряженных точек. Вторая задача – построение любого количества точек квадрики, заданной 9 точками. Отличительная особенность первой задачи состоит в том, что у нее отсутствует аналитическое решение, поскольку нет универсального алгебраического уравнения, описывающего квадратичное кременова преобразование. Аналитическое решение второй задачи сводится к решению системы 9 линейных уравнений. Геометрически точное решение обеих задач, предложенное в настоящей работе, может быть конструктивно реализовано только с помощью специализированного программного средства “Универсальный компьютерный коникограф” [2].

5. Литература

[1] Андреев К.А. О геометрических соответствиях в применении к вопросу о построении кривых линий / К.А. Андреев – М.: Издание Московского Математического общества, состоящего при Императорском Московском Университете, 1879. – 168 с.
 [2] Короткий В.А. Универсальный компьютерный коникограф / В.А. Короткий, Л.И. Хмарова // Труды 26-й Международной научной конференции GraphiCon 2016 (19-23 сентября 2016), НИГАСУ, Нижний Новгород, Россия. – С. 347-351.

Об авторе

Короткий Виктор Анатольевич, к.т.н., доцент кафедры инженерной и компьютерной графики архитектурно-строительного факультета Южно-Уральского государственного университета. E-mail: ospolina@mail.ru.

Распознавание шрифтов по изображениям на основе анализа структуры букв

А.Л. Липкина¹, Н.А. Ломов¹, Л.М. Местецкий¹

lipkina96@mail.ru|nikita-lomov@mail.ru|mestlm@mail.ru

¹Московский государственный университет имени М.В. Ломоносова, Москва, Россия

Рассматривается задача определения шрифта и его начертания на цифровых изображениях, таких как сканированные изображения и фотографии. Предложен метод выделения типов начертания шрифта: прямого, наклонного и с засечками. В основе предлагаемого метода лежит контурное и скелетное представление фигуры. Наклон текста на изображении можно определить, основываясь на информации о наклонах рёбер контурного представления текста. Определить наличие засечек на изображённой букве можно по количеству рёбер агрегированного скелета буквы. Этот метод используется для предварительной фильтрации базы шрифтов, в результате которой в качестве кандидатов рассматриваются только шрифты в нужном начертании. Дальнейшим признаковым описанием изображения служит совокупность морфологических моментов – функций, отражающих пространственное распределение ширины отдельных символов. Для идентификации шрифта строится взвешенная композиция метрических классификаторов, связанных с этими функциями. Эксперименты показывают, что предварительное определение начертания, а также учёт геометрических особенностей букв повышают качество распознавания по сравнению с методом на основе морфологической ширины.

Ключевые слова: цифровой шрифт, наклонное начертание, засечки, граничное представление, агрегированный скелет, морфологические моменты.

Font Recognition from Images Based on Letter Structure Analysis

We consider the problem of the recognition of the typeface and its style on the digital images, such as scanned images and photographs. A method is proposed for detection of font style: normal, italic and serif. The proposed method is based on the contour and skeleton representations of the figure. The slope of the text in the image can be determined using information about the slopes of the edges of the contour representation of the text. It is possible to determine the presence of serifs on the represented letter by the number of edges of its aggregated skeleton. This method is used to pre-filter the font database, resulting in the fact that only the fonts in the desired style are considered as candidates. A further feature description of the image is the set of morphological moments, i.e. functions, which reflect the spatial distribution of the width of individual symbols. To identify the font, a weighted composition of the metric classifiers associated with these functions is constructed. Experiments show that the preliminary style detection, as well as the consideration of the geometric features of letters, increase the quality of recognition in comparison with the method based on the morphological width.

Keywords: digital font, italic style, serifs, boundary representation, aggregated skeleton, morphological moments.

1. Введение

Типографика является неотъемлемой частью графического дизайна. У графических дизайнеров часто возникает желание определить шрифты, с которыми они сталкиваются в реальной жизни, с целью их дальнейшего использования. Хотя они могут сфотографировать образец шрифта, вызвавшего особый интерес, и попросить эксперта определить шрифт, процесс распознавания «вручную» оказывается очень утомительным и подверженным ошибкам. Поэтому существует потребность в методах автоматической идентификации шрифта по изображению, причём нужный шрифт нужно будет определить из тысяч кандидатур, ведь шрифтов, используемых в современной типографике, – великое множество. Тем не менее работ, посвящённых этой задаче и предполагающих взаимодействие с обширной (более 1000 шрифтов) базой, очень немного, последние из них основаны на формировании кодовой книги из векторов локальных признаков, в качестве которых может использоваться HOG [7] или SIFT [8]. Известен также метод, который по аналогии с методом собствен-

ных лиц можно назвать «методом собственных букв» [10]. Самым успешным оказался метод на основе глубинного обучения [11], внедрённый в редактор Adobe Photoshop в виде инструмента Match Font.

Ряд веб-сайтов позволяет пользователям совершать поиск по шрифтовой базе и ранжировать шрифты по степени соответствия запросу. Среди наиболее популярных из таких сервисов – WhatTheFont [2] и Identifont [1]. Хотя алгоритмы, лежащие в основе этих сервисов, не раскрываются, сам характер их взаимодействия с пользователем говорит о том, что принципы их работы кардинально различаются. Сервис WhatTheFont получает на вход картинку с образцом текста, набранного интересующим пользователя шрифтом, и делает его побуквенную разметку: выделяет и распознаёт на изображении отдельные символы. Это позволяет сделать вывод, что сходство шрифтов оценивается на основе сходства их отдельных символов. При работе с сайтом Identifont картинка не требуется – вместо этого пользователю предлагается ответить на ряд вопросов, касающихся как общего облика шрифта (на-

пример, есть ли в нём засечки), так и особенностей отдельных букв (например, какой формы хвостик у буквы Q).

В предыдущих работах [3, 4] был предложен метод распознавания шрифтов, основанный на анализе ширины штрихов в символах шрифта. В качестве признакового описания при этом использовалась диаграмма морфологической ширины, описывающая площадь дискового покрытия фигур в зависимости от размера покрывающих дисков. Общая схема метода выглядит следующим образом:

- выделение и распознавание отдельных букв в сканированном тексте на основе сегментации исходного изображения;
- построение диаграмм морфологической ширины для каждой буквы;
- сравнение диаграмм морфологической ширины букв с диаграммами эталонных описаний шрифтов (метрика сходства букв);
- вычисление обобщенной меры сходства шрифта текста с эталонными описаниями шрифтов (метрика сходства шрифтов);
- выбор наиболее похожих эталонных шрифтов.

Метод показал хорошие результаты, ошибки составили 4–9% в зависимости от качества сканированного текста. Вместе с тем, эксперименты показали, что оценивание сходства шрифтов только лишь на основе морфологической ширины оказывается недостаточным. Имеются примеры шрифтов, которые при общем сходстве диаграмм ширины существенно отличаются с точки зрения человеческого (дизайнерского) представления. Например, похожие распределения ширины могут иметь шрифты, отличающиеся наклоном букв или наличием штрихов-засечек.

В настоящей статье рассматривается подход к уменьшению ошибок такого рода за счёт учёта структурных признаков шрифта: наклона и засечек. Кроме того, вводится понятие морфологических моментов фигуры, обобщающее понятие морфологической ширины и позволяющее использовать не только распределение ширины штрихов, но также учесть более детальное распределение «массы» в символах шрифта. Это существенно расширяет возможности для классификации шрифтов с похожими диаграммами морфологической ширины.

Вычислительный метод, обеспечивающий получение соответствующих признаков, основан на модели непрерывного медиального представления бинарных растровых изображений [5]. Метод развивает идею, описанную в работе [3].

Структура оставшейся части статьи такова. В разделе 2 описан метод классификации шрифтового начертания по типу наклона – прямое либо курсивное. В разделе 3 представлен метод классификации типа шрифта

по наличию штрихов-засечек. В разделе 4 описывается метод построения признакового описания распределения «массы» в символах шрифта, так называемые морфологические моменты. Там же представлена идея вычисления моментов на основе непрерывного медиального представления дискретных изображений. В разделе 5 описаны общие решающие правила классификации шрифта по сводному набору признаков: морфологические моменты, наклон, засечки. В разделе 6 представлены результаты вычислительных экспериментов.

2. Классификация текста по типу наклона

На вход методу подаётся незашумлённое бинарное изображение некоторого текста. На всех стадиях работы метода мы будем оперировать непрерывным представлением формы. С точки зрения такого представления бинарное изображение рассматривается как непрерывная сцена – набор фигур, каждая из которых является многоугольником с многоугольными дырами. На этапе определения наклона шрифта мы будем работать со сценой как с единым целым, дальнейшие этапы потребуют разметки текста – указания, какие буквы присутствуют в тексте и из каких фигур состоит каждая буква. Приведём необходимые определения.

Фигура – замкнутая область на плоскости, ограниченная конечным числом непересекающихся замкнутых жордановых кривых.

Контурное представление фигуры – аппроксимация границ фигуры многоугольниками минимального периметра.

Пустой круг фигуры – круг, целиком лежащий в фигуре.

Вписанный круг фигуры – пустой круг фигуры, не содержащийся ни в каком другом пустом круге фигуры.

Медиальное представление формы – совокупность скелета (множества центров всех вписанных пустых кругов фигуры) и радиальной функции (функции, которая каждой точке скелета ставит в соответствие радиус вписанного круга).

В действительности скелет представляет собой планарный граф (рис. 1). Вершинами графа являются центры вписанных пустых кругов, имеющих либо одну, либо три общих точки с границей фигуры, а рёбрами – линии из центров вписанных пустых кругов, касающихся границы ровно в 2 точках.

Методы аппроксимации формы многоугольниками фигурами и построения скелета подробно изложены в [5].



Рис. 1. Фигура (слева) и её скелет (справа).

2.1. Подготовка данных

Предлагаемый алгоритм классификации опирается только на контурное представление фигуры. Идея этого алгоритма состоит в том, что наклон ребер контурного представления фигуры в среднем достаточно хорошо описывает наклон фигуры.

Каждой связной компоненте бинарного изображения соответствует непрерывная фигура. Для каждой фигуры (чаще всего это вся буква целиком) на непрерывной сцене можно получить явное описание всех её контуров. Совокупность внешнего и внутренних контуров связной компоненты будем называть *полной границей* компоненты.

Каждый контур представляется набором вершин, лежащих в порядке обхода многоугольника по или против часовой стрелки. По соседним вершинам можно получить описание ребра: его длину l_i и угол наклона $\alpha_i \in [-\pi, \pi)$ относительно оси абсцисс. Угол наклона для удобства приводится к отрезку $[0, \pi)$ по формуле:

$$\alpha_i = \begin{cases} \alpha_i, & \alpha_i \in [0, \pi); \\ \alpha_i + \pi, & \alpha_i \in [-\pi, 0). \end{cases}$$

Таким образом, после обработки углов наклона ребер каждая полная граница представляется набором пар (l_i, α_i) .

2.2. Построение дескриптора наклона контурного представления

Для каждой полной границы на изображении мы построим гистограмму распределения длин ребер по углам наклона из n равных столбцов. В полученной гистограмме нас будет интересовать значение высоты столбца, отвечающего за угловой интервал, содержащий точку $\frac{\pi}{2}$. Так как углы, которые являются равными $\frac{\pi}{2}$, на самом деле измеряются с некоторой погрешностью за счёт аппроксимации границы многоугольником, стоит обеспечить, чтобы для малых ε углы $\frac{\pi}{2} - \varepsilon$ и $\frac{\pi}{2} + \varepsilon$ попали в одну и ту же ячейку гистограммы. Для решения этой проблемы мы будем рассматривать только чётные n и отобразим углы $\alpha_i \in [\pi - \frac{\pi}{2n}, \pi) \rightarrow [-\frac{\pi}{2n}, 0)$ за счёт вычитания из них π . Вследствие такого преобразования все углы будут принадлежать отрезку $[-\frac{\pi}{2n}, \pi - \frac{\pi}{2n})$, который мы разобьём на n равных частей при построении гистограммы. Столбец гистограммы, соответствующий отрезку $[\frac{\pi}{2} - \frac{\pi}{2n}, \frac{\pi}{2} + \frac{\pi}{2n})$, будем называть показательным. Вклад каждого ребра в гистограмму примем равным его длине. По результатам экспериментов n полагается равным 32.

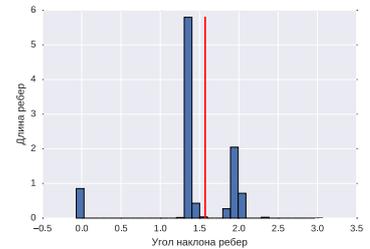
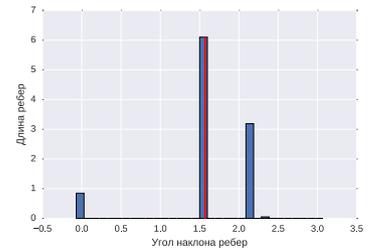


Рис. 2. Гистограмма распределения длине ребер по углам для прямого начертания буквы «N» в шрифте Arial (сверху) и для её наклонного начертания (снизу).

2.3. Определение типа наклона буквы

Найдём показательный столбец и рассмотрим такие столбцы, которые отвечают за углы в окрестности $U = [\frac{\pi}{4}, \frac{3\pi}{4})$.

Обозначим за U_i^k номера таких столбцов гистограммы, которые лежат не более, чем на расстоянии k столбцов от i -го столбца, но не считая самого i -го столбца.

Будем говорить, что столбец является *локальным пиком*, если $h_i > \gamma \sum_{j \in U_i^k} h_j$ для некоторых фиксированных k и γ .

В данном исследовании на основании проведённых экспериментов γ берётся равным 0,5.

Определим, есть ли какой-либо пик в окрестности U . Для этого найдём максимальный столбец из U (обозначим его номер за m) и проверим, является ли он локальным пиком при $k = 1$. Если является, и этот столбец совпадает с показательным столбцом, то начертание буквы однозначно классифицируется как прямое. Если он является локальным пиком, и содержит углы, меньшие $\frac{\pi}{2}$, то начертание однозначно классифицируется как наклонное. Иначе определить тип начертания невозможно.

Как видно на рис. 2, в верхней части явно выраженный пик гистограммы приходится на значение $\frac{\pi}{2}$ (красная линия), а в нижней части пик лежит левее красной линии. Это свидетельствует о том, что буква «N» в шрифте Arial является представительной. В то же время, из рис. 3 видно, что у «O» нет каких-либо явно выраженных пиков – это говорит о том, что по этой букве нельзя достоверно определить, является ли она наклонной или нет.

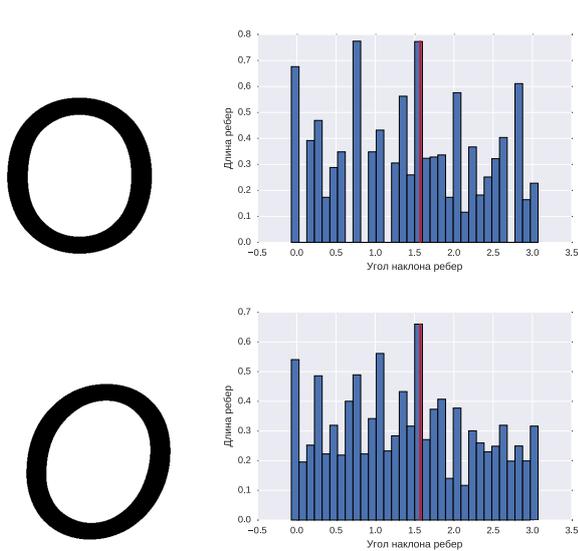


Рис. 3. Гистограмма распределения длин рёбер по углам для прямого начертания буквы «О» в шрифте Arial (сверху) и для её наклонного начертания (снизу).

2.4. Решающее правило

Пусть $M_r = \frac{i_r}{d_r + i_r}$, где d_r – число букв, начертание которых классифицировано как прямое, i_r – как наклонное. Зададим некоторый порог $\theta \in [0, 1]$. Если $M_r > \theta$, то говорим, что шрифт имеет наклонное начертание, иначе – прямое. Эксперименты показали, что подходящим является значение $\theta = 0,4$.

2.5. Классификация текста по наличию засечек

Введём понятие *агрегированного скелета* – это граф, вершинами которого являются вершины скелета фигуры степени 1 и 3, а рёбрами – цепи, соединяющие вершины степени 1 и 3 и проходящие по вершинам степени 2 (рис. 4). Алгоритм выявления засечек опирается только на агрегированный скелет фигуры. Идея алгоритма состоит в том, что количество рёбер агрегированного скелета фигуры в среднем хорошо описывает наличие в ней засечек.

Скелет связанной фигуры и последующая его стрижка (регуляризация) делаются стандартным способом, описанным в [5]. Экспериментальным путём было выяснено, что оптимальным параметром стрижки скелета является 7% общей высоты фигуры. Число рёбер E агрегированного скелета можно определить по теореме Эйлера: $E = V + F - 1$, где V – число вершин агрегированного скелета (число вершин исходного скелета степени 1 и 3), F – число дыр в фигуре (число внутренних контуров). Исключением из этого правила является случай, когда скелет состоит из единственного цикла, что возможно, например, для буквы «О». В этом случае мы по определению считаем, что агрегированный скелет содержит одну вершину и одно ребро-петлю.

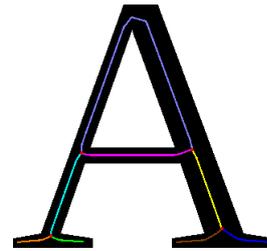


Рис. 4. Выделенные разным цветом рёбра (цепи) агрегированного скелета.

Интуитивно понятно, что у буквы с засечками количество рёбер агрегированного скелета должно быть больше, чем у такой же буквы из той же гарнитуры без засечек. Но порог, по которому можно явно отнести букву к одному из классов, естественно, для каждой буквы свой. В качестве порога для каждой буквы можно взять среднее число рёбер в шрифтах с засечками (c_i^{serif}) и без (c_i^{sans}):

$$\delta_i = \frac{c_i^{serif} + c_i^{sans}}{2}.$$

Напомним, что на вход этой задаче подается *размеченное* изображение буквы. Это означает, что известно, какая буква представлена на изображении.

Для каждой буквы по обучающей выборке можно определить, какую долю среди шрифтов, у которых число рёбер не больше порогового, составляют шрифты без засечек:

$$p(s = sans | E_i \leq \delta_i) = \frac{\sum_{j: E_i^j \leq \delta_i} [s_j = sans]}{\sum_{j: E_i^j \leq \delta_i} 1},$$

где s_j – стиль j -го шрифта, E_i^j – количество рёбер в агрегированном скелете его i -й буквы. Аналогичным образом можно рассчитать величины $p(s = serif | E_i \leq \delta_i)$, $p(s = sans | E_i > \delta_i)$, $p(s = serif | E_i > \delta_i)$.

Тогда, если

$$\prod p(s = serif | E_k \leq \delta_k)^{[E_k \leq \delta_k]} p(s = serif | E_k > \delta_k)^{[E_k > \delta_k]} > \prod p(s = sans | E_k \leq \delta_k)^{[E_k \leq \delta_k]} p(s = sans | E_k > \delta_k)^{[E_k > \delta_k]},$$

то начертание надписи классифицируется как начертание с засечками (индекс k пробегает только номера букв, присутствующих на изображении), иначе – как начертание без засечек.

Заметим, что и в этом случае для надёжной классификации необходимо присутствие в слове «представительных» букв, вид которых в шрифтах с засечками и без значительно отличается.

3. Идентификация шрифта по морфологическим моментам

3.1. Понятие морфологических моментов

Морфологические моменты развивают концепцию морфологической ширины фигуры, введённую в [3]. Напомним необходимые определения:

Дисковое r -покрытие фигуры – объединение всех пустых кругов радиуса r .

r -площадь фигуры – площадь её дискового r -покрытия.

Морфологическая ширина фигуры – это r -площадь фигуры как функция от r .

Идея морфологических моментов заключается в том, что вычисление площади фигуры можно рассматривать как взятие интеграла от функции $f(x, y) = 1$ по рассматриваемой фигуре. В случае моментов рассматривают функции более общего вида: $f(x, y) = x^p y^q$. В результате моменты позволяют описать не только то, как меняется «масса» покрытия с ростом радиуса, но и то, как меняется её пространственное распределение.

Морфологическим моментом порядков (p, q) фигуры X называется функция, описывающая зависимость геометрического момента, вычисленного по дисковому r -покрытию, от размера примитива:

$$\mathfrak{M}_{pq}(X, r) = \iint_{(x,y) \in X_r} x^p y^q dx dy,$$

где X_r – дисковое r -покрытие фигуры X .

Нетрудно заметить, что в таком виде морфологические моменты оказываются чувствительными к выбору системы координат. Чтобы обеспечить их инвариантность к параллельному переносу фигуры, моменты можно центрировать, поместив начало координат в центр масс фигуры:

$$\mathfrak{M}_{pq}^*(X, r) = \iint_{(x,y) \in X_r} (x - \bar{x})^p (y - \bar{y})^q dx dy,$$

(\bar{x}, \bar{y}) – координаты центра масс, которые находятся по формулам

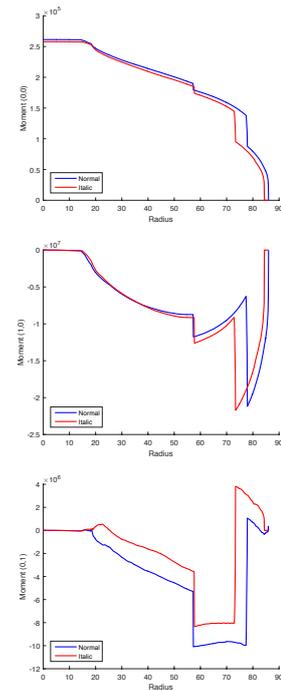
$$\bar{x} = \mathfrak{M}_{10}(X, 0) / \mathfrak{M}_{00}(X, 0), \bar{y} = \mathfrak{M}_{01}(X, 0) / \mathfrak{M}_{00}(X, 0).$$

При этом центральные моменты по-прежнему остаются чувствительными к масштабированию и повороту фигуры. Пользуясь особенностями задачи, эту проблему можно решить следующим образом: установим систему координат, в которой начало совпадает с центром масс буквы, ось OX параллельна базовой линии шрифта, а масштаб выбран так, что высота заглавной буквы шрифта оказывается равной фиксированному значению.

Как показывает верхний график (рис. 5), ширинные характеристики символа часто не меняются при смене начертания с прямого на курсивное. Однако эти символы могут быть различимы по моментам более высоких порядков (два нижних графика).



(a)



(b)

Рис. 5. Прямое и курсивное начертание символа (a). Графики моментов нулевого и первого порядка (b); синий – прямое, красный – курсивное.

3.2. Процедура вычисления

Как и морфологическую ширину, морфологические моменты можно рассчитать с высокой точностью в результате высокоэффективной вычислительной процедуры, использующей медиальное представление формы. Дисковое r -покрытие представляет собой геометрическую фигуру (или совокупность фигур), ограниченную отрезками прямых и выпуклыми дугами окружностей. В ходе работы алгоритма [3] дисковое покрытие получает явное представление в виде объединения и исключения простых геометрических фигур (рис. 6), среди которых:

- собственные области рёбер скелета (выделены светло-серым);
- внешние сектора меньших концевых кругов усечённых рёбер (тёмно-серый);
- линзы в пересечении внешних секторов (красный).

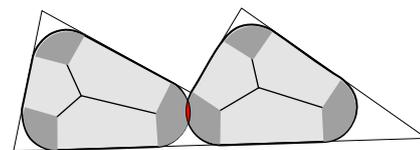


Рис. 6. Составные части дискового покрытия многоугольной фигуры.

Площадь покрытия складывается из суммы площадей собственных областей и секторов за вычетом суммы площадей линз. Но ничто не мешает нам сохранить общий вид алгоритма, заменив на третьем этапе вычисление площадей составляющих фигуру областей на вычисление их моментов.

Любая область первого типа представляет собой простой многоугольник, второго – круговой сектор, третьего – объединение двух областей, каждая из которых – это дополнение треугольника до кругового сектора. Таким образом, нам достаточно уметь интегрировать функцию $f(x, y) = x^p y^q$ по многоугольнику и круговому сектору. Это можно сделать аналитически, точные формулы выведены в работе [9].

3.3. Задание метрики

Для сравнения шрифтов мы вычислим диаграммы моментов – значения функций для аргументов r_i , взятых с определённым шагом. Обозначим код j -го символа на изображении c_j , его диаграммы моментов – t_{pq}^j , эталонные моменты символа i -го шрифта с кодом j – f_{pq}^{ji} . Расстоянием от тестового образца $\mathbf{t} = \{t_1, \dots, t_n\}$ до шрифта \mathbf{f}_k будем считать взвешенную сумму расстояний до эталонов по моментам отдельных символов:

$$d(\mathbf{f}_k, \mathbf{t}) = \sum_{pq} \alpha_{pq} \sum_{j=1}^n \beta_j \|f_{pq}^{c_j k}, t_{pq}^j\|_{L_1}. \quad (1)$$

Здесь α_{pq} – вес момента порядка (p, q) , а β_j – вес j -й буквы. В результате распознавания имеющиеся в базе шрифты ранжируются по мере сходства с представленным образцом, т.е. в порядке убывания расстояния, что позволяет получить как количественную (расстояния), так и качественную (номер в упорядоченном списке) оценку сходства.

4. Построение классификатора

Пусть $\mathbf{t} = (t_1, t_2, \dots, t_n)$ – множество символов, выделенных в исходном тексте на этапе сегментации. В это множество входят все выделенные символы, в том числе это могут быть изображения одних и тех же букв. Их растровые представления могут отличаться из-за особенностей сканирования, бинаризации и т.п. Обозначим

$C(\mathbf{t})$ – результат классификации типа наклона текста \mathbf{t} : для прямого шрифта $C(\mathbf{t}) = 0$, а для наклонного $C(\mathbf{t}) = 1$.

$S(\mathbf{t})$ – результат классификации текста по наличию штрихов-засечек: если $S(\mathbf{t}) = 0$, то засечек нет, а если $S(\mathbf{t}) = 1$, то есть.

F – библиотека эталонных описаний шрифтов, содержащая K описаний. В ней для каждого шрифта \mathbf{f}_i , $i = 1, \dots, K$ заданы априори известные признаки наклона и засечек C_i и S_i соответственно.

Для всех шрифтов выделен набор символов, которые могут быть использованы для распознавания шрифтов. Этот набор из M символов включает прописные и строчные буквы латиницы и кириллицы, а также

цифры. В каждом шрифте этот набор символов представлен своим набором глифов $Z_i = z_j^{(i)}$, $j = 1, \dots, M$.

Эталонное описание шрифта включает в себя набор из $\frac{(R+1)(R+2)}{2}$ функций моментов $\{f_{pq}^{ji}\}$, $p, q \geq 0$, $p+q \leq R$ для каждого символа $j = 1, \dots, M$ каждого шрифта $i = 1, \dots, K$. Число R может быть небольшим, достаточно рассмотреть $R = 1$ – это означает, что мы принимаем во внимание только площадь и центр масс дискового покрытия символа.

Общая схема работы классификатора для текста \mathbf{t} выглядит так.

Шаг 1. Вычисление признаков наклона $C(\mathbf{t})$ и засечек $S(\mathbf{t})$ для текста \mathbf{t} .

Шаг 2. Отбор подмножества эталонных шрифтов $F_{\mathbf{t}}$ из библиотеки F , у которых $C_i = C(\mathbf{t})$ и $S_i = S(\mathbf{t})$.

Шаг 3. Вычисление сходства фрагмента текста \mathbf{t} с эталонными шрифтами из $F_{\mathbf{t}}$.

Шаг 4. Отбор $top(\mathbf{t})$ – множества эталонных шрифтов, наиболее близких к шрифту текста \mathbf{t} . Размер списка $top(\mathbf{t})$ определяется, исходя из потребностей пользователя. По экспертной оценке шрифтовых дизайнеров этот список может включать 5–10 наиболее похожих шрифтов.

5. Эксперименты

Для проведения экспериментов была расширена база данных, уже использованная в работе [4]. Новая база содержит изображения 52 символов латиницы (26 строчных и 26 прописных букв) из 2555 шрифтовых начертаний коллекции цифровых шрифтов компании Паратайп [6]. Ранее коллекция содержала 1884 шрифта. Для экспериментов по идентификации начертания использовались только такие шрифты, начертание которых можно установить однозначно, – в частности, декоративные шрифты были исключены из коллекции.

Для каждого шрифта из букв в низком разрешении были получены изображения 1000 наиболее популярных английских слов (средняя длина слова составила 5,33 символа, в наборе были и очень короткие слова, например, $\mathbb{1}$, $\mathbb{b}\epsilon$), случайные 40% из которых были переведены в верхний регистр. Эти изображения и их сегментация использовались как запросы для идентификации шрифта.

5.1. Классификация по типу наклона

Для проведения эксперимента из коллекции было отобрано 1760 шрифтов: 967 в прямом начертании и 793 в наклонном. Размер шрифта был подобран так, чтобы высота заглавной буквы \mathbb{H} составила 50 пикселей.

В результате применения предложенного метода точность классификации по слову составила 97,14% для прямого начертания и 92,71% – для наклонного.

Матрица ошибок представлена в таблице 1:

	Прямое	Наклонное	Не определено
Прямое	939 338	26 242	1 420
Наклонное	57 380	735 202	418

Таблица 1. Матрица ошибок при распознавании наклонного начертания по буквам из слова.

Далее для этих же шрифтов в прямом и наклонном начертании были сгенерированы картинки с текстом «the quick brown fox jumps over the lazy dog». Эта фраза примечательна тем, что в ней содержатся все буквы латинского алфавита. Высота заглавной буквы Н снова составила 50 пикселей.

В результате применения предложенного метода точность классификации прямого начертания составила 97,72%, а наклонного – 98,49%. Примечательно, что, так как в тексте заведомо присутствуют буквы, у которых можно определить тип наклона, метод больше не выдаёт отказы от классификации.

Более подробные результаты представлены в нижеприведённой матрице ошибок:

	Прямое	Наклонное
Прямое	945	22
Наклонное	12	781

Таблица 2. Матрица ошибок при распознавании наклонного начертания по всем буквам латинского алфавита.

5.2. Классификация по наличию засечек

Для проведения эксперимента из коллекции был отобран 2101 шрифт: 1285 с засечками и 816 без засечек, по которым был определён порог δ_i для каждой буквы. Высота заглавной буквы при растеризации символов составила 50 пикселей.

Результаты показали, что у некоторых букв (например, «А», «Н») показатели среднего количества рёбер для начертания с засечками и без засечек сильно различаются, а у некоторых (например, «о», «d») – практически не отличаются (рис. 7). Понятно, почему так получается: начертания каких-то букв с засечками и без засечек могут отличаться незначительно (например, у буквы «о» вообще нет засечек в обоих случаях).

В ходе проведённого эксперимента для тех букв, где отличие между средним количеством рёбер для начертаний с засечками и без достаточно большое, точность классификации составила примерно 95%. Некоторые буквы оказались совершенно не информативными для алгоритма. Средняя точность классификации по одной букве превысила 80%, а по одному слову составила 94,8%. Можно сделать вывод: достаточно наличия в тексте одной характерной буквы, чтобы определить, имеет ли шрифт засечки.

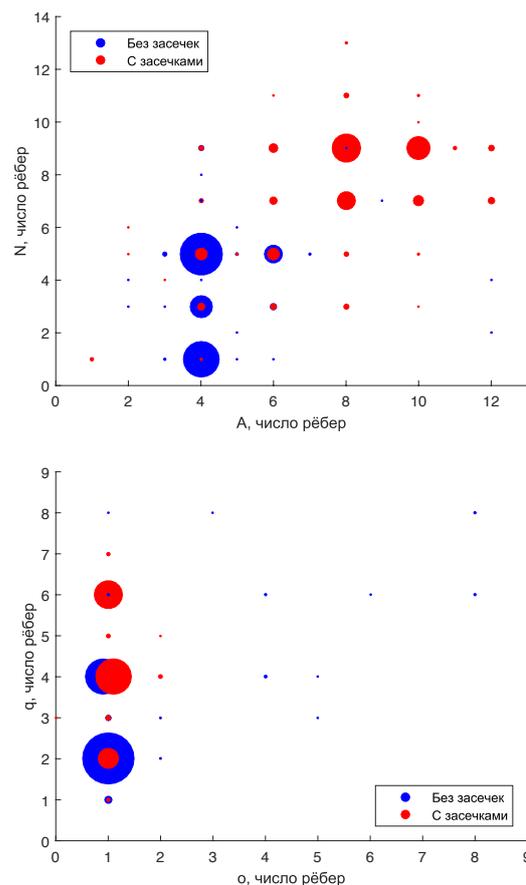


Рис. 7. Распределение числа рёбер агрегированного скелета в шрифтах с засечками и без для букв «А» и «N» (сверху), «о» и «q» (снизу). Площадь каждой отметки прямо пропорциональна количеству шрифтов с данной комбинацией числа рёбер.

5.3. Определение шрифта по морфологическим моментам

Для эталонных изображений были получены диаграммы морфологических моментов методом, описанным в статье [9]. Для этого каждый символ был отрисован в виде бинарного растрового изображения в таком масштабе, чтобы высота прописной буквы Н составила 1000 пикселей. По этим изображениям были построены непрерывные скелеты методом, описанным в [5], и на их основе вычислены центральные морфологические моменты нулевого и первого порядков с шагом радиуса в 0,5 пикселя.

Для этих же шрифтов были получены изображения символов в более низком разрешении, так чтобы высота буквы Н составила 100 и 50 пикселей. Для этих символов также были построены диаграммы моментов с тем же шагом радиуса. При этом все координаты в скелетно-граничном представлении предварительно были увеличены в 10 раз для размера 100 и в 20 раз для размера 50, так что, хотя качество аппроксимации формы было значительно хуже эталонного, масштаб сим-

волов был тем же, и диаграммы моментов можно было сравнивать.

При сравнении шрифтов вес момента нулевого порядка был принят равным 1, моментов первого порядка – 0,01. Такая разница объясняется тем, что с ростом порядка момента значения в диаграммах резко возрастают. Вес каждой буквы был взят пропорциональным качеству классификации, которое даёт распознавание по одной этой букве.

Результаты эксперимента показали, что средний уровень правильного распознавания шрифта по слову при использовании только момента нулевого порядка (т.е. морфологической ширины) составил 89,3% при размере 100 и 63,3% при размере 50, добавление центральных моментов первого порядка позволило повысить точность распознавания до 94,1% и 84,6% соответственно.

Таким образом, проведенный эксперимент подтвердил работоспособность предложенного метода и показал его эффективность на практической задаче сравнения большого количества изображений ($2555 \times 2555 \times 52$) с весьма высоким качеством распознавания.

6. Заключение

В данной работе были предложены методы классификации начертаний изображений текста: по прямому начертанию/наклонному начертанию/невозможности определить начертание, а также по наличию засечек у буквы. Для идентификации шрифта предложен метод на основе морфологических моментов, обобщающий понятие морфологической ширины.

Алгоритмы для распознавания начертания показывают хорошие результаты на изображениях отдельных слов, особенно, если на них присутствуют буквы, сильно меняющие облик в зависимости от начертания. Применяя эти алгоритмы, можно успешно отфильтровать исходную выборку и исключить из рассмотрения шрифты, начертание которых отличается от установленного.

Экспериментально показано, что добавление моментов первого порядка к диаграммам морфологической ширины повышает качество распознавания. При этом требования к качеству входных данных достаточно невысоки, что позволяет работать с изображениями, полученными в результате обычного сканирования.

7. Благодарности

Работа выполнена при поддержке РФФИ (грант № 17-01-00917).

8. Литература

- [1] URL: www.identifont.com. – Identifont – 2000–2017.
 [2] URL: www.myfonts.com/WhatTheFont. – WhatTheFont! – 1999–2017.

- [3] Ломов Н.А., Местецкий Л.М. Площадь дискового покрытия – дескриптор формы изображения // Компьютерная оптика, 40(4): 516–525, 2016.
 [4] Ломов Н.А., Местецкий Л.М. Идентификация цифровых шрифтов на основе морфологической ширины // GraphiCon-2016, Нижний Новгород, с. 89–94.
 [5] Местецкий Л.М. Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры // М.: ФИЗМАТЛИТ, 2009.
 [6] Якупов Э., Петрова И., Фридман Г., Королькова А., Левин Б. 2008–2014. ПАРАТАЙП. Цифровые шрифты. Каталог // Москва: ParaType, 2014.
 [7] Bui T., Collomosse J. Font finder: Visual recognition of typeface in printed documents // Conference on Image Processing (ICIP), 2015 – pp. 3926-3930.
 [8] Chen G., Yang J., Jin H., Brandt J., Shechtman E., Agarwala A., Han T. Large-Scale Visual Font Recognition // Conference on Computer Vision and Pattern Recognition, 2014. – pp. 3598–3605.
 [9] Lomov N., Sidyakin S. Morphological Moments of Binary Images // ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W4, 2017. – pp. 19–25.
 [10] Solli M., Lenz R. FyFont: Find-your-Font in Large Font Databases // Proceedings of the 15th Scandinavian conference on Image analysis, 2007. – pp. 432–441.
 [11] Wang, Z., Yang, J., Jin H., Shechtman E., Agarwala A., Brandt J. Huang, T. S. DeepFont: Identify Your Font from An Image // Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 2015. – pp. 451–459.

Виртуальный испытательный полигон для роботизированных охранных комплексов

И.О. Михайлов¹, И.С. Стародубцев^{1,2}, В.С. Провков¹
igormich88@gmail.com|starodubtsevIS@robolab.tk|provkovv@mail.ru

¹Уральский Федеральный Университет, Екатеринбург, Россия;

²Институт математики и механики им. Н. Н. Красовского Уральского отделения РАН,
Екатеринбург, Россия

В докладе будет представлена концепция Виртуального Испытательного Полигона (ВИП). Описаны задачи, которые необходимо решить, а также перечислены преимущества ВИП. Будут приведены реальные задачи, для которых ВИП разрабатывается.

Ключевые слова: робототехника, виртуальная реальность, испытательный полигон, визуализация, компьютерное моделирование, техническое зрение.

Virtual test area for robotic security complex

I.O. Mikhailov¹, I.S. Starodubtsev^{1,2}, V.S. Provkov¹
igormich88@gmail.com|starodubtsevIS@robolab.tk|provkovv@mail.ru

¹Ural Federal University, Yekaterinburg, Russia;

²Krasovskii Institute of Mathematics and Mechanics UB RAS, Yekaterinburg, Russia

The report will be presented to the concept of Virtual Test Area (VTA). Describes the tasks to be resolved and the benefits of the VTA. Will be given real tasks for which VTA is being developed.

Keywords: robotic, VR, testing area, visualisation, computer modeling, computer vision.

1. Введение

Испытание и тестирование нового оборудования – это весьма трудоемкий процесс, отнимающий много времени. Тестирование программного обеспечения для автономных робототехнических комплексов (РТК), с дополнительным оборудованием, добавляет дополнительные расходы на развитие и разработку алгоритмов управления, технического зрения, взаимодействия с окружающей средой.

Эти расходы могут быть значительно сокращены за счет ускорения и упрощения технологического процесса, используя контролируемую окружающую среду виртуального испытательного полигона для разработки методов и алгоритмов. Использование контролируемой параметрической среды позволяет разработать различные сценарии взаимодействия РТК на различных уровнях (робот – человек, робот – робот, группа – группа).

Моделирование различных датчиков позволяет выбрать конфигурацию, которая подходит для особых условий использования, а также дает возможность разработать алгоритмы без необходимости приобретения дорогостоящего оборудования.

Виртуальный испытательный полигон включает в себя следующие субъекты:

1. Виртуальные камеры (включая установленные на мобильных платформах).
2. Система датчиков (инфракрасные, ультразвуковые, микроволновые, датчики пересечения луча);
3. Модель испытательного полигона.
4. Программное обеспечение, имитирующие людей, как законопослушных, так и нарушителей.
5. Программное обеспечение, соответствующее роботу и роботизированной системе;
6. Система сценариев для обучения и испытания РТК.

2. Методы

2.1 Модульный подход

При разработке виртуального испытательного полигона, модульный подход выбирается для достижения наибольшей универсальности, благодаря возможности выбрать необходимые модули, предназначенные для достижения конкретных целей. Динамическая загрузка и обновление модулей позволяет изменить режим визуализации (помехи, освещенность, погодные условия) и поведение модели без перезагрузки всей системы. Также стоит отметить инструмент «Debug pause», позволяющий остановить виртуальный испытательный полигон с возможностью просмотра и редактирования свойств каждого из модулей, а затем продолжить работу.

2.2 Распределенные вычисления

Распределенная визуализация с синхронизацией между несколькими вычислительными узлами (Рис.1). Присутствует возможность визуализировать изображение на большое количество виртуальных камер в режиме реального времени. В этом случае, один из вычислительных узлов выполняет физические расчеты и контролирует модули, затем сведения о состоянии полигона переносятся на другие вычислительные узлы, которые отображают изображения, полученные с камер, и значения, полученные с датчиков. Потенциально, физические расчеты и логика также могут быть распределены между несколькими узлами.

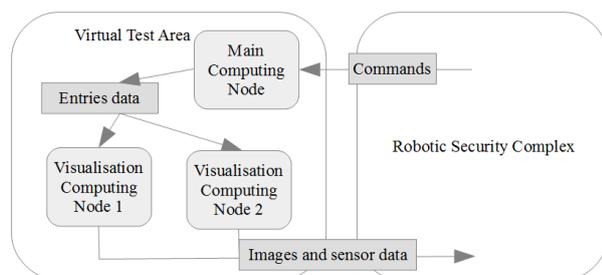


Рис. 1. Организация виртуального испытательного полигона и общения с РТК.

3. Результаты

3.1 Система визуализации

Процесс визуализации реализуется благодаря данным с каждой из виртуальной камеры, включая и реально подключенную. Отрисованное изображение сохраняется в памяти, а затем, по необходимости, передаются в систему распознавания. Различные параметры визуализации позволяют имитировать не только течение времени. В то же время когда освещенность снижается (что возможно сделать командой управления), камеру можно переключить в ночной режим. Присутствует поддержка устройств kinest-визуализации, основаны на глубине буфера. Также возможно имитировать помехи и потери сигнала с камер.

3.2 Система датчиков

Использование специальных материалов позволяет интерполировать данные для объемных и температурных датчиков. Следует отметить, что в большинстве случаев изображения с низким разрешением вполне достаточно для приемлемого уровня интерполяции. При работе с датчиками, как и при работе с камерами, возможно добавить шум и ложные срабатывания.

Расчет данных для инфракрасных датчиков осуществляется с помощью специальных материалов, яркость которого соответствует интенсивности теплового излучения поверхности. Сначала, полигон визуализируется с этими материалами, в тот момент, пока видимая область соответствует данным, полученным с датчиков. Затем, полученное изображение преобразуется согласно типу датчика. В случае простого инфракрасного датчика (Рис.2), изображение делится на две зоны и разница в сумме яркости (соответствует интенсивности теплового излучения) является результатом.

Для ультразвуковых и микроволновых датчиков, используются специальные материалы, яркость которых зависит от скорости объекта и прозрачности, которая обратно пропорциональна коэффициенту отражения объекта для волн данной частоты. Таким образом, эффект Доплера для датчика интерполируется.

Датчик пересечения луча реализуется путем создания виртуальной модели лучей в физический движок и обнаружение столкновения объектов с ними.



Рис. 2. Различные режимы рендеринга (1) – нормальная, (2) – инфракрасный свет.

3.3 Система датчиков

Существует два типа модулей в системе: манекен, имитирующий поведение человека и робот, управляемый с помощью РТК. Поведение манекенов может быть установлено как единый сценарий, или более комплексная индивидуальная модель поведения. В этом случае можно двигаться как вдоль указанной траектории так и существует возможность произвести расчет перемещения с учетом физической модели. Однако, во втором случае, описание некоторых действий ботов (например, перелезание через забор) является более сложным. Кроме того, неестественное поведение, например столкновение со статическими элементами, в испытательном полигоне, необходимо избегать.

Роботы, управляемые роботизированными комплексами, могут использовать камеры и вариативные датчики для обратной связи, при этом для удобства реализации их поведение находится на достаточно высоком уровне («Двигаться вперед со скоростью 1 метр в секунду», «повернуть на 45 градусов» и т.д.). При необходимости, возможно неточное выполнение команд либо непосредственным вводом ошибки, либо благодаря неоднородности поверхности, по которой робот перемещается (трение, перекося, неровности). Также в виртуальном испытательном полигоне присутствует упрощенная модель робота, движущегося по рельсам. Для них физическая модель не используется. Важно отметить, что робототехника – это возможность целиком возложить выполнение задачи на робототехнический комплекс.

4. Выводы

В текущей реализации воплощено большинство требований для специализированных виртуальных полигонов. Система позволяет тестировать программное обеспечение РТК без необходимости покупки или разработки аппаратных средств. Кроме того, виртуальный испытательный полигон позволяет анализировать различные сценарии в короткие сроки и оценивать эффективность РТК.

4.1 Сравнение

В сравнении с другими системами виртуальных полигонов, можно выделить следующие особенности:

1. Кросс-платформенное решение без перекомпиляции с помощью Java.
2. Распределенная визуализация на несколько камер.
3. Различные режимы визуализации и эмуляции датчиков.

4.2 Перспективы

Виртуальный испытательный полигон, рассмотренный выше, реализует готовый набор, но он может быть улучшен путем реализации следующих функций:

1. Коммуникация между роботами методом point-to-point.
2. Базовый ИИ «для чайников» чтобы избежать столкновений.
3. Виртуальные адаптеры для эмуляции конкретной камеры или конкретного датчика модели.

5. Литература

- [1] A. B. Degtyarev, Yu. V. Logvinenko and Yu. I. Nechaev, “Conception of design of multiagent systems for complex self-organization model of virtual testbed functioning”, in Information-measuring and Control Systems, Vol. 7, No. 2. (Publishing house "Radio engineering", Moscow, 2016), pp. 50–53.
- [2] L. A. Finalgin and D. M. Loginov, “Universal Virtual Poligon for Robotics”, in Virtual Modelling, Prototyping and Industrial Design – III International Scientific and Practical Conference Proceedings. (Tambovsk State Technical University, Tambov, 2016), pp. 393–396.
- [3] M. V. Mikhaylyuk and M. A. Torgashev, “Virtual Imitation and Training Poligon for Robotics”, in International Conference Resilience2014 of the International Center for Nuclear Safety of the Institute of Physical and Technical Informatics. (Autonomous Non-profit Organization “Institute of Physical and Technical Informatics”, Protvino, 2015), pp. 34–40.
- [4] Y. H. Tseng, S. C. Kang, Y. S. Su, C. H. Lee and J. R. Chang. “Strategies for Autonomous Robot to Inspect Pavement Distresses”, in IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 – Conference Proceedings. (Taipei, 2010), pp. 1199–1201.
- [5] E. V. Umnikov, “Virtual robotic test area from robot kinematics modeling systems to a virtualized environment managed simulate robots changing environment and natural processes”, in News of The Institute of Engineering Physics, Vol. 4, No. 42. (Interregional public institution "Institute of Engineering Physics", Serpukhov, 2016), pp. 50–53.

Геометро-графическая подготовка в рамках специалитета в русле всемирной инициативы CDIO

В.А. Семенов, С.Н. Абросимов, Д.Е. Тихонов-Бугров
semenov.1947@inbox.ru|abr-sn@yandex.ru|tbde@mail.ru

Балтийский государственный технический университет «Военмех», Санкт-Петербург, РФ

Аннотация: Рассматривается пример геометро-графической подготовки инженеров на базе САПР надувных конструкций. Показана концепция построения данной САПР, её место в учебном процессе. Среди достоинств отмечается междисциплинарность предложенного учебного процесса, ознакомление студентов с элементами инженеринговой деятельности, что обеспечивается обучением в русле всемирной инициативы CDIO.

Ключевые слова: геометро-графическая подготовка, САПР, надувные конструкции, CDIO.

Geometrical and graphic preparation within the framework of the speciality in the mainstream of the worldwide CDIO initiative

V.A. Semenov, S.N. Abrosimov, D.E. Tikhonov-Bugrov
Semenov.1947@inbox.ru|abr-sn@yandex.ru|tbde@mail.ru

Baltic State Technical University "Voenmeh", Saint-Petersburg, Russian Federation

Abstract : An example of geometrographic training of engineers based on CAD of inflatable structures is considered. The concept of building this CAD is presented, its place in the educational process is shown. Among the merits of such training is its interdisciplinarity and familiarization of the student with elements of engineering activity, which is provided by work in the mainstream of the worldwide CDIO initiative.

Keywords: geometrographic preparation, CAD, inflatable construction, CDIO.

В настоящее время, благодаря усилиям высокотехнологичных отраслей промышленности, РАН, ведущих вузов страны, приостановлен процесс сворачивания подготовки в рамках специалитета. Это обстоятельство позволяет организовать учебный процесс, располагая такими ресурсами, которых никогда не будет при обучении бакалавров.

Всемирная инициатива CDIO направлена на подготовку всесторонне образованных инженеров способных: планировать, проектировать, производить и применять сложные инженерные системы и процессы высокой добавленной стоимости [2]. Только в условиях подготовки специалистов (инженеров) возможно эффективно моделировать стадии: осмысления и планирования; проектирования; производства; применения.

Причём, данное моделирование в купе с интеграцией содержания разных дисциплин возможно в рассматриваемой ситуации уже на ранней стадии. Попутно отметим, что такая концепция обучения согласуется с известными трендами Д.В. Мантурова [7].

Базой для такого учебного процесса могут стать реальные САПР. В нашем случае такой базой является САПР надувных конструкций, разработанная авторами и реализуемая в довольно широком спектре номенклатуры изделий [9]. Рассмотрим подход к созданию данной САПР.

Надувные конструкции занимают достойное место в современной инженерной практике.

Из класса надувных конструкций можно выделить так называемые надувные каркасы, обеспечивающие форму изделия: надувные каркасы мобильных госпиталей, ангаров и т.д.; стартовые и финишные створы на трассах соревнований; баллоны надувных лодок и т.п. Примеры таких конструкций представлены на рис. 1–4.

Их проектирование является актуальной инженерной задачей, особенно с использованием систем автоматизированного проектирования.

В рамках учебного процесса разработка таких прикладных систем является хорошим примером подготовки современного технического специалиста.

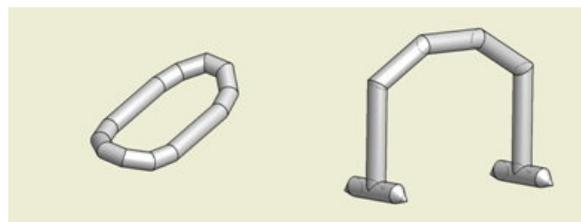


Рис. 1. Корпус рафта.

Рис. 2. Арка финишного створа.

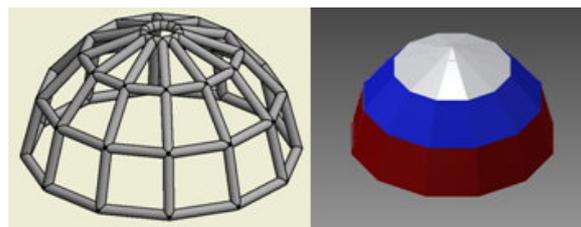


Рис. 3. Надувной каркас «полусфера».

Рис. 4. Каркас с тентом.

При проектировании надувных каркасов можно выделить следующие основные этапы:

- выбор конструкции каркаса и задание его характеристик;
- формирование трёхмерной модели изделия;
- исследование свойств модели;
- оптимизация параметров с возможным изменением структуры изделия;
- разработка технологических данных для автоматического раскроя материала и выпуск конструкторской документации.

Сама разработка таких приложений может использовать несколько подходов:

- создание библиотек трехмерных моделей и их фрагментов;

- разработка и использование шаблонов библиотек 2D и 3D элементов с таблицей параметров EXCEL;
- использование средств программирования, например C++, Python;
- разработка API – интерфейсов.

Приведённые на рисунках примеры изделий состоят из звеньев, каждое из которых – цилиндр или конус вращения, т.е. развёртываемая поверхность второго порядка. Эти поверхности целесообразно подбирать так, чтобы их сочетание подчинялось теореме Монжа. В этом случае сопрягаемые поверхности должны быть описаны вокруг одной сферы. Линия пересечения поверхностей – плоская кривая (эллипс). На плоскость симметрии линия пересечения проецируется в отрезок прямой. Плоскость симметрии – плоскость, заданная пересекающимися осями сопрягаемых поверхностей. Плоскость сечения, перпендикулярна плоскости симметрии.

Для задания исходных данных достаточно задать координаты центров сфер описываемых поверхностями и их радиусы. Если у изделия есть плоскости симметрии, исходные данные можно ограничить координатами сфер между этими плоскостями. Если радиусы сфер одного звена равны, то имеет место цилиндр вращения, если не равны – конус вращения. Эти данные удобно свести в таблицу, или несколько таблиц. Для примера рассмотрим исходные данные финишной арки приведённой на рис. 2.

Данные верхней части сведены в таблицу 1; данные нижней части – в таблицу 2. Графическая иллюстрация исходных данных приведена на рис. 5.

Точка	X	Y	Z	R
1	0	0	3600	200
2	950	0	3600	200
3	2500	0	2700	200
4	2500	0	0	200

Таблица 1. Табличные данные верхней части модели.

Точка	X	Y	Z	R
1	2500	1225	-150	0
2	2500	750	0	500
3	2500	0	0	500

Таблица 2. Табличные данные нижней части модели.

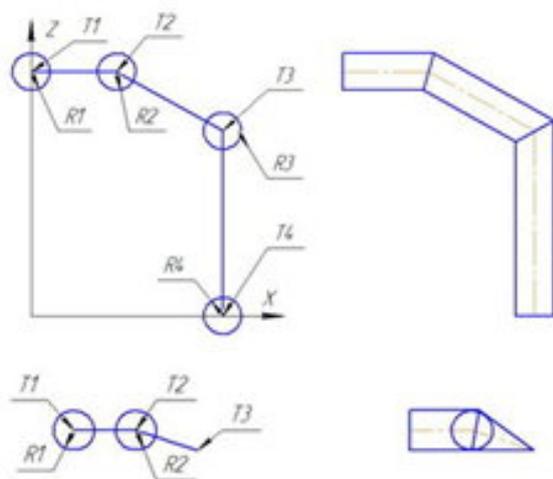


Рис. 5. Исходные данные для построения модели арки финишного створа.

При формировании трехмерной модели изделия исходные данные могут быть внесены в компьютер в виде электронных таблиц или непосредственно в виде параметров модели. На основании этих данных создаётся трёхмерная параметризованная модель изделия.

Для этого можно использовать практически любой пакет трёхмерного моделирования (например, Компас-3D, Inventor, SolidWorks и др.). С полученной трёхмерной моделью можно провести необходимые исследования:

расчёт снеговой нагрузки, ветроустойчивость, прочностные расчёты и т.д. Эти программные модули, практически представляют самостоятельные САПР.

При необходимости осуществляется корректировка исходных данных. Для этого в некоторых случаях достаточно изменить значения параметров в таблицах исходных данных и провести перестроение трёхмерной модели. Если этого недостаточно, то можно изменить структуру сооружения – создать новые таблицы исходных данных и новую трёхмерную модель. Результатом такого итерационного процесса будет трёхмерная модель, удовлетворяющая заданным требованиям.

Конечным этапом проектирования является выпуск конструкторской документации, включающий построение развёрток (выкроек) отдельных звеньев. Этот процесс также может быть автоматизирован. Поскольку отдельные звенья – развёртываемые поверхности, большинство графических пакетов позволяет получить развёртки непосредственно с трёхмерной модели. Для этого нужно выделить отдельное звено, обозначить место разреза и выдать команду на построение развёртки. При наличии раскrojного центра модель развёртки в большинстве случаев является исходным данным для её непосредственного изготовления.

Опыт построения надувных каркасов различных изделий позволяет провести некоторое объединение проектов или их элементов в отдельные группы, схожие по своей структуре. Например: для построения верхней части арки финишного створа рис. 5 и вертикальных стоек надувного каркаса полусферы рис. 3 строятся модели трёхзвенных конструкций. Как видно, по своей структуре они совпадают с исходными данными для верхней части финишного створа. Для их построения можно использовать одну и ту же модель с корректировкой исходных данных. Для рафта используется четырёхзвенная модель (при наличии двух вертикальных плоскостей достаточно построить модель четверти рафта). Тем не менее, наличие структурных отличий незначительно усложняет задачу. В таблицу 3 с исходными данными, кроме числовых изменений, добавляется одна строка дополнительных исходных данных (рис. 6).

Точка	X	Y	Z	R
1	0	880	0	300
2	1320	880	0	300
3	2000	600	140	300
4	2200	220	320	300
5	2200	0	320	300

Таблица 3. Табличные данные для построения модели рафта.

В работах [1, 3, 10] рассматривается возможность использования проектного обучения с элементами инжиниринга. Значительный интерес представляет опыт болгарских коллег [6].

Однако, обучение на ранней стадии с использованием реально практикуемой САПР, нам неизвестно.

При обучении на базе данной САПР имеется возможность интеграции следующих дисциплин: начертательной геометрии; инженерной и компьютерной графики; сопротивления материалов (прочностные расчёты с учётом свойств материалов и нагрузок, в том числе – ветровых);

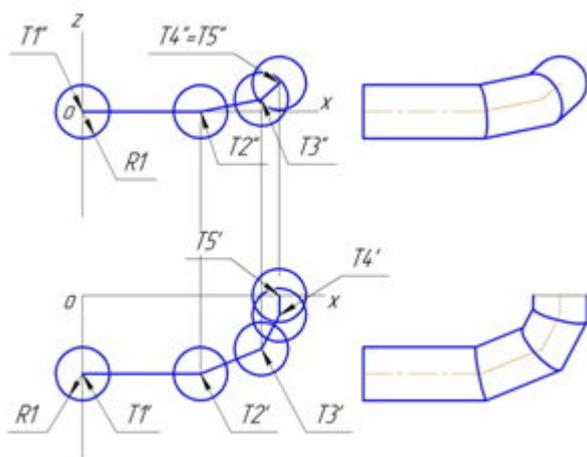


Рис.6. Исходные данные для построения модели рафта

материаловедения (подбор материалов, клеев с учётом процессов старения); основ автоматизированного проектирования; технологии (раскрой, склеивание, сварка).

Минимальный потребный временной ресурс – семестр.

Основными особенностями проектного подхода является более высокий уровень заинтересованности, комплексность задания и поэтому выделяет студентов, занимающихся по этой траектории по отношению к традиционному (предметному) подходу. Общая оценка за работу складывается из оценки вклада студента в реализацию задания и оценки защиты работы.

Данный проект, как и другие, является авторским и осуществляется востребованными специалистами в своей области, что требует больших трудозатрат. Сюда входит и написание методических пособий и разработка гибких педагогических сценариев.

При соответствующей договорённости с профилирующей кафедрой, возможен перенос части учебного процесса в рамки курса «Основы САПР».

Заключение

Важно отметить, что речь не идёт о формальном использовании готовой САПР. Моделируется среда, в которой студент изучает назначение и потребность в данной продукции, создаёт техническое задание, осмысливает геометрию изделия, алгоритмы проектирования, вносит свои коррективы, проводит выбор материалов, получает модель объекта, реализует продукт (чаще в виде маломасштабной модели или распечатки на 3D принтере, а в ряде случаев, привлекается к производству реального продукта), обсуждается процесс утилизации. Важным моментом учебного процесса является изучение соответствующих стандартов [4, 5, 8].

Литература

- [1] Бойков А.А. Компьютерные средства поддержки учебных курсов графических дисциплин //Геометрия и графика, 2013. Т.1. Вып.2.
- [2] Всемирная инициатива CDIO стандарты. Изд. ТПУ, 2011.
- [3] Вышнепольский В.И., Сальков Н.А. Цели и методы обучения графическим дисциплинам //Геометрия и графика. 2013. Т.1. Вып.2.
- [4] ГОСТ 34.003-90 «Информационная технология. Комплекс стандартов на автоматизированные системы. Термины и определения».

- [5] ГОСТ 23501.101-87 «Системы автоматизированного проектирования. Основные положения».
- [6] Лепаров М.Н., Попов М.Х. Состояние и тенденции геометро-графической подготовки как компоненты инженерного образования в Болгарии //Геометрия и графика. 2014. Т.2. Вып.1.
- [7] Мантуров Д.В. Инженеры становятся специалистами – универсалами. www.enginrussia.ru/Manturov-lektzia.pdf
- [8] РД 250-680-88 «Методические указания. Автоматизированные системы. Основные положения»
- [9] Семёнов В.А., Тихонов-Бугров Д.Е., Дюмин В.А. Геометро-графические элементы САПР надувных конструкций. Динамика систем, механизмов и машин. 2016 Т.4, №1.
- [10] Столбова И.Д. Актуальные проблемы графической подготовки студентов в технических вузах //Геометрия и графика. 2014. Т.2. Вып.1.

Об авторах

Семенов Виктор Алексеевич, доцент кафедры «Инженерная и машинная геометрия и графика». Балтийского государственного технического университета «Военмех», г. Санкт-Петербург. Его e-mail semenov.1947@inbox.ru

Абросимов Сергей Николаевич, к.т.н., профессор кафедры «Инженерная и машинная геометрия и графика». Балтийского государственного технического университета «Военмех», г. Санкт-Петербург. Его e-mail abr-sn@yandex.ru

Тихонов-Бугров Дмитрий Евгеньевич, к.т.н., профессор, зав. кафедрой «Инженерная и машинная геометрия и графика». Балтийского государственного технического университета «Военмех», г. Санкт-Петербург. Его e-mail tbde@mail.ru

Интегративная модель геометро-графической подготовки студентов

И.Д. Столбова, А.Б. Шахова

stolbova.irina@gmail.com|sab_perm@mail.ru

Пермский национальный исследовательский политехнический университет, г. Пермь, Россия

В статье рассматривается проблема разработки интегративной дисциплины «Начертательная геометрия, инженерная и компьютерная графика», реализуемой в рамках базовой геометро-графической подготовки студентов в Пермском национальном исследовательском политехническом университете. Обсуждаются предпосылки применения интегративного подхода в соответствии с инновациями в инженерном проектировании и современными требованиями к проектной культуре будущих выпускников. Показаны разнообразные проявления интегративного подхода при инновационной организации учебного процесса, создающие условия функционирования учебной среды для формирования у студентов профессиональных компетенций, соответствующих запросам реальной проектно-конструкторской деятельности.

Ключевые слова: интегративный подход, геометро-графическая подготовка, компьютерная графика, практико-направленная деятельность, профессиональные компетенции.

An integrative model of the geometro-graphic training of students

I.D. Stolbova, A.B. Shahova

stolbova.irina@gmail.com|sab_perm@mail.ru

Perm National Research Polytechnic University, Perm, Russia

The article considers the problem of developing an integrative discipline "descriptive geometry, engineering and computer graphics", implemented in the framework of the basic geometric and graphic training in Perm National Research Polytechnic University. Prerequisites for the application of an integrative approach in accordance with the innovations in engineering design and modern requirements to the project culture of future graduates are discussed. Diverse manifestations of an integrative approach to innovation of the educational process are given. They create conditions for the functioning of learning environments shape students' professional competence, relevant to the demands of the real design activities.

Keywords: integrative approach, geometro-graphic training, computer graphics, practice-oriented activity, professional competence.

1. Введение

Высокие темпы мирового научного и технического прогресса требуют ускоренного перевода экономики России на инновационный путь развития, концентрации усилий на наукоёмких высокотехнологичных отраслях. В силу этого важнейшей задачей высшего технического образования является подготовка конкурентоспособных специалистов, способных разрабатывать и осваивать новые наукоёмкие технологии, участвовать в инженерно-инновационной деятельности [16]. Концепция развития инженерного образования «будущего» предполагает развитие системного мышления обучающихся и формирование опыта решения профессионально ориентированных задач на основе применяемых в практической жизни компьютерных технологий [11]. В свою очередь, для создания условий перехода российской экономики к новому (пятому по общепринятой периодизации) технологическому укладу при реализации профессиональных образовательных программ технической направленности необходимо опираться на интеграцию различных составляющих содержания подготовки инженеров, обеспечивающую формирование у выпускников конструкторско-технологических компетенций [4].

Идеологической основой действующих образовательных стандартов высшего образования является компетентностный подход, методологические основы которого на сегодняшний день имеют обширные наработки. Компетенция — это интегральное качество индивидуума, выражающееся в способности осуществлять продуктивные действия при решении поставленной задачи. Профессиональная компетенция — это способность успешно действовать на основе практического опыта, умения и знаний при решении задач профессиональной

сферы. При формировании компетенций (организация учебного процесса) следует опираться на использование активных и интерактивных форм обучения студентов. Наличие компетенции (оценка успешности обучения) выражается в интегральном стиле мышления, в способности совмещать теорию с практикой, интегрировать знания в практическую деятельность [8]. Очевидно поэтому, что реализация компетентностно-ориентированной образовательной программы предполагает организацию интегративного обучения.

Стремительное развитие современного производства и автоматизированных цифровых технологий обостряет проблему подготовки специалистов инженерного профиля, способных при минимальных сроках профессиональной адаптации осуществлять свою деятельность на уровне последних достижений в области науки и техники. В этой связи очевидно, что проблемно-ориентированное и проектно-организованное обучение является одной из наиболее эффективных инновационных форм обновления учебного процесса [5].

Интегративный подход является максимально эффективным для формирования у выпускников профессиональных компетенций как целостных новообразований, затрагивающих когнитивную, деятельностьную и личностную сферы. Любая профессиональная задача по сути своей является интегральной, требует системного анализа, построения целостной модели ее решения. От специалиста востребованы комплексные действия, актуализация некоторой совокупности профессиональных компетенций и личностных качеств.

В связи с вышесказанным предметное обучение по дисциплинам общетехнического и профессионального циклов должно осуществляться в среде и формах, наиболее близких к реальной профессиональной деятельности,

происходить как своеобразное погружение в будущую сферу производственных интересов, что и будет способствовать эффективному овладению практическими навыками и связанным с ними теоретическим материалом. Построение учебного процесса на интегративной основе содержит значительно больше технологических возможностей, и именно интегративный подход должен стать определяющим при разработке содержания образовательных программ и выборе форм проведения учебной деятельности.

2. Инновации в инженерном проектировании

В настоящее время цифровые технологии 3d-печати, 3d-сканирования, 3d-моделирования и 3d-визуализации становятся основой высокотехнологичного производства. Они позволяют сократить производственные затраты и сроки проектирования, обойти технологические ограничения и повысить качество и конкурентоспособность выпускаемой продукции. Информационные технологии принципиально изменили проектно-конструкторскую деятельность, они внесли значительные коррективы в разработку проектно-конструкторской документации, включая ее создание и контроль.

Динамичное развитие проектирования как вида инженерной деятельности объясняется совершенствованием возможностей современных компьютерных технологий и САПР, обновлением функционала инженера-проектировщика, возросшими требованиями к проектной культуре специалиста, системному подходу к созданию и инженерно-техническому обеспечению всех этапов жизненного цикла технических объектов [3]. На всех стадиях жизненного цикла изделий присутствуют информационные модели, в число которых входят 3D геометрические электронные модели, представляющие набор данных, однозначно определяющих форму, структуру и размеры изделия. Наличие 3D-модели значительно повышает производительность и качество процесса проектирования, его вариативность и наглядность.

Среди моделей, используемых в конструкторских САПР в настоящее время, можно выделить модели пространственных объектов и модели чертежей. Более эффективные модели объектов (твердотельные) используются в современных системах трехмерного проектирования.

Огромным преимуществом современных САПР является возможность параметризации создаваемых объектов. Наличие условий для создания параметрической модели заложено в возможности системы проектирования, а параметрическое описание объекта становится идеологической базой современных конструкторских разработок. Данный подход, кроме рационализации процесса создания объекта и использования его в дальнейшем в качестве прототипа, позволяет проектировщику за короткое время производить очень большое количество итераций при поиске оптимальной формы модели, а с применением современных расчетных модулей также получить и оптимальные значения таких физических (и не только) параметров как вес, прочность, обтекаемость, гибкость, долговечность, экологичность, эргономичность и др. [7].

Учитывая мировые тенденции в сфере проектной деятельности и исходя из вышесказанного, можно констатировать, что обучение сегодняшних студентов инженерного профиля методам моделирования, основанных на параметризации, просто необходимо для повышения уровня конкурентоспособности будущих выпускников.

Все перечисленное говорит о необходимости разработки новых технологий обучения, которые должны быть практико-ориентированными и проектно-направленными, организовывать учебную среду, приближенную к реальной деятельности конструктора-проектанта, и создавать условия для формирования у студентов востребованных компетенций проектно-конструкторской деятельности [8].

3. О развитии проектной культуры в ходе графической подготовки

Несомненно, что развитие проектной культуры у студентов необходимо начинать при изучении дисциплины «Начертательная геометрия, инженерная и компьютерная графика», которая относится к профессиональному циклу, но осваивается студентами уже на первом-втором курсах учебы в вузе. Поэтому очевидна сложность в организации прикладного проектно-конструкторского обучения графическим дисциплинам [9], и в силу этого требуется применение адекватных педагогических технологий для продуктивности учебного процесса.

К эффективным технологиям продуктивного обучения для дисциплин профессионального цикла относят *деятельностные* технологии обучения [6]:

- методы, основанные на моделировании профессиональной деятельности в учебном процессе;
- развивающие, проблемные формы и методы, интенсифицирующие процесс обучения;
- технологии, обеспечивающие способность личности к саморазвитию и самосовершенствованию за счет осознанных и самостоятельных действий.

Для геометро-графического обучения студентов в рамках базовой инженерной подготовки на младших курсах вуза особенно актуальны вопросы внедрения метода проектов. Во-первых, сам профиль такой подготовки предполагает практико-направленную профессиональную деятельность, Во-вторых, использование этого метода позволит приблизить образовательные технологии к профессиональной деятельности проектировщика, которая в последнее время претерпела значительные инновационные изменения. В этой связи необходима разработка проектно-ориентированных проблемных заданий профессиональной направленности, а также их эффективная и последовательная реализация в рамках учебного процесса [2].

Еще одной сложностью в организации проектного обучения при массовом переходе на бакалавриат является ограничение выделяемого в учебном плане времени на графическую подготовку. Отсюда и возникает интерес к интегративным приемам и методам, интенсифицирующим процесс обучения графическим дисциплинам [9, 10].

Особенно значимый эффект в интенсификации обучения дисциплине «Начертательная геометрия, инженерная и компьютерная графика» можно достичь путем интеграции двух ее базовых разделов — инженерной и компьютерной графики, объединив их в единый курс — геометрическое моделирование [1]. Такая интеграция позволяет также частично решать проблему внедрения элементов «метода проектов» на всех этапах обучения, приближая учебные технологии к профессиональным условиям и способствуя усилению практической направленности графической подготовки студентов.

4. Формы интеграции в рамках графической подготовки

Традиционно вузовская базовая геометро-графическая подготовка была представлена в учебных планах технических специальностей дисциплинами (разделами) «Начертательная геометрия» (теории построения проекционного чертежа), «Инженерная графика» (основы разработки конструкторской документации) и «Компьютерная графика» (инструментальная подготовка для автоматизации выпуска конструкторской документации). Они закладывали основы проектной культуры будущих выпускников и изучались последовательно. С развитием компьютерных технологий, в том числе компьютерной графики, были запущены интеграционные процессы взаимопроникновения разделов графической подготовки, что особенно ускорило с распространением технологий проектирования в формате 3d [15]. Возможности 3d-моделирования многократно повышают эффективность учебного процесса, позволяют создавать новые механизмы для понимания сложной к усвоению графической информации, развивают пространственное мышление у обучаемых и учитывают индивидуальные особенности студентов.

С переходом на ФГОС 3-го поколения в ПНИПУ базовая графическая подготовка студентов осуществляется на общеобразовательной кафедре «Дизайн, графика и начертательная геометрия» в рамках интегративного курса «Начертательная геометрия, инженерная и компьютерная графика», при разработке которого можно выделить различные вариации применения интегративного подхода.

1. Интеграция требований к результатам геометро-графического обучения на основании требований различных стандартов по направлениям подготовки в области техники и технологий.

В качестве общеобразовательной кафедра участвует в реализации около 80 образовательных программ очной формы обучения и 60 – заочной. При переходе на стандарты 3 поколения встал вопрос о необходимости унификации компетентностно-ориентированного обучения графическим дисциплинам и, как следствие, разработки уровневых программ освоения предметной области. Интегративность в решении этого вопроса позволила оптимизировать процесс модернизации требуемых программ предметного обучения и при этом обеспечить формирование предметных компетенций на таком уровне, который гарантировал бы достижение компетентностной модели выпускника по каждой из образовательных программ в соответствии с ФГОС ВО [13].

2. Интеграция отдельных разделов дисциплины при организации различных видов учебной деятельности.

Интегративная модель графической подготовки приведена на рис. 1. При разработке инновационного курса графической подготовки, компенсирующего при переходе на бакалавриат снижение как общего количества часов предметного обучения, так и часов, отведенных на теоретическую подготовку, особое внимание было уделено освоению практических приемов использования компьютерной графики. В программу обучения введен лабораторный практикум, где рассматриваются вопросы не только автоматизированного выпуска конструкторской документации (синтез инженерной и компьютерной графики), но и решаются практико-ориентированные задачи геометрического моделирования, сочетающие теоретические основы начертательной геометрии и современного инструментария САПР. Практикум по геометрическому моделированию содержит задачи – современные аналоги метрических и позиционных задач – различного уровня сложности, развивающие

пространственное воображение обучаемых, демонстрирующие необходимость освоения основ начертательной геометрии и одновременно позволяющие совершенствовать навыки использования инструментария 3d моделирования [1].



Рис. 1. Интегративная модель курса «Начертательная геометрия (НГ), инженерная (ИГ) и компьютерная графика (КГ)».

3. Концентрация учебного материала на основе использования статической и динамической визуализации подготовленных 3d-моделей.

Техническая графическая информация связана с представлением данных о пространственных объектах, важными ее характеристиками являются точность и наглядность. А одной из главных и в то же время сложных задач графического образования считается развитие пространственного мышления у обучаемых. Возможности компьютерного трехмерного геометрического моделирования качественно изменили уровень представления учебного материала. Мультимедийная технология обучения с помощью программных и технических средств дает уникальную возможность имитировать образы реальных объектов, явлений и процессов. Это позволяет при обучении подходить к абстрактным понятиям, не только следуя логике разума, но также через чувства и ощущения, которые возникают при рассмотрении виртуальных моделей на экране компьютера. Такой подход способствует целостному восприятию материала, увеличению скорости и качества его усвоения [10].

4. Разработка комплексного методического сопровождения учебного процесса, охватывающего различные виды учебной деятельности.

Важным фактором успешности применяемых образовательных технологий является методически проработанное сопровождение образовательного процесса, технологичность и ясность воспроизведения запланированных учебных заданий. Современный интегративный учебно-методический комплекс дисциплины (УМКД) создает эффективную среду обучения, оптимально организует учебные действия студентов, задает алгоритм для самостоятельного выполнения проектных заданий, обеспечивает формирование всей совокупности инженерно-графических компетенций на требуемом уровне и с учетом

индивидуальных особенностей отдельных студентов [10, 12]. Электронная форма УМКД позволяет эффективно пользоваться учебным материалом. Так, например, разработанные мультимедийные лекции применяются в двух вариантах: как современное дидактическое средство на лекционных занятиях при разборе нового материала и как учебное пособие для самостоятельного изучения или повторения разделов дисциплины студентами всех форм обучения.

Как было сказано выше, важнейшей задачей организации современного учебного процесса является создание условий для практико-направленной и проектно-организованной деятельности обучаемых. Для реализации проектно-конструкторской функции учебной среды, необходимо максимальное приближение ее к реальной обстановке профессиональной деятельности конструктора. Учебные задания проектной направленности должны предусматривать использование алгоритмов геометрического моделирования при создании твердотельных моделей как структурных составляющих технического изделия, так и сборочной единицы в целом; разработку ассоциативных чертежей отдельных деталей и цельной конструкции; сборку и получение реального макета объекта с использованием 3d принтера. При выполнении таких заданий студент должен активно использовать дополнительную справочную информацию электронных ресурсов Интернета или специально создаваемых библиотек стандартных изделий и параметризованных деталей.

Результатом выполнения проектной работы является, как правило, полный комплект электронной документации, содержащий достаточные сведения для изготовления и контроля проектируемого изделия. Необходимо дополнительно предусматривать и задачи исследовательского типа разного уровня сложности с анализом проекционных, конструкторских, технологических ситуаций. Такие задания по своей направленности и практической значимости приближаются к реальным производственным задачам, а компьютерные технологии позволяют интенсифицировать практический этап обучения, укладываясь в рамки выделяемых учебной программой часов.

Интегративные формы учебной деятельности позволяют также создавать малые студенческие конструкторские группы для коллективной работы над проектным заданием, совершенствуют навыки принятия коллективных решений, развивают коммуникативные приемы поведения и комплексно формируют профессиональные компетенции у студентов.

На рис. 2 приведен пример задания на проектирование реальной модели изделия. Данное задание относится к завершающему этапу базовой графической подготовки студентов. Работа над заданием демонстрирует достигнутый за время графической подготовки уровень сформированности у студентов проектно-конструкторских компетенций как совокупности освоенных знаний, практических умений, навыков и личностных качеств.

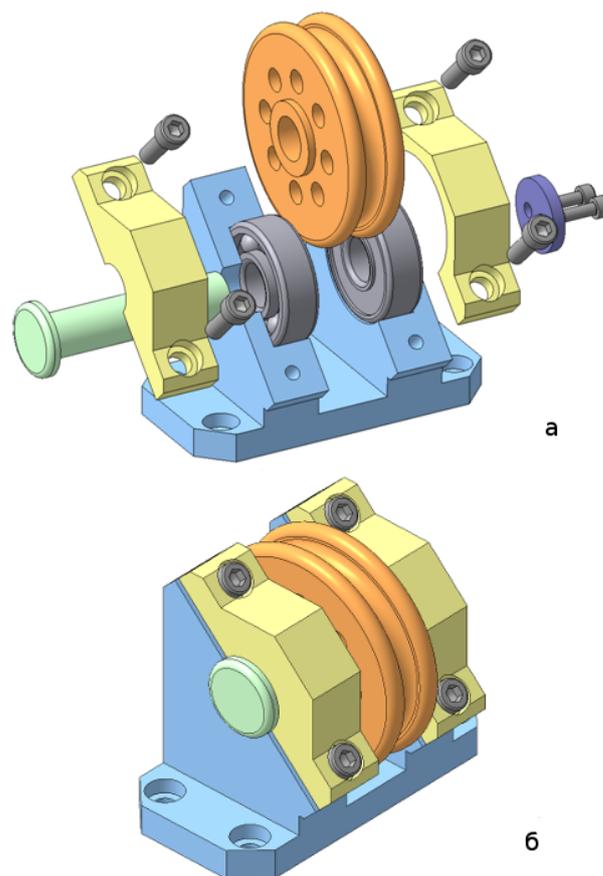


Рис. 2. 3D-модель блока роликового: а – модель в разнесенном виде; б – модель в собранном виде.

5. Заключение

Для инновационного развития экономики России необходима подготовка кадров новой формации, способных осваивать передовые наукоемкие технологии, а также создавать и выпускать конкурентоспособную отечественную продукцию. Актуальной является задача формирования готовности выпускников к проектно-конструкторской деятельности, соответствующей запросам сегодняшнего дня.

Авторами предложена инновационная модель интегративной базовой графической подготовки студентов технического университета, закладывающей основы современной проектной культуры. Интегративные процессы позволяют разработать образовательную среду, создающую условия для формирования у обучаемых профессиональных компетенций как целостных новообразований, затрагивающих когнитивную, деятельностную и личностную сферы.

Оптимизирует образовательный процесс и дает качественный результат интеграция отдельных разделов графической подготовки: начертательной геометрии и компьютерной графики; инженерной и компьютерной графики. Комплексные задания — учебные проекты создают имитацию реальной проектно-конструкторской деятельности и стимулируют студентов к овладению технологиями современного геометрического моделирования.

6. Литература

- [1] Александрова Е.П., Носов К.Г., Столбова И.Д. Геометрическое моделирование как инструмент повышения качества графической подготовки студентов // Открытое образование. 2014. № 5 (106). С. 20-27.
- [2] Александрова Е.П., Носов К.Г., Столбова И.Д. Практическая реализация проектно-ориентированной деятельности студентов в ходе графической подготовки // Открытое образование. 2015. № 5. С. 55-62.
- [3] Амирджанова И.Ю., Виткалов В.Г. Современное состояние развития геометро-графической культуры и компетентности будущих специалистов // Вектор науки ТГУ. 2015. No 2 (32-2). С. 26-31.
- [4] Гилязова С.Р., Старшинова Т.А. Интегративный подход в преподавании графических дисциплин // Высшее образование в России. 2013. № 1. С.99-104.
- [5] Дульзон А.А. Опыт проблемно-ориентированного и проектно-организованного обучения // Высшее образование в России. 2010. No 10. С. 42—48.
- [6] Дюмин В.А., Тихонов-Бугров Д.Е. Хорошо забытое старое или проектно-конструкторское обучение инженерной графике // V Международная интернет-конференция КПП-2015. URL: <http://dngn.pstu.ru/conf2015/papers/18/> (дата обращения: 1.06.2017).
- [7] Иванов А.Д., Ямникова О.А. Применение параметризации при разработке конструкторской документации // Известия ТулГУ. Технические науки. 2014. №11-2. С.524-528.
- [8] Матушкин Н.Н., Столбова И.Д. Роль междисциплинарного компонента образовательных программ, реализующих компетентностную парадигму // Инновации в образовании. 2010. № 11. С. 5-17.
- [9] Минин М.Г., Захарова А.А., Сафьянников И.А., Вехтер Е.В. Организация процесса подготовки бакалавров техники и технологии к проектно-конструкторской деятельности // Высшее образование в России. 2013. № 5. С. 106-113.
- [10] Михелькевич В.Н., Москалева Т.С., Пузанкова А.Б. Инженерно-графическая подготовка студентов на базе электронного учебно-методического комплекса // Вектор науки ТГУ. 2014. №3. С. 314-317.
- [11] Петрунева Р.М., Топоркова О.В., Васильева В.Д. Учебное инженерное проектирование в структуре подготовки студентов технического вуза // Высшее образование в России. 2015. №7. С. 30-36. Автор В. Источник.
- [12] Столбова И.Д., Александрова Е.П., Крайнова М.Н., Кочурова Л.В. О создании учебно-методического комплекса для сопровождения графической подготовки студентов // Геометрия и графика. 2015. Т. 3. № 2. С. 29-37. DOI: 10.12737/1216.
- [13] Столбова И.Д., Александрова Е.П., Крайнова М.Н. Позиции интегративности при технологизации предметной подготовки // Геометрия и графика. 2014. Т. 2. № 2. С. 21-26.
- [14] Столбова И.Д., Александрова Е.П., Носов К.Г. Метод проектов в организации графической подготовки // Высшее образование в России. 2015. № 8-9. С. 22-31.
- [15] Тельной В.И. Новые подходы к изучению дисциплины «Инженерная графика» с использованием современных информационных технологий // Вестник МГСУ. 2013. №8. С.168-176.
- [16] Шитов С.Б. Подготовка креативных компетентных специалистов-исследователей в обществе знания // Высшее образование сегодня. 2015. № 8. С. 22-25.

Об авторах

Столбова Ирина Дмитриевна, д.т.н., зав. кафедры «Дизайн, графика и начертательная геометрия» Пермского национального исследовательского университета. E-mail: stolbova.irina@gmail.com.

Шахова Алевтина Бруновна, доцент кафедры «Дизайн, графика и начертательная геометрия» Пермского национального исследовательского университета. E-mail: sab_perm@mail.ru.

Авторский указатель

A

Adzhiev V. 79
Alexandrov V. 17
Arseev S. 151
Averbukh V.L. 69, 71

B

Bakhterev M.O. 69
Barkov F.L. 74
Benois-Pineau J. 237
Bogdanov N.N. 37, 42
Burdin V.V. 74

C

Catheline G. 237
Chayka Z.S. 69

D

Denisov E.U. 37

E

Espinosa J. 17

F

Forghani M. 71
Frolov V.F. 28

G

Galaktionov V.A. 28
Garanin A.I. 74
Gonakhchyan V.I. 32

K

Khvostikov A. 237
Kishalov A.A. 42
Klimenko S.V. 18
Konin Yu.A. 74
Konstantinov Yu.A. 74
Konushin A.S. 151, 154, 158, 175
Konushin V. 154
Krylov A. 237
Kukleva A. 154

L

Latkin K.P. 74

M

Malikova E. 79
Malyshev A. 319, 323
Morozov F. 158

P

Pasko A. 79
Pervadchuk V.P. 74
Pilyugin V. 79
Poskrebyshchev M.M. 74
Potemin I.S. 37, 42

R

Rudomin I. 17

S

Sangarov V.S. 28
Shalnov E.V. 175
Sidorov K.A. 21, 162
Slyadnev S. 169, 319, 323
Smirnov A.S. 74
Sokolov V.G. 37

T

Turlapov V. 169, 319, 323

V

Valuiskaia I.A. 175
Vasev P.A. 69, 71
Vladimirova D.B. 74

Z

Zhdanov A.D. 37, 42
Zhdanov D.D. 37, 42

A

Абросимов С.Н. 370
Авербух В.Л. 87
Авербух Н.В. 87
Андреев С.В. 92
Аноприенко А.Я. 290
Аношина Н.А. 243
Аристов Г.В. 180

B

Багутдинов Р.А. 97
Бакенов А.С. 328
Балюба И.Г. 332
Баранов Д.А. 138
Барладян Б.Х. 47
Белобородов Д.С. 184
Белокаменская А.А. 275
Белоусов К.И. 138
Бехтольд О.В. 101
Бобков В. 189
Бойков А.А. 335
Бондарев А.Е. 92, 103
Бондаренко А.В. 92
Боревич Е.В. 51
Брылкин Ю.В. 295, 339
Будак В. 55

B

Варшанина Т.П. 298
Василюк Н.Н. 342
Васин Д.Ю. 20, 303
Васин Ю.Г. 212
Вехтер Е.В. 23, 116

Визильтер Ю.В. 92
Волобой А.Г. 22, 47, 228
Волошинов Д.В. 346
Выголов О.В. 125

G

Гайнияров И. 193
Галактионов В.А. 22, 47, 92, 103
Гергет О.М. 258
Гнатюк В.С. 269
Голубятников Т.В. 248
Горбацевич В.С. 125
Горбачев В.Н. 108
Горевой А.В. 197
Гудков А.В. 92
Гудков В.Ю. 101
Гурьянов Ф.А. 253

D

Данилов В.В. 258
Дебелов В.А. 58
Долотов Е.А. 202
Дудаков Н.Ю. 308
Дураков А.В. 25

E

Евсютин О.О. 264
Ерёмченко Е.Н. 19, 290

J

Жданов Д.Д. 22
Желтов В. 55
Желтов С.Ю. 92
Жилина Н.Д. 350
Жуков В.Т. 92

Z

Завьялов Д.А. 112
Захарова А.А. 23, 97, 112, 116
Знаменская И.А. 121
Зуева Г.А. 147

I

Иловайская Е.Б. 92
Исайчев С.А. 121

K

Казначеева Е.С. 346
Кайнарова Е.М. 108
Клевцов Д.Г. 125
Клестов Р.А. 208
Клеценков А.Б. 223
Клименко С.В. 290
Клиорин Н.И. 248
Клюев А.В. 180
Князь В.А. 92

Князь В.В. 47, 125
Коваленко Д.А. 269
Ковернинский И.В. 47
Конопацкий Е.В. 332
Конушин А.С. 217
Коротеева Е.Ю. 121
Короткий В.А. 355
Крылов А.С. 253, 279
Кудряшов А. 189, 312
Кузьмичева Е.А. 121
Кустикова В.Д. 202
Кушнер К.Г. 58

Л

Лагунова М.В. 350
Лачинов Д.А. 275
Лебедев Л.И. 212
Липкина А.Л. 359
Ломов Н.А. 359
Люттов В.С. 217

М

Макаров Н.Л. 20
Манаков Д.В. 129
Мануковский К.В. 92
Мачихин А.С. 197
Мельникова А.А. 279
Мерзляков И.Н. 308
Местецкий Л.М. 184, 359
Метелёв И.К. 108
Мецерыков Р.В. 264
Мецерыков С.В. 51
Мизгинов В.А. 125
Михайличенко А.А. 223
Михайлов И.О. 367
Мошкова Т.В. 20, 350

Н

Небаба С.Г. 97
Новикова Н.Д. 92
Новоторцев Л.В. 228

О

Обабков И. 193
Осипов М.П. 212
Ососков М.В. 92

П

Павлов И.Н. 133
Полякова М.А. 108
Пономарев Н.С. 315
Попов Е.В. 20
Поринев С.В. 248
Провков В.С. 367

Р

Разумков А.В. 26
Расковская И.Л. 133
Ращупкина А.В. 264
Редькин М.А. 303
Ринкевичюс Б.С. 133
Роменский С.А. 20, 350
Ротков С.И. 20, 350
Рябинин К.В. 138, 142

С

Сагиндыков Т.Б. 243
Семенов В.А. 370
Силаев Н.Ж. 92
Скирневский И.П. 258
Смышчёр М.М. 350
Соловьёв И. 312
Солоделов Ю.А. 47
Сорокин Д.В. 24, 243, 284
Стародубцев И.С. 147, 367
Столбов В.Ю. 208
Столбова И.Д. 373

Т

Тарасенко А.С. 315
Тихонов-Бугров Д.Е. 370
Толкачев А.В. 133
Томчинская Т.Н. 308

Турлапов В.Е. 275
Тюрина В.А. 20, 350

У

Утешева Т.Ш. 315

Ф

Федоренко В.В. 125
Федотов А.М. 335
Феодоритова О.Б. 92
Фролов В.А. 47, 63

Х

Хайбарахманова Е.С. 346
Ханнанов А.Р. 308
Ханыков И.Г. 232
Харинов М.В. 232
Хахалин А.В. 121
Хлебников Н. 193

Ч

Чембаев В. 55
Чепкасов В.Л. 20
Черноризов А.М. 121
Чуприна С.И. 142

Ш

Шатино Л.З. 47, 103
Шахова А.Б. 373
Шишаков В.В. 121
Шкляр А.В. 23, 112, 116

Щ

Щербаков А.С. 63

Я

Янчус В.Э. 51

Научное издание

GraphiCon 2017

27-я Международная конференция
по компьютерной графике
и машинному зрению

Труды конференции

Издается в авторской редакции
Компьютерная вёрстка: *К.В. Рябинин, Д.Н. Мухутдинова, И.А. Горяева*

Подписано в печать 02.09.2017. Формат 60×84/8.
Усл. печ. л. 44,18. Тираж 220 экз. Заказ ____

Издательский центр
Пермского государственного национального исследовательского университета
614990, г. Пермь, ул. Букирева, 15

Кунгурская Типография, г. Кунгур