Метод построения цифрового аналога реального окружения с использованием нейронных сетей

М.И. Сорокин [1], Д.Д. Жданов [1], А.Д. Жданов [1]

[1] Университет ИТМО, Кронверкский проспект, дом. 49, литер А, Санкт-Петербург, 197101, Россия

Аннотация

Проблема формирования естественных условий освещения виртуальных объектов и взаимодействия виртуальных объектов с реальным миром еще полностью не решена. Целью данного исследования является исследование возможности устранения причин конфликтов зрительного восприятия систем смешанной реальности за счет замены реальных объектов их виртуальными аналогами. Основная идея – это перевод всех объектов реального мира в виртуальное окружение. Это решение переводит систему смешанной реальности в систему виртуальной реальности и обеспечивает корректное физическое и световое взаимодействие объектов виртуального и реального миров. Такое решение открывает новые возможности для систем смешанной реальности, например, в темном помещении можно добавить виртуальный источник света и осветить помещение, или появляется возможность увидеть отражение виртуальных объектов в реальных зеркалах. Для построения виртуальных аналогов объектов реального мира используются средства машинного обучения и нейросетевые технологии, которые успешно применяются в задачах компьютерного зрения для решения проблем ориентирования в пространстве и реконструкции окружения. Для создания цифровой копии реального окружения в данной работе представлен алгоритм с использованием RGB-D датчиков и Z-буфера для получения облака точек. Данный подход позволяет не только выделять предметы интерьера из сцены, но и определять их положение и размеры. Для аннотирования и сегментирования облака точек была использована архитектура нейронной сети «PointNet», обученная на наборе данных ScanNet. Для построения треугольной сетки использовалась нейронная сеть «Total3D understanding». В результате в данной работе была реализована система реконструкции реального помещения, которая использует в качестве входных данных RGB фотографии и облака точек. Приводится пример реконструкции простого интерьера помещения и оценка качества полученной реконструкции.

Ключевые слова

Нейронные сети, смешанная реальность, 3D сканирование, реконструкция окружения.

Method for Constructing a Digital Analogue of a Real World Environment Using Neural Networks

M.I. Sorokin [1], D.D. Zhdanov [1], A.D. Zhdanov [1]

[1] ITMO University, Kronverksky Pr. 49, bldg. A, St. Petersburg, 197101, Russia

Abstract

The problem of forming natural lighting conditions for virtual objects and interaction between real and virtual objects is not yet solved. The main goal of this study is to investigate the possibility of eliminating the causes of the mixed reality visual perception conflicts by replacing real-world objects with their virtual counterparts. The main idea is to transfer all of the real-world objects into a virtual environment. This solution transforms the mixed reality system into a virtual reality system and ensures the correct physical and light interaction between objects of different worlds. This will

ГрафиКон 2022: 32-я Международная конференция по компьютерной графике и машинному зрению, 19-22 сентября 2022 г., Рязанский государственный радиотехнический университет им. В.Ф. Уткина, Рязань, Россия
EMAIL: vergotten@gmail.com (М.И. Сорокин); ddzhdanov@mail.ru (Д.Д. Жданов); andrew.gtx@gmail.com (А.Д. Жданов)
ORCID: 0000-0001-9093-1690 (М.И. Сорокин); 0000-0001-7346-8155 (Д.Д. Жданов); 0000-0002-2569-1982 (А.Д. Жданов)

© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

open up new possibilities for mixed reality systems, e.g., illuminating a dark room with a virtual light source, or seeing the reflection of virtual objects in a real mirror. This paper presents an algorithm that uses machine learning tools and neural network technologies along with RGB-D sensors and a Z-buffer to obtain a real-world point cloud. This approach allows not only to select scene items but also to determine their position and size. The PointNet network architecture trained on the ScanNet dataset was used to annotate and segment scene data. The "Total3D understanding" network was used to build a triangular grid. Finally, a real-world room reconstruction system was implemented using RGB images and point clouds as input parameters. An example of the reconstruction of a simple room interior and reconstruction quality assessment is presented.

Keywords
Neural networks, mixed reality systems, 3D scanning, environment reconstruction.

1. Введение

Несмотря на то, что с точки зрения технической реализации смешанная реальность может быть похожа на виртуальную и дополненную реальности, она имеет ряд принципиальных отличий. В виртуальной реальности человек полностью перемещается в виртуальный мир и у него полностью теряется представление о реальном мире, он не может корректно взаимодействовать с объектами реального мира. В системах дополненной реальности человек остается в реальном мире, а виртуальный мир — это информационное пространство в виде текста, указателей, схем, видеоконтента и т. п. То есть виртуальный мир не воспринимается в виде части реального мира и не вызывает дискомфорта зрительного восприятия.

Другое дело системы смешанной реальности. В данных системах виртуальный мир должен восприниматься естественным образом как часть реального мира. Это условие вызывает ряд проблем, связанных не только с тем, как физически корректно отобразить объекты виртуального мира, но то, как их корректно вписать в реальный мир. В зависимости от типа систем смешанной реальности визуальное взаимодействие объектов реального и виртуального миров выполняется различными способами.

В оптически прозрачных устройствах смешанной реальности человек воспринимает реальный мир через прозрачную оптику, которая одновременно является зеркалом (голографическим, интерференционным, микрорастровым и т. п.), проецирующим изображение виртуального мира в глаз человека. При этом виртуальный объект будет прозрачен и виден только за счет повышения яркости его изображения по отношению к остальным объектам реального мира (как правило, его яркость превышает яркость объектов реального мира на один порядок). Взаимодействие объектов реального и виртуального миров ограничено возможностью визуального экранирования виртуальных объектов, находящихся позади реальных. Обратное экранирование объектов реального мира виртуальными объектами, а также корректное освещение виртуальных объектов источниками света реального мира и формирование теней от виртуальных невозможно для данного типа устройств смешанной реальности.

В видео прозрачных системах смешанной реальности реальный мир транслируется с видеокамер на LCD дисплей, который является источником изображения виртуального мира. Поскольку реальный и виртуальный миры оказываются на одном пространстве LCD дисплея, то область их возможного взаимодействия становится значительно шире. Виртуальный объект может затенять реальный мир, может быть корректно освещен источниками света реального мира и отбрасывать тени на объекты реального мира. Однако программная реализация сложного физически корректного светового взаимодействия объектов виртуального и реального миров является технически сложной задачей. Например, практически невозможно реализовать корректное вторичное освещение реальных объектов объектами виртуального мира и наоборот, или обеспечить корректное отражение виртуальных объектов в реальных зеркалах.

Основная причина дискомфорта систем смешанной реальности заключается в том, что объекты виртуального и реального миров находятся в разных пространствах, используют разные средства обработки и отображения информации. Этот дискомфорт часто ведет к появлению головной боли, тошноты и даже увечьям [1]. Поэтому устранение этого конфликта является важной задачей. Одним из решений, позволяющим снизить дискомфорт зрительного

восприятия, вызванный неестественным световым взаимодействием объектов реального и виртуального миров, это перевод всех объектов в единое пространство. Поскольку перевод виртуальных объектов в реальный мир невозможен, то предлагается перенести объекты реального мира в виртуальный.

Для переноса объектов реального мира в виртуальный мир необходимо добиться максимальной точности реконструкции реального окружения, извлекать как можно больше данных о среде для построения виртуального аналога реального мира. Например, использовать сканеры глубины и современные технологии в сфере восстановления геометрии объектов. К камерам глубины можно отнести RGB-D камеры, которые позволяют получить не только RGB изображение, но и расстояние от камеры до объектов съемки. RGB-D камеры находят широкое применение в компьютерной графике, системах компьютерного зрения, робототехнике, системах виртуальной, дополненной и смешанной реальностей [2]. Глубина изображения, как правило, хранится в Z-буфере изображения. При визуализации глубина изображения варьируется по яркости, чем дальше точка объекта от камеры, тем более светлая она на экране. Пример RGB изображения сцены и ее Z-буфера представлен на рисунке 1.

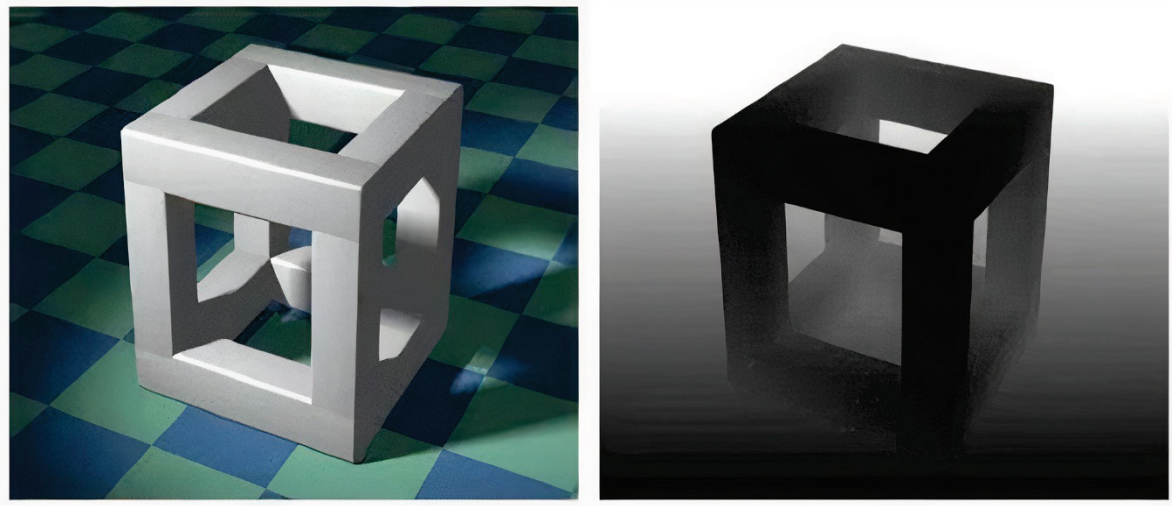

Рисунок 1 – Пример RGB-D изображения сцены. а) RGB изображение исходной сцены. б) Z-буфер исходной сцены

Как было отмечено выше, основой для построения виртуального аналога реального мира может служить его RGB-D изображение. Работа с RGB-D камерами и обработка результатов ее работы для получения информации о реальном мире использовалась в большом количестве приложений. Изначально это была игровая индустрия [3]. Со временем применение RGB-D камер распространилось на другие сферы, такие как робототехника [4], сельское хозяйство [5], автономное вождение [6], распознавание действий человека [7], распознавание объектов [8] и реконструкция 3D-сцены [9, 10]. Применение RGB-D в различных прикладных задачах позволило накопить большо опыт по реконструкции моделей сцен. Что может быть очень полезно при создании виртуальных аналогов реальных трехмерных объектов сцен. Необходимо различать реконструкцию статических и динамических объектов сцен. Несмотря на то, что есть много наработок по восстановлению динамичных и модифицируемых объектов сцен и большие наборы RGB-D данных, включающих в себя информацию о жестах и позах [11], действиях человека [12], жестикуляции рук [13], построение корректных динамических моделей, работающих в реальном времени, на данный момент представляется нереальным. Данное исследование ограничено решениями, нацеленными на восстановление только статических объектов сцены. В ряде случаев такое ограничение является оправданным, поскольку человек погружается в виртуальный мир, часть которого адекватно повторяет его реальное окружение, и он уже не ощущает различия между этими мирами. И несмотря на то, что остается конфликт вергенции-аккомодации, присущий всем системам, связанным с реалистичным восприятием

виртуальных объектов, полностью уходит проблема конфликтов светового взаимодействия реального и виртуального миров.

2. Анализ существующих решений

Как правило, производители устройств смешанной реальности имеют в своем распоряжении программные системы для построения геометрической модели реального окружения. Без использования данной модели применение устройств смешанной реальности становится бессмысленным. И это не зависит от того, какой тип системы смешанной реальности используется, оптически или видео прозрачный. Без геометрии реального мира нет никакой возможности корректно встроить виртуальную модель в реальное окружение. Например, компания Microsoft представила Windows Mixed Reality для Windows [14]. Данный продукт обрабатывает изображения реального мира, анализирует окружающую среду и выдает цифровой аналог помещения, который транслируется на очки виртуальной или смешанной реальностей. Windows Mixed Reality запускается на двух основных типах устройств: оптически и видео прозрачных. Microsoft HoloLens является примером оптически прозрачной системы с голографическим проекционным устройством, а примером видео прозрачного устройства может служить Samsung HMD Odyssey+.

Однако, данное решение далеко от совершенства. Отзывы пользователей показывают, что программное обеспечение еще сырое и содержит множество недостатков, связанных с некорректным восстановлением геометрии. Например, существует проблема, связанная с неспособностью определять вертикальные белые стены. Однако, с точки зрения зрительного восприятия основная проблема заключается в двойном представлении геометрии реального мира. Одна модель – это реальный мир, видимый непосредственным образом, например в системе Microsoft HoloLens, а другая модель – восстановленная геометрия, с которой взаимодействует виртуальный объект. При наложении этих моделей возникает конфликт, например, часть виртуального объекта может быть видна за непрозрачной стеной, или объект начинает появляться из-за стены на некотором удалении от нее. Даже небольшие погрешности в модели реального мира могут вызвать значительный дискомфорт зрительного восприятия. С другой стороны, если реальный мир устраняется в процессе визуализации, что возможно для системы Samsung HMD Odyssey+, то дискомфорт полностью уходит, поскольку человек видит только виртуальную модель реального мира и виртуальный объект. То, что реальный мир несколько отличается от его виртуальной модели, человек не ощущает и дискомфорт отсутствует.

2.1. Нейронная сеть “PointNet”

Возникает вопрос, как получить виртуальную модель реального мира. Как правило, на помощь приходят нейросетевые решения, основанные на предположении, что имеется RGB облако точек, полученное RGB-D камерой с одного или нескольких ракурсов. Это облако точек существует в мировом пространстве и положение устройства смешанной реальности также определено в этом пространстве. На первом шаге необходимо выполнить распознавание объектов сцены, то есть классификацию, сегментацию и семантический анализ сцены. Для решения данной задачи наиболее подходящим будет использование сети PointNet [15]. Архитектура нейронной сети PointNet используется для распознавания объектов интерьера по облаку точек. Разработчики из Стэндфордского университета предложили данную архитектуру, которая на вход принимает неупорядоченный набор точек. Данная сеть обеспечивает унифицированную архитектуру для приложений, и ее области применения варьируются, начиная от классификации объектов, сегментации частей и заканчивая семантическим анализом сцены. Данная архитектура соблюдает инвариантность перестановок точек на входе и напрямую обрабатывает облако точек как входные данные.

Для работы с неупорядоченным набором входных данных используется единственная симметричная функция «max pooling» (взятие максимума в области). Сеть учится набору функций (критериев) оптимизации, которые выбирают информативные точки и кодируют

причину их выбора. Полносвязные слои объединяют эти значения в глобальный дескриптор для всей формы или используются для прогнозирования меток по точкам. Поскольку каждая точка трансформируется независимо, то можно применить к входным данным жесткие или аффинные преобразования. Пример сегментации по облаку точек представлен на рисунке 2.

Рисунок 2 – Пример сегментации по облаку точек. Архитектура PointNet [15]

2.2. Нейронная сеть “Total3DUnderstanding”

Следующим шагом является реконструкция геометрии распознанных объектов сцены. Для реконструкции всего помещения выбрана нейронная сеть «Total3D understanding» [16], соединяющая в себе объектно-ориентированную реконструкцию с пониманием 3D-сцены, что позволяет совместно использовать в ходе реконструкции фотографии комнаты, положение камеры, подписи объектов и ограничивающие прямоугольные параллелепипеды.

Данная сеть подчиняется принципу «коробка в коробке» и состоит из трех модулей:

1. Сеть оценки компоновки LEN (layout estimation network);
2. Сеть обнаружения 3D-объектов ODN (Object detection network);
3. Сеть генерации сетки MGN (Mesh Generation Network). Визуализация алгоритма представлена на рисунке 3.

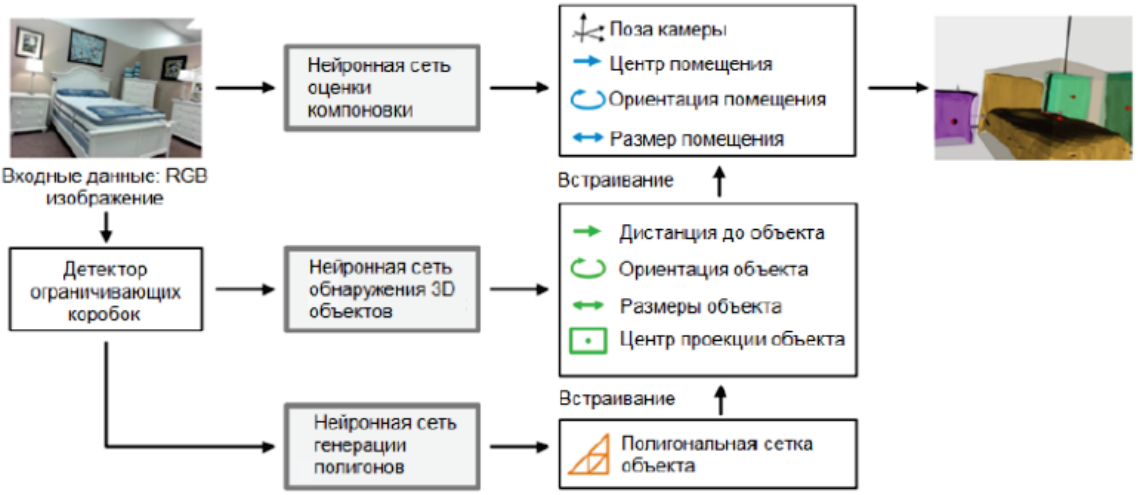

Рисунок 3 – Визуализация алгоритма работы нейронной сети Total3DUnderstanding [16]

По изображению изначально строятся ограничивающие прямоугольники и вычисляются координаты положения камеры. Модуль ODN принимает на вход плоские ограничивающие квадраты и положения о камере и генерирует трехмерные ограничивающие прямоугольные параллелепипеды в системе камеры. Модуль MGN генерирует полигональные сетки объектов. После этого объекты помещаются в одну общую сцену и выравниваются. В результате получается готовая 3D сцена.

3. Алгоритм восстановления сцены

Для создания виртуальной копии реального помещения необходима не только информация о геометрии помещения и объектах, которые в нем находятся, но и информация об оптических свойствах объектов реального мира (коэффициенты отражения, пропускания, текстуры). Для решения данной задачи был предложен следующий алгоритм, который начинает свою работу с обработки исходного изображения RGB-D сцены:

1. Получение изображения окружения с помощью камеры. Это необходимо для извлечения текстур.
2. Получение Z-буфера при помощи RGB-D сенсоров.
3. Получение облака точек при помощи Z-буфера. Пример представлен на рисунке 4. Необходимо отметить, что разрешение и качество RGB изображения и его D аналога могут отличаться, т.е. не для каждой точки RGB изображения определен его D аналог.

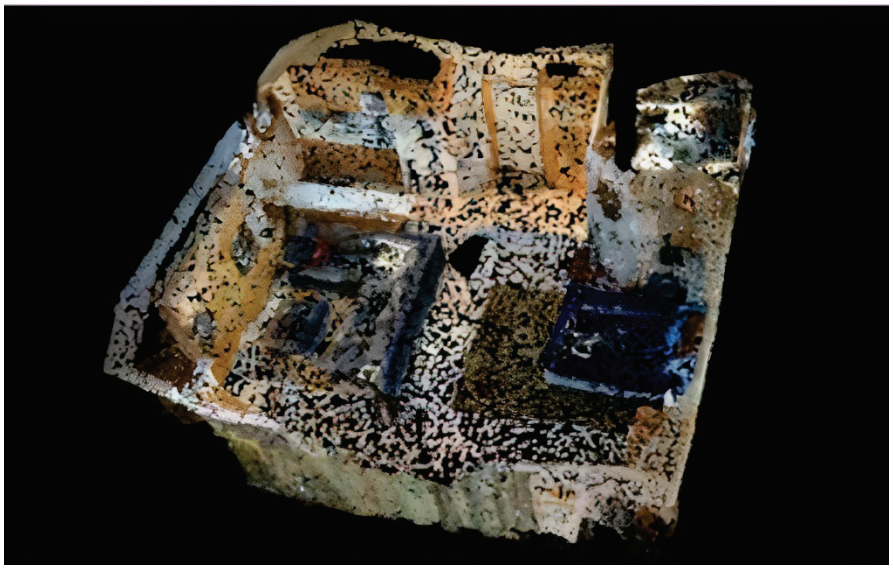

Рисунок 4 – Пример RGB облака точек помещения

4. Выделение объектов в ограничивающие объемы (англ. “bounding box”) на основе сформированного облака точек. Это позволит выделить объекты и определить их положение и размеры. Пример определения ограничивающих объемов представлен на рисунке 5.
5. Получение меток для каждого выделенного объекта сцены. При этом окружение сцены (пол стены, потолок) могут представлять из себя независимые объекты. Таким образом, на данном этапе решается задача распознавания объектов по облаку точек. Каждому объекту присваивается тип, например, стол, стул, шкаф и т. п.
6. После выделения и определения типа объекта решается задача выделения его частей, то есть выделяются его сегменты. Для шкафа это могут быть дверца, полки, стенка. А для стула спинка, сиденье, ножки. Пример выделенного объекта (стул) и трех его частей (спинка, сиденье, ножки) представлен на рисунке 6.

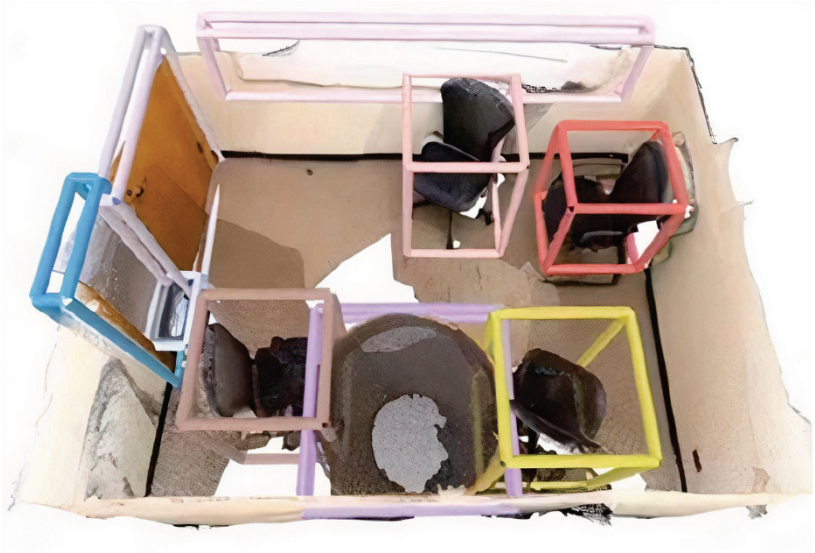

Рисунок 5 – Пример обрамления объектов сцены ограничивающими прямоугольными параллелепипедами

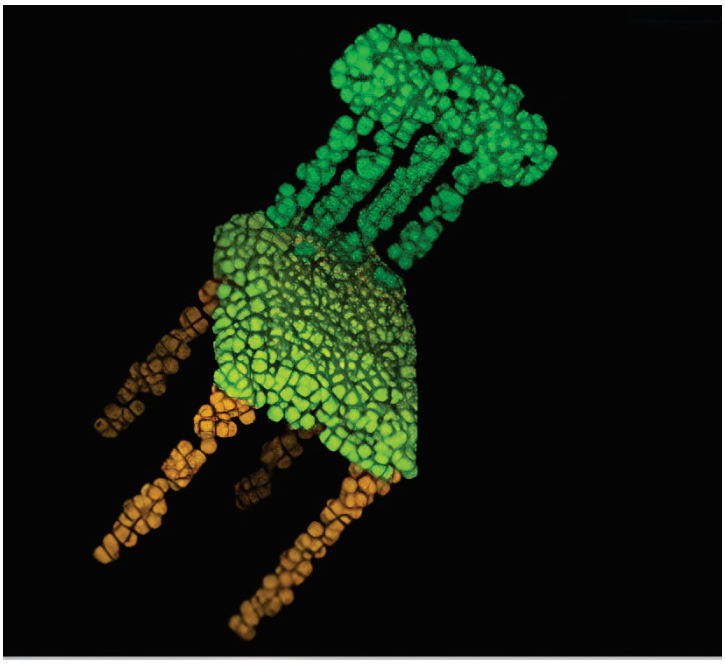

Рисунок 6 – Пример сегментации частей объекта: темно-зеленый цвет – спинка, светло-зелёный – сиденье, а коричневый – ножки стула

7. Создается текстовое описание объектов интерьера, которое бы дало комплексную аннотацию объекту и / или его частям. Это описание можно использовать для поиска CAD (Computer-aided design) моделей в базе данных.
8. CAD модель используется как виртуальный аналог объекта окружения. Чем точнее текст аннотации, тем более похожая модель будет выбрана. Из базы данных извлекается эта модель и выравнивается поверх заменяемого объекта.
9. Поверх CAD модели накладывается текстура, придавая реализм виртуальному объекту в реальном окружении.
10. Если по текстовому описанию объекта не было возможности найти подходящий аналог в базе CAD моделей, то используется решение нейронной сети «Total3D understanding», которое позволяет построить треугольную сетку MGN (Mesh Generation Network) по выделенному и сегментированному объекту сцены.

4. Результаты

Для демонстрации работоспособности предложенного алгоритма использовались достаточно простые синтезированные сцены интерьера помещения. Сцены были получены с помощью программного комплекса Blender. Пример изображения сцены, которое использовалось на входе нейронной сети, и карта глубин в виде изображения Z-буфера в псевдоцветах представлены на рисунке 7. При этом дальние области соответствуют синему цвету, а ближние – красному.

Рисунок 7 – Исходное изображение сцены и ее Z-буферное представление

Необходимо отметить, что целью данного исследования является реконструкция сцены в виде, обеспечивающем пользователю максимально комфортное пребывание в виртуальной копии реального помещения. При этом абсолютное соответствие реального и виртуального миров не требуется. Виртуальная модель реального мира должна обеспечить человеку возможность свободно перемещаться и ориентироваться в виртуальном пространстве, а также обеспечивать естественное встраивание виртуальных объектов в новый виртуально-реальный мир.

Следующие этапы — это обработка RGB-D изображения сцены. Первый этап – формирование RGB облака точек. Полученное RGB облако точек для разных положений камеры представлено на рисунке 8. При этом синие области на изображении сцены, которые выглядят как тени, соответствуют областям пространства, для которых геометрия сцены не определена в облаке точек.

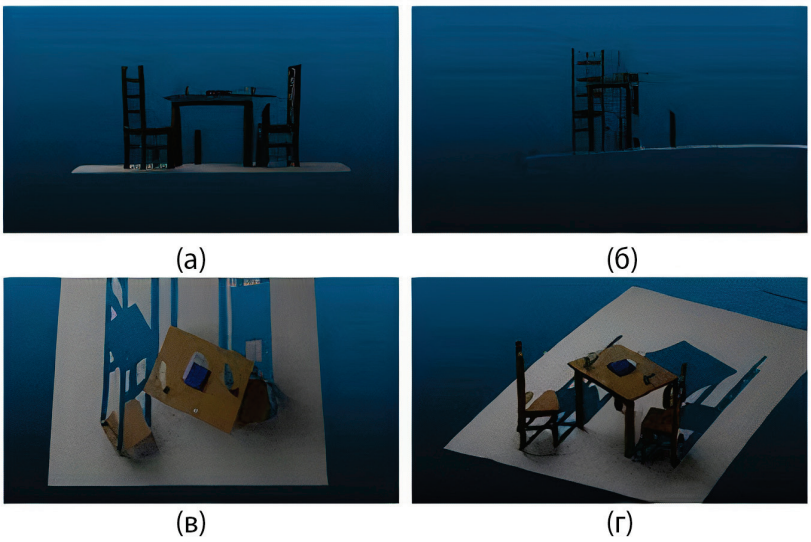

Рисунок 8 – Облако точек, созданное из исходной сцены с различных ракурсов. (а) спереди, (б) справа), (в) сверху, (г) произвольное положение

После выполнения процедуры сегментации и аннотирования объектов интерьера было выделено 6 независимых объектов: два стула, стол, две чашки и книга. Полученные сегментированные объекты и их граничные объемы представлены на рисунке 9.

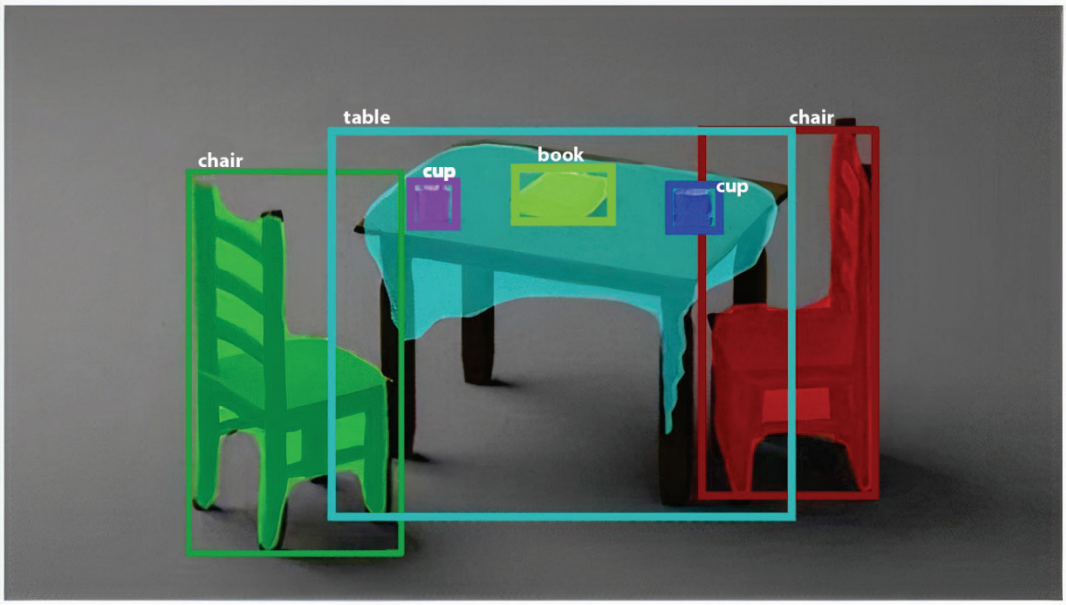

Рисунок 9 – Пример выделения в прямоугольники объектов сцены. Цифры означают вероятность соотнесения данного объекта к классу

Поскольку авторы не располагали аннотированной базой данных CAD объектов, то восстановление геометрии происходило с использованием нейронной сети «Total3D understanding», которая позволила построить треугольную сетку по выделенным и сегментированным объектам сцены. На рисунке 10 можно увидеть и сравнить изображения исходного стола и его реконструированной модели. Изображения представлены для ряда одинаковых ракурсов. Рисунок 11 показывает аналогичные изображения для исходного стула и его реконструированной модели.

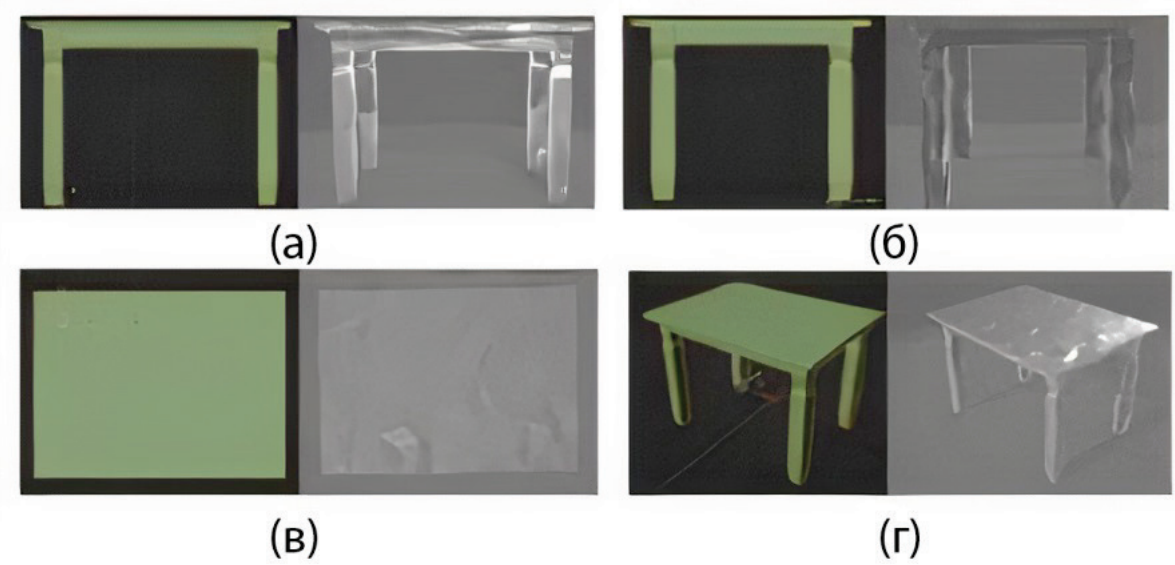

Рисунок 10 – Сравнение исходного стола (слева) и восстановленного (справа) с различных ракурсов. (а) спереди, (б) справа), (в) сверху, (г) произвольное положение

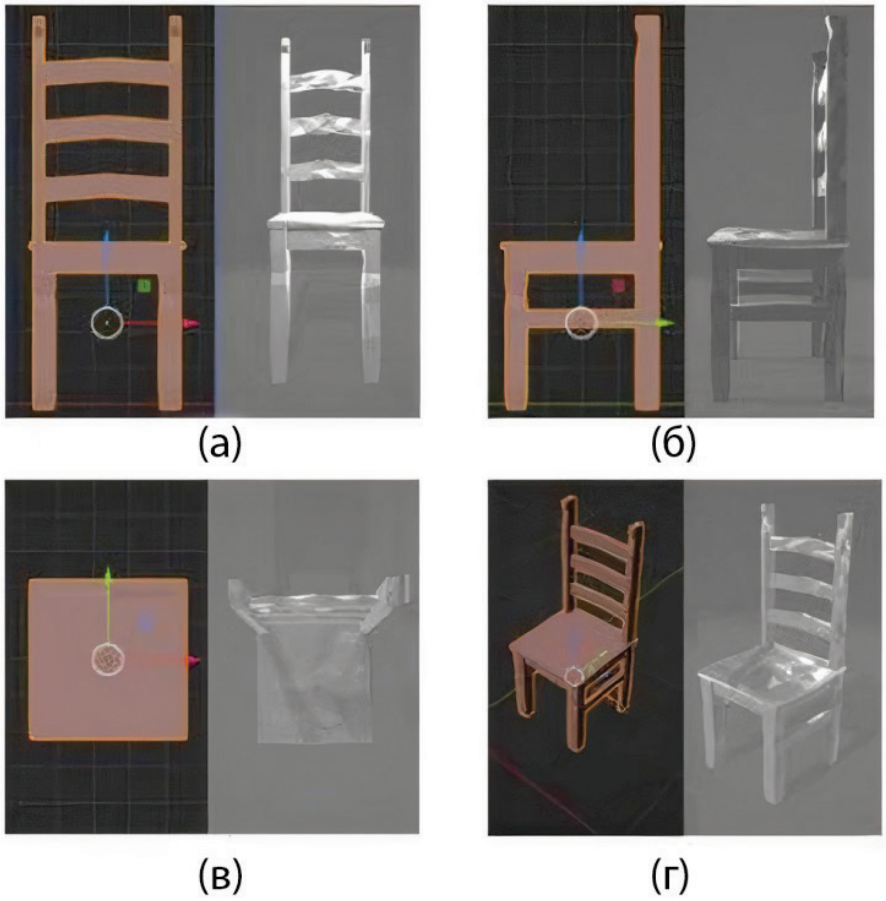

Рисунок 11 – Сравнение исходного стула (слева) и восстановленного (справа) с различных ракурсов. (а) спереди, (б) справа), (в) сверху, (г) произвольное положение

На рисунках 12 и 13 представлена реконструированная сцена с различных ракурсов после наложения соответствующих текстур.

Рисунок 12 – Пример реконструированной сцены с наложением текстур

Рисунок 13 – Пример реконструированной сцены с другого ракурса

Поскольку получение точной модели объекта реального мира не является главной задачей исследования (задачей является построить похожую модель реального мира для корректной ориентации человека внутри этой модели и «естественном» ее визуальном восприятии), то численная оценка качества восстановленной геометрии и ее текстуры не имеет практического смысла. Более подходящим решением будет экспертная оценка, которая была получена в результате опроса. В опросе участвовало 238 человек. Был представлен ряд изображений и поставлена задача выставить оценку по десятибалльной шкале. Для оценки использовались два критерия оценить качество реконструкции с точки зрения естественности зрительного восприятия и оценить схожесть реконструированного помещения (пример реконструированного изображения представлен на рисунках 12 и 13) с исходным изображением сцены, представленным на рисунке 7. Результаты представлены на рисунке 14.

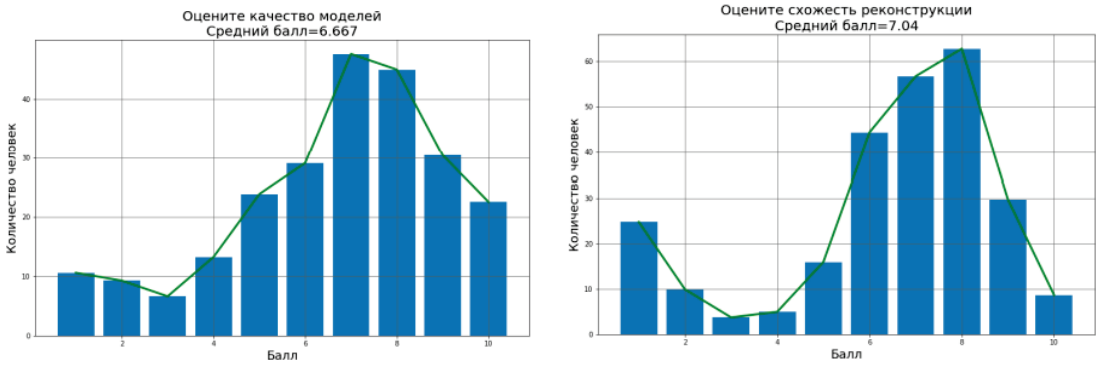

Рисунок 14 – Результаты опроса по оценке качества реконструкции моделей

5. Заключение

В результате данной работы была разработана и реализована система реконструкции модели реального помещения на основе RGB облака точек, полученного по изображению сцены и его карте глубин. Поскольку целью исследования был выбран максимальный комфорт пребывания человека в виртуальной копии реального помещения, то для эмпирической оценки полученных результатов качества реконструкции был проведен опрос. Результаты опроса показали удовлетворительное качество реконструкции, что свидетельствует об успешном построении виртуальных аналогов реальных объектов сцены для создания систем смешанной реальности, свободных от дискомфорта зрительного восприятия.

6. Благодарности

Работа была выполнена при финансовой поддержке Российского Научного Фонда, грант № 22-11-00145.

7. Список источников

[1] Virtual Reality Sickness and Challenges Behind Different Technology and Content Settings. / Guna, J., Geršak, G., Humar, I. et al. // Mobile Netw Appl 25, 1436–1445. 2020. URL: https://doi.org/10.1007/s11036-019-01373-w.

[2] A Novel Method for Extrinsic Calibration of Multiple RGB-D Cameras Using Descriptor-Based Patterns. / Hang Liu, Hengyu Li, Xiahua Liu, Jun Luo, Shaorong Xie, and Yu Sun. // Image and Video Processing (eess.IV); Robotics (cs.RO). 2018. https://doi.org/10.48550/arxiv.1807.07856

[3] Zhang, Z. Microsoft kinect sensor and its effect. IEEE Multimed. 2012, 19, 4–10. URL: https://doi.org/ 10.1109/MMUL.2012.24.

[4] Comparative Study of Registration Methods for RGB-D Video of Static Scenes. / Morell Gimenez, V., Saval-Calvo, M., Azorin-Lopez, J., Garcia-Rodriguez, J., Cazorla, M., OrtsEscolano, S., Fuster-Guillo, A. A. // Sensors 2014, 14, 8547–8576. URL: https://doi.org/ 10.3390/s140508547.

[5] An Approach to the Use of Depth Cameras for Weed Volume Estimation. / Andújar D, Dorado J, Fernández-Quintanilla C, Ribeiro A. // Sensors (Basel). 2016 Jun 25; 16(7):972. doi:10.3390/s16070972. PMID: 27347972; PMCID: PMC4970024.

[6] SECOND: Sparsely Embedded Convolutional Detection. / Yan, Yan, Yuxing Mao, and Bo Li. // Sensors 18. 2018. no. 10: 3337. URL: https://doi.org/ 10.3390/s18103337.

[7] A Unified Deep Framework for Joint 3D Pose Estimation and Action Recognition from a Single RGB Camera. / Huy-Hieu Pham, Houssam Salmane, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A. Velastin. // Computer Vision and Pattern Recognition. 2019. URL: https://doi.org/10.48550/arXiv.1907.06968.

[8] An Effective 3D Shape Descriptor for Object Recognition with RGB-D Sensors. / Liu Z, Zhao C, Wu X, Chen W. // Sensors (Basel). 2017 Feb 24;17(3):451. doi: 10.3390/s17030451. PMID:28245553; PMCID: PMC5375737.

[9] A Fast and Robust Extrinsic Calibration for RGB-D Camera Networks. / Su, Po-Chang, Ju Shen, Wanxin Xu, Sen-Ching S. Cheung, and Ying Luo. // Sensors 2018, no. 1: 235.

[10] Texture Mapping for 3D Reconstruction with RGB-D Sensor. / Fu, Yanping, Qingan, Yan, Long, Yang, Jie, Liao, and Chunxia, Xiao. // In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018. (pp. 4645–4653). URL: https://doi.org/ 10.1109/CVPR.2018.00488.

[11] First-Person Hand Action Benchmark with RGB-D Videos and 3D Hand Pose Annotations. / Guillermo Garcia-Hernando, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim. // CoRR abs/1704.02463. 2017. URL: https://doi.org/10.48550/arXiv.1704.02463.

[12] Scene Semantic Reconstruction from Egocentric RGB-D-Thermal Videos. / R. Luo, O. Sener, and S. Savarese. // In 2017 International Conference on 3D Vision (3DV) IEEE Computer Society, 2017. (pp. 593-602). URL: https://doi.org/10.1109/3DV.2017.00073.

[13] Understanding Everyday Hands in Action from RGB-D Images. / Rogez, Grégory, James S., Supancic, and Deva, Ramanan. In 2015 IEEE International Conference on Computer Vision (ICCV) 2015. (pp. 3889-3897). URL: https://doi.org/10.1109/ICCV.2015.443.

[14] The impact of virtual, augmented and mixed reality technologies on the customer experience. / Carlos Flavián, Sergio Ibáñez-Sánchez, Carlos Orús. // Journal of Business Research. Volume 100, 2019, Pages 547-560, ISSN 0148-2963. URL: https://doi.org/10.1016/j.jbusres.2018.10.050.

[15] PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. / Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. // CoRR abs/1612.00593. 2016.

[16] Total3DUnderstanding: Joint Layout, Object Pose and Mesh Reconstruction for Indoor Scenes from a Single Image. / Yinyu Nie, Xiaoguang Han, Shihui Guo, Yujian Zheng, Jian Chang, and Jian-Jun Zhang. // CoRR abs/2002.12212. 2020. URL: https://doi.org/10.48550/arXiv.2002.12212.