УДК 004.032.26, 004.852, 514.48

ГЕНЕРАЦИЯ ТРЕХМЕРНЫХ СИНТЕТИЧЕСКИХ ДАТАСЕТОВ

В. В. Кугуракова[1, [0000-0002-1552-4910]], В. Д. Абрамов[2, [0000-0003-1465-5238]],
Д. И. Костюк[3, [0000-0002-6542-6980]], Р. А. Шараева[4, [0000-0002-2359-1873]],
Р. Р. Газизов[5, [0000-0002-8349-264X]], М. Р. Хафизов[6, [0000-0001-7275-9102]]

[1–6] *Институт информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета*

[1]vlada.kugurakova@gmail.com, [2]vitaly.d.abramov@gmail.com, [3]xdxnxkx@gmail.com, [4]r.sharaeva3496@gmail.com,[5]starkindustries14579@gmail.com,[6]murkorp@gmail.com

Аннотация

Работа посвящена описанию процесса разработки универсального инструментария для генерации синтетических данных для обучения разных нейронных сетей. Используемый подход показал свою успешность и эффективность в решении различных задач, в частности, обучения нейросети для распознавания покупательского поведения внутри магазинов через камеры наблюдения и пространств устройствами дополненной реальности без использования вспомогательных инфракрасных камер. Обобщающие выводы позволяют спланировать дальнейшее развитие технологий генерации трехмерных синтетических данных.

Ключевые слова*: синтетические данные, датасет, искусственный интеллект, нейронные сети, машинное обучение, компьютерное зрение, трехмерные модели, metahuman, игровые движки, Unreal Engine*

ВВЕДЕНИЕ

В настоящий момент трудно представить нашу жизнь без нейронных сетей. Их используют в большом количестве областей, где особенно необходим не алгоритмический четко структурированный подход в «лоб», а более деликатный – творческий. Это важно для тех сред, в которых происходят постоянные изменения, где существует достаточно высокая степень энтропии как исходных, так и конечных данных, а также где необходимы постоянное обучение и «переобучение» алгоритмов. Рассмотрим достаточно узкоспециализированные нейронные сети, которые занимаются машинным обучением, в частности, машинным зрением.

Одна из основных проблем таких сетей – потребность большого количества данных для их обучения. Для этого нужны качественные датасеты, и если в случае обучения текстовых сетей текстовых данных и датасетов в целом достаточно, то в случае с визуальными данными у сообщества исследователей ощущается острая их недостача [1]. Рассмотрим вопросы генерации синтетических трехмерных данных для обучения таких нейронных сетей.

Фокус внимания представленной работы нацелен на то, как такие данные генерируются из трехмерных виртуальных пространств, имитирующих интерьеры строений реального мира. Внутри этих пространств располагаются разные объекты, созданные на основе реальных объектов, а именно: мебели, интерьерных элементов, малых архитектурных форм и т. п. По этому реалистичному пространству перемещаются сгенерированные персонажи разных рас, полов, роста, веса и пр. Кроме того, всё пространство и объекты в нём могут изменяться в зависимости от выбранных параметров. В итоге открывается возможность сформировать датасет из изображений этого пространства, причем, что особенно важно, все изображения уже будут обладать всей необходимой метаинформацией, что устраняет необходимость в ручной разметке данных.

СВЯЗАННЫЕ РАБОТЫ

Первые научные работы, затрагивающие создание трехмерных синтетических данных, появились примерно в одно и то же время со становлением трехмерных движков. Исследовательские группы Д. Хеегера в 1987 г. [2] и Д. Барона в 1994 г. [3] использовали виртуальные сцены для количественной оценки методов оптического потока. Мак Кейн и другие [4] собирали изображение объектов на фоне, варьируя сложность сцен и движений, чтобы понять, как эти изменения скажутся на точности оценки. Похожая на эту работу была создана Отте и Нагелем [5], они записывали реальную сцену с упрощенной геометрией через виртуальную камеру, настройки которой можно было задавать вручную через параметры. В [6] была получена реалистичная сцена путем воссоздания её трехмерной модели. Датасет Middlebury flow [7] содержит как реальные, так и сгенерированные данные. В [8] представлены синтетические последовательности для сравнительного анализа оценки потока сцены в качестве бенчмарка. Где-то между синтетическими и реальными находятся датасеты, такие, как представленный в [9], где

были созданы коллажи из реальных фотографий, чтобы получить обучающие данные, например, для обнаружения локаций. Последние синтетические датасеты от UCL [10] и чуть более крупный от Sintel benchmark [11] использовали программу для создания трехмерных сцен и моделей Blender, которая позволила им добиться плотного и реалистичного визуального потока в сложных вычисленных сценах (рендер-сценах). В исследовании точности оптического потока при автовождении [12] использовали графический редактор Maya для визуализации набора данных.

Подходы разных авторов различаются и по технологическим предпочтениям. Например, в [13] появляется возможность процедурной генерации с помощью уже готовых игровых ассетов, созданных другими разработчиками. Подобный технологический подход использовался также в [14] для визуализации сложных визуальных эффектов, таких как зеркальные отражения. Игровые движки используют для создания тестовых данных для систем видеонаблюдения [15] и виртуального автомобильного симулятора [16], где необходимо создание больших, масштабируемых данных. В [11] предложен другой подход к созданию датасетов – на основе кинофильма с открытой лицензией.

Несколько недавних и параллельных работ были сосредоточены на создании крупномасштабных наборов синтетических обучающих данных для различных задач компьютерного зрения. Эти датасеты разделяются на широкую выборку сценариев, например, автомобильные виртуальные полигоны: Virtual KITTI [17], SYNTHIA [18] и датасет Рихтера [19]; экстерьерные сцены: SceneNet [20], SceneNet RGB D [21], ICL-NUIM [22] и SUNCG [23]; отдельные объекты: ModelNet [24] и ShapeNet [25]; распознавание человеческого поведения [26].

Вопрос, какие свойства синтетических обучающих данных помогают обобщению реальных данных, пока остается мало проанализированной областью. Авторы [27] использовали просчитанные данные для обучения нейронной сети оценки удаленности объекта. Их результаты показывают лучшую производительность в реальных сценах, когда обучающие данные содержат реальные текстуры фона и изменяющееся освещение. В [28] исследовано, как качество освещения в отрендеренном изображении влияет на производительность нейронной сети при оценке точки обзора, и обнаружено, что в этом помогает более реалистичное

освещение. Рассмотрено также влияние качества синтетических данных на семантическую сегментацию [29].

В целом большинство исследований, в которых так или иначе затрагивается оценка качества синтетических данных, выделяет следующие ключевые моменты:

- Многоэтапное обучение на нескольких отдельных наборах данных работает лучше, чем то, которое использует один тип данных; кроме того, оно работает и лучше, чем просто смешивание всех данных.
- Не выявляется прямая корреляция между усложнением рендеринга света и увеличением качества обучения. Тем не менее, датасеты, не использующие рендер света, негативно влияют на качество обучения. Необходима более точная и методологически выверенная оценка рендеринга света как фактора, влияющего на качество обучения нейронной сети.
- Моделирование искажения реальной камеры во время обучения улучшает производительность сети, которая в дальнейшем будет работать на изображениях с таких камер.

ТРЕБОВАНИЯ К ГЕНЕРАЦИИ СИНТЕТИЧЕСКИХ ДАННЫХ

Правильный сбор требований к системам программного обеспечения – одна из самых важных частей разработки программного обеспечения. Классический подход PMBOK учит нас тому, что ошибка на этом этапе может стоить увеличению стоимости проекта в два раза. Множество научных и индустриальных коллективов нуждается в качественных синтетических трехмерных данных для решения своих задач. Нам удалось поработать с двумя международными заказчиками, проводящими научно-исследовательские и опытно-конструкторские работы по машинному обучению. Ниже в оценке результатов будет описано различие в их задачах и полученных выводах.

Чтобы создать датасет, необходимо создать трехмерные данные, для чего потребуется некоторый интерпретатор (программная среда), который сможет преобразовывать низкоуровневый код в векторную компьютерную графику. Есть два основных пути. Первый – создать свой собственный интерпретатор, что в целом не займет много времени, если вы хотите создать простую компьютерную

графику без поддержки сложных шейдеров, текстур и высокополигональных моделей. В противном случае подобный процесс занимает огромное количество человеко-часов, что может, по своей сути, вылиться в отдельные увлекательные исследования: как правило, разработку современных движков компьютерной графики (их также можно называть игровыми движками) могут позволить себе большие IT компании с крупным штатом разнонаправленных разработчиков. Именно использование готовых игровых движков является вторым путем, который был выбран нами. На рынке существует достаточно большое количество игровых движков, позволяющих работать с графикой реального времени и характеризующихся тремя основными правилами: поддержка сложных шейдеров, сложных текстур и высокополигональных моделей. Как дополнительные критерии необходимо ввести опыт работы команды разработки с движком, а также «популярность», которая характеризуется в большом объеме знаний как внутри сообщества разработчиков, так и в исходной документации. Под данные критерии попадают два основных мировых игровых движка [30]:

- Unreal Engine [31] – игровой движок, разрабатываемый и поддерживаемый компанией Epic Games;
- Unity [32] – межплатформенная среда разработки компьютерных игр, разработанная американской компанией Unity Technologies.

На первый взгляд, между этими игровыми движками нет большой разницы, но в Unreal Engine есть много важных возможностей, что позволило сделать выбор в пользу него.

Итак, средства для реализации генератора синтетических данных выбраны, определим теперь, какие требования необходимы для его успешного функционирования. Под успешным функционированием подразумевается возможность правильного обучения на этих данных целевых нейросетей. Поскольку речь идет о машинном обучении, одним из самых важных требований является валидность данных, следующее требование – правильная разметка датасета.

Валидность – это мера соответствия методик и результатов исследования поставленным задачам. Для нашего исследования это означает, что такие данные, во-первых, могут быть использованы для обучения, во-вторых, нейросеть,

обученная на этих данных, сможет успешно детектировать или распознавать события, сущности и объекты в соответствии с заявленными требованиями.

Первые требования к генератору трехмерных данных от заказчиков и экспертов звучали в довольно неформализованной форме:

- генерация реалистичных данных – т. е. данные должны быть максимально похожи на фотографическое изображение действительности;
- возможность установки входных параметров симуляции перед стартом генерации (случайность генерации в данном случае определяется зерном генерации – начальным значением генератора псевдослучайных чисел, который можно использовать повторно в случае понравившихся данных);
- возможность простого изменения виртуального пространства (например, расстановкой объектов, выбором текстур, а также регулировкой параметров освещения и настройкой постэффектов [33]);
- работа с виртуальными камерами, настройка точек установки;
- возможность изменять параметры трехмерных моделей людей (такие как раса, вес, пол, рост, форма тела, тон кожи и т. п.);
- возможность изменять и рандомизировать одежду в соответствии с внешним видом персонажа;
- создание полноценного интерьерного (внутреннего) пространства со сложной архитектурой, мебелью и привычными аксессуарами.

Очевидно, что также необходимы некоторые технические требования:

- возможность выбора папки для сохранения генерируемых данных;
- возможность запуска на компьютерах под управлением Windows;
- возможность работы с генератором через разные интерфейсы (из командной строки, с помощью интерфейса, через API).

Требования были собраны as is, представлены в формате «чистых мыслей» экспертов. А этап их формализации стал первой решаемой задачей создания универсального генератора и будет подробно освещен ниже.

ТРЕБОВАНИЯ К ГЕНЕРАТОРУ ДАННЫХ

Генератор данных состоит из нескольких модулей, каждый из которых должен представлять собой программный код на языках C++ или Blueprint, способный к запуску в программной скриптовой среде Unreal Engine.

Рис. 1. Интерфейс приложения: выбор параметров

Для *модуля запуска приложения* определяется как требование то, что интерфейс запуска может быть осуществлен через командную строку или пользовательское приложение (см. рис. 1), а в качестве параметров выступают: зерно генерации; параметры уровня, с помощью которых происходят изменения текстур трехмерных моделей и геометрии трехмерных моделей; параметры освещения; количество персонажей на уровне; требуемое количество сгенерированных секвенций; количество кадров в каждой секвенции; выбор сезона одежды; папка хранения сгенерированных файлов; настройки камеры.

Для *модуля генерации уровня* определяется как требование генерация модели уровня с переданными параметрами: типы объектов; количество зон на уровне; этажность здания; текстуры пола в зонах; текстуры стен в зонах. Пример генерации разных зон можно наблюдать на рис. 2.

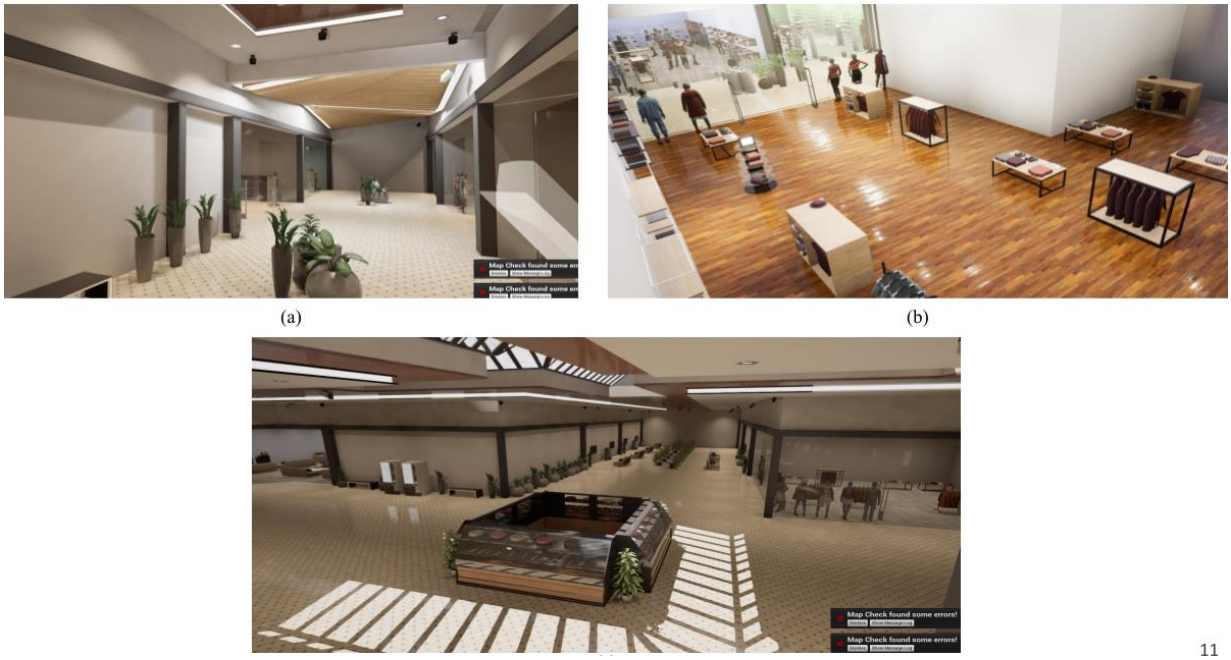


Рис. 2. Примеры генерации разных частей уровня и разных точек установки виртуальных камер

Для *модуля виртуальных камер* определяется как требование возможность настройки виртуальных камер с уточнением параметров: высота и ширина матрицы; фокусное расстояние; величина диафрагмы.

Для *модуля объектов на уровне* определяются статические и динамические объекты. И для тех, и для других объектов параметрами являются заданные точки их расстановки, но в отличие от статических, которые расставляются единожды, запуск динамических объектов происходит каждую секвенцию. Динамические объекты перемещаются на заданное расстояние или изменяют свое состояние (например, дисплей с разными изображениями) каждый тик секвенций.

Для *модуля персонажей* необходимо иметь возможность генерации до 100000 уникальных персонажей (см. рис. 3), персонажи должны варьироваться по группе параметров: пол; форма тела; рост; прическа; цвет волос; цвет глаз; тон кожи. Понятно, что «параметров» у реальных людей намного больше, и включение в будущем в генерацию персонажа технологии Metahuman [34], учитывающей огромное количество параметров лица и тела человека, позволит усилить эту сторону генератора.

Второе требование: одежда должна автоматически подстраиваться под физические параметры персонажа, а в качестве параметров уточняются: тип

одежды верхней части тела; тип одежды нижней части тела; тип верхней одежды; тип обуви; тип специальной одежды, которая заполняет два слота; набор из типов аксессуара.

Рис. 3. Результат работы модуля генерации персонажей

Третье требование: позы должны быть естественными, наиболее применимыми для каждодневной рутины человека (см. рис. 4), а параметрами являются типы поз на каждый тик.

Четвертое требование относится к расстановке получившихся персонажей по уровню в заданных зонах. Персонажи могут как находиться в этих зонах, так и перемещаться в них и между ними.

Модуль освещения регламентирует положение и тональность источников света, что для большинства из них задается через параметры запуска.

Рис. 4. Трехмерные персонажи во время тестирования анимаций

ТРЕБОВАНИЯ К ГЕНЕРАЦИИ УРОВНЯ

На данный момент существует множество способов создания сложных высокополигональных моделей, особенно в области симуляции ткани, волос и органических объектов в целом. Тем не менее, все это в конечном счете в машинном представлении сводится к обычному мешу, состоящему из треугольных полигонов.

Кроме создания трехмерных моделей необходимо создание трехмерной локации, или уровня, что должно быть реализовано полуавтоматизированно через процедурный подход. Уровень должен состоять из блоков, которые могут заменяться в зависимости от изменяющихся требований.

В качестве образцов (референсов) для создания уровня были использованы привычные общественные пространства – торговые галереи, встречаемые в молах (см. рис. 5).

Рис. 5. Пример торговой галереи реального мола

Почему были выбраны именно такие пространства:

- такие пространства имеют, с одной стороны, большие и открытые галереи, а с другой, поддаются четкой формализации по составляющим (коридоры, точки входа, помещения), что удобно для генератора уровня;
- в таких пространствах находится одновременно большое количество людей, значит, сгенерированные данные с большим количеством персонажей для целевых датасетов будут валидны реальным данным;
- в таких пространствах встречаются сложные архитектурные элементы, которые могут действовать в совокупности: например, стеклянные стены и отражающие поверхности – а детекция объектов и нивелирование негативных факторов от таких объектов являются важной проблемой алгоритмов машинного зрения;
- системы освещения таких пространств могут включать множество источников разных типов освещения: естественных – от окон и мансард, искусственных – от ламп или динамических световых объектов (дисплеев, световых лент и т. д.).

В качестве схемы уровня может быть взят план реального мола (см. рис. 6), причем коридоры, отмеченные зеленым, будут неизменяемыми трехмерными

локациями, а зоны, отмеченные чёрным, будут зонами с динамическими и статическими генерируемыми объектами.

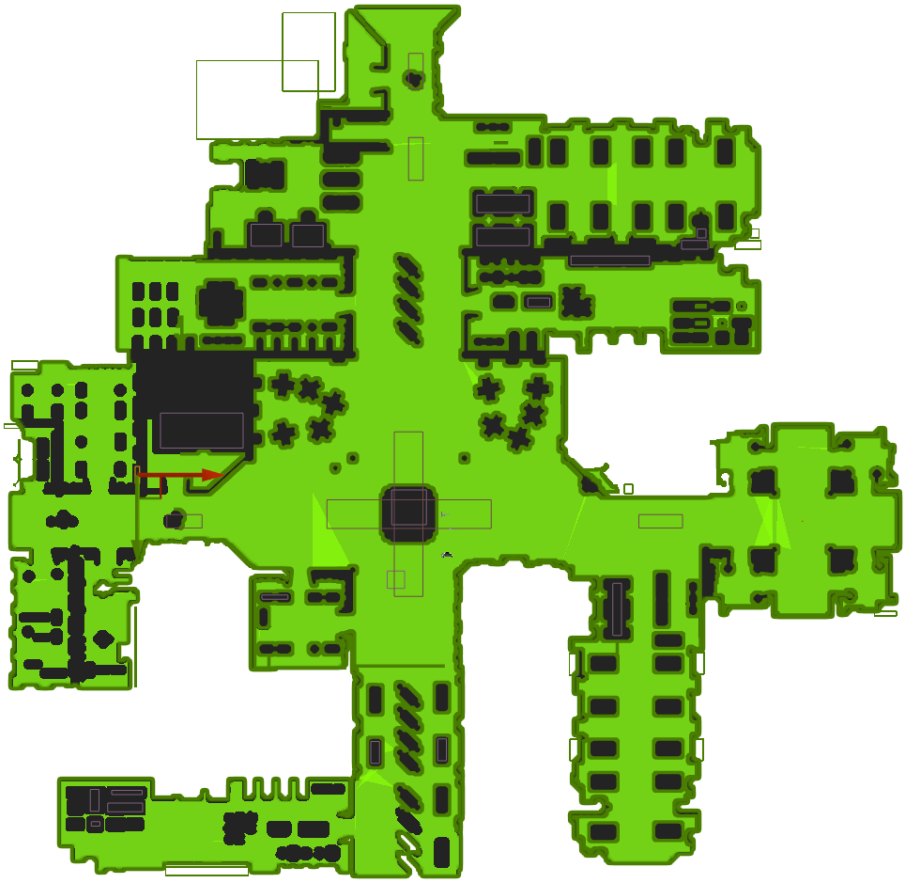


Рис. 6. План реального мола для схемы уровня

В данных зонах могут быть расположены разные объекты, ассоциирующиеся с торговым центром, а именно, зоны рекреации, разные типы магазинов, кафе и пр.

ТРЕБОВАНИЯ К МЕТАИНФОРМАЦИИ

Одним из самых важных этапов обучения нейронных сетей является правильная разметка данных. Для данного типа визуальных данных может существовать несколько типов вспомогательных данных или метаданных.

Первый тип — визуальные метаданные. Это могут быть тепловые карты, подсветка или выделение определенных областей, вывод текстовой информации. Поскольку для одной нейросети, участвовавшей в данном исследовании, требовалась оценка удаленности окружения и объектов окружения, было решено

использовать карту глубины (см. рис. 7), где черные области были бы наиболее отдалёнными от нас, а белые области были бы наиболее приближенными к виртуальной камере. Области измерялись градиентным переходом монохромного цвета с шагом увеличения интенсивности в каждый сантиметр.

Рис. 7. Пример карты глубины

Второй тип – это текстовые метаданные. Такие данные могут описывать разные сущности и объекты, расположенные на виртуальной сцене и ассоциированные с конкретными камерами или изображениями с этих камер.

Одна из нейросетей, используемая для оценки валидности синтетических данных, использовала детекцию трехмерных персонажей, чтобы создать рабочую ассоциацию между визуальным изображением и объектом на нем. Для этого собиралась такая информация: расположения ячеек объектов по сетке координат на сгенерированном изображении, где нулем является нижний левый угол изображения; расположения ячеек объектов по сетке координат в пространстве уровня; порядковый номер персонажа; параметры используемых камер; параметры уровня (зерно генерации, температура освещения, сезон одежды и т. п.); информация по персонажам.

Рис. 8. Детекция габаритов персонажей

Габаритная рамка (в англоязычной литературе используются термины bounding boxes, или object outlines) выделяет конкретных персонажей (см. рис. 8).

ОСНОВНОЙ АЛГОРИТМ РАБОТЫ

Ключевой особенностью базового алгоритма (см. рис. 9) должна стать оптимизированная работа с высокой степенью вариативности.

Со стороны пользователя работа приложения должна выглядеть как черный ящик, в который поставляются определенные параметры, а на выходе получаются совокупности изображений, разделенных на папки по времени (так называемые секвенции). Если рассматривать это с приближением к реальному взаимодействию по времени, то можно определить это как формулу, где (в каждый момент времени) происходит съемка всего пространства с каждой из камер.

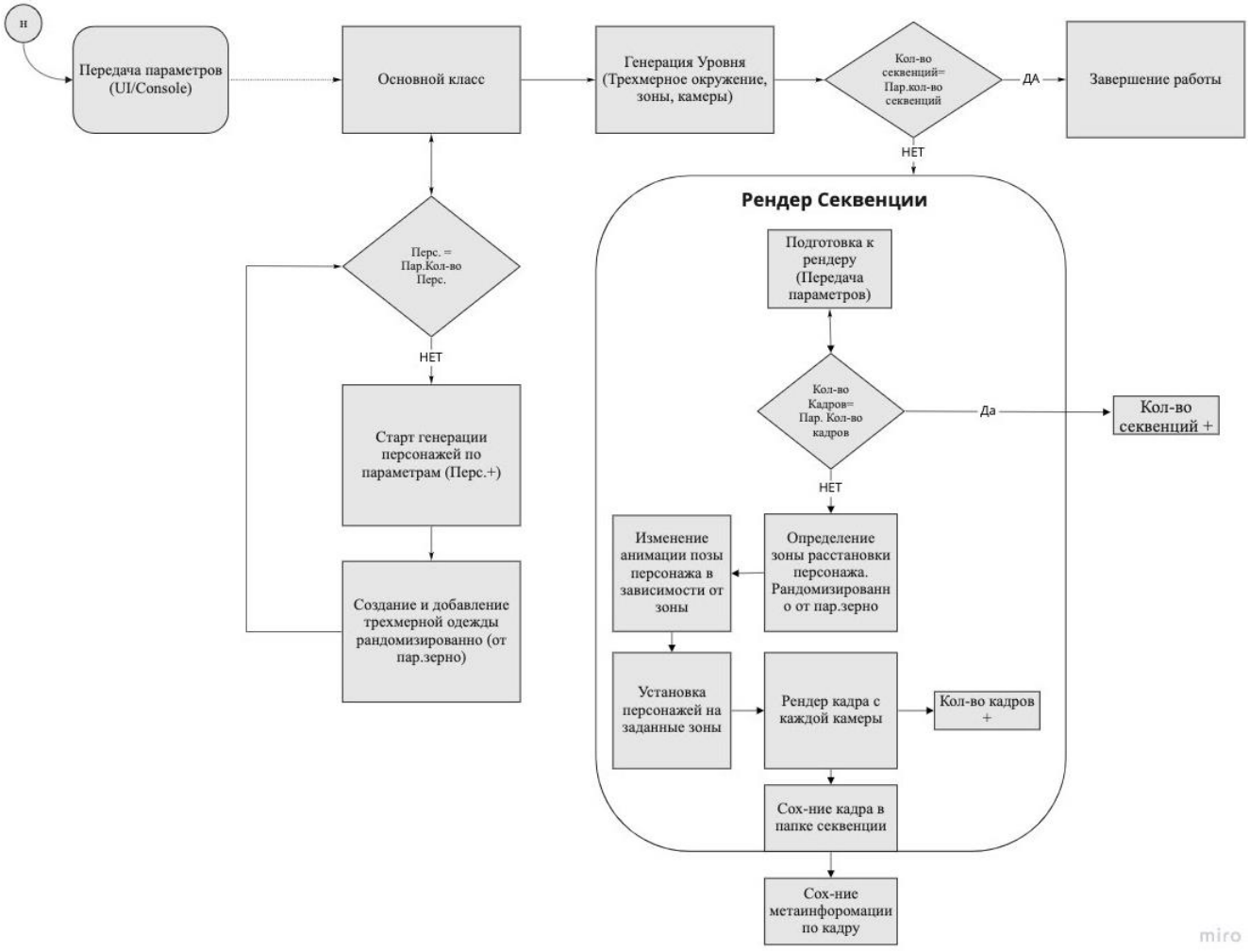


Рис. 9. Упрощенная блок-схема алгоритма работы приложения

После того как сформированы все кадры для данного момента времени, происходит формирование кадров для следующего момента, и так, пока не будет достигнуто требуемое количество. Разница между моментами совпадает с тиками и подразумевает перемещения объектов внутри виртуального пространства. Тем самым был получен набор реалистичных изображений, имитирующих «жизнь» торгового центра, где перемещаются люди, а их перемещения фиксируют виртуальные камеры.

РЕЗУЛЬТАТЫ РАБОТЫ

В первую очередь, важно поделиться основными и ключевыми метриками результатов работы – готовый датасет (см. рис. 10) – это необходимое количество секвенций jpeg-изображений.

Рис. 10. Набор изображения для датасета в папке секвенций

Рассмотрим результаты подробнее с примерами из работ по подготовке датасетов для каждого из заказчиков. Хронологически первые данные были сгенерированы для команды «Б».

Рис. 11. Общественное здание

В рамках данной разработки система была менее гибкой, но, с другой стороны, имела более реалистичные настройки графики. В качестве базового уровня использовалась виртуальная среда, напоминающая крупное офисное или торговое общественное здание (см. рис. 11). Ключевыми особенностями данной виртуальной среды были:

- наличие сложной многоэлементной архитектуры с использованием разных типов материалов для текстур трехмерных поверхностей;
- наличие большого количества поверхностей с большими коэффициентами отражения и преломлениями света (полы, стены, стекла и т. д.);
- использование разных типов освещения:
  - статического естественного освещения (от световых окон, дверей и т. п.);
  - статического неестественного освещения (от светильников, ламп, световых панелей и т. п.);
  - динамического: от световых табло и экранов;
- гибкая настройка параметров текстур при запуске (см. рис. 12).

Рис. 12. Пример изменения текстур изображения

Команда «Б» занималась обучением нейросетей для устройств дополненной реальности, по сути эти нейросети предлагают иной подход к локализации и распознаванию пространства, в отличие от устоявшихся SLAM, причем без использования вспомогательных инфракрасных камер. Именно поэтому при рендеринге виртуальной камеры использовались настройки, близкие к требуемому устройству, что также отмечается как эффективный подход к созданию датасетов, см. выше. Сущность SLAM заключается в объединении данных сканирования пространства и отслеживания местоположения [35].

При обучении алгоритмов локализации не менее важно дать им не только визуальную информацию, но также информацию о расстоянии объектов относительно точки наблюдения. Наиболее оптимизированный и эффективный формат в данном случае – это использование карт глубин, монохромных изображений, где интенсивность одного из цветов обозначает степень удаления от точки обзора, и наоборот, интенсивность другого цвета обозначает приближённость к точке обзора. Линейное смешение этих цветов даёт в свою очередь информацию о конкретной дальности, вплоть до миллиметра. Единственная метаинформация, которую поставлялась для данного синтеза, была черно-белая карта глубины (см. рис. 13).

Рис. 13. (a) конечный рендер с красно-белыми калибровочными шкалами для тестов; (b) метаинформация в виде карты глубины

В результате работы этой версии генератора получилась совокупность jpeg-изображений, сделанных с заданного пролёта виртуальной камеры, с метаинфор-

мацией для каждого изображения в виде карты глубины. Команда «Б», к сожалению, в силу корпоративной специфики не смогла поделиться большим количеством подробностей, кроме довольно очевидных выводов:

1. Подход с использованием синтетических данных успешно показал себя в обучении нейронной сети.
2. Синтетических данных, сгенерированных в этом решении, не хватало, использовались ещё и другие датасеты как реальные, так и синтетические, что также считается эффективным способом.
3. Необходимо большее разнообразие сгенерированного трехмерного уровня, это позволит разнообразить различные ситуации в датасете.
4. Использование карты глубины положительно влияет на обучение нейронной сети.

В целом результат полученного датасета для команды «Б» можно считать успешным, сгенерированные данные помогли им переосмыслить подходы к SLAM.

Нейросеть команды «А» имела немного другую задачу, главной целью были детекция и «анонимное» распознавание покупательского поведения внутри любого типа магазинов через камеры наблюдения. Командой «А» было решено не использовать достаточно распространенные нейросети для детекции и распознавания человеческих образов, так как они не давали интерпретации информации о конкретном поведении покупателя (например, посмотрел на конкретную полку / задержался у интерактивного стенда / взял такой-то продукт).

Генератор синтетических данных для команды «А» является продолжением наработанных идей и технических подходов с командой «Б», но с рядом ключевых отличий:

1. Необходимо большее разнообразие персонажей как по физическим параметрам (тон кожи, рост, пол, форма тела), так и по параметрам одежды. Так как нейросеть будет работать с разными регионами и нужно, чтобы она была обучена на разной выборке людей, изначально планировалось создание ста тысяч уникальных персонажей.
2. Необходимо было создать реалистичное окружение, состоящее из разных типов магазинов, с разными типами полок, товаров и т. д.

3. В качестве настроек виртуальной камеры необходимо было использовать параметры реальных камер наблюдения, применяемых в качестве систем наблюдения.
4. Для метаинформации в данном случае самым оптимальным форматом стало создание текстового файла, в котором указывались как глобальные настройки уровня, так и информация о трехмерных персонажах их реального кадра.

Подобные требования позволили модифицировать систему, сделав её более гибкой. Пример сгенерированного изображения на рис. 14.

Рис. 14. Трехмерные персонажи находятся в магазине одежды, они варьируются по параметрам и стоят в естественных позах

Команда «А» ещё занимается тестированиям и обучением своей нейросети, но некоторыми выводами они уже готовы поделиться:

- сгенерированного датасета уже достаточно для обучений нейросети детекции людей;
- генератор позволяет генерировать более 100 тысяч изображений в течение 20 минут;

- на сцене может размещаться до 10 тысяч персонажей, хотя это и не имеет смысла, т. к. хватает от 10 до 200 персонажей;
- настройки камеры положительно влияют на распознавание с видеопотока реальной камеры;
- желаемым следующим шагом будет добавление не только статичного рендера, но также и видеорендера.

В целом, мы ещё ожидаем финального результата от команды «А», чтобы сформировать выводы по этому кейсу окончательно.

ЗАКЛЮЧЕНИЕ

Разработанный подход для создания синтетических датасетов, состоящих из двумерных секвенций, получаемых из процедурно генерируемой трехмерной локации, включающей в себя специальные модули (генерации персонажей, генерации уровня, сбора метаинформации и т. д.), успешно использован для обучения двух различных нейронных сетей, одна из которых обучалась распознаванию и локализации в пространстве (SLAM), а другая обучалась выявлению и распознаванию жестов конкретных людей в людных общественных местах. Результаты проведенной работы были также представлены на отраслевой бизнес-конференции MIXR (Россия, Москва) в июне 2021 года в рамках доклада «Синтетические датасеты для улучшения XR-алгоритмов». Основные выводы, которые можно сделать по результатам проведенных разработок:

- генератор синтетических данных вполне успешно позволяет формировать датасеты, приближенные к реальной жизни;
- использование параметров реальных камер на виртуальных – это способ увеличения реалистичности датасета;
- необходимо более глубокое изучение влияния разных параметров симуляции (освещения, цветовой коррекции, сглаживания, пост обработки и т. д.) на работу обученной нейросети;
- использование игрового движка позволяет серьезно сократить трудозатраты на разработку генератора данных, так как это снимает необходимость в разработке собственного рендера и физических движков, кроме того, игровые

движки позволяют давать большой выбор готовых ассетов для создания виртуального трехмерного окружения;

- гибкость системы, заключающаяся как в наличии большого количества разнообразных ассетов, так и в увеличении количества программных модулей, позволит использовать генератор для различных нейросетей с разными задачами.

В качестве развития наполнения локаций датасета реалистичными персонажами, кроме логичного использования всех возможностей Metahuman, предполагаем рандомизацию лиц [36] и гибридный подход использования реальных фотографий для генерации трехмерных моделей по единственному изображению [37].

Синтетические данные интересны, и эта тема имеет простор для изучения. Их генерация может облегчить жизнь множеству исследователей со всего мира. Главными плюсами синтетических данных можно назвать возможность их быстрого сбора и генерации, в отличие от сбора и разметки натуралистичных датасетов.

СПИСОК ЛИТЕРАТУРЫ

1. AI Training Dataset Market Size, Share & Trends Analysis Report By Type (Text, Image/Video, Audio), By Vertical (IT, Automotive, Government, Healthcare, BFSI), By Region, And Segment Forecasts, 2020–2027 // Grand View Research. 2020. 100 p. URL: https://www.grandviewresearch.com/industry-analysis/ai-training-dataset-market

2. *Heeger D.J.* A model for the extraction of image flow // Proceedings of the Optical Society of America Topical Meeting on Computer Vision. 1987. P. 151–154.

3. *Barron J.L., Fleet D.J., Beauchemin S.S.* Performance of optical flow techniques // Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1992. P. 236–242.

4. *McCane B., Novins K., Crannitch D., Galvin B.* On benchmarking optical flow. // Computer Vision and Image Understanding. 2001. V. 84. P. 126–143.

5. *Otte M., Nagel H.H.* Estimation of optical flow based on higher-order spatiotemporal derivatives in interlaced and non-interlaced image sequences // Artificial Intelligence. 1995. V. 78. P. 5–43.

6. *Meister S., Kondermann D.* Real versus realistically rendered scenes for optical flow evaluation // 14th ITG Conference on Electronic Media Tech. 2011. P. 1–6.

7. *Baker S., Roth S., Scharstein D., Black M.J., Lewis J.P., Szeliski R*. A database and evaluation methodology for optical flow // IEEE 11th International Conference on Computer Vision. 2007. P. 1–8.

8. *Vaudrey T., Rabe C., Klette R., Milburn J.* Differences between stereo and motion behaviour on synthetic and real-world stereo sequences // 23rd International Conference Image and Vision Computing. 2008. P. 1–6.

9. *Dwibedi D., Misra I., Hebert M.* Cut, paste and learn: Surprisingly easy synthesis for instance detection // The IEEE International Conference on Computer Vision. 2017. P. 1–12.

10. *Mac Aodha O., Brostow G.J., Pollefeys M.* Segmenting video into classes of algorithm-suitability // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010. P. 1054–1061.

11. *Butler D.J., Wulff J., Stanley G.B., Black M.J.* A naturalistic open source movie for optical flow evaluation // ECCV 2012: Computer Vision – ECCV. 2012. P. 611–625.

12. *Onkarappa N., Sappa A.D.* Speed and texture: An empirical study on optical-flow accuracy in ADAS scenarios // IEEE Transactions on Intelligent Transportation Systems. 2014. V. 15. No. 1. P. 136–147.

13. *Qiu W., Yuille A.L.* UnrealCV: Connecting computer vision to Unreal Engine // Computer Vision – ECCV 2016. 2016. Workshops. P. 909–916.

14. *Zhang Y., Qiu W., Chen Q., Hu X., Yuille A.L.* UnrealStereo: A synthetic dataset for analyzing stereo vision // ArXiv Preprint arXiv:1612.04647. 2016. P. 1–10.

15. *Taylor G.R., Chosak A.J., Brewer P.C.* OVVV: Using virtual worlds to design and evaluate surveillance systems // 007 IEEE Conference on Computer Vision and Pattern Recognition. 2007. P. 1–8.

16. *Dosovitskiy A., Ros G., Codevilla F., Lopez A., Koltun V.* Carla: An open urban driving simulator // Conference on Robot Learning. 2016. P. 1–16.

17. *Gaidon A., Wang Q., Cabon Y., Vig E.* Virtual worlds as proxy for multi-object tracking analysis // IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 4340–4349.

18. *Ros G., Sellart L., Materzynska J., Vazquez D., Lopez A.M.* The Synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 3234–3243.

19. *Richter S.R., Hayder Z., Koltun V.* Playing for benchmarks // International Conference on Computer Vision. 2017. Iss. 8237505.

20. *Handa A., Pătrăucean V., Badrinarayanan V., Stent S., Cipolla R.* Understanding realworld indoor scenes with synthetic data // IEEE Conference on Computer Vision and Pattern Recognition. 2016. Iss. 7780811. P. 4077–4085.

21. *McCormac J., Handa A., Leutenegger S., Davison A.J.* Scenenet RGB-D: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation? // The IEEE International Conference on Computer Vision. 2017. Iss. 8237554. P. 2697–2706.

22. *Handa A., Whelan T., McDonald J., Davison A.* A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM // IEEE International Conference on Robotics and Automation. 2014. Iss. 6907054. P. 1524–1531.

23. *Song S., Yu F., Zeng A., Chang A.X., Savva M., Funkhouser T.* Semantic scene completion from a single depth image // IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 190–198.

24. *Wu Z., Song S., Khosla A., Yu F., Zhang L., Tang X., Xiao J.* 3D shapenets: A deep representation for volumetric shapes // IEEE Conference on Computer Vision and Pattern Recognition. 2015. Iss. 7298801. P. 1912–1920.

25. *Chang A.X., Funkhouser T., Guibas L., Hanrahan P., Huang Q., Li Z., Savarese S., Savva M., Song S., Su H., Xiao J., Yi L., Yu F.* ShapeNet: An Information-Rich 3D Model Repository // Tech. Rep. ArXiv preprint arXiv:1512.03012. 2015.

26. *de Souza C.R., Gaidon A., Cabon Y., Peña A.M.L.* Procedural generation of videos to train deep action recognition networks // 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 2594–2604

27. *Su H, Qi C.R., Li Y., Guibas L.J.* Render for CNN: viewpoint estimation in images using CNNs trained with rendered 3D model views // IEEE International Conference on Computer Vision. 2015. Iss. 7410665. P. 2686–2694.

28. *Movshovitz-Attias Y., Kanade T., Sheikh Y.* How useful is photo-realistic rendering for visual learning? // ECCV Workshops. 2016. P. 1–16.

29. *Zhang Y., Song S., Yumer E., Savva M., Lee J.Y., Jin H., Funkhouser T.* Physically-based rendering for indoor scene understanding using convolutional neural networks // IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 5057–5065.

30. Абдурайимов Л.Н., Халилова З.Э. Краткий обзор популярных движков для создания игровых приложений под операционную систему Android // Информационно-компьютерные технологии в экономике, образовании и социальной сфере. 2018. С. 80–86.

31. Unreal Engine // URL: https://www.unrealengine.com/

32. Unity // URL: https://unity.com/

33. *Magdics M., Sauvaget C., García R.J., Sbert M.* Post-Processing NPR Effects for Video Games // 12th ACM International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI). 2013. P. 147–156.

34. Metahuman Creator // URL: https://www.unrealengine.com/en-US/digital-humans

35. *Кугуракова В.В., Зыков Е.Ю., Касимов А.В., Ситдиков А.Г., Скобелев А.А., Шайхутдинова Е.Ф.* In situ двухдиапазонная 3D-дефектоскопия стенописей архитектурных памятников // Электронные библиотеки. 2016. Т. 19. №6. С. 538–558.

36. *Тарасов А.С., Кугуракова В.В.* Реконструкция трехмерной модели человека по единственному изображению // Электронные библиотеки. 2021. Т. 24, № 3. С. 485–504.

GENERATION OF THREE-DIMENSIONAL SYNTHETIC DATASETS

V. V. Kugurakova[1,[0000-0002-1552-4910]], V. D. Abramov[2,[0000-0003-1465-5238]], D. I. Kostyuk[3,[0000-0002-6542-6980]], R. A. Sharaeva[4,[0000-0002-2359-1873]], R. R. Gazizov[5,[0000-0002-8349-264X]], M. R. Khafizov[6,[0000-0001-7275-9102]]

[1–6] *The Institute of Information Technology and Intelligent Systems of Kazan Federal University*

[1]vlada.kugurakova@gmail.com, [2]vitaly.d.abramov@gmail.com, [3]xdxnxkx@gmail.com, [4]r.sharaeva3496@gmail.com,[5]starkindustries14579@gmail.com, [6]murkorp@gmail.com

Abstract

The work is devoted to the description of the process of developing a universal toolkit for generating synthetic data for training various neural networks. The approach used has shown its success and effectiveness in solving various problems, in particular, training a neural network to recognize shopping behavior inside stores through surveillance cameras and training a neural network for recognizing spaces with augmented reality devices without using auxiliary infrared cameras. Generalizing conclusions allow planning the further development of technologies for generating three-dimensional synthetic data.

Keywords*: synthetic data, synth data, dataset, artificial intelligence, AI, neural networks, NN, machine learning, ML, computer vision, three-dimensional models, 3D, metahuman, game engine, unreal engine, UE*

REFERENCES

1. AI Training Dataset Market Size, Share & Trends Analysis Report By Type (Text, Image/Video, Audio), By Vertical (IT, Automotive, Government, Healthcare, BFSI), By Region, And Segment Forecasts, 2020–2027 // Grand View Research. 2020. 100 p. URL: https://www.grandviewresearch.com/industry-analysis/ai-training-dataset-market

2. *Heeger D.J.* A model for the extraction of image flow // Proceedings of the Optical Society of America Topical Meeting on Computer Vision. 1987. P. 151–154.

3. *Barron J.L., Fleet D.J., Beauchemin S.S.* Performance of optical flow techniques // Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1992. P. 236–242.

4. *McCane B., Novins K., Crannitch D., Galvin B.* On benchmarking optical flow. // Computer Vision and Image Understanding. 2001. V. 84. P. 126–143.

5. *Otte M., Nagel H.H.* Estimation of optical flow based on higher-order spatio-temporal derivatives in interlaced and non-interlaced image sequences // Artificial Intelligence. 1995. V. 78. P. 5–43.

6. *Meister S., Kondermann D.* Real versus realistically rendered scenes for optical flow evaluation // 14th ITG Conference on Electronic Media Tech. 2011. P. 1–6.

7. *Baker S., Roth S., Scharstein D., Black M.J., Lewis J.P., Szeliski R*. A database and evaluation methodology for optical flow // IEEE 11th International Conference on Computer Vision. 2007. P. 1–8.

8. *Vaudrey T., Rabe C., Klette R., Milburn J.* Differences between stereo and motion behaviour on synthetic and real-world stereo sequences // 23rd International Conference Image and Vision Computing. 2008. P. 1–6.

9. *Dwibedi D., Misra I., Hebert M.* Cut, paste and learn: Surprisingly easy synthesis for instance detection // The IEEE International Conference on Computer Vision. 2017. P. 1–12.

10. *Mac Aodha O., Brostow G.J., Pollefeys M.* Segmenting video into classes of algorithm-suitability // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010. P. 1054–1061.

11. *Butler D.J., Wulff J., Stanley G.B., Black M.J.* A naturalistic open source movie for optical flow evaluation // ECCV 2012: Computer Vision – ECCV. 2012. P. 611–625.

12. *Onkarappa N., Sappa A.D.* Speed and texture: An empirical study on optical-flow accuracy in ADAS scenarios // IEEE Transactions on Intelligent Transportation Systems. 2014. V. 15. No. 1. P. 136–147.

13. *Qiu W., Yuille A.L.* UnrealCV: Connecting computer vision to Unreal Engine // Computer Vision – ECCV 2016. 2016. Workshops. P. 909–916.

14. *Zhang Y., Qiu W., Chen Q., Hu X., Yuille A.L.* UnrealStereo: A synthetic dataset for analyzing stereo vision // ArXiv Preprint arXiv:1612.04647. 2016. P. 1–10.

15. *Taylor G.R., Chosak A.J., Brewer P.C.* OVVV: Using virtual worlds to design and evaluate surveillance systems // 007 IEEE Conference on Computer Vision and Pattern Recognition. 2007. P. 1–8.

16. *Dosovitskiy A., Ros G., Codevilla F., Lopez A., Koltun V.* Carla: An open urban driving simulator // Conference on Robot Learning. 2016. P. 1–16.

17. *Gaidon A., Wang Q., Cabon Y., Vig E.* Virtual worlds as proxy for multi-object tracking analysis // IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 4340–4349.

18. *Ros G., Sellart L., Materzynska J., Vazquez D., Lopez A.M.* The Synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 3234–3243.

19. *Richter S.R., Hayder Z., Koltun V.* Playing for benchmarks // International Conference on Computer Vision. 2017. Iss. 8237505.

20. *Handa A., Pătrăucean V., Badrinarayanan V., Stent S., Cipolla R.* Understanding realworld indoor scenes with synthetic data // IEEE Conference on Computer Vision and Pattern Recognition. 2016. Iss. 7780811. P. 4077–4085.

21. *McCormac J., Handa A., Leutenegger S., Davison A.J.* Scenenet RGB-D: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation? // The IEEE International Conference on Computer Vision. 2017. Iss. 8237554. P. 2697–2706.

22. *Handa A., Whelan T., McDonald J., Davison A.* A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM // IEEE International Conference on Robotics and Automation. 2014. Iss. 6907054. P. 1524–1531.

23. *Song S., Yu F., Zeng A., Chang A.X., Savva M., Funkhouser T.* Semantic scene completion from a single depth image // IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 190–198.

24. *Wu Z., Song S., Khosla A., Yu F., Zhang L., Tang X., Xiao J.* 3D shapenets: A deep representation for volumetric shapes // IEEE Conference on Computer Vision and Pattern Recognition. 2015. Iss. 7298801. P. 1912–1920.

25. *Chang A.X., Funkhouser T., Guibas L., Hanrahan P., Huang Q., Li Z., Savarese S., Savva M., Song S., Su H., Xiao J., Yi L., Yu F.* ShapeNet: An Information-Rich 3D Model Repository // Tech. Rep. ArXiv preprint arXiv:1512.03012. 2015.

26. *de Souza C.R., Gaidon A., Cabon Y., Peña A.M.L.* Procedural generation of videos to train deep action recognition networks // 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 2594–2604

27. *Su H, Qi C.R., Li Y., Guibas L.J.* Render for CNN: viewpoint estimation in images using CNNs trained with rendered 3D model views // IEEE International Conference on Computer Vision. 2015. Iss. 7410665. P. 2686–2694.

28. *Movshovitz-Attias Y., Kanade T., Sheikh Y.* How useful is photo-realistic rendering for visual learning? // ECCV Workshops. 2016. P. 1–16.

29. *Zhang Y., Song S., Yumer E., Savva M., Lee J.Y., Jin H., Funkhouser T.* Physically-based rendering for indoor scene understanding using convolutional neural networks // IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 5057–5065.

30. *Abdurajimov L.N., Halilova Z.E.* Kratkij obzor populyarnyh dvizhkov dlya sozdaniya igrovyh prilozhenij pod operacionnuyu sistemu Android // Informacionno-komp'yuternye tekhnologii v ekonomike, obrazovanii i social'noj sfere. 2018. S. 80–86.

31. Unreal Engine // URL: https://www.unrealengine.com/

32. Unity // URL: https://unity.com/

33. *Magdics M., Sauvaget C., García R.J., Sbert M.* Post-Processing NPR Effects for Video Games // 12th ACM International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI). 2013. P. 147–156.

34. Metahuman Creator // URL: https://www.unrealengine.com/en-US/digital-humans

35. *Kugurakova V.V., Zykov E.Yu., Kasimov A.V., Sitdikov A.G., Skobelev A.A., Shaykhutdinova E.F*. In situ two-diagnostic 3d-defectoscopy of the frescoes architectural monuments // Russian Digital Libraries Journal. 2016. V. 19, N0. 6. P. 538–558.

36. *Tarasov A.S., Kugurakova V.V. R*econstruction of a three-dimensional human model from a single image // Russian Digital Libraries Journal. 2021. V. 24, No. 3. P. 485–504.

СВЕДЕНИЯ ОБ АВТОРАХ

КУГУРАКОВА Влада Владимировна — кандидат технических наук, доцент Института информационных технологий и интеллектуальных систем Казанского федерального университета. Область научных интересов: расширенная реальность, разработка игр.

Vlada Vladimirovna KUGURAKOVA – PhD (tech. science), Associate Professor of the Institute of Information Technologies and Intelligent Systems, Kazan Federal University. Research interests: mixed reality, game development.

Email: vlada.kugurakova@gmail.com

ORCID 0000-0002-1552-4910

АБРАМОВ Виталий Денисович — старший преподаватель Института информационных технологий и интеллектуальных систем Казанского федерального университета. Область научных интересов: расширенная реальность, разработка игр, синтетические данные.

Vitaly Denisovich ABRAMOV – head teacher at the Institute of Information Technologies and Intelligent Systems, Kazan Federal University. Research interests: mixed reality, game development, synth data.

Email: vitaly.d.abramov@gmail.com

ORCID 0000-0003-1465-5238

КОСТЮК Даниил Иванович — старший преподаватель Института информационных технологий и интеллектуальных систем Казанского федерального университета. Область научных интересов: программирование, компьютерная графика.

Daniil Ivanovich KOSTIUK – head teacher at the Institute of Information Technologies and Intelligent Systems, Kazan Federal University. Research interests: computer programming, computer graphic.

Email: xdxnxkx@gmail.com

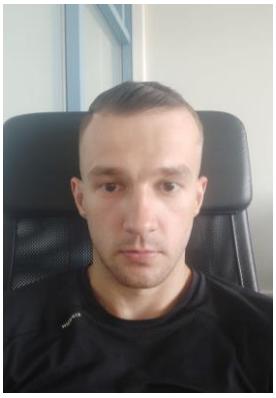
ORCID 0000-0002-6542-6980

ШАРАЕВА Регина Айратовна — инженер Института информационных технологий и интеллектуальных систем Казанского федерального университета. Область научных интересов: проектный менеджмент.

Regina Airatovna SHARAEVA – engineer of the Institute of Information Technologies and Intelligent Systems, Kazan Federal University. Research interests: project management.

Email: r.sharaeva3496@gmail.com

ORCID 0000-0002-2359-1873



ГАЗИЗОВ Рим Радикович — ассистент Института информационных технологий и интеллектуальных систем Казанского федерального университета. Область научных интересов: компьютерная графика, 3D моделирование, визуализация.

Rim Radikovich GAZIZOV – assistant of the Institute of Information Technologies and Intelligent Systems, Kazan Federal University. Research interests: computer graphic, 3D modeling, animation, visualization.

Email: starkindustries14579@gmail.com

ORCID 0000-0002-8349-264X

ХАФИЗОВ Мурад Рустэмович — старший преподаватель Института информационных технологий и интеллектуальных систем Казанского федерального университета. Область научных интересов: разработка игр, виртуальная реальность, синтетические данные.

Murad Rustemovich KHAFIZOV – head teacher at the Institute of Information Technologies and Intelligent Systems, Kazan Federal University. Research interests: game development, virtual reality, synth data.

Email: murkorp@gmail.com

ORCID 0000-0001-7275-9102

Материал поступил в редакцию 19 июля 2021 года