УДК 004.93

ПОСТРОЕНИЕ ДИНАМИЧЕСКОЙ СЦЕНЫ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

BUILDING A DYNAMIC SCENE USING NEURAL NETWORK TECHNOLOGIES

Хисамов Георгий Эдуардович
магистрант,
Уфимский государственный нефтяной технический университет,
Khisamovg2@mail.ru

Чиганова Наталья Викторовна
доцент,
Кандидат физико-математических наук,
Уфимский государственный нефтяной технический университет,
Natali-TH@yandex.ru

Аннотация. В работе формулируется идея по построению динамической сцены на основе видеоизображения в потоковом формате. Главный акцент делается на технологиях распознавания объектов, с учетом ряда ограничительных условий, формирующих наиболее легковесную систему для интеграции в уже функционирующие предприятия. На основе сравнительного анализа существующих реализаций выдвигаются основные аспекты существующих технологий, опираясь на которые, реализовывается прототип обозначенной идеи системы динамической реконструкции 3D-сцен в реальном времени с использованием нейросетей.

Ключевые слова: Иммерсивное приложение, нейросети, компьютерное зрение, 3-D реконструкция, разработка прототипа, Unity.

Khisamov Georgy Eduardovich
Masters student,
Ufa State Petroleum Technological University,
Khisamovg2@mail.ru

Chiganova Natalia Viktorovna
Associate Professor,
Candidate of Physical and Mathematical Sciences,
Ufa State Petroleum Technological University,
Natali-TH@yandex.ru

Annotation. This paper formulates an idea for constructing a dynamic scene based on streaming video. The main focus is on object recognition technologies, taking into account a number of restrictive conditions that form the most lightweight system for integration into already functioning enterprises. Based on a comparative analysis of existing implementations, the main aspects of existing technologies are put forward, based on which a prototype of the outlined idea of a system for dynamic reconstruction of 3D scenes in real time using neural networks is being implemented.

Keywords: immersive app, neural networks, computer vision, 3-D reconstruction, prototype development, Unity.

Современные технологии дополненной и виртуальной реальности подразумевают воссоздание статичной сцены, выполняющей роль фундамента для различных иммерсивных приложений [1], таких как тренажеры по технике безопасности или интерфейсы операторов роботов. Чтобы обеспечить возможность приложения моделирования в заранее неизвестной среде, необходимо разобраться с проблемой построения динамической сцены.

Для решения данной проблемы необходимо построение динамической трехмерной сцены в режиме реального времени. Рассмотрим реализацию задачи как распознавание окружающего пространства на основе видеопотока.

Исходя из поставленной цели, были сформулированы следующие задачи:

– необходимо исследовать существующие методы реализации компьютерного зрения, в частности, системы распознавания изображений;

– нужно провести анализ существующих инструментов, необходимых для реализации тех или иных аспектов прикладного приложения;

– реализация прототипа готового решения, наглядно демонстрирующей сформированную идею.

В рамках выполнения задач, можно выделить следующие ключевые аспекты, выгодно выделяющие разработку на фоне прямых и косвенных аналогов:

– работоспособность без использования GPU – в современных реалиях повсеместного кризисы рынка комплектующих, необходимо, чтобы разрабатываемое решение было работоспособным без видеоускорителя;

– скорость вычислений при помощи алгоритмов должна быть адаптирована под потоковую обработку изображений, в формате видео с минимальными задержками;

– обеспечение максимальной гибкости – возможность конфигурации под конкретную задачу и специфику, без существенных затрат по времени и ресурсам.

Распознавание объектов может осуществляться при помощи классических алгоритмов, основанных на пороговой бинаризации и морфологической обработке [2], что показывает высокое качество и скорость обработки, подходящей в том числе для использования в обработке изображений [3]. Однако, при изменении условий среды, изменению освещенности и добавлении шумов на изображении, значительно падает точность распознаваемых объектов, что накладывает существенные ограничения для использования в целях анализа динамической среды.

После рассвета технологий глубокого обучения, для множества привычных алгоритмов были найдены способы существенной оптимизации тех или иных существующих решений, сфера компьютерного зрения не стала исключением. Главным образом на развитие технологий компьютерного зрения повлияли сверточные нейронные сети (CNN), эффективно работающих с данными топологической структуры, которыми и являются изображения [4]. Важным фактором при работе с нейросетями является необходимость в крупных (в сравнении с классическими алгоритмами), изначальных затратах на осуществление подготовку данных и осуществление обучения. Однако, подобная система, взамен, предлагает большую точность и устойчивость к изменяющимся условиям окружения, критически важным при обработке видеопотока.

Анализ работы Пономарева Е.С. подтверждает то, насколько существенной является проблема недостатка GPU при использовании нейросетевых технологий для осуществления распознавания объектов [5]. В то же время автор подмечает эффективность и сравнительно высокие показатели качества при использовании модели YOLO, в том числе в целях дальнейшего предобучения на самостоятельно подобранных данных.

В целях осуществления построения 3D сцены на основе видеопотока, использование нейросетевых технологий является самым приемлемым, за счет критической важности к изменению окружения. Повышенные затраты на осуществления обучения нивелируются возможностью дальнейшей эксплуатации на low-end решениях, доступных даже небольшим предприятиям.

Рассмотрим реализацию задачи, используя готовую модель CNN нейросети YOLO.

В рамках исследования технологии распознавания объектов особенно важными являются такие параметры, как:

– перечень классов распознавания – какой конкретный перечень объектов подразумевается для обработки тем или иным алгоритмом, в рамках данной работы система использует предобученную модель YOLOv8n, содержащей в себе 80 классов объектов;

– устойчивость к трансформациям и искажениям – распознаваемый объект должен распознаваться при осуществлении манипуляций, не меняющих его принадлежность к одному из заданных классов объектов, YOLOv8 устойчив к масштабу в диапазоне 0.3–3 от исходного и обучен на повернутых объектах;

– точность и надежность – представляют собой перечень количественных критериев, олицетворяющих, насколько достоверным является прогноз по распознаванию объекта алгоритмом, в рамках работы этот показатель составляет порядка ~50–55% [6];

– показатель производительности – насколько быстро алгоритм выполняет свою работу, при данной реализации захват кадра составляет ~1–2 мс, детекция при помощи YOLO ~80–120 мс, таким образом оптимальное общее время на обработку кадра составляет порядка 2–3 секунд, чего достаточно для достижения достаточен для эффективного near-realtime (30 FPS) сканирования рабочего места;

– способность к обобщению данных – показывает, то, насколько алгоритм приспособлен к работе с материалами, выходящими за пределы обучающей выборки, система, на данный момент, требовательна к обучению расширению количества определяемых ею классов, не выходя за их перечень;

– требования к вводным данным – подразумевают то, какое количество и с каким заданным качеством информации должен обработать алгоритм как для осуществления своей задачи при эксплуатации, необходимый минимум составляют наличие камеры, поддерживающей разрешение 480p, равномерное освещение, без сильных теней и бликов, желательно использование в помещении;

Программа выполнена в формате модульной структуры, с возможностью доработки и изменению существующих методов, а также добавления различного функционала для дальнейшего построения сцены [7].

В рамках выполнения проекта был учтен каждый из пунктов задач и ключевых моментов, за счет чего созданный прототип отличается быстротой под заданные нужды обработки видеопотока, сравнительно низкими системными требованиями. Тестирование проводилось на устройстве без графического ядра, что также было заявлено как одно из требований к выполнению задания.

Стоит выделить также возможности как для дальнейшего масштабирования функциональных способностей на базе модели YOLO, так и создание и обучение собственной нейросети, если того потребует задача.

Несмотря на приведенное сравнение, подкрепленное самостоятельной реализацией, продемонстрированной на рисунке 1, следует отметить, что она не является окончательной для решения поставленной цели.

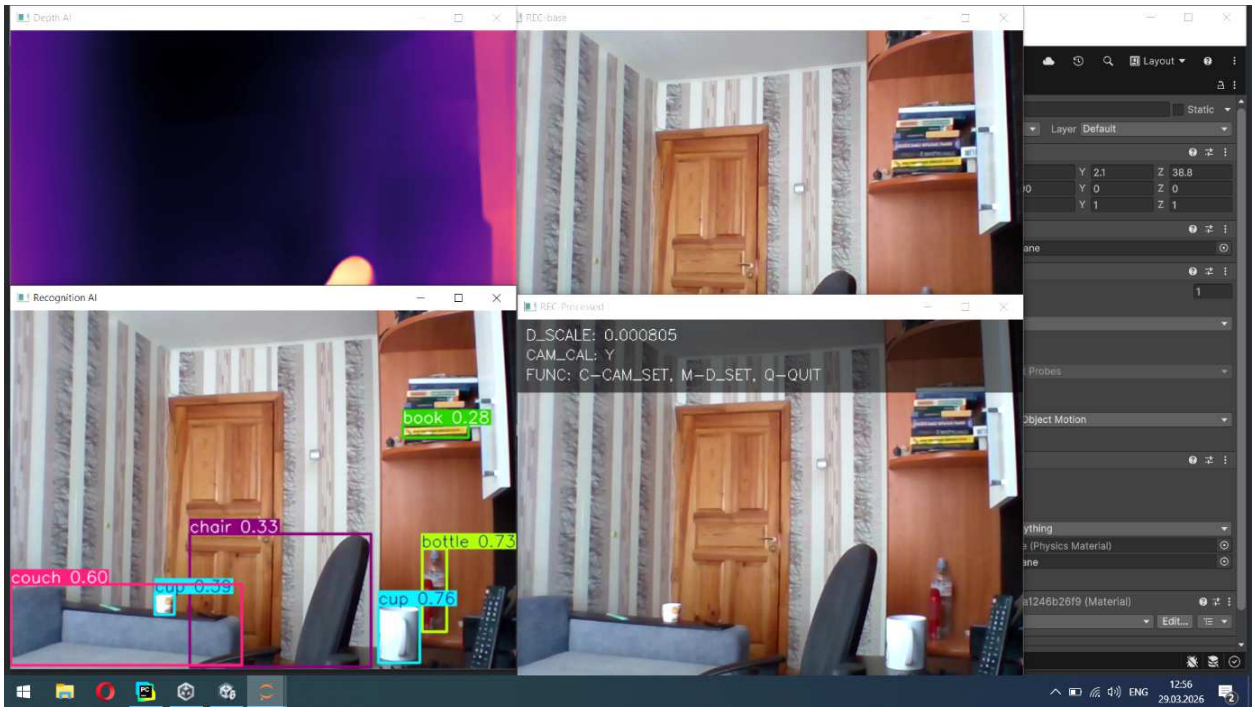

Рисунок 1 – Демонстрация работы компьютерного зрения

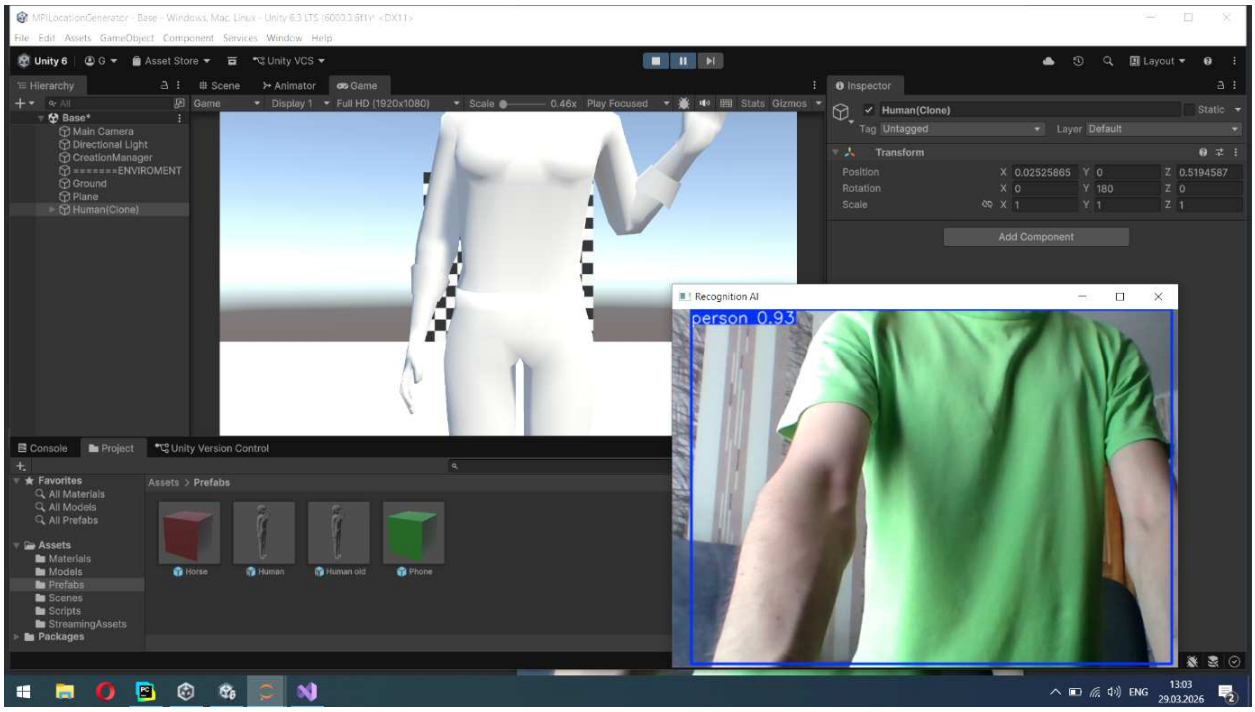

Рисунок 2 – Демонстрация сцены в Unity

Главной проблемой, из-за которой построение аутентичной сцены в Unity (рис. 2), всецело передающей реальное положение дел в окружающем пространстве, не представляется возможным, является отсутствие физических величин. Программе, состоящей лишь из распознавания объектов, не хватает данных об относительном положении, на основе которых должна осуществляться расстановка объектов в виртуальном пространстве.

В рамках выполненной работы был осуществлен анализ существующих решений для реализации задачи по построению динамической трехмерной сцены в режиме реального времени. На основе анализа были выдвинуты основные критерии для реализации, с учетом которых, на базе модели YOLOv8n была создана программа, корректно распознающая окружающее пространство на основе видеопотока.

Список литературы:

1. Глотова М.И. Основы разработки приложений виртуальной реальности: учебное пособие. – Оренбургский государственный университет, 2024. – 258 с.
2. OpenCV documentation // OpenCV. – URL : https://docs.opencv.org/4.x/ (дата обращения: 25.03.2026).
3. Практическое сравнение методов компьютерного зрения и глубокого обучения в задаче бинарной классификации изображений // Научная электронная библиотека «КиберЛенинка». – URL : https://cyberleninka.ru/article/n/prakticheskoe-sravnenie-metodov-kompyuternogo-zreniya-i-glubokogo-obucheniya-v-zadache-binarnoy-klassifikatsii-izobrazheniy/viewer (дата обращения: 26.03.2026).
4. Khan S. Transformers in Vision: A Survey // arXiv.org – международный архив электронных препринтов и постпринтов научных статей. – URL : https://arxiv.org/abs/2101.01169 (дата обращения: 25.03.2026).
5. Ускорение, сжатие и усовершенствование нейросетевых алгоритмов классификации и распознавания объектов на изображении и в видеопотоке // disserCat – электронная библиотека диссертаций. – URL : https://www.dissercat.com/content/uskorenie-szhatie-i-usovershenstvovanie-neirosetevykh-algoritmov-klassifikatsii-i-raspoznava (дата обращения: 27.03.2026).
6. Explore Ultralytics YOLOv8 // Ultralytics YOLO Docs. – URL : https://docs.ultralytics.com/models/yolov8/?spm = a2ty_o01.29997173.0.0.76555171nuPzyw#supported-tasks-and-mode (дата обращения: 28.03.2026).
7. Рочев К. В. Информационные технологии. Анализ и проектирование информационных систем: учебное пособие для вузов. – 4-е изд., стер. изд. – Издательство «Лань», 2025. – 128 с.

List of references:

1. Glotova M.I., Fundamentals of Developing Virtual Reality Applications: A Tutorial. – Orenburg State University, 2024. – 258 p.
2. OpenCV documentation // OpenCV. – URL: https://docs.opencv.org/4.x/ (Accessed: March 25, 2026).
3. Practical comparison of computer vision and deep learning methods in binary image classification // CyberLeninka Scientific Electronic Library. – URL : https://cyberleninka.ru/article/n/prakticheskoe-sravnenie-metodov-kompyuternogo-zreniya-i-glubokogo-obucheniya-v-zadache-binarnoy-klassifikatsii-izobrazheniy/viewer (Accessed: March 26, 2026).
4. Khan S. Transformers in Vision: A Survey // arXiv.org – international archive of electronic preprints and postprints of scientific articles. – URL : https://arxiv.org/abs/2101.01169 (date of access: 03/25/2026).
5. Acceleration, compression, and improvement of neural network algorithms for classification and recognition of objects in images and video streams // disserCat – electronic library of dissertations. – URL : https://www.dissercat.com/content/uskorenie-szhatie-i-usovershenstvovanie-neirosetevykh-algoritmov-klassifikatsii-i-raspoznava (date of access: 03/27/2026).
6. Explore Ultralytics YOLOv8 // Ultralytics YOLO Docs. – URL : https://docs.ultralytics.com/models/yolov8/?spm = a2ty_o01.29997173.0.0.76555171nuPzyw#supported-tasks-and-mode (accessed: March 28, 2026).
7. Rochev K.V. Information Technology. Analysis and Design of Information Systems: A Textbook for Universities. – 4th ed., reprinted. – Lan Publishing House, 2025. – 128 p.