

Методы для генерации 3D-сцен по одному изображению с использованием карт глубины и сегментации

Анастасия Р. Попова
Национальный
исследовательский
университет ИТМО
renoridoru@gmail.com

Алёна С. Сырых
Национальный
исследовательский
университет ИТМО
alyoshca.syrh@mail.ru

Глеб О. Бондаренко
Санкт-Петербургский
государственный
электротехнический
университет «ЛЭТИ»
им. В.И. Ульянова (Ленина)
Olakola9@gmail.com

Егор В. Патока
Национальный
исследовательский
университет ИТМО
yegor-patoka@mail.ru

Евгений И. Гейченко
Санкт-Петербургский
государственный
электротехнический
университет «ЛЭТИ»
им. В.И. Ульянова (Ленина)
geychenko.1995@mail.ru

Владислав С. Павлюк
Санкт-Петербургский
государственный
электротехнический
университет «ЛЭТИ»
им. В.И. Ульянова (Ленина)
vlad.pavluk03.@mail.ru

Аннотация. В данной работе представлены методы, предназначенные для генерации трехмерной сцены на основе одного входного изображения. Предлагаемый подход сочетает в себе методы предсказания карты глубины и семантической сегментации, что позволяет повысить точность пространственного восстановления и структурной согласованности сцены. Архитектура состоит из модулей глубинного анализа, семантической декомпозиции и пространственной регрессии, взаимодействующих для формирования согласованной 3D-реконструкции.

Ключевые слова: 3D-реконструкция, карта глубины, семантическая сегментация

I. ВВЕДЕНИЕ

Исследования задач реконструкции 3D-сцены из одного 2D-изображения, проводящиеся последние 8 лет, отражают стремление современных технологий к созданию систем, которые могли бы эффективно и точно восстанавливать пространственные характеристики объектов, основываясь на минимальном объеме входных данных. Это направление является чрезвычайно важным, так как открывает перспективы в таких сферах, как виртуальная и дополненная реальность, робототехника, компьютерная графика и автоматизированный анализ изображений.

Система, способная решать эту задачу, должна обладать высокой точностью восстановления глубины и структурных характеристик, быть устойчивой к шумам и артефактам входных изображений, а также обеспечивать универсальность, позволяя работать с разнообразными источниками данных. Кроме того, важными характеристиками являются быстроедействие и возможность масштабирования, что критично для обработки больших объемов данных или работы в

режиме реального времени. Поддержка современных форматов экспорта и интеграция с популярными 3D-редакторами делают такую систему особенно ценной для разработки приложений и контента.

Исследование в этой области имеет ключевое значение, так как задача преобразования 2D-изображения в 3D-модель до сих пор остаётся сложной и недостаточно изученной. Существующие решения часто ограничены по функциональности или требуют большого количества дополнительных данных, таких как стереопары или видео. Создание универсальных и точных систем не только расширяет научные и технологические горизонты, но и помогает удовлетворить растущие потребности в автоматизации процессов и генерации высококачественного трёхмерного контента.

Разработка системы для преобразования одного изображения в 3D-сцену, которая содержит в себе много объектов или стилизацию является крайне актуальной в условиях современных тенденций компьютерного зрения, виртуальной и дополненной реальности, а также требований индустрий, связанных с игровым, архитектурным и художественным дизайном. С ростом потребности в генерации качественных 3D-моделей из минимального объема исходных данных, предлагаемая система демонстрирует ряд преимуществ по сравнению с существующими аналогами

II. ЛИТЕРАТУРНЫЙ ОБЗОР

Существующие методы генерации трёхмерной сцены из одного изображения разнообразны и основаны на различных подходах, каждый из которых извлекает признаки и интерпретирует их по-своему.

Одним из основных методов являются подходы, основанные на глубинных картах, где цель заключается в оценке глубины сцены. Для этого используются глубокие нейронные сети, такие как MiDaS или DenseDepth, которые извлекают признаки, отражающие пространственное расположение объектов и их относительную глубину. После генерации карты глубины она применяется для создания облака точек или реконструкции поверхности. Этот метод прост в реализации и эффективен для сцен с явной геометрией, однако он имеет ограничения в однородных или текстурированных областях изображения [1].

Другим подходом являются методы, направленные на изучение геометрии сцены, которые позволяют моделировать полную трёхмерную геометрию, а не только глубину. Сети, такие как Pix2Vox или SceneNet, обучаются на изображениях для интерпретации их в контексте трёхмерного пространства. В результате создаётся латентное представление сцены, которое затем преобразуется в трёхмерную модель [2]. Эти методы хорошо подходят для сложных сцен, но требуют больших объемов данных для обучения.

Методы многовидового обучения используют информацию о сходстве объекта с его проекциями из других ракурсов. Даже если доступен только один вид, обученные модели могут предполагать недостающие данные. Такие подходы применяют данные, извлечённые из сетей, предобученных на многовидовых наборах, например ShapeNet или AtlasNet. Это позволяет восстанавливать скрытые поверхности, однако такие методы плохо работают с незнакомыми или уникальными объектами [3].

Гибридные подходы объединяют извлечение глубины, геометрии и текстуры, создавая более универсальные системы. Они используют как низкоуровневые признаки (глубина, края, градиенты), так и высокоуровневые (формы, текстуры, симметрия), что позволяет создавать точные трёхмерные модели. Однако такие методы сложны в реализации и требуют калибровки [4].

Современные подходы, такие как методы на основе нейронных представлений, например, Neural Radiance Fields (NeRF), обучают модели, которые кодируют плотность и цвет сцены в трёхмерном пространстве. Эти методы обеспечивают реалистичность и подходят для создания анимаций, однако отличаются высокой вычислительной сложностью.

III. МЕТОДЫ

В рамках обзора данных методов и анализа требований DenseDepth является предпочтительным выбором для решения задачи генерации трёхмерной сцены по одному изображению в условиях разнообразной стилистики, количества объектов, их геометрической сложности и разнообразия по нескольким ключевым причинам. На рис. 1 изображен результат предсказания карты глубины с помощью архитектуры, основанной на DenseDepth.



Рис. 1. Результат предсказания карты глубины

Для улучшения точности карты глубины после её создания применяется постобработка, в том числе сглаживание и фильтрация для удаления артефактов и шумов, которые могут возникать в процессе генерации глубины. Пусть для каждого пикселя (u, v) с глубиной $D(u, v)$ 3D координаты (X, Y, Z) вычисляются через обратную проекцию (1):

$$\begin{cases} X = (u - c_x) \cdot \frac{D(u, v)}{f_x}; \\ Y = (v - c_y) \cdot \frac{D(u, v)}{f_y}; \\ Z = D(u, v). \end{cases} \quad (1)$$

где: (f_x, f_y) – фокусные расстояния; (c_x, c_y) – оптический центр камеры.

Далее карта глубины используется для построения облака точек. Это облако точек представляет собой набор точек в 3D-пространстве, каждая из которых соответствует пикселю изображения с определенной глубиной. Для точной реконструкции важно правильно откалибровать камеру, поскольку неточные параметры камеры могут привести к искажениям в 3D-модели. На рис. 2 и 3 показаны результаты генерации облака точек на основе карты глубины вид полной трехмерной сцены и фрагментов соответственно.



Рис. 2. Результат предсказания карты глубины трехмерной сцены



Рис. 3. Результат предсказания карты глубины трехмерной сцены (фрагменты)

После получения облака точек применяется метод реконструкции поверхности, например Poisson Surface Reconstruction. Этот метод преобразует облако точек в сплошную 3D-сетку, вычисляя нормали для точек, что помогает в корректном построении поверхности модели. Его работа сильно зависит от точности нормалей, и, если они были вычислены неверно, это может привести к дефектам в реконструированной модели. Также алгоритм может быть чувствителен к шуму и выбросам в данных, что повлияет на качество восстановленной поверхности [5].

Для получения отдельных объектов на трехмерной сцене предложено провести сегментацию изначального изображения на отдельные сектора, а полученную маску сегментации использовать для задачи сегментации облака точек. Для уточнения сегментации предварительно необходимо удалить статистические выбросы из облака точек. Точка p_i рассматривается как выброс, если (2):

$$\frac{1}{k} \sum_{j=1}^k \|p_i - p_j\| > \mu + \alpha \cdot \sigma \quad (2)$$

где: μ – среднее расстояние до соседей; α – пороговый коэффициент; σ – стандартное отклонение.

Удаление статистических выбросов из облака точек является важным этапом предварительной обработки данных, особенно в задачах, связанных с анализом, сегментацией и визуализацией трёхмерных объектов. В данном случае используется метод, основанный на оценке плотности ближайших соседей, для идентификации и удаления аномальных точек. Сначала задаются два ключевых параметра: количество ближайших соседей и множитель стандартного отклонения.

Метод вычисляет среднее расстояние между каждой точкой облака и её ближайшими соседями. Затем оценивается глобальное среднее значение таких расстояний для всего облака точек. Точки, у которых расстояния до соседей существенно превышают это среднее значение, классифицируются как выбросы. Это позволяет учитывать локальную плотность точек и учитывать вариативность структуры облака точек.

На рис. 4 представлены результаты без удаления статистических выбросов и с удалением соответственно.



Рис. 4. Результат предсказания карты глубины трехмерной сцены без удаления и с удалением статистических выбросов

Сегментация облака точек опирается на два ключевых источника данных: сегментированное изображение, на котором области интереса уже классифицированы по категориям, и карту глубины, которая предоставляет информацию о расстоянии от камеры до каждой точки сцены.

На первом этапе проводится сегментация изображения, что может быть достигнуто с использованием современных алгоритмов машинного обучения, таких как модели на основе глубоких сверточных сетей, например, DeepLab или Mask R-CNN.

На рис. 5 изображена детекция с классификацией объектов, а на рис. 6 – сегментация объектов сцены и их отделение от фона изображения (в частности стен, окна и картины).



Рис. 5. Результат детекции и классификации исходного изображения

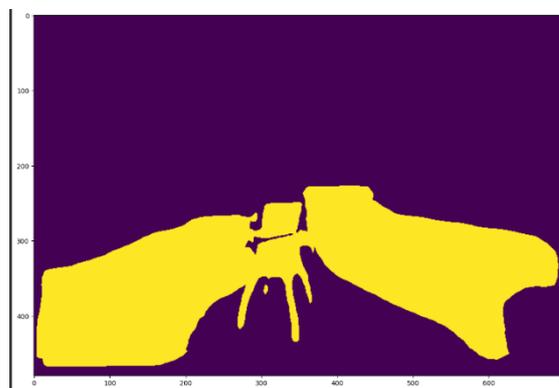


Рис. 6. Результат сегментации исходного изображения

Этот этап позволяет выделить регионы, соответствующие различным объектам или областям сцены, с точным определением их границ в двухмерном пространстве. Параллельно генерируется карта глубины, которая может быть получена с помощью сенсоров.

На следующем этапе данные из сегментированного изображения и карты глубины объединяются для восстановления облака точек с дополнительной семантической информацией. Каждому пикселю карты глубины соответствует точка в трёхмерном пространстве, координаты которой вычисляются на основе значений глубины и калибровки камеры. Эта информация преобразуется в облако точек, где каждая точка наследует метку сегмента из соответствующего пикселя сегментированного изображения. Таким

образом, облако точек обогащается семантическими данными, что позволяет разделять его на области, соответствующие различным объектам или категориям.

Сегментация облака точек, основанная на этом подходе, имеет множество преимуществ. Она позволяет использовать высокую точность и контекстуальную информацию, доступную в двумерной сегментации, в трёхмерных данных, что особенно важно для задач, требующих высокой детализации и точности, таких как робототехника, автономное вождение или создание цифровых моделей окружающей среды. Кроме того, совместное использование карты глубины и сегментированного изображения уменьшает влияние шумов, характерных для облаков точек, полученных исключительно из сенсоров глубины. Этот подход также является вычислительно эффективным, поскольку обработка на уровне изображения выполняется быстрее и с меньшими требованиями к памяти, чем обработка трёхмерных данных.

В результате сегментированное облако точек позволяет проводить точный анализ объектов, включая их локализацию, классификацию и реконструкцию, что делает этот метод важным инструментом в современных системах обработки пространственных данных.

IV. ЗАКЛЮЧЕНИЕ

В ходе выполнения работы были исследованы и изучены существующие методы для прогнозирования признаков для дальнейшей генерации трёхмерной сцены по одному изображению с помощью нейронных сетей, и был предложен вариант алгоритма для генерации трёхмерной сцены, которая будет генерироваться по изображению, имеющему большое количество объектов, стилистические и художественные приемы за быстрое время и с минимальным количеством дефектов меша, связанного с коллизиями между объектами, и плохого предсказания их формы.

Также были выявлены недостатки предложенных методов в решении, но с помощью методов аугментации

изображений и добавления нескольких обучающих слоев в архитектуру модели для оценки маски глубины и подбора автоматического камеры для метода получения облака точек данные недостатки стали не столь явными.

В дальнейшем планируется осуществить реконструкцию с помощью восстановления недостающих частей сегментированного облака точек, так как облака точек зачастую содержат данные только с одной стороны объектов. Это будет осуществляться с помощью симметричных предположений, обучения моделей на основе данных других объектов и логической интерполяции между существующими точками. После получения полной модели объекта будет проводиться постобработка, включающая упрощение сетки для снижения вычислительных затрат, сглаживание поверхностей для устранения артефактов и наложение текстур на основе исходных изображений, а также сборка всей сцены.

СПИСОК ЛИТЕРАТУРЫ

- [1] Siyu Hu, Xuejin Chen, Xin Tong, "Point sets joint registration and co-segmentation", *The Visual Computer*, vol.35, no.12, pp.1841, 2019.
- [2] Liwen Huang, Lan Tang, Chenglei Wang, Yun Chen, "Ultra-short-term Prediction of Small-sample Photovoltaic Power Based on WGAN-GP and BiLSTM-NGO", *2024 IEEE 2nd International Conference on Power Science and Technology (ICPST)*, pp.1587-1593, 2024.
- [3] K. Wang, J. Zou, S. Lu, J. Wang and B. Zhou, "Wind Power Output Scene Reconstruction Method Considering Wind Speed Autocorrelation and Source Load Cross-Correlation," *2022 5th International Conference on Power and Energy Applications (ICPEA)*, Guangzhou, China, 2022, pp. 687-692
- [4] Y. Ren and J. Liu, "Single Outdoor Image Depth Map Generation Based on Scene Classification and Object Detection," *2014 7th International Conference on Ubi-Media Computing and Workshops*, Ulaanbaatar, Mongolia, 2014, pp. 91-94
- [5] Y. Zhao, G. Lin, J. Zheng, Y. You and H. Fu, "Single Domain Generalization For Scene Classification Using Style-Oriented Data Augmentation," *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, Athens, Greece, 2024, pp. 8811-8815