

УДК 004.93

Сравнительный анализ методов восстановления 3D-геометрии по фотографиям

Иванов Алексей Сергеевич

ivanovas@student.bmstu.ru

Гагарин Юрий Евгеньевич

gagarin_ye@bmstu.ru

КФ МГТУ им. Н.Э. Баумана, Калуга, Россия

Представлен сравнительный анализ методов восстановления 3D-геометрии объектов по фотографиям, включая классическую фотограмметрию, нейросетевые подходы (NeRF, Mesh R-CNN, неявные представления) и гибридные методы. Рассмотрены их преимущества, недостатки и применимость в практических задачах.

Ключевые слова: цифровизация, компьютерное зрение, фотограмметрия, искусственный интеллект, нейросети

Введение. Восстановление 3D-геометрии объектов по фотографиям является одной из ключевых задач компьютерного зрения, находящей широкое применение в архитектуре, реставрации и создании виртуальных сред. С развитием технологий появились несколько категорий методов 3D-реконструкции объектов по изображениям. Целью данной статьи является их сравнительный анализ с целью нахождения рационального подхода к этой задаче с учетом различных метрик, таких как требования к исходным данным, производительность, точность, затраты вычислительных ресурсов и сложность

Обзор существующих методов. На данный момент существуют три основных подхода к восстановлению 3D-геометрии по изображениям.

1. Классическая фотограмметрия. Фотограмметрия — дисциплина, занимающаяся определением геометрических характеристик объектов по их изображениям. Классический подход объединяет под собой несколько алгоритмов восстановления 3D-геометрии, основными из которых являются:

- стереозрение. Имитирует человеческое бинокулярное зрение, используя два изображения одной сцены, снятых с разных точек. Основной принцип — триангуляция, с помощью которой на основе расстояния между камерами и их параметров определяется глубина каждой точки;

- Structure from Motion (SfM). Метод восстановления 3D-структуры сцены и положения камеры на основе набора неупорядоченных изображений, снятых с разных ракурсов. Извлекаются ключевые точки различных изображений и их дескрипторы, и между ними устанавливается соответствие; затем оценивается движение камеры (например, алгоритмом «восемь точек») и с помощью триангуляции строится облако точек;

- Multi-View Stereo (MVS). Расширение SfM, направленное на создание более плотных 3D-моделей. От исходного алгоритма берется начальная оценка положения камер, затем вычисляется плотная карта глубины для каждого изображения, и в результате облака точек объединяются в единую модель.

Преимущество классического подхода — высокая точность [1] и отсутствие обучения. Но, как отмечают исследователи Шонбергер и Фрам [2], нужны десятки изображений хорошего качества с разных ракурсов и, иногда, параметры камеры. Производительность зависит от данных: обработка тысяч изображений занимает часы или дни, особенно в MVS. Сложность реализации снижается из-за наличия готовых библиотек, таких как OpenCV и COLMAP.

2. Нейросетевые методы. Задействование искусственного интеллекта происходит с целью использования машинного обучения для предсказания геометрии и даже текстур. Существуют следующие основные подходы:

- NeRF (Neural Radiance Fields): моделирует сцену как непрерывное поле излучения и плотности, предсказывая цвет и плотность для каждой точки пространства;

- Mesh R-CNN: нейросетевая модель, которая определяет на основе R-CNN регионы изображения, где находятся объекты, предсказывает их грубую маску и инициализирует базовую 3D-форму, которую путем деформаций приводит в соответствие объектам на изображении;

- неявные представления: группа методов, моделирующая 3D-формы как непрерывные функции, определяющие их поверхности. Примерами представителей этой группы являются DeepSDF и PIFu.

Главное преимущество нейросетевых методов — гибкость: они работают с одним или несколькими изображениями, а результат генерируется за минуты без больших вычислительных мощностей. Однако обучение требует тысяч изображений с 3D-данными (карты глубины, сетки) [3], что занимает время и ресурсы. Точность варьируется: например, оценка глубины имеет ошибки 10–20 % [3], NeRF точен для конкретных сцен, но ограничен в обобщении [4]. Сложность реализации высока и требует знаний в области машинного обучения, даже с использованием таких библиотек, как PyTorch и TensorFlow.

3. Гибридные методы. Представляют собой комбинацию фотограмметрии и нейросетей. Успешным примером можно считать DeepSfM [5] — модификацию SfM, заменяющую традиционные алгоритмы поиска соответствий и Bundle Adjustment на нейросетевые аналоги, сохраняя при этом геометрическую основу классического подхода.

Гибридные методы требуют несколько изображений, схожих с требуемыми классическим подходом, но нейросети могут снизить их количество [6]. Производительность переменна, часто выше классических благодаря оптимизациям вроде ускорения поиска соответствий [5]. Они сочетают точность классики с адаптивностью нейросетей, достигая потенциально большей точности, хотя она варьируется. Сложность реализации высока и требует знания обеих групп методов, а обучение нейросетей занимает значительные ресурсы.

Сравнение методов 3D-реконструкции. Сравнение рассмотренных выше подходов к восстановлению 3D-геометрии может быть представлено в виде следующей таблицы.

Сравнение методов 3D-реконструкции объектов по изображениям

Критерий	Классическая фотограмметрия	Нейросетевые методы	Гибридные методы
Требования к исходным данным	Множество изображений, могут быть нужны параметры камеры	1+ изображений для вывода, большие наборы для обучения	Несколько изображений, схожие с классическими, с улучшениями нейросетей
Производительность	Медленный вывод (часы-дни)	Быстрый вывод (минуты), долгое обучение (часы-дни)	Переменная, потенциально быстрее классических
Точность	Высокая, сильно зависит от качества данных	Переменная, зависит от данных и метода	Потенциально выше, чем у иных методов
Затраты вычислительных ресурсов	Умеренные–высокие	Высокие (обучение), умеренные (вывод)	Переменные, зависят от подхода
Сложность реализации	Умеренная, есть библиотеки	Высокая, но есть готовые решения	Высокая, требуется экспертиза в обеих областях

Заключение. На основе сравнения классическая фотограмметрия обеспечивает высокую точность, но требует много ресурсов и качественных данных. Нейросетевые методы гибки, работают даже с одним изображением, но точность зависит от данных, а обучение ресурсоемко. Гибридные подходы (DeepSfM) сочетают их преимущества, повышая точность при меньших затратах данных, хотя сложнее в реализации.

Таким образом, выбор метода зависит от задачи: классическая фотограмметрия предпочтительна для высокоточных реконструкций с достаточным количеством данных, нейросетевые методы подходят для быстрых решений с ограниченными ресурсами, а гибридные методы представляют собой перспективное направление для сложных сценариев, требующих баланса между точностью, скоростью и гибкостью.

Литература

- [1] Чибуничев А.Г. *Фотограмметрия*. Москва, Изд-во МИИГАиК, 2022, 328 с.
- [2] Schönberger J.L., Frahm J.M. Structure-from-Motion Revisited. CVPR, 2016.
- [3] Eigen D., Puhrsch C., Fergus R. Depth map prediction from a single image using a multi-scale deep network. NIPS, 2014, pp. 2366–2374.
- [4] Mildenhall B., Srinivasan P.P., Tancik M., Barron J.T., Ramamoorthi R., Ng R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. CVPR, 2020, arXiv:2003.08934v2.

- [5] Wei X., Zhang Y., Li Z., Fu Y., Xue X. DeepSfM: Structure from Motion via Deep Bundle Adjustment. ECCV, 2020. <https://doi.org/10.48550/arXiv.1912.09697>
- [6] Yi K.M., Trulls E., Sali V., Fua P., Lepetit V. Learning to find good correspondences. CVPR, 2018. <https://doi.org/10.48550/arXiv.1711.05971>

Comparative analysis of methods of 3D geometry reconstruction from photographs

Ivanov Alexey Sergeevich

ivanovas@student.bmstu.ru

Gagarin Yuriy Evgenievich

gagarin_je@bmstu.ru

Kaluga Branch of BMSTU, Kaluga, Russia

The article demonstrates a comparative analysis of methods of 3D geometry reconstruction from photographs, including classic photogrammetry, neural network approaches (NeRF, Mesh R-CNN, implicit representations), and hybrid methods. Their advantages, downsides, and applicability in practical tasks have been considered.

Keywords: digitalization, computer vision, photogrammetry, artificial intelligence, neural networks