

Алгоритм генерации трехмерных моделей местности в монокулярном случае с использованием моделей глубокого обучения

Д.Е. Усенко, Т.А. Кильчуков

Московский авиационный институт, Москва

Аннотация: Статья посвящена разработке алгоритма трехмерной реконструкции местности на основе одиночных спутниковых снимков. Предлагаемый метод основан на алгоритмическом формировании трехмерных моделей на основе выходных данных двух моделей глубокого обучения для решения задач восстановления высот и инстанс-сегментации соответственно. В работе также представлены методы обработки больших спутниковых изображений моделями глубокого обучения. Предложенный в рамках работы алгоритм позволяет значительно снизить требования к входным данным в задаче трехмерной реконструкции.

Ключевые слова: трехмерная реконструкция, глубокое обучение, компьютерное зрение, восстановление высот, сегментация, определение глубины, аппроксимация контуров.

Цифровые модели местности (далее ЦММ) играют ключевую роль в различных областях, таких, как геодезия, оборонная промышленность и урбанистика, обеспечивая трехмерное представление земной поверхности для точной навигации и планирования. Традиционные методы создания ЦММ, включая фотограмметрию, многовидовые подходы [1] и лазерное сканирование, хотя и обеспечивают высокую точность, требуют значительных затрат времени и ресурсов. В свете текущего всплеска интереса к исследованиям в области искусственного интеллекта, существует потребность в разработке новых технологий для упрощения и удешевления процессов создания ЦММ. В данной статье предложен новый подход к созданию ЦММ на основе одиночных спутниковых снимков, используя технологии глубокого обучения, что может радикально изменить динамику работы в этой области, сократив затраты и время на производство моделей.

Предлагаемый алгоритм базируется на комбинации двух ключевых этапов, реализуемых с использованием методов глубокого обучения. Первым этапом является восстановление карты высот, где модель глубокого обучения анализирует спутниковый снимок и генерирует двумерное изображение,

каждому пикселю которого соответствует значение высоты в метрах. Второй этап заключается в инстанс-сегментации объектов на спутниковом снимке, с использованием отдельной модели глубокого обучения для выделения контуров зданий.

Сочетание полученной карты высот и контуров зданий позволяет построить трехмерную модель местности в формате 3D Wavefront OBJ. Данный формат широко применяется в сфере трехмерного моделирования и обеспечивает хранение геометрической информации о модели, включая вершины, ребра и грани.

На рис. 1 продемонстрирован общий принцип предлагаемого алгоритма, где представлены исходный спутниковый снимок, полученная карта высот и выделенные контуры зданий, на основе которых генерируется трехмерная модель местности.

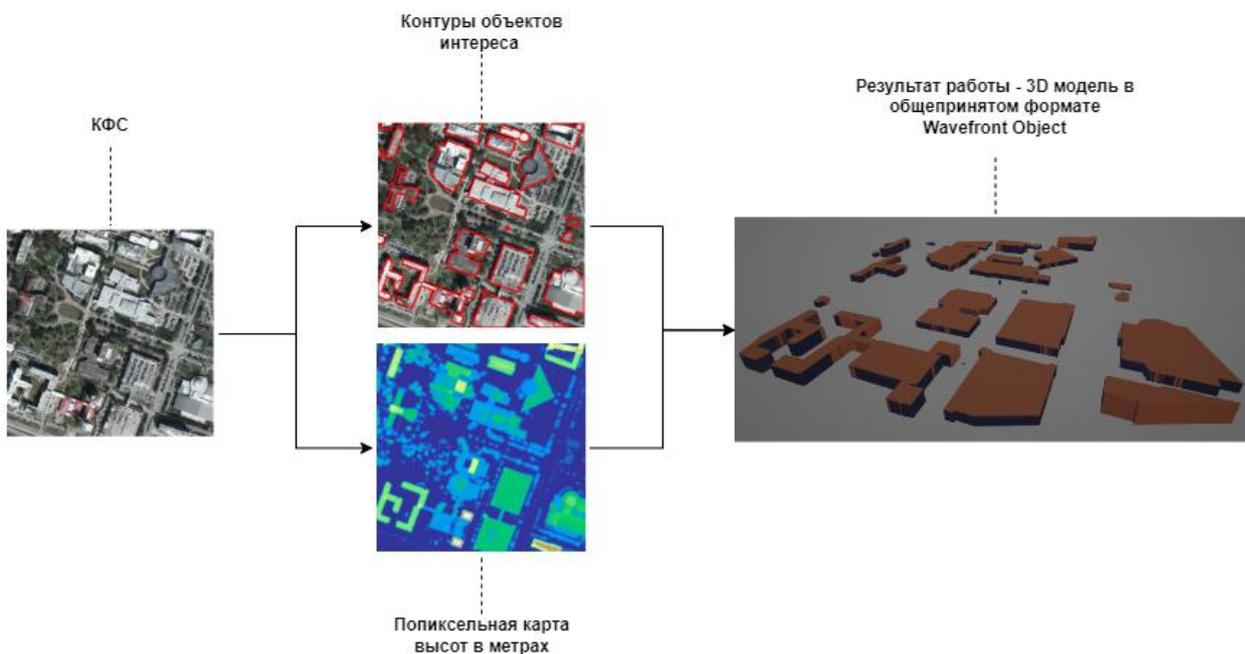


Рис. 1. — Предлагаемая методология алгоритма

В рамках исследования использовался набор данных DFC2018, включающий мультиспектральные, гиперспектральные изображения и данные LiDAR, такие как цифровые модели поверхности (далее ЦМП) и

цифровые модели рельефа (далее ЦМР). Пространственное разрешение исходных данных было снижено до одного метра на пиксель и сформированы нормализованные цифровые модели поверхности (далее нЦМП), которые будут использоваться для обучения модели восстановления высот по одиночному спутниковому снимку в видимом спектре. Для обучения модели инстанс-сегментации была произведена ручная разметка набора данных по инстансам.

Изначально задача восстановления высот с помощью моделей глубокого обучения решалась классической архитектурой сверточных нейронных сетей [2]. В рамках текущего исследования выдвигалась гипотеза о способности моделей на основе архитектуры трансформер [3], изначально проектировавшихся под решение задачи определения глубины, успешно решать аналогичную задачу регрессии — восстановления высот. Так, в текущем исследовании на основе предварительного анализа для обучения и последующего сравнения для выбора наиболее предпочтительной в рамках работы алгоритма были выбраны две модели для решения задачи восстановления высот — DPT [4] и MiDaS [5]. Для решения задачи инстанс-сегментации и проведения были выбраны модели Mask2Former [6] и YOLOv8 [7], зарекомендовавшие себя ранее в контексте обработки спутниковых снимков [8].

При обучении моделей восстановления высот использовалась функция потерь $SI Loss$ (1), впервые введенная в работе «Depth Map Prediction from a Single Image using a Multi-Scale Deep Network» [9], которая позволила снизить значение ошибки в процессе обучения на 16%, в среднем добившись понижения ошибки до 2.1 метра для модели *DPT*, в то время, как при использовании классической *MSE* функции потерь ошибка не опускалась ниже 2.5 метров.

$$D(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (\log y_i - \log \hat{y}_i + a(y_i, \hat{y}_i))^2, \#(1)$$

где \hat{y} — истинное значение карты высот, y — предсказание модели, $a(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (\log \hat{y}_i - \log y_i)$ — значение, которое минимизирует ошибку для данной пары y, \hat{y}_i , n — количество пикселей.

Для любого прогноза y , e^a является масштабом, который лучше всего соответствует истинному значению, при этом все масштабные множители y имеют одинаковую ошибку, что обеспечивает масштабную инвариантность.

Для оценки качества моделей для решения задачи восстановления высот использовались метрики MSE (2) и MAE (3).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (3)$$

В задаче восстановления высот целесообразно использовать обе эти метрики, поскольку MAE менее чувствительна к выбросам по сравнению с MSE , так как абсолютные значения не усиливают эффект от выбросов.

Для оценки качества моделей для решения задачи восстановления моделей использовалась метрика *mean Average Precision* (4).

$$mean Average Precision = \frac{1}{n} \sum_{n=1}^n \int_0^1 p(r) dr \quad (4)$$

где n — количество классов, $p(r)$ — функция Precision от Recall.

Здесь $\int_0^1 p(r) dr$ является Average Precision. Данная метрика объединяет в себе все ключевые показатели в задачах инстанс-сегментации, детекции и классификации.

При этом важно отметить, что зачастую в тестировании и оценке точности моделей инстанс-сегментации приводится mAP с разными пороговыми значениями IoU (Intersection over Union - Пересечение над

объединением), так, например, если приводится метрика mAP50, это означает, что в оценке участвовали объекты с перекрытием не менее 50%, а mAP50-95 с перекрытием от 50 до 95%.

Результаты тестирования обученных моделей на наборе данных DFC2018 приведены в таблицах 1 и 2.

Таблица 1

Тестирование и оценка точности моделей инстанс-сегментации

Метрика	mAP50	mAP50-95
Mask2Former	0.71	0.55
YOLOv8	0.76	0.51

Таблица 0

Тестирование и оценка точности моделей восстановления высот

Метрика	MSE, метров	MAE, метров
DPT	1.5	2.3
MiDaS	1.78	2.8

Таким образом, на основе проведенных экспериментов при дальнейшей разработке алгоритма принято решение использовать модели Mask2Former и DPT для решения задач восстановления высот и инстанс-сегментации соответственно.

При создании трехмерных моделей местности особое внимание уделяется аппроксимации контуров, полученных от моделей инстанс-сегментации, для снижения их сложности и последующего использования в специализированных расчетах. Для этой цели были имплементированы три алгоритма аппроксимации контуров: Рамера-Дугласа-Пеккера [10], итеративный и многоугольниками высокого порядка [11].

Пример результатов аппроксимации различными алгоритмами приведен на рис. 2.

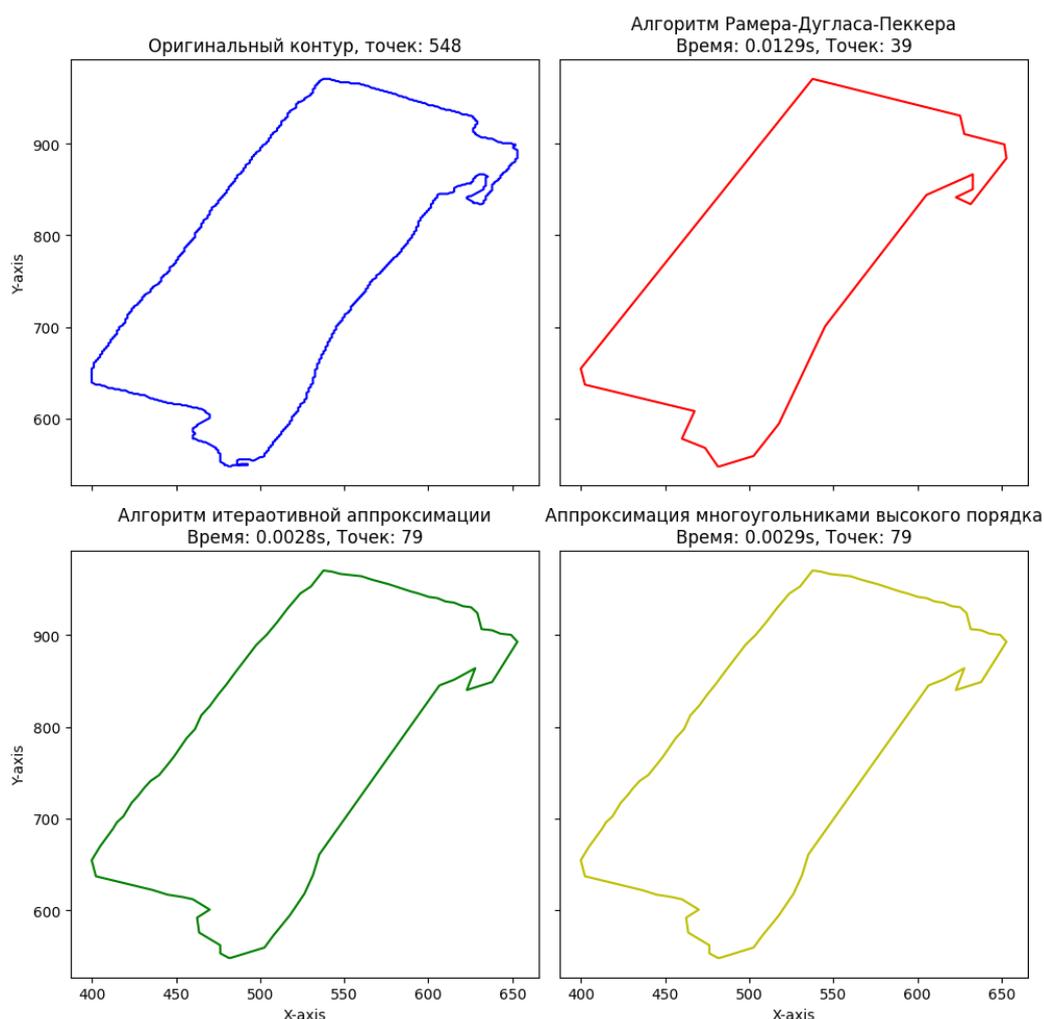


Рис. 2. — Пример аппроксимации контуров различными алгоритмами.

В связи с наибольшим упрощением контура при сохранении общей структуры здания, что выявлено эмпирической оценкой, предлагается в дальнейшем использовать в рамках алгоритма метод Рамера-Дугласа-Пеккера.

Важным аспектом при работе моделей глубокого обучения со спутниковыми снимками является непосредственная их обработка, поскольку модели не рассчитаны на обработку изображений разрешением в несколько тысяч пикселей необходимо выработать алгоритм обработки таких изображений.

Относительно задачи инстанс-сегментации решение становится не тривиальным, поскольку необходимо оптимизированно обрабатывать

смежные и перекрывающиеся контуры окон, осуществляя поиск соседних контуров, которые вместе образуют единый объект.

Для решения этой задачи был разработан алгоритм, представленный на рис. 3.

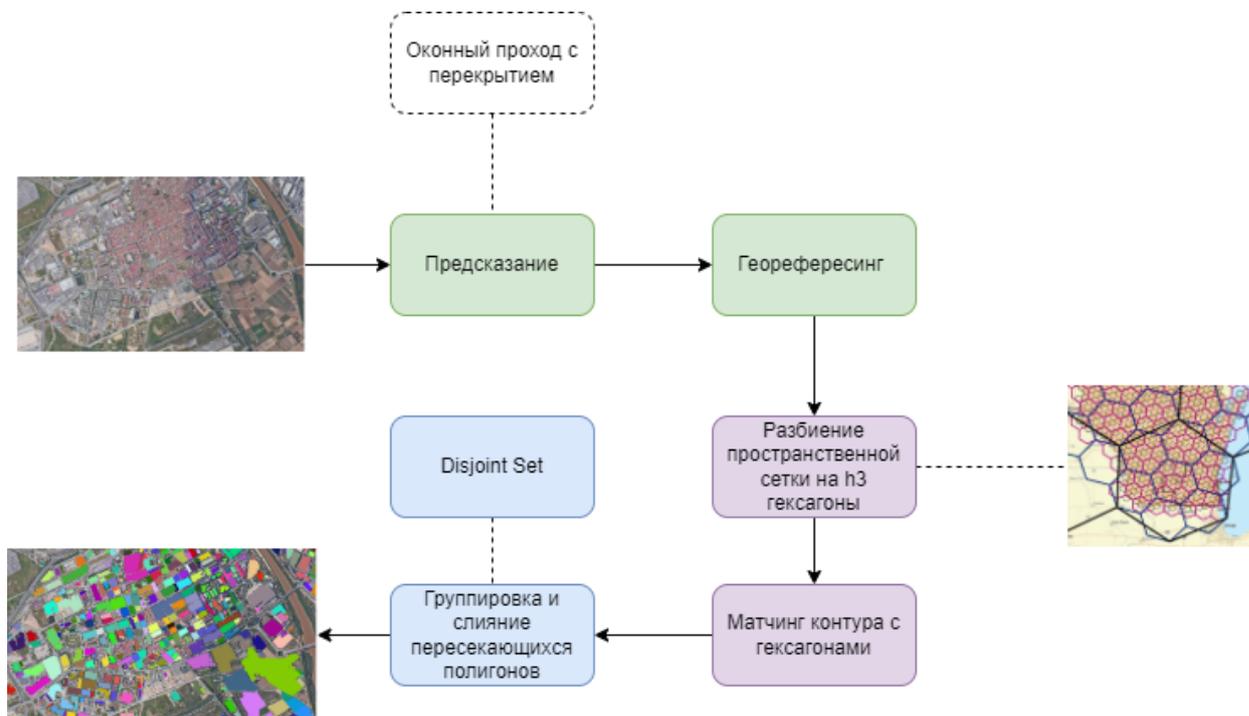


Рис. 3. — Алгоритм обработки спутниковых снимков моделями инстанс-сегментации

Концепция алгоритма заключается в делении снимка на меньшие участки с помощью методики оконного прохода для обработки частями, что полезно при ограниченных ресурсах и обеспечивает целостность предсказаний благодаря перекрытию сегментов. На каждом участке модель инстанс-сегментации идентифицирует и выделяет контуры объектов, затем предсказания переводятся в глобальные географические координаты для использования в ГИС и интеграции с другими данными. Система гексагональной индексации $h3$ организует данные, а структура данных *Disjoint Set* группирует смежные полигоны, уменьшая избыточность и формируя целостные объекты для упрощения анализа и визуализации.

Для решения задачи предсказания на больших изображениях моделью восстановления высот было адаптировано под работу системы библиотеки *EMPatches*, позволяющее выполнять оконный проход со смещением по изображению, получать предсказание модели и устранять граничный эффект за счёт различных фильтров, позволяющих определять логику расчета значений на границах окон.

Заключительный этап создания трехмерных моделей местности включает в себя генерацию моделей LOD 1-го уровня [12], основываясь на контурной и высотной информации о зданиях. Процесс начинается с триангуляции многоугольников ушным методом [13], определения трехмерных точек и создания вершин для каждого уровня здания. Затем, для обеспечения визуализации, проводится ориентация нормалей и добавление вершин в формат *.obj*.

Таким образом, после рассмотрения всех подмодулей системы, предлагается алгоритм генерации трехмерных моделей местности на основе одиночных спутниковых снимков, представленный на рис 4.

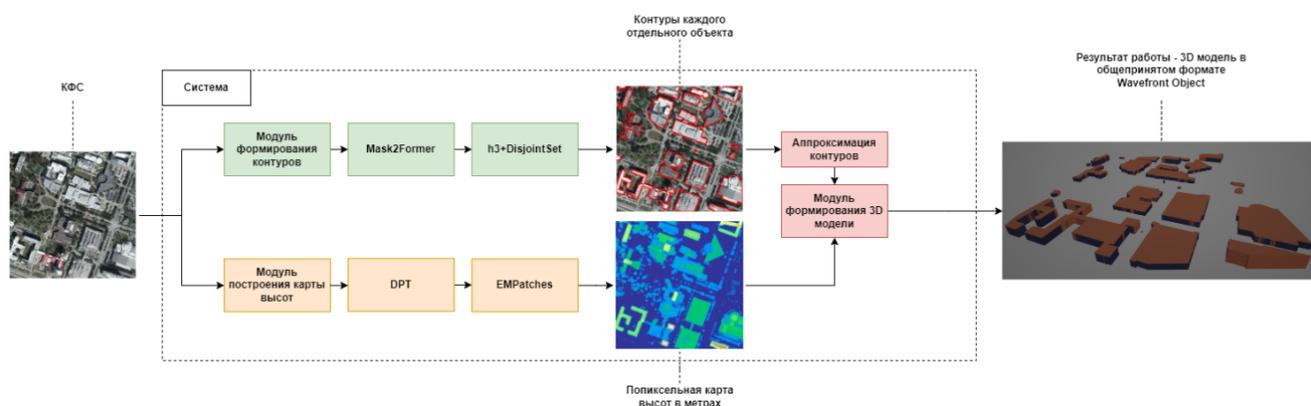


Рис. 4. — Общий алгоритм генерации трехмерных моделей местности на основе одиночных спутниковых снимков

Разработанная на основе его система подразумевает под входными данными одиночный геопривязанный спутниковый снимок в видимом спектре.

Полученный входной снимок отправляется в два независимых подмодуля для извлечения с помощью моделей глубокого обучения

информации о контурах зданий и карты высот спутникового снимка. При этом в каждый из подмодулей интегрирован алгоритм для обработки больших спутниковых снимков.

В конечном итоге, полученная из подмодулей выходная информация объединяется для формирования полноценной трехмерной сцены.

Таким образом, предложенный алгоритм позволяет осуществлять генерацию трехмерных моделей местности на основе одиночных спутниковых снимков, убирая требование большинства современных алгоритмов на наличие многовидовых снимков. Из основных направлений дальнейшего развития является повышение инвариантности моделей глубокого обучения за счет модификации процесса обучения и значительного расширения выборки, а также повышения уровня детализации генерируемых моделей.

Литература (References)

1. Luo H., Zhang J., Liu X., Zhang L., Liu J. Large-Scale 3D Reconstruction from Multi-View Imagery: A Comprehensive Review. URL: [mdpi.com/2072-4292/16/5/773#B1-remotesensing-16-00773](https://www.mdpi.com/2072-4292/16/5/773#B1-remotesensing-16-00773).
2. Najaf M., Arefi H., Amini Amirkolae H., Farajelahi B. Monocular Depth Estimation of Google Earth Images Using Convolutional Neural Networks.
3. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention Is All You Need. URL: arxiv.org/abs/1706.03762.
4. Ranftl R., Bochkovskiy A., Koltun V. Vision Transformers for Dense Prediction. URL: arxiv.org/abs/2103.13413.
5. Ranftl R., Lasinger K., Hafner D., Schindler K., Koltun V. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. URL: arxiv.org/abs/1907.01341.



6. Cheng B., Misra I., Schwing A.G., Kirillov A., Girdhar R. Masked-attention Mask Transformer for Universal Image Segmentation. URL: arxiv.org/abs/2112.01527.
7. Reis D., Kupec J., Hong J., Daoudi A. Real-Time Flying Object Detection with YOLOv8. URL: arxiv.org/abs/2305.09972.
8. Guo S., Yang Q., Xiang S., Wang S., Wang X. Mask2Former with Improved Query for Semantic Segmentation in Remote-Sensing Images. URL: mdpi.com/2227-7390/12/5/765.
9. Eigen D., Puhrsch C., Fergus R. Depth Map Prediction from a Single Image using a Multi-Scale Deep Network.
10. Douglas D., Peucker T. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature.
11. Mathur P.C., Bhattacharya D.D. High Order Polygonal Approximation of Closed Curves
12. Biljecki F., Ledoux H., Stoter J. An Improved LOD Specification for 3D Building Models. URL: sciencedirect.com/science/article/abs/pii/S0198971516300436
13. Eberly D. Triangulation by Ear Clipping URL: geomerictools.com/Documentation/TriangulationByEarClipping.pdf

Дата поступления: 9.05.2024

Дата публикации: 22.06.2024