

Санкт-Петербургский государственный университет

Кафедра системного программирования

Группа 21.Б10-мм

Сравнение алгоритмов плотной реконструкции на задаче аэрофотограмметрии

Громова Арабелла Михайловна

Отчёт по учебной практике
в форме «Эксперимент»

Научный руководитель:
Старший преподаватель кафедры СП, к.т.н. Ю. В. Литвинов

Консультант:
Инженер-исследователь Mobile Robotics Lab, Skoltech А. В. Корнилова

Санкт-Петербург
2024

Оглавление

1. Введение	3
2. Постановка задачи	5
3. Обзор	6
3.1. Алгоритмы плотной реконструкции	6
3.2. Алгоритмы без обучения	6
3.3. Алгоритмы с обучением	9
3.4. Датасеты	11
3.5. Метрики	12
3.6. Вывод	13
4. Унификация интерфейса запуска алгоритмов	14
5. Ассоциация между пикселями и точками	16
5.1. Алгоритм	16
5.2. Реализация	18
6. Эксперимент	19
6.1. Тестовый датасет	19
6.2. Условия эксперимента	20
6.3. Результаты эксперимента	21
6.4. Исследование этапа ассоциации между пикселями и точками	25
6.5. Вывод	28
Заключение	29
Список литературы	30

1. Введение

Фотограмметрия — это подход к построению трехмерной реконструкции объекта с помощью его снимков с разных ракурсов. Данный подход находит свое применение в архитектуре, в медицине, мониторинге окружающей среды. Задача создания 3D-модели актуальна и для территориального планирования. Построение трехмерных реконструкций городов позволяет оптимизировать управление ими, прогнозировать события, выявлять риски. На данный момент существуют 3D-модели Сингапура¹, Хельсинки² и некоторых других городов.

Процесс построения 3D-модели чаще всего состоит из разреженной (sparse) реконструкции (structure from motion), плотной (dense) реконструкции (multi view stereo), добавления текстур. Structure from Motion – это процесс реконструкции трехмерной структуры объекта на основе набора двумерных пересекающихся изображений. Результатом работы разреженной реконструкции является разреженное облако из ключевых точек. Multi view stereo – метод построения 3D реконструкции объекта путем интерполяции поверхности между точками разреженного облака. Результат работы плотной (dense) реконструкции – плотное облако точек.

Задача построения трехмерной модели с аэрофотоснимков не нова, но большинство современных алгоритмов в основном нацелены на построение реконструкций архитектурных объектов. На данный момент существует множество методов, направленных на улучшение этапа плотной реконструкции к архитектурным объектам. Однако применение большинства данных методов не исследовано по отношению к реконструкции по аэрофотоснимкам, используются устаревшие подходы к созданию плотного облака точек [57, 16, 55]. В связи с этим предлагается исследовать применение современных методов для улучшения этапа плотной реконструкции по аэрофотоснимкам, увеличить

¹3D-модель Сингапура: <https://www.sla.gov.sg/geospatial/gw/virtual-singapore> (дата обращения 2023-09-10)

²3D модель Хельсинки: <https://www.hel.fi/en/decision-making/information-on-helsinki/maps-and-geospatial-data/helsinki-3d> (дата обращения 2023-09-10)

их точность.

Семантическая сегментация является важной задачей в построении 3D реконструкций городов. Разделение 3D модели на семантические примитивы, такие как здания, дороги, деревья и другие объекты, открывает широкие возможности для эффективного управления городской инфраструктурой. Существуют алгоритмы сегментации облаков без опоры на снимки [38, 46, 37], но таких методов в открытом доступе гораздо меньше, чем методов сегментации аэрофотоснимков. Также существуют алгоритмы сегментации облаков точек, основанные на сегментации снимков и на установлении соответствия между точками 3D реконструкции и семантическими классами [26]. В связи с вышеизложенным предлагается реализовать ассоциацию между пикселями снимков и облаком точек, а также протестировать на задаче переноса семантической маски на 3D реконструкцию.

В рамках данной работы планируется провести обзор алгоритмов 3D-реконструкции объектов, провести сравнение методов плотной реконструкции на аэрофотоснимках, в том числе на наборе данных, предоставленном лабораторией, реализовать ассоциацию между точками облака и пикселями снимков, провести апробацию на задаче переноса семантической маски.

2. Постановка задачи

Целью работы является сравнение алгоритмов плотной реконструкции на задаче аэрофотограмметрии и реализация алгоритма ассоциации между точками облака и пикселями снимков. Для достижения данной цели были поставлены следующие задачи.

1. Выполнить обзор существующих алгоритмов плотной реконструкции объектов, а также связанных с этой задачей датасетов и метрик.
2. Реализовать воспроизводимый запуск решений с открытым исходным кодом.
3. Провести сравнение алгоритмов плотной реконструкции на целевых датасетах.
4. Реализовать ассоциацию между пикселями и облаком точек, провести апробацию на задаче переноса семантической маски.

3. Обзор

В данном разделе рассматриваются существующие алгоритмы плотной реконструкции, датасеты, которые упоминаются в работах, посвященных multi view stereo, также метрики с помощью которых оцениваются алгоритмы плотной реконструкции.

3.1. Алгоритмы плотной реконструкции

Существуют методы как традиционные, так и основанные на глубоком обучении. Согласно [11] традиционные алгоритмы можно поделить на четыре вида: основанные на вокселях, на эволюции поверхности, на точках, на картах глубины. Подходы с глубоким обучением подразделяются на основанные на вокселях и на картах глубины.

3.2. Алгоритмы без обучения

Рассмотрим более детально традиционные алгоритмы плотной реконструкции.

3.2.1. Подход, основанный на вокселях

Одним из методов плотной реконструкции, основанных на вокселях, является воксельная раскраска [45]. Данный подход восстанавливает объект не по форме, а по цвету. В воксельной раскраске используется условие согласованности по цветам, с помощью этого правила можно определить, лежит ли точка на поверхности или нет. Входными данными этого алгоритма являются изображения объекта. Первоначально трехмерное пространство разбивается на воксели, также предполагается, что пространство содержит только ламбертовы поверхности – поверхности, которые одинаково яркие со всех углов обзора. Далее воксели проецируются в пиксели изображений. Алгоритм обходит воксели и окрашивает только те, которые являются частью поверхности. Также в данном методе необходима определенная конфигурация камер,

чтобы воксели не перекрывали друг друга. Авторы работы [51] предложили оптимизацию алгоритма воксельной раскраски с помощью двуправленного распределения коэффициента отражения и согласованности ориентации [19]. Данный подход позволяет вычислять нормали для каждого вокселя, что, в свою очередь, увеличивает точность реконструкции мелких деталей объекта.

Используя метод раскраски вокселей, авторы исследования [23] разработали алгоритм Space Carving. Основное отличие Space Carving от воксельной раскраски заключается в том, что в данном алгоритме несогласованные воксели удаляются. Однако реконструкция, получившаяся в результате работы данного метода, может содержать несогласованные по цвету воксели, а также ошибочно вырезанные воксели. В работе [9] было предложено для минимизации ошибочно вырезанных вокселей в методе Space Carving назначать каждому вокселю вероятность того, что он является согласованным. Также на основе раскраски вокселей были разработаны алгоритмы [2, 1]. Важно отметить, что методы основанные на вокселях имеют проблему с восстановлением мелкозернистых поверхностей.

3.2.2. Подход, основанный на эволюции поверхности

Подход эволюции поверхности основан на постепенном уточнении поверхности. Первым шагом алгоритмов данного подхода обычно является нахождение начальной поверхности, близкой к оптимальной. Далее итеративно уменьшается функция стоимости [53, 12, 14]. В алгоритмах, использующих наборы уровней, минимизируется набор дифференциальных уравнений, определенных на поверхности [13, 47]. Стоит отметить, что к алгоритмам, основанных на эволюции, можно отнести Space Carving и его вариации [24, 5, 41, 32]. В качестве начальной поверхности используется воксельная сетка, на каждой итерации раскрашиваются согласованные воксели и удаляются несогласованные. Точность реконструкции зависит от выбора первоначальной поверхности.

3.2.3. Подход, основанный на точках

Алгоритмы данной категории [15, 27] обычно состоят из следующих основных шагов:

1. Извлечение признаков.
2. Сопоставление ключевых точек.
3. Интерполяция поверхности между ключевыми точками.

В данном подходе для построения плотного облака точек первоначально строится разреженная реконструкция с помощью алгоритма *Structure from motion*, впервые предложенного в работе [52]. Авторы работы [33] реконструируют поверхность с помощью триангуляции, вершинами треугольников являются ключевые точки. В работах [21, 28] также используют триангуляцию для восстановления поверхности. В основном интерполяция с помощью триангуляции используется в алгоритмах, строящих меш-реконструкции (полигональная сетка). Важно отметить, что для методов, основанных на точках, не требуется инициализирующая поверхность, как в подходах, основанных на вокселях и на эволюции поверхности. Точность плотных облаков точек зависит от построения разреженной реконструкции, от способа интерполяции между ключевыми точками.

3.2.4. Подход, основанный на картах глубины

Большинство современных алгоритмов основаны на точках и картах глубины [40, 16, 44, 20, 22, 34, 3]. В методах данной категории вычисляются карты глубины и нормали. Далее с помощью их объединения строится плотное облако точек. В большинстве современных алгоритмов [35, 34, 3, 16, 20] используется *PatchMatch* [7] для вычисления карт глубины. *PatchMatch* для карт глубины состоит из следующих шагов:

1. Случайная инициализация карты, каждому пикселю присваивается значение глубины.

2. Распространение значений глубины на близлежащие пиксели, сопоставление с пикселями на других картах.
3. Уточнение глубины пикселей.

В работе Silvano Galliani и др. [16] был представлен алгоритм вычисления карт глубины, авторы сопоставляют несколько снимков, что увеличивает точность карт, также в данном исследовании была предложена схема распространения глубины пикселей в виде шахматной доски. Johannes L. Schönberger и др. предложили алгоритм COLMAP [40]. Авторы этой работы увеличили точность карт глубины и нормалей, используя геометрическую и фотометрическую согласованность. Qingshan Xu и др. предложили несколько алгоритмов плотной реконструкции: АСМН [35], АСММ [35], АСМР [54] и АСММР [34]. АСМН использует шахматную схему распространения глубины пикселей для увеличения точности карт глубины. Основой АСММ и АСМР является АСМН, в АСММ авторы применяют геометрическую согласованность для более точной оценки карт, в АСМР улучшают оценку карт глубины для кадров низкого разрешения. АСММР – алгоритм, состоящий из АСММ и АСМР. В работе Yuesong Wang и др. [3] увеличивают точность сопоставления, деформируя участок, содержащий пиксели с неоднозначно определенным значением глубины. Авторы исследования [55] адаптировали COLMAP для аэрофотоснимков с геопривязкой в виде RPC (Rational Polynomial Coefficient).

3.3. Алгоритмы с обучением

Рассмотрим более детально алгоритмы плотной реконструкции с обучением.

3.3.1. Подход, основанный на картах глубины

Методы этой категории [31, 43, 10, 18] строят плотную реконструкцию объединением карт глубины. Одним из популярных нейросетевых

алгоритмов данного подхода является MVSNet [31]. Авторы этой нейронной сети используют 3D CNN для регуляризации и регрессии исходной карты глубины. Позже была предложена нейросеть RMVSNet [43], в которой вместо 3D-CNN используют GRU для снижения потребления памяти. MVSNet является основой для множества методов [43, 10, 18]. Xiaodong Gu и др. разработали нейросетевой алгоритм CasMVSNet [10] на основе MVSNet. Авторы увеличивают точность MVSNet, используя каскад. Zhe Zhang и др. представили нейронную сеть GeoMVSNet [18], основанную на CasMVSNet. В данной работе учитывают геометрию объектов, тем самым увеличивая точность предсказания значений глубины. Важно отметить, что GeoMVSNet тестировался на аэрофотоснимках. Fangjinhua Wang и др. разработали нейросетевой алгоритм PatchMatchNet [39] на основе PatchMatch [7]. Jin Liu и др. представили нейронную сеть SatMVS [17] для построения DSM (Digital surface model) с использованием аэроснимков с геопривязкой. В работе был предложен нейросетевой алгоритм вычисления карт глубины для аэрофотоснимков REDNet [29], данный метод основан на архитектуре Recurrent Encoder-Decoder. Стоит отметить, что REDNet не строит плотную реконструкцию. Corinne Stucker и др. разработали нейросеть [48] для построения DSM (Digital surface model) с использованием аэроснимков с геопривязкой.

3.3.2. Подход, основанный на вокселях

В отличие от нейронных сетей, основанных на картах глубины, нейросетевые алгоритмы на вокселях менее распространены. Mengqi Ji и др. представили нейронную сеть SurfaceNet [49], основанную на CNN. SurfaceNet предсказывает принадлежность вокселя поверхности. Despoina Paschalidou и др. представили нейронную сеть [42], вычисляющую карты глубины. Авторы данной работы тестируют нейросеть на аэрофотоснимках.

3.4. Датасеты

Рассмотрим наиболее распространенные датасеты для алгоритмов плотной реконструкции.

- DTU [25] – набор данных, содержащий облака точек, меш реконструкции, параметры камеры и изображения объектов. Датасет состоит из 124 сцен. Для построения каждой реконструкции используется 49-64 фотографий. Сцены включают широкий спектр объектов от домов до волейбольных мячей. Однако в данном датасете отсутствуют аэрофотоснимки.
- BlendedMVS [6] – синтетический набор данных, содержащий 113 сцен, более 17000 изображений, карты глубины, параметры камеры. Датасет включает объекты архитектуры, детские игрушки, синтетические аэрофотоснимки и др.
- Tanks and Temples [50] – набор данных, включающий помещения и архитектурные объекты. Датасет содержит изображения, параметры камеры, облака точек, полученные с помощью LiDAR-сканера.
- ETH3D [36] – набор данных, включающий здания и помещения. Датасет содержит изображения, параметры камеры, карты глубины, облака точек, полученные LiDAR сканером.
- WHU-MVS [29] – синтетический набор данных, состоящий из карт глубины и аэрофотоснимков уезда Мэйтань провинции Гуйчжоу. Датасет содержит 1776 кадров с разрешением 5376×5376 .
- Аэрофотоснимки Москвы – набор данных, предоставленный лабораторией. Датасет содержит более тридцати трех тысяч аэрофотоснимков района Москвы. Кадры были сняты двумя камерами VisionMap A3 EDGE.
- UrbanScene3D [30] – набор данных, состоящий из шестнадцати городских сцен, четырнадцать из которых являются синтетически-

ми. Набор данных содержит аэрофотоснимки, карты глубины, параметры камеры, меш реконструкции.

3.5. Метрики

Оценить качество работы алгоритмов плотной реконструкции можно как визуально, так и с использованием метрик. Наиболее распространенные метрики для оценки точности плотных облаков точек перечислены ниже:

1. Accuracy показывает отношение корректных точек к общему числу всех предсказанных точек и определяется следующей формулой:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \text{ где}$$

TP – True Positive, TN – True Negative, FP – False Positive, FN – False Negative.

2. F_1 – метрика, объединяющая *Precision* и *Recall*, определяется следующим образом:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \text{ где}$$

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}$$

3. Surface Density – метрика, показывающая плотность облака на квадратный метр:

$$Surface\ Density = \frac{N}{\pi R^2}, \text{ где}$$

N – количество соседей точки внутри сферы с радиусом R , R – радиус сферы.

4. Volume Density – метрика, показывающая плотность облака на кубический метр:

$$Volume\ Density = \frac{N}{\frac{4}{3}\pi R^3}, \text{ где}$$

N – количество соседей точки внутри сферы с радиусом R , R – радиус сферы.

5. Условная емкость – метрика, показывающая максимальное количество фотографий, которое может обработать алгоритм на определенной машине.
6. Время работы алгоритма.

3.6. Вывод

Из обзора следует, что, алгоритмы, основанные на картах глубины, выглядят наиболее перспективными для запуска на аэрофотоснимках. Данные методы позволяют реконструировать мелкие детали, учитывать освещение и тени. Важно отметить, что некоторые алгоритмы этой категории были адаптированы для аэрофотоснимков, также карты глубины используются для построения DSM (Digital surface model). В то же время методы на основе вокселей и эволюции поверхности менее точны, а также подходят лишь для реконструкции геометрических фигур, объектов без мелких деталей и теней. Стоит подчеркнуть, что большинство state-of-the-art алгоритмов используют карты глубины для построения плотной реконструкции.

4. Унификация интерфейса запуска алгоритмов

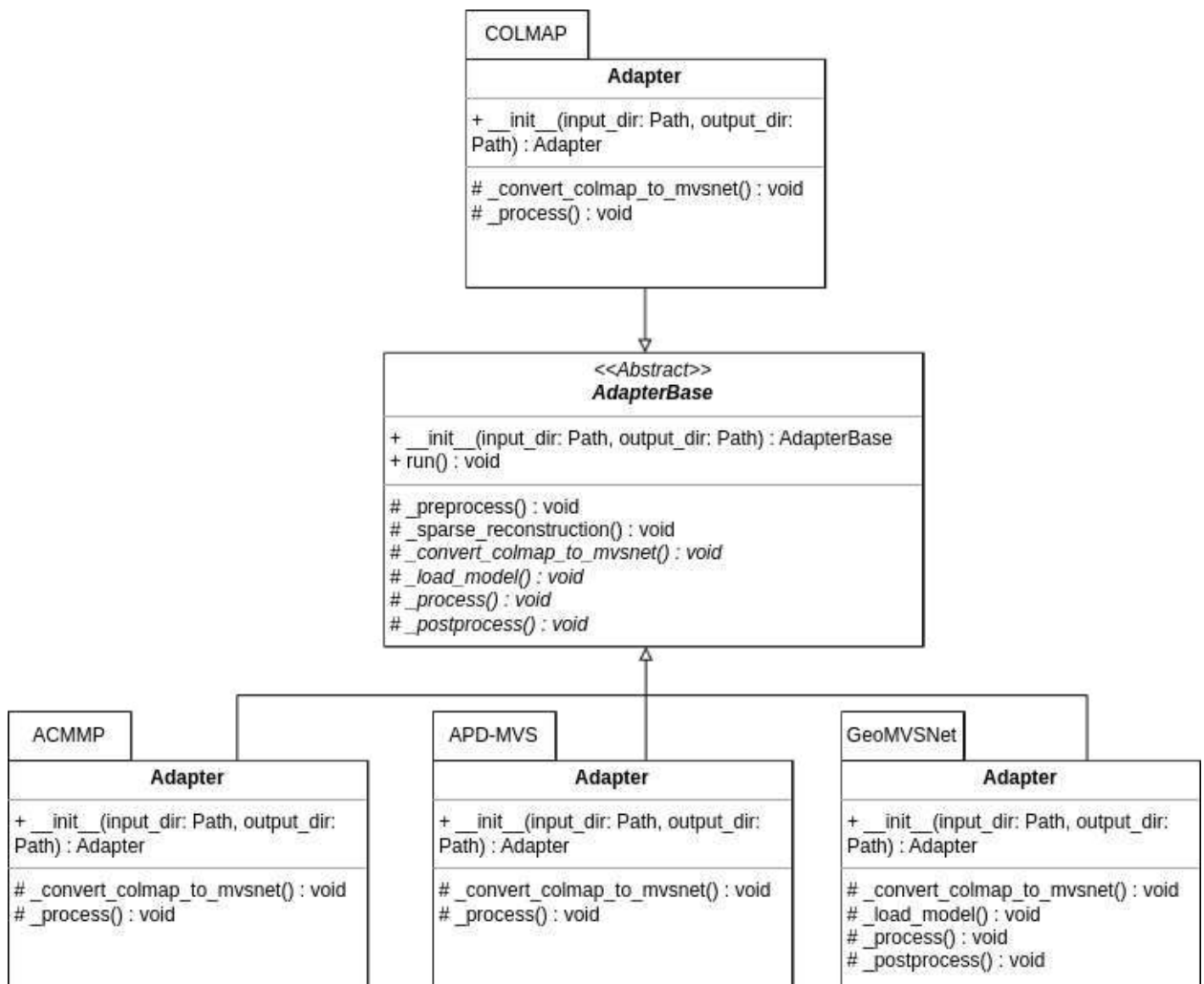


Рис. 1: Диаграмма классов адаптеров, реализованных для унифицированного интерфейса запуска алгоритмов.

Для проведения сравнения алгоритмов и дальнейшей работы необходимо наличие унифицированного интерфейса запуска алгоритма.

Из алгоритмов с открытым исходным кодом удалось произвести запуск ACMMP [34], MVSNet [31], R-MVSNet [43], REDNet [29], APD-MVS [3], GeoMVSNet [18] и CasMVSNet [10]. CasMVSNet, MVSNet, R-MVSNet были запущены на тестовой подпоследовательности аэрофотоснимков Москвы и на данных, предоставленных авторами работ, были построены реконструкции для оригинальных данных, однако для те-

стовой подпоследовательности построить плотное облако точек не получилось. REDNet был запущен на тестовой подпоследовательности, но построить плотную реконструкцию также не удалось. Для набора алгоритмов ACMMP, APD-MVS, GeoMVSNet удалось построить реконструкцию для тестовой подпоследовательности. Были написаны адаптеры для ACMMP, APD-MVS, GeoMVSNet. Для унификации интерфейса запуска этих алгоритмов и получения плотных реконструкций были реализованы классы-адаптеры (рис. 1).

Данные классы задают параметры, необходимые для работы алгоритмов, такие как директории входных данных и выходных данных. Эти классы предоставляют интерфейс запуска алгоритма посредством публичного метода `run()`.

Для этих алгоритмов были созданы инструкции Dockerfile для уменьшения зависимости от платформы и среды выполнения.

Также был написан адаптер для COLMAP, адаптеры были расширены на датасеты, не содержащие информацию о внутренних и внешних параметрах камеры.

5. Ассоциация между пикселями и точками

В данном разделе представлен алгоритм ассоциации между пикселями и облаком точек.

5.1. Алгоритм

Процесс сопоставления между пикселями и точками состоит из трех основных этапов: преобразование координат облака точек, удаление скрытых точек и проекция 3D точек в 2D.

5.1.1. Преобразование координат

Каждая точка облака обладает тремя координатами, определяющими ее местоположение в глобальной системе отсчета. Для ассоциации между пикселями фотографии и облаком точек необходимо преобразовать координаты точек из мировой системы координат в систему, связанную с камерой. Это требует использования внешних параметров камеры таких, как 3×3 матрица поворота, 3×1 вектор переноса. Внешние и внутренние параметры камеры являются результатом работы разреженной (sparse) реконструкции. Преобразование координат точки в систему координат камеры определяется следующим образом:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^\top & 1 \end{bmatrix}^{-1} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

– произведение обратной матрицы внешних параметров камеры на вектор-столбец координат точки в мировой системе отсчета, где

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

– матрица поворота,

$$t = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^\top$$

– вектор переноса.

5.1.2. Удаление скрытых точек

После вышеописанного этапа необходимо удалить скрытые точки из облака. Точка называется скрытой, если она не попадает в поле зрения при просмотре облака с заданной точки обзора. Точкой обзора является крепление камеры. Цель данного этапа – удалить точки в 3D карте, которые заведомо не видны с заданной точки обзора.

После преобразования облака в координатную систему камеры и установки точки обзора вызывается функция `hidden_point_removal` из библиотеки Open3D [56], удаляющая скрытые точки.

Описанные действия проводятся для облака точек относительно каждого снимка последовательности.

5.1.3. Проекция 3D-точек в 2D

После преобразования облака точек в систему координат, связанных с камерой, следует этап установления соответствия между 3D-точкой облака и 2D-пикселем изображения. Для данного шага необходима матрица внутренних параметров камеры:

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

где f_x , f_y соответствуют фокусному расстоянию, u_0 , v_0 – координатам x и y оптического центра. Также следует нормализовать координаты точки облака путем деления на координату Z :

$$X = \frac{X_c}{Z_c}$$

$$Y = \frac{Y_c}{Z_c}$$

$$Z = 1$$

Проекция 3D-точки облака в 2D-пиксель определяется следующим образом:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

– произведение матрицы внутренних параметров камеры на вектор-столбец координат точки облака. После выполнения вышеописанных операций следует округлить в большую сторону значения координат пикселей.

5.2. Реализация

Для реализации алгоритма был выбран язык программирования Python. Этот выбор обусловлен удобством работы с 3D данными, а также читаемостью кода, быстротой его написания. Для хранения и обработки облаков точек применяется библиотека Open3D [56]. Open3D представляет собой набор алгоритмов для оптимизированной параллельной обработки и визуализации трехмерных данных. Также используются библиотеки NumPy [4], OpenCV [8].

Входными данными алгоритма являются плотная реконструкция и фотографии с калибровкой. Изображения с калибровкой являются результатом работы разреженной (sparse) реконструкции. Для иллюстрации работы алгоритма проводится его апробация на задаче переноса семантической маски на облако точек.

6. Эксперимент

Чтобы понять, какие алгоритмы плотной реконструкции наиболее применимы для аэрофотоснимков, было проведено экспериментальное сравнение качества работы алгоритмов.

6.1. Тестовый датасет

В качестве одного из датасетов для тестирования было выбрано 44 фотографии деревни Клишева Московской области из набора данных с аэроснимками Москвы. Аэрофотоснимки имеют размер 4005×2662 . Данный выбор был обусловлен наличием на этих фотографиях различных объектов, таких как жилые дома, дороги, деревья и водоем. Пример изображений из датасета приведен на рисунке 2.



Рис. 2: Примеры изображений из датасета.

Также для тестирования использовался набор данных UrbanScenes3D, из которого было выбрано 28 снимков из городской сцены Polytech.

Тестовый датасет включает аэрофотоснимки городского здания и его окружающей территории. Кадры имеют разрешение 6000×4000 . Данные снимки отличаются от аэрофотоснимков Москвы высокой частотой кадров, а также были сделаны с более низкой высоты. Это позволяет построить более детальную плотную реконструкцию по сравнению с реконструкцией Москвы. Пример изображений из набора данных приведен на рисунке 3.



Рис. 3: Примеры изображений из датасета.

6.2. Условия эксперимента

Запуск алгоритмов на тестовых датасетах был произведен на машине, имеющей следующие характеристики:

- CPU — AMD Ryzen Threadripper 2990WX 32-Core Processor 3GHz, RAM 32 GB
- GPU — NVIDIA GeForce RTX 2080 Ti 11GB,

- OS — Ubuntu 22.04.1 LTS.

В тестовом датасете отсутствуют референсные значения, поэтому качество облаков точек было оценено визуально, замерено время работы алгоритмов и с помощью метрик Surface Density, Volume Density, условная емкость алгоритма. Для проведения сравнения было использовано пять алгоритмов построения плотной реконструкции: Agisoft Metashape³, APD-MVS [3], ACMMP [34], COLMAP [40], GeoMVSNet [18]. Замер метрик проведен с помощью инструмента CloudCompare⁴.

6.3. Результаты эксперимента

Алгоритм	Mean Surface Density (points/m ²)	Mean Volume Density (points/m ³)	Время (минуты)	Условная емкость
COLMAP	2 209 556.25	331 433 472	33.36	108
Agisoft Metashape	5 612 190	841 828 608	-	-
ACMMP	22 059 734	3 308 960 512	55.865	202
APD-MVS	21 650 402	3 247 560 448	112.604	210
GeoMVSNet	15 061 715	2 259 257 600	10.313	190

Таблица 1: Сравнение облаков точек между алгоритмами, где $R = 0.005$, R – параметр метрик 3, 4.

Результаты сравнения алгоритмов на аэрофотоснимках Москвы по плотности, условной емкости, времени представлены в таблице 1. Для инструмента Agisoft Metashape не удалось замерить время и емкость на вышеописанной машине, так как возникли проблемы с лицензией. Облако точек, построенное ACMMP, имеет самую высокую плотность. В то же время реконструкция COLMAP имеет самую низкую плотность.

³Agisoft Metashape: <https://www.agisoft.com/> (дата обращения 2023-12-12)

⁴CloudCompare: <https://www.danielgm.net/cc/> (дата обращения 2023-12-12)

Фрагменты реконструкций методов представлены на рисунке 4. На рисунке видно, что реконструкция COLMAP более разреженная по сравнению с другими методами. Облако точек, построенное GeoMVSNet, не имеет четких очертаний объектов. Стоит отметить, что Agisoft Metashape лучше реконструирует деревья, чем ACMMP, APD-MVS. В то время как ACMMP, APD-MVS более точно восстанавливают дома. Agisoft Metashape генерирует более разреженное облако по сравнению с ACMMP, APD-MVS. Реконструкция ACMMP визуально более точна и плотна, чем облако точек APD-MVS.

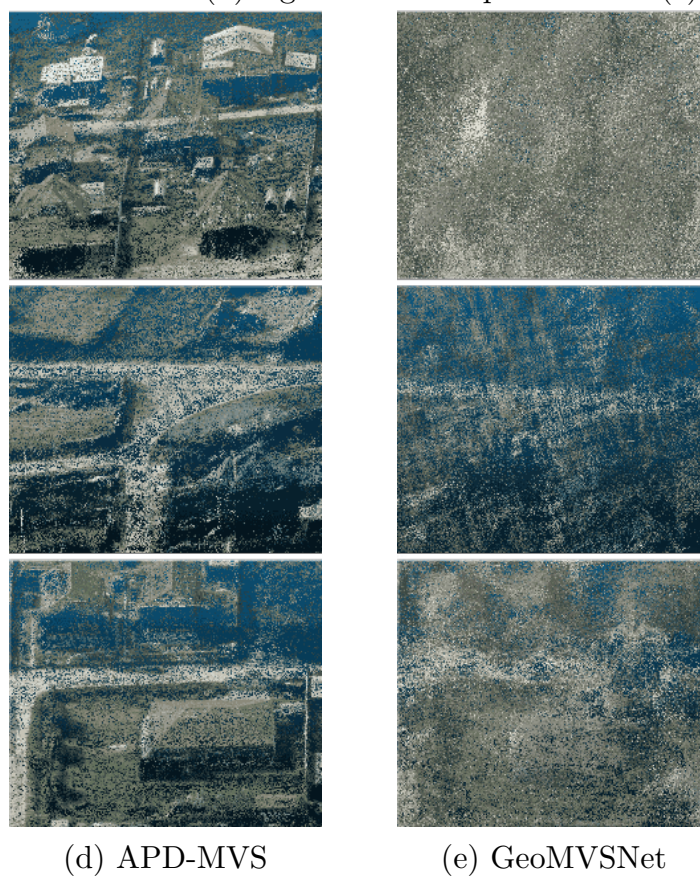
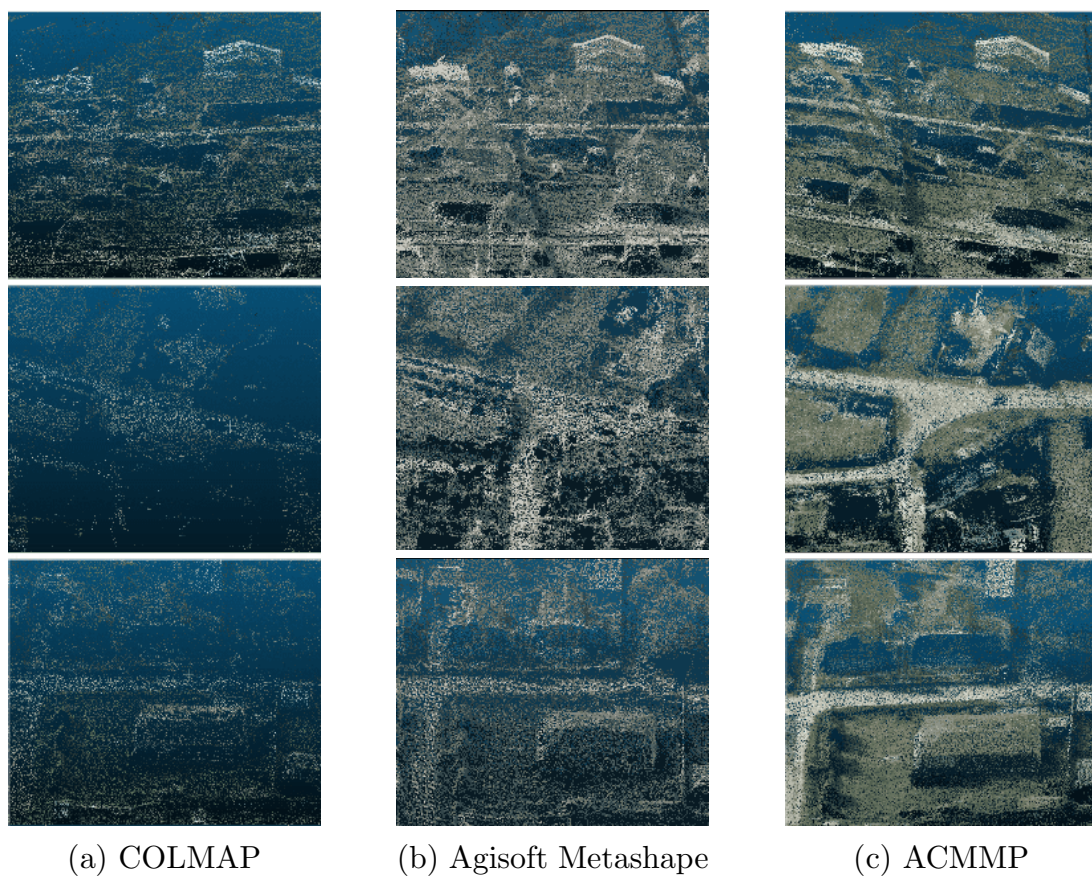


Рис. 4: Визуальное сравнение облаков точек между алгоритмами на тестовом наборе данных

Алгоритм	Mean Surface Density (points/m ²)	Mean Volume Density (points/m ³)	Время	Условная емкость
COLMAP	92 386	13 857 853	18.603	80
Agisoft Metashape	346 915	52 037 224	-	-
ACMMP	121 213	18 181 942	31.152	181
APD-MVS	236 628	35 494 248	156.08	190
GeoMVSNet	127 492	19 123 836	5.154	175

Таблица 2: Сравнение облаков точек между алгоритмами, где $R = 0.005$, R – параметр метрик 3, 4.

Результаты сравнения алгоритмов на аэрофотоснимках городской сцены Polytech по плотности, условной емкости, времени представлены в таблице 2. Облако точек, построенное Agisoft Metashape, имеет самую высокую плотность. В то же время реконструкция COLMAP имеет самую низкую плотность.

Фрагменты реконструкций методов представлены на рисунке 5. На рисунке видно, что реконструкция COLMAP более разреженная по сравнению с другими методами. В облаке точек, построенном GeoMVSNet, отсутствуют части объектов. Agisoft Metashape не восстанавливает боковые стороны зданий. APD-MVS, ACMMP визуально более точно строят реконструкцию данной сцены.

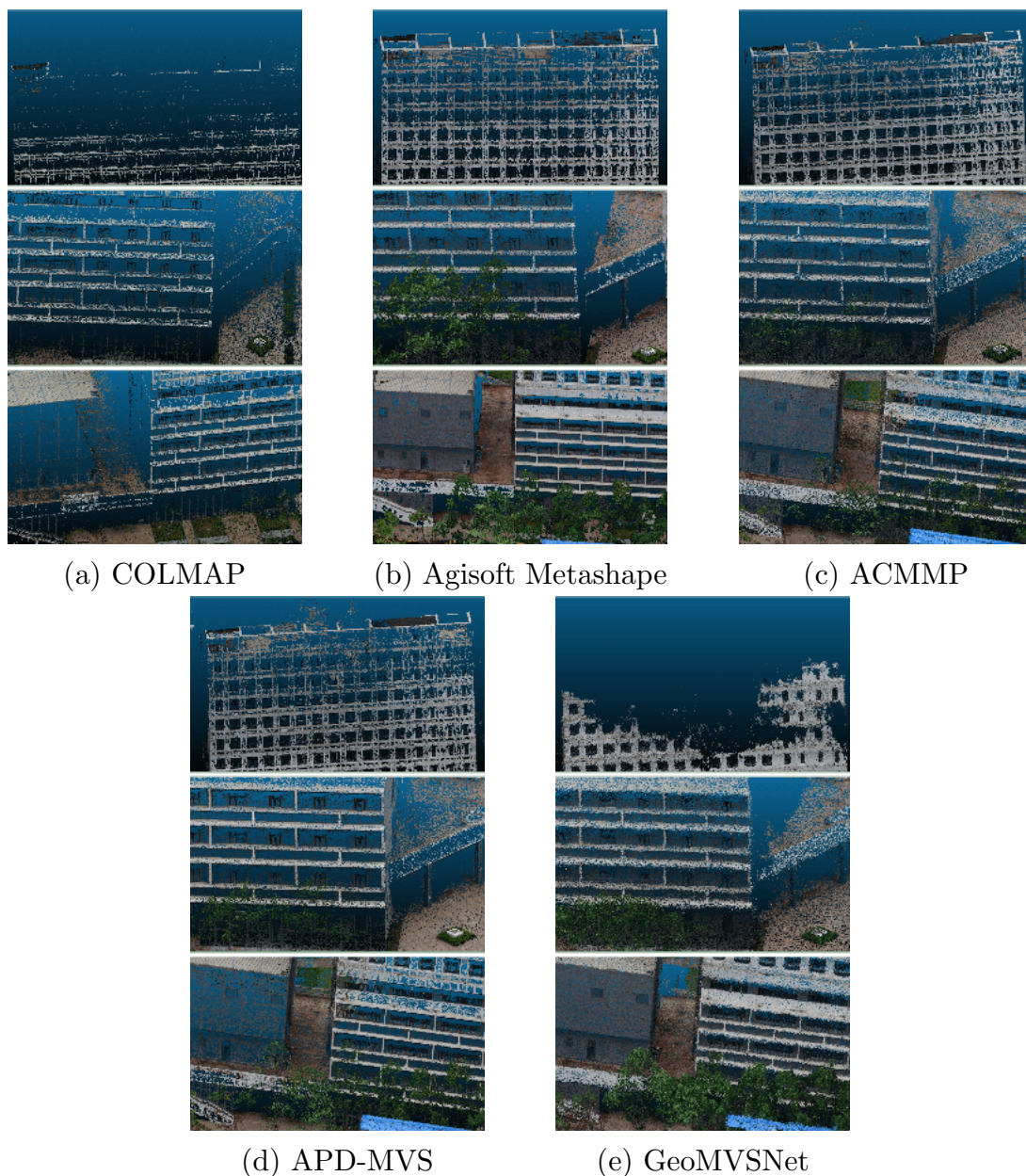


Рис. 5: Визуальное сравнение облаков точек между алгоритмами на тестовом наборе данных

6.4. Исследование этапа ассоциации между пикселями и точками

Эксперимент был проведен на облаке точек, построенном ACMMP. Для визуальной оценки работы алгоритма в качестве покрывающих снимков были взяты четыре семантических маски аэрофотоснимков Москвы, представленные на рисунке 6.

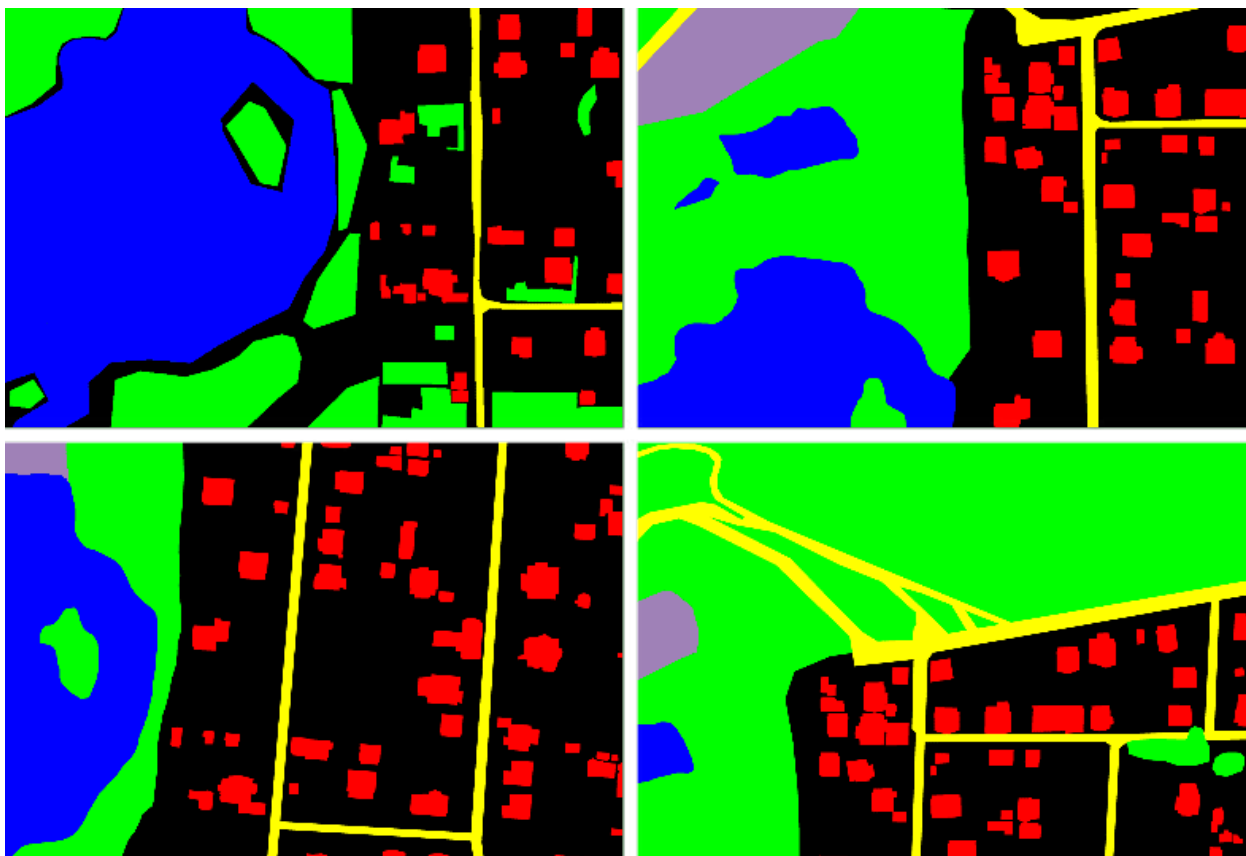


Рис. 6: Семантические маски аэрофотоснимков Москвы.

Точки облака были ассоциированы с пикселями снимков в соответствии с пунктом 5.1.3. Результат раскраски точек представлен на рисунке 7. Как можно наблюдать, при ассоциации пикселей с точками есть скрытые точки.

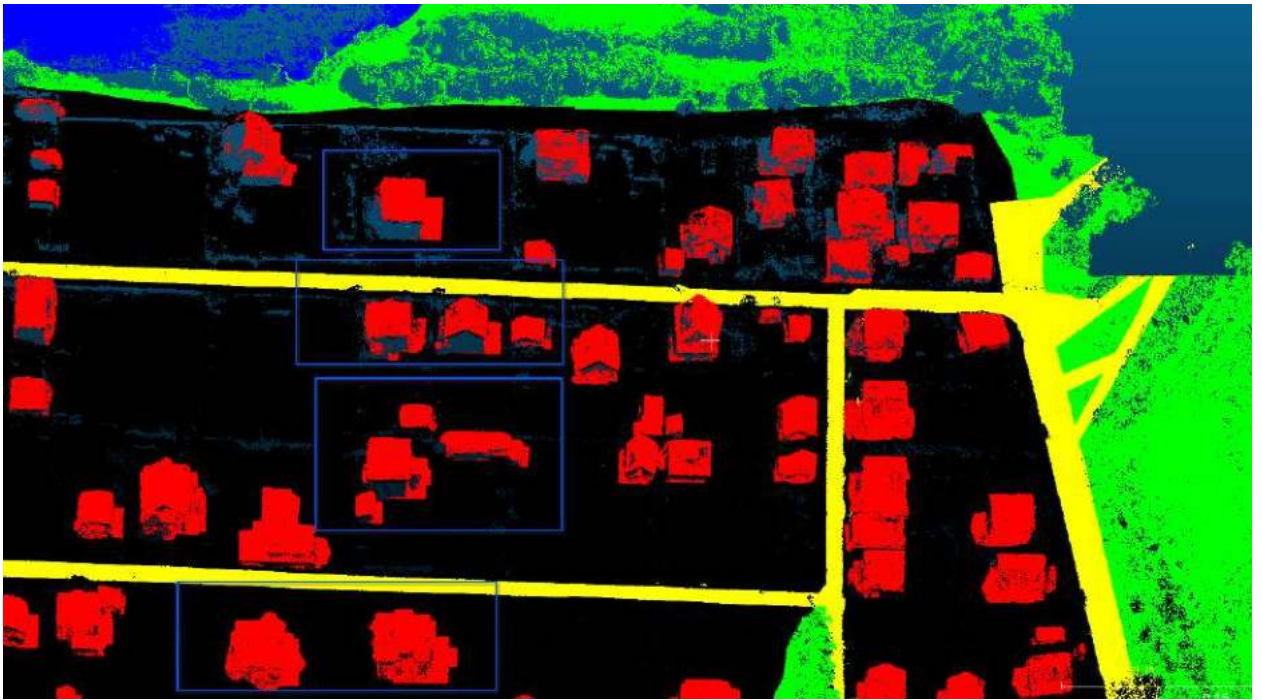


Рис. 7: Ассоциация между пикселями и облаком точек.

Для удаления скрытых точек был применен метод, описанный в пункте 5.1.2. Результаты алгоритма ассоциации пикселей с точками и удалением скрытых точек представлен на рисунке 8.

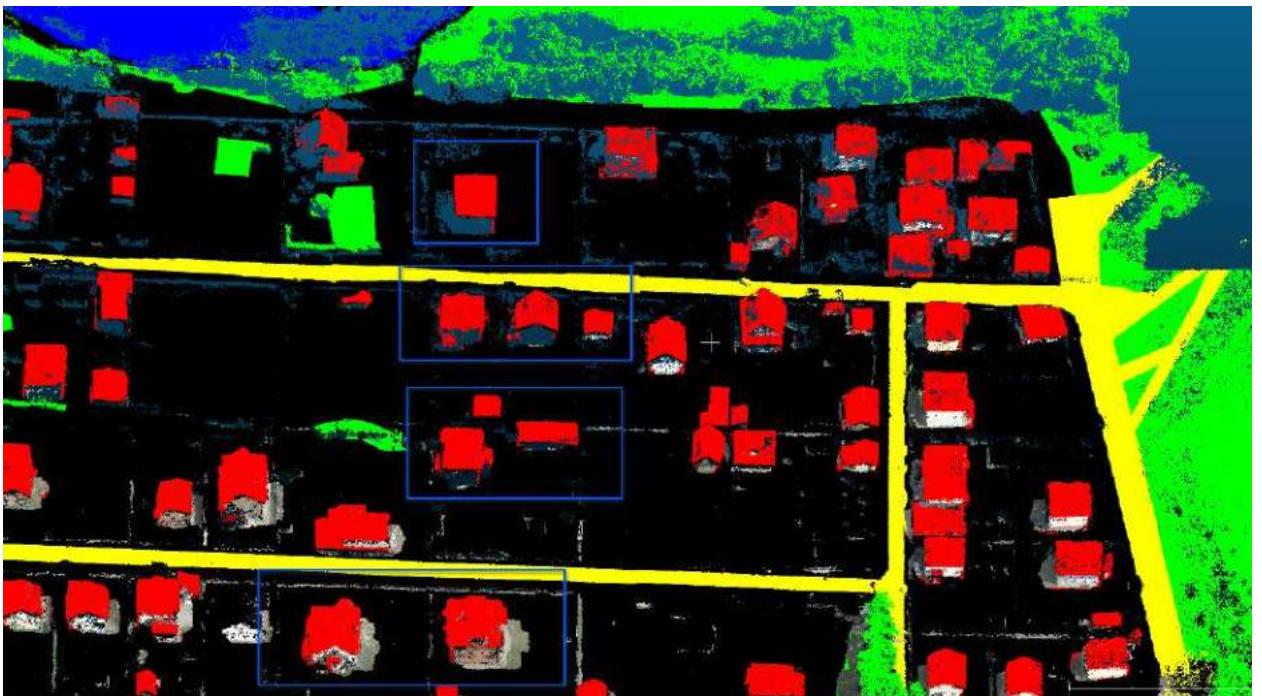


Рис. 8: Ассоциация между пикселями и облаком точек.

6.5. Вывод

Из экспериментального сравнения видно, что более точная реконструкция была построена алгоритмами ACMMP, APD-MVS. Данные алгоритмы относятся к методам без обучения. Важно отметить, что нейронная сеть GeoMVSNet, которая сгенерировала плотное облако без четких очертаний объектов, была обучена на DTU [25]. Данный датасет не содержит аэрофотоснимков. Возможно, дообучив GeoMVSNet на аэрофотоснимках Москвы, можно добиться лучших результатов.

Также был реализован алгоритм ассоциации между пикселями и точками. Данное решение можно доработать удалением шумовых точек и кластеризацией облака на классы, также увеличением радиуса видимости для удаления скрытых точек.

Заключение

В ходе работы было проведено исследование и визуальное сравнение существующих решений для плотной реконструкции объектов, а также изучение их возможности применения на задаче аэрофотограмметрии. В этом семестре были выполнены следующие задачи:

1. Выполнен обзор существующих алгоритмов плотной реконструкции объектов, а также связанных с этой задачей датасетов и метрик.
2. Реализован воспроизводимый запуск для ACMMP, GeoMVSNet, APD-MVS, COLMAP⁵.
3. Проведено сравнение алгоритмов плотной реконструкции на целевых датасетах. Установлено, что ACMMP позволяет построить наиболее точную и плотную реконструкцию на аэрофотоснимках.
4. Реализована ассоциация между пикселями и облаком точек, проведена апробация на задаче переноса семантической маски.

⁵<https://github.com/rongirl/multi-view-stereo-dockers> (дата обращения: 2023-12-15)

Список литературы

- [1] 3-D object reconstruction using spatially extended voxels and multi-hypothesis voxel coloring / E. Steinbach, B. Girod, P. Eisert, A. Betz // Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. — Vol. 1. — 2000. — P. 774–777 vol.1.
- [2] 3-D reconstruction of real-world objects using extended voxels / E. Steinbach, B. Girod, P. Eisert, A. Betz // Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101). — Vol. 1. — 2000. — P. 569–572 vol.1.
- [3] Adaptive Patch Deformation for Textureless-Resilient Multi-View Stereo / Yuesong Wang, Zhaojie Zeng, Tao Guan et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2023. — June. — P. 1621–1630.
- [4] Array programming with NumPy / Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt et al. // [Nature](#). — 2020. — sep. — Vol. 585, no. 7825. — P. 357–362. — URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- [5] Bhotika Rahul, Fleet David J., Kutulakos Kiriakos N. A Probabilistic Theory of Occupancy and Emptiness // Computer Vision — ECCV 2002 / Ed. by Anders Heyden, Gunnar Sparr, Mads Nielsen, Peter Johansen. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2002. — P. 112–130.
- [6] BlendedMVS: A Large-scale Dataset for Generalized Multi-view Stereo Networks / Yao Yao, Zixin Luo, Shiwei Li et al. // Computer Vision and Pattern Recognition (CVPR). — 2020.
- [7] Bleyer Michael, Rhemann Christoph, Rother Carsten. PatchMatch Stereo - Stereo Matching with Slanted Support Windows. — Vol. 11. — 2011. — 01. — P. 14.1–14.11.

- [8] Bradski G. The OpenCV Library // Dr. Dobb's Journal of Software Tools. — 2000.
- [9] Broadhurst A., Drummond T.W., Cipolla R. A probabilistic framework for space carving // Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. — Vol. 1. — 2001. — P. 388–393 vol.1.
- [10] Cascade cost volume for high-resolution multi-view stereo and stereo matching / Xiaodong Gu, Zhiwen Fan, Siyu Zhu et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2020. — P. 2495–2504.
- [11] A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms / S.M. Seitz, B. Curless, J. Diebel et al. // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). — Vol. 1. — 2006. — P. 519–528.
- [12] Esteban C.H., Schmitt F. Silhouette and stereo fusion for 3D object modeling // Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. — 2003. — P. 46–53.
- [13] Faugeras O., Keriven R. Variational principles, surface evolution, PDEs, level set methods, and the stereo problem // IEEE Transactions on Image Processing. — 1998. — Vol. 7, no. 3. — P. 336–344.
- [14] Fua P., Leclerc Y. G. Object-Centered Surface Reconstruction: Combining Multi-Image Stereo and Shading // International Journal of Computer Vision. — Vol. 16. — 1995. — P. 35–56.
- [15] Furukawa Yasutaka, Ponce Jean. Accurate, Dense, and Robust Multiview Stereopsis // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2010. — Vol. 32, no. 8. — P. 1362–1376.
- [16] Galliani Silvano, Lasinger Katrin, Schindler Konrad. Massively Parallel Multiview Stereopsis by Surface Normal Diffusion // The IEEE International Conference on Computer Vision (ICCV). — 2015. — June.

- [17] Gao Jian, Liu Jin, Ji Shunping. Rational Polynomial Camera Model Warping for Deep Learning Based Satellite Multi-View Stereo Matching // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). — 2021. — October. — P. 6148–6157.
- [18] GeoMVSNet: Learning Multi-View Stereo With Geometry Perception / Zhe Zhang, Rui Peng, Yuxi Hu, Ronggang Wang // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2023. — P. 21508–21518.
- [19] Hertzmann A., Seitz S.M. Shape and materials by example: a photometric stereo approach // 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. — Vol. 1. — 2003. — P. I–I.
- [20] Hierarchical Prior Mining for Non-local Multi-View Stereo / Chunlin Ren, Qingshan Xu, Shikun Zhang, Jiaqi Yang // Proc. IEEE/CVF International Conference on Computer Vision. — 2023. — October. — P. 3611–3620.
- [21] Image unprojection for 3D surface reconstruction: A triangulation-based approach / Min-Hyuk Sung, Hwasup Lim, Hyoung-Gon Kim, Sang Chul Ahn // 2013 IEEE International Conference on Image Processing. — 2013. — P. 161–165.
- [22] Kuhn Andreas, Lin Shan, Erdler Oliver. Plane Completion and Filtering for Multi-View Stereo Reconstruction // Pattern Recognition / Ed. by Gernot A. Fink, Simone Frintrop, Xiaoyi Jiang. — Cham : Springer International Publishing, 2019. — P. 18–32.
- [23] Kutulakos K.N., Seitz S.M. A theory of shape by space carving // Proceedings of the Seventh IEEE International Conference on Computer Vision. — Vol. 1. — 1999. — P. 307–314 vol.1.
- [24] Kutulakos Kiriakos N. Approximate N-View Stereo // Computer Vision - ECCV 2000. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2000. — P. 67–83.

- [25] Large Scale Multi-view Stereopsis Evaluation / Rasmus Jensen, Anders Dahl, George Vogiatzis et al. // 2014 IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — P. 406–413.
- [26] [Learning 3D Semantic Segmentation with only 2D Image Supervision](#) / Kyle Genova, Xiaoqi Yin, Abhijit Kundu et al. // 2021 International Conference on 3D Vision (3DV). — 2021. — P. 361–372.
- [27] Lhuillier M., Quan L. A quasi-dense approach to surface reconstruction from uncalibrated images // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2005. — Vol. 27, no. 3. — P. 418–433.
- [28] Litvinov Vadim, Yu Shuda, Lhuillier Maxime. 2-manifold reconstruction from sparse visual features // 2012 International Conference on 3D Imaging (IC3D). — 2012. — P. 1–8.
- [29] Liu Jin, Ji Shunping. A Novel Recurrent Encoder-Decoder Structure for Large-Scale Multi-view Stereo Reconstruction from An Open Aerial Dataset. — 2020. — 03.
- [30] Liu Yilin, Xue Fuyou, Huang Hui. UrbanScene3D: A Large Scale Urban Scene Dataset and Simulator. — 2021. — 2107.04286.
- [31] MVSNet: Depth Inference for Unstructured Multi-view Stereo / Yao Yao, Zixin Luo, Shiwei Li et al. // European Conference on Computer Vision (ECCV). — 2018.
- [32] Methods for Volumetric Reconstruction of Visual Scenes / Gregory G. Slabaugh, W. Bruce Culbertson, Thomas Malzbender et al. // International Journal of Computer Vision. — Vol. 57. — 2004. — P. 179–199.
- [33] Morris D.D., Kanade T. Image-consistent surface triangulation // Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662). — Vol. 1. — 2000. — P. 332–338 vol.1.

- [34] Multi-Scale Geometric Consistency Guided and Planar Prior Assisted Multi-View Stereo / Qingshan Xu, Weihang Kong, Wenbing Tao, Marc Pollefeys // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2022.
- [35] Multi-Scale Geometric Consistency Guided and Planar Prior Assisted Multi-View Stereo / Qingshan Xu, Weihang Kong, Wenbing Tao, Marc Pollefeys // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2023. — Vol. 45, no. 4. — P. 4945–4963.
- [36] A Multi-view Stereo Benchmark with High-Resolution Images and Multi-camera Videos / Thomas Schöps, Johannes L. Schönberger, Silvano Galliani et al. // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — P. 2538–2547.
- [37] Özdemir E, Remondino F. Segmentation of 3D photogrammetric point cloud for 3D building modeling // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2018. — Vol. 42. — P. 135–142.
- [38] Özdemir E, Remondino F, Golkar A. Aerial point cloud classification with deep learning and machine learning algorithms // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2019. — Vol. 42. — P. 843–849.
- [39] Wang Fangjinhua, Galliani Silvano, Vogel Christoph et al. PatchmatchNet: Learned Multi-View Patchmatch Stereo. — 2021.
- [40] Pixelwise View Selection for Unstructured Multi-View Stereo / Johannes L. Schönberger, Enliang Zheng, Jan-Michael Frahm, Marc Pollefeys // Computer Vision – ECCV 2016 / Ed. by Bastian Leibe, Jiri Matas, Nicu Sebe, Max Welling. — Cham : Springer International Publishing, 2016. — P. 501–518.
- [41] Progressive surface reconstruction from images using a local prior / Gang Zeng, S. Paris, L. Quan, F. Sillion // Tenth IEEE International

Conference on Computer Vision (ICCV'05) Volume 1. — Vol. 2. — 2005. — P. 1230–1237 Vol. 2.

- [42] RayNet: Learning Volumetric 3D Reconstruction with Ray Potentials / Despoina Paschalidou, Ali Osman Ulusoy, Carolin Schmitt et al. // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — IEEE Computer Society, 2018.
- [43] Recurrent MVSNet for High-resolution Multi-view Stereo Depth Inference / Yao Yao, Zixin Luo, Shiwei Li et al. // Computer Vision and Pattern Recognition (CVPR). — 2019.
- [44] Romanoni Andrea, Matteucci Matteo. TAPA-MVS: Textureless-Aware PatchMatch Multi-View Stereo // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). — 2019. — 03.
- [45] Seitz S.M., Dyer C.R. Photorealistic scene reconstruction by voxel coloring // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — 1997. — P. 1067–1073.
- [46] Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers / Martin Weinmann, Boris Jutzi, Stefan Hinz, Clément Mallet // ISPRS Journal of Photogrammetry and Remote Sensing. — 2015. — Vol. 105. — P. 286–304.
- [47] Shape Reconstruction from 3D and 2D Data Using PDE-Based Deformable Surfaces / Ye Duan, Liu Yang, Hong Qin, Dimitris Samaras // Computer Vision - ECCV 2004 / Ed. by Tomáš Pajdla, Jiří Matas. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2004. — P. 238–251.
- [48] Stucker Corinne, Schindler Konrad. ResDepth: A deep residual prior for 3D reconstruction from high-resolution satellite images // ISPRS Journal of Photogrammetry and Remote Sensing. — 2022. — Vol. 183. — P. 560–580.

- [49] SurfaceNet: An End-To-End 3D Neural Network for Multiview Stereopsis / Mengqi Ji, Juergen Gall, Haitian Zheng et al. // Proceedings of the IEEE International Conference on Computer Vision (ICCV). — 2017. — P. 2307–2315.
- [50] Tanks and temples: benchmarking large-scale scene reconstruction / Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, Vladlen Koltun // ACM Transactions on Graphics. — 2017. — 07. — Vol. 36. — P. 1–13.
- [51] Treuille Adrien, Hertzmann Aaron, Seitz Steven M. Example-Based Stereo with General BRDFs // Computer Vision - ECCV 2004 / Ed. by Tomás Pajdla, Jiří Matas. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2004. — P. 457–469.
- [52] Ullman S. The Interpretation of Structure from Motion // Proceedings of the Royal Society of London. Series B, Biological Sciences. — 1979. — Vol. 203, no. 1153. — P. 405–426.
- [53] Variational stereovision and 3D scene flow estimation with statistical similarity measures / Pons, Keriven, Faugeras, Hermosillo // Proceedings Ninth IEEE International Conference on Computer Vision. — 2003. — P. 597–602 vol.1.
- [54] Xu Qingshan, Tao Wenbing. Planar Prior Assisted PatchMatch Multi-View Stereo // AAAI Conference on Artificial Intelligence (AAAI). — 2020.
- [55] Zhang Kai, Sun Jin, Snavely Noah. Leveraging Vision Reconstruction Pipelines for Satellite Imagery // IEEE International Conference on Computer Vision Workshops. — 2019.
- [56] Zhou Qian-Yi, Park Jaesik, Koltun Vladlen. Open3D: A Modern Library for 3D Data Processing // arXiv:1801.09847. — 2018.
- [57] An automatic and modular stereo pipeline for pushbroom images / C. de Franchis, E. Meinhardt-Llopis, J. Michel et al. // ISPRS An-

nals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2014. — Vol. II-3. — P. 49–56.