

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
**“САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ,
МЕХАНИКИ И ОПТИКИ”**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**ПОВЫШЕНИЕ КАЧЕСТВА ПОСТРОЕНИЯ 3D МОДЕЛЕЙ ДЛЯ СЦЕН
ВНУТРИ ПОМЕЩЕНИЙ**

Автор Никулин Даниил Сергеевич _____
(Фамилия, Имя, Отчество) _____
(Подпись) _____

Направление подготовки (специальность) 01.04.02
(код, наименование)
Прикладная математика и информатика

Квалификация магистр
(бакалавр, магистр)*

Руководитель ВКР Буздалов Максим Викторович (к.т.н.), руководитель подразделения
Лаборатория "Эволюционные вычисления"; сотрудник подразделения МЛ КТ;
член совета УС МФ ТИиТ; доцент ФИТиП; научный сотрудник ФИТиП, ИТМО

(Фамилия, И., О., ученое звание, степень) _____
(Подпись) _____

К защите допустить

Руководитель ОП _____
(Фамилия, И.О., ученое звание, степень) _____
(Подпись) _____

“ ____ ” 20 ____ г.

Санкт-Петербург, 20 ____ г.

Студент Никулин Даниил Сергеевич
(Фамилия, И. О.)

Группа M42362 Факультет ИТиП

Направленность (профиль), специализация Разработка программного обеспечения / Software Engineering

Консультант (ы):

a) Подольская Анна Владимировна (эффективно отсутствует), отсутствует
(Фамилия, И., О., ученое звание, степень)

_____ (Подпись)

ВКР принята “ ____ ” 20 ____ г.

Оригинальность ВКР _____ %

ВКР выполнена с оценкой _____

Дата защиты “ ____ ” 20 ____ г.

Секретарь ГЭК _____
(ФИО) _____ (подпись)

Листов хранения _____

Демонстрационных материалов/Чертежей хранения _____

* за исключением направления подготовки 27.04.08 Управление интеллектуальной собственностью (Магистр. Инженер-патентовед), специальностей 12.05.01 Электронные и оптико-электронные приборы и системы специального назначения (Инженер), 38.05.02 Таможенное дело (Специалист таможенного дела).

СОДЕРЖАНИЕ

Введение	5
1 Постановка задачи	6
1.1. Метрики качества	6
1.2. Выбор метода	7
2 Выбор и тестирование алгоритмов “дорисовки” карт глубин.	9
2.1. Выбор алгоритмов	9
2.2. Выбор датасета для тестирования	9
2.3. Генерация типичных входных данных	12
2.4. Метрики ошибок	12
2.5. Результаты	13
3 Встраивание	14
3.1. Выбор систем 3D реконструкции для тестирования	14
3.2. Встраивание в системы 3D реконструкции	14
4 Тестирование.	19
4.1. Выбор датасета для тестирования	19
4.2. Описание структуры датасета и схемы тестирования	20
4.3. Создание системы автоматического сопоставления реконстру- ированных и эталонных 3D моделей.	22
4.4. Вычисление метрик	26
4.5. Полученные результаты	28
Заключение	31
Список литературы	32
Приложения	37

Введение

Существуют системы, способные по набору фотографий построить 3D модель. Задача построения 3D модели из некоторого набора исходных данных называется задачей 3D реконструкции. В данной работе будет рассматриваться в первую очередь реконструкция из набора фотографий. Т.е. на вход системе подается некоторое множество фотографий, а на выходе должна получиться 3D модель того, что изображено на этих фотографиях.

Под 3D моделью обычно понимают либо плотное облако точек (*dense point cloud*), либо же полигональную сетку (*mesh*), полученную, как правило, путем триангуляции плотного облака.

Для решения задачи реконструкции большинство систем [15, 19, 21–23, 32, 36] с помощью методов Structure from Motion (SfM) для каждой фотографии оценивают положения и ориентации камеры в общих 3D координатах. Полученная информация используется в методах Multi-View Stereo (MVS) [8, 9, 15, 18, 21–23] для построения плотного облака точек и полигональной сетки (*mesh*).

Иногда модель содержит дыры, когда большие части оригинальной 3D сцены не представлены в реконструированной модели. Методы MVS имеют настраиваемые параметры, которые способны увеличить плотность конечной 3D модели. Однако изменение параметров, необходимое для сколь-либо заметного повышения плотности приводит к резкому понижению точности или даже к полному расхождению алгоритма построения 3D модели.

В данной работе представлено решение, позволяющее автоматически избавляться от дыр с сохранением точности построенной модели для большого подмножества типов сцен – сцены внутри помещений.

1. Постановка задачи

1.1. Метрики качества

Для сравнения качества различных реализаций систем 3D реконструкции, необходима некоторая численная метрика качества.

На датасете “Tanks and Temples” [34] проходит соревнование между системами реконструкции. В этом соревновании для сравнения моделей используются метрики полноты и точности.

Введем обозначения:

- \mathcal{G} – множество точек эталонной модели
- \mathcal{R} – множество точек реконструкции
- $dist(\mathbf{a}, \mathbf{b})$ – метрическое расстояние между объектами \mathbf{a} и \mathbf{b} .
- $e_{\mathbf{a} \rightarrow \mathcal{B}} = \min_{\mathbf{b} \in \mathcal{B}} dist(\mathbf{a}, \mathbf{b})$ – минимальное расстояние от объекта \mathbf{a} до множества \mathcal{B}

Тогда точность (**Precision**, $P(d)$) и полноту (**Recall**, $R(d)$) определим так:

$$P(d) = \frac{\text{card}\{\mathbf{r} \in \mathcal{R} : e_{\mathbf{r} \rightarrow \mathcal{G}} < d\}}{\text{card}\{\mathcal{R}\}} \quad R(d) = \frac{\text{card}\{\mathbf{g} \in \mathcal{G} : e_{\mathbf{g} \rightarrow \mathcal{R}} < d\}}{\text{card}\{\mathcal{G}\}} \quad (1.1)$$

Где **card** – это мощность множества.

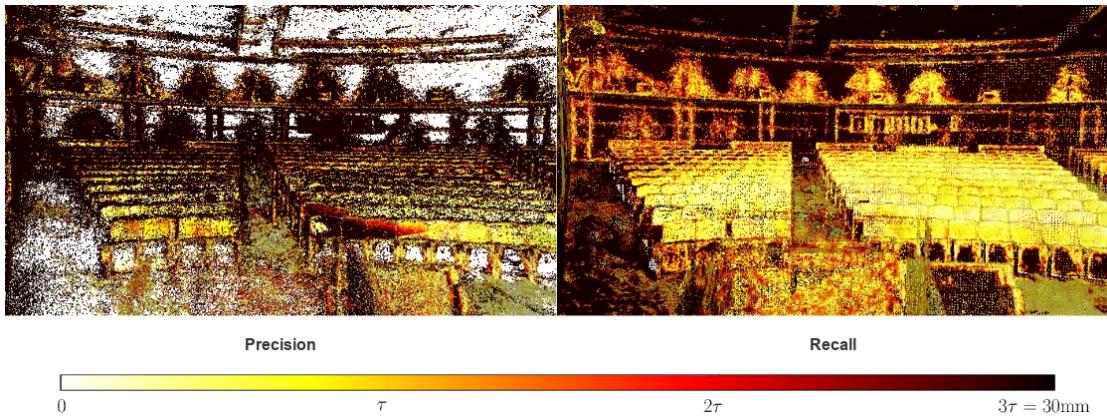


Рис. 1: Вычисленные метрики реконструкции. [34]

На Рис. 1 представлены вычисленные метрики качества для реконструкции, созданной системой COLMAP [23] в рамках соревнований на датасете “Tanks and Temples” [34]. Слева – реконструированная 3D модель в виде облака точек. Цветом для каждой точки модели показано расстояние до ближайшей точки эталонной модели. Справа – эталонная модель и расстояния до реконструированного облака.

На Рис. 1 можно заметить дыры, о которых уже упоминалось ранее. Произведя анализ результатов соревнования на датасете [34], можно заметить, что на момент написания диплома все представленные в соревновании системы допускают подобные “дыры” для большинства сцен, внутри помещений (*Auditorium, Courtroom, Museum*). Эти дыры приводят к существенному снижению метрик полноты по отношению к метрикам точности: у лидеров для сцены *Auditorium* 90% точности и 60% полноты.

Таким образом, если удастся избавиться от подобных дыр для всего множества сцен, снятых внутри помещений, и сохранить при этом значения метрики точности конечной 3D модели, то это приведет к значимому улучшению качества.

1.2. Выбор метода

Основная причина появления дыр – отсутствие надежных значений на участках карт глубин, проецирующих пропущенные участки сцены. Значит, если удастся произвести коррекцию карт глубин, предоставив надежные значения для указанных участков, то это позволит избавиться от образовавшихся дыр.

По этой причине было решено сосредоточить усилия на методах оценки (depth map estimation) и фильтрации карт глубин.

Пара из цветной фотографии и соответствующей ей карте глубины иногда ещё называется RGB-D изображением, так как может быть представлена единым изображением из 4-х каналов.

В научной области, посвященной фильтрации RGBD изображений, полученных от специальных сенсоров, имеются методы “дорисовки” карт глубин (depth map inpainting), которые решают задачу оценки значений карт

глубин для пропущенных участков [10, 12, 14, 16, 39] и [33]. Модификации данных методов уже применяются в приложениях дополненной реальности компаний Google [7] и Facebook [11] для генерации плотных карт глубин в реальном времени. В то же время, в системах 3D реконструкции, методы “дорисовки” карт глубин пока нигде не применялись.

Было принято решение провести исследования применимости методов “дорисовки” для избавления от дыр получаемой 3D модели путем встраивания их в существующие системы 3D реконструкции.

Так как наблюдаемые дыры характерны в первую очередь для сцен помещений, а также то, что из-за технических ограничений RGBD сенсоры применяются также исключительно внутри помещений с искусственным освещением, а значит, вместе с ними и методы “дорисовки”, то было решено ограничиться в данной работе исключительно сценами данного типа.

2. Выбор и тестирование алгоритмов “дорисовки” карт глубин.

2.1. Выбор алгоритмов

Так как рассматриваемый набор данных (карты глубин на выходе MVS алгоритмов) весьма специфичен и не характерен для типичных данных, используемых в методах “дорисовки”, то было принято решение взять несколько лидеров области и самостоятельно произвести тестирование для сравнения и выявления лучшего.

В качестве методов “дорисовки” были выбраны следующие:

1. GFMM [10]
2. CBF [12]
3. Sparse-to-Dense [16]
4. ALC [14]
5. DeepDepth [39]

2.2. Выбор датасета для тестирования

Типичная схема тестирования: берется RGB фотография и соответствующая ей эталонная карта глубины. Далее часть карты глубины маскируется, данные о ней удаляются, и метод пытается восстановить исходные значения.

Для тестирования необходим набор эталонных данных. В нашем случае это должны быть фотографии помещений с эталонными картами глубин для каждой фотографии.

Долгое время для получения эталонной пары RGB-фотографии и карты глубины к ней использовали высококачественный RGB-D сенсор. Однако данные, полученные с сенсоров не идеальны. На Рис. 2 представлен характерный результат работы сканера: далекие, прозрачные и бликующие поверхности не представлены в отснятом материале. Это приводит к тому, что для целого класса часто встречающихся артефактов невозможно оце-

нить качество методов “дорисовки”. С этим приходилось мириться и высчитывать метрики только для тех участков, для которых сенсор смог оценить глубину.

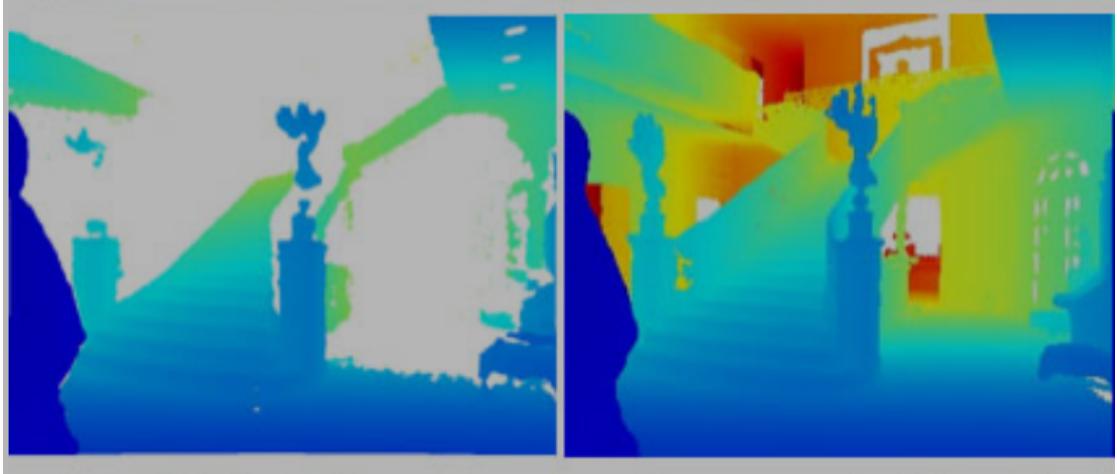


Рис. 2: Карта глубин, полученной от сенсора (слева), и эталонная карта (справа).

Подобных проблем можно избежать при использовании синтетических датасетов. В данный момент имеется значительное количество синтетических датасетов помещений с высоким уровнем качества и детализации, имитирующих характерные шумы камер и сенсоров. Например, SunCG [31] или SceneNet [30]. Однако, синтетические датасеты, к сожалению, на данный момент не в состоянии заменить реальные данные, имеющие слабо предсказуемые особенности. Иногда именно различия в обработке данных особенностей приводят к наиболее существенным различиям в точности. По этой причине синтетики нельзя использовать при тестировании.

В 2018 году в работе DeepDepth [39] для тренировки и тестирования предложили альтернативную схему. Как и ранее сначала снимался датасет с помощью RGB-D сенсора. Далее из этого датасета с помощью методов RGB-D реконструкции строили 3D модель отнятого помещения. После создания 3D модели, задача стала идентична тестированию на синтетиках, когда карты глубин можно генерировать для любого положения камеры. Полученная

модель также имела артефакты для прозрачных и зеркальных поверхностей, однако проблема бликов и дальнодействия сенсоров была снята.

Было решено последовать идеи из [39] и в качестве эталонных данных использовать проекции построенной 3D модели. Так как самые объемные из высококачественных RGB-D датасетов, имеющих в своем составе собранную 3D модель сцены, были уже использованы для обучения системы DeepDepth (Matterport3D [17] и ScanNet [28]), то был взят менее объемный аналог – SceneNN [29].

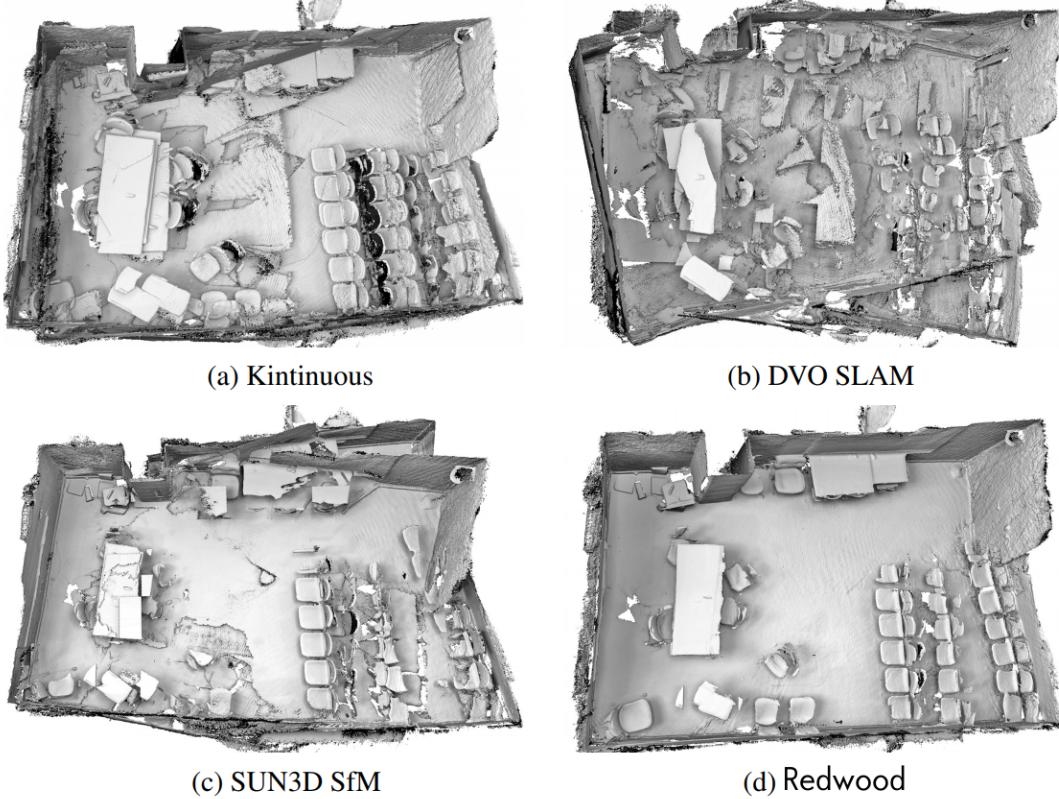


Рис. 3: Сравнение методов RGB-D реконструкции

Имеется более объёмный аналог – Sun3D [37], однако система реконструкции *SUN3D SfM*, используемая при создании 3D моделей сцен в датасете [37], значительно уступает системе Redwood [5], модификация которого была использована при составлении датасета SceneNN. Это можно видеть на

Рис. 3, где сравниваются 3D модели, созданные разными системами RGB-D реконструкции.

2.3. Генерация типичных входных данных

После получения эталонных RGBD изображений, необходимо создать входные данные, которые были бы типичны для рассматриваемой задачи.

Семейство алгоритмов, строящих плотные карты глубин, имея на входе фотографии и их взаимные положения, называется Multi-View Stereo (MVS). Большинство современных алгоритмов данного семейства генерируют карты глубин с одинаковой типичной структурой.

В [16] уже имеется код, генерирующий типичную для алгоритмов MVS разряженную карту глубин из эталонной. Именно этот код с небольшими модификациями и был взят для генерации входных данных.

2.4. Метрики ошибок

Для тестирования методов “дорисовки” карт глубин уже имеется набор метрик качества, которой используется в большинстве существующих на данный момент работ.

1. RMSE – корень из среднего квадрата отклонения в метрах
2. MAE – средняя абсолютная ошибка в метрах
3. δ_d - процент предсказанных пикселей, для которых относительная ошибка больше, чем порог d .

$$\delta_d = \frac{\text{card} \left\{ y_{pred} : \max \left(\frac{y_{pred}}{y_{gt}}, \frac{y_{gt}}{y_{pred}} \right) > d \right\}}{\text{card} \{ y_{gt} \}} \quad (2.1)$$

Где y_{gt} и y_{pred} соответственно являются эталонным и предсказанным значениями глубины. **card** – мощность множества. Чем меньше δ , тем результат лучше.

Реализация вычисления данных метрик была также взята из [16].

2.5. Результаты

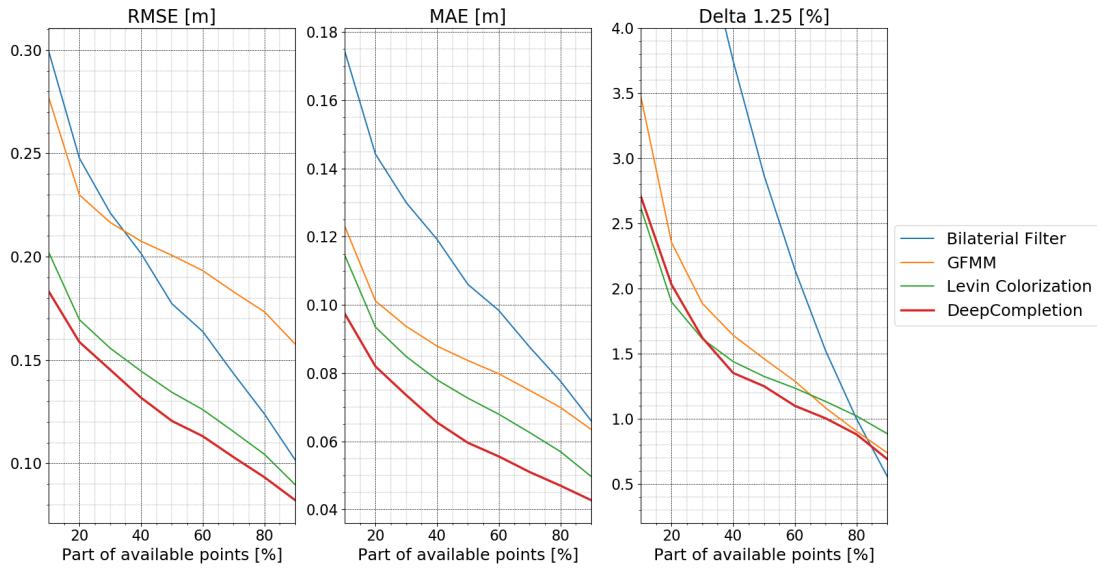


Рис. 4: Результаты тестирования методов “дорисовки” карт глубин.

Как можно заметить, на Рис. 4 не представлен один из рассматривавшихся методов, а именно Sparse-to-Dense [16]. Это связано с тем, что что метрики ошибок для [16] оказались настолько велики, что не могли быть выведены в одном масштабе с остальными алгоритмами. Поэтому данный алгоритм даже не представлен на конечных графиках.

Результаты тестирования, позволяют заключить что лучшей системой “дорисовки” для любой плотности является DeepDepth, причем вне зависимости от выбранной метрики качества. По этой причине именно DeepDepth и будет применяться далее.

3. Встраивание

3.1. Выбор систем 3D реконструкции для тестирования

При выборе системы реконструкции, на которой проводилась бы проверка применимости подхода “дорисовки” для повышения качества, в первую очередь принимались во внимание результаты в соревновании “Tanks and Temples” [34]. Специфика датасета из [34] заключается в том, что он разделен на два множества сцен: *intermediate* и *advanced*. Подмножество *intermediate* содержит сцены вне помещений (*outdoor scenes*) с единственным реконструируемым объектом и легко отделяемым от него задним фоном. *Advanced* же содержит крупномасштабные сцены самих помещений (*indoor scenes*), на которых большинство алгоритмов, успешно справившихся с *intermediate*, расходятся.

Для *advanced* набора на момент написания диплома лидером являлся COLMAP [23]. COLMAP имеет открытый исходный код, а этапы его работы разделены и независимы, что позволяет легко встроить DeepDepth в процесс реконструкции между ними.

Чтобы продемонстрировать универсальность подхода “дорисовки” для повышения качества, было решено произвести модификацию и тестирование ещё одной системы. В качестве такой системы была выбрана коммерческая система Photoscan [21], которая, хоть и не участвовала в соревновании [34], занимает лидирующие позиции на рынке. Процесс построения модели в ней также разделен на этапы, что упрощает модификацию.

В процессе работы, оказалось, что ещё две MVS системы выбились в лидеры на *advanced* наборе данных, опередив COLMAP. Это ACMN [38] и R-MVSNET [24].

3.2. Встраивание в системы 3D реконструкции

На Рис. 5 представлена обобщенная схема работы систем с учетом модификации.

Процесс реконструкции в обоих системах (COLMAP и PhotoScan) состоит из отдельных этапов. Для данной работы было важно, что генерация

карт глубин и построение плотного облака точек для рассматриваемых систем разделены. Это делает возможным осуществление модификации без изменения исходных текстов программ.

Каждый отдельный этап работы рассматриваемых систем реконструкции оставляет после себя некоторые артефакты, которые сохраняются в файловой системе. По этой причине необходимо было только модифицировать артефакты, получившиеся в результате работы этапа MVS. После чего запустить все оставшиеся этапы и получить готовую 3D модель с модифицированными картами глубин.

Для PhotoScan данными артефактами служили только сами карты глубин, для COLMAP помимо карт глубин дополнительно пересчитывались ещё карты нормалей. В остальном же обе системы остались без изменений.

Предфильтрация Большинство методов “дорисовки”, и DeepDepth [39] тоже, чувствительны к выбросам. Данные методы разрабатывались чтобы дополнять пропущенные сенсором участки. В большинстве случаев сенсоры имеют достаточное качество для того, чтобы выбросы были настолько редки, чтобы их можно было не принимать во внимание.

По этой причине избавление от выбросов выполнялось самостоятельно в качестве предварительной фильтрации. Лучший способ найти выбросы – это вычислить метрику ошибки для каждого пикселя и отрезать по порогу. Данная операция для плотных карт глубин ($> 10\%$ заполненности) требует больших вычислительных ресурсов и не всегда позволяет получить достоверный результат, так как в метрике ошибки также может содержаться ошибка. В частности COLMAP позволяет задать пороговое значение для ошибки репроекции для этапа MVS: все значения в картах глубин, для которых ошибка больше порога будут обнулены. Тем не менее, даже малые

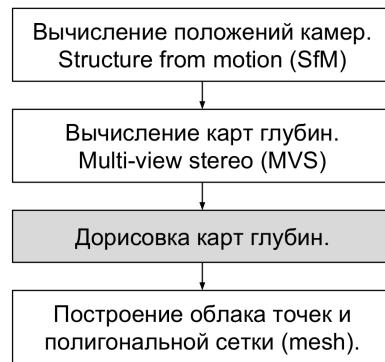


Рис. 5: Обобщенная схема работы.

значения порога не спасают от всех выбросов, в то время как плотность карт уменьшается.

Вместо этого была выдвинута гипотеза, что выбросы в основном представляют собой небольшую группу точек, полностью отделенных от остальных пикселей неизвестными областями, или же резко от них отличающихся. В рамках предположения все области карт глубин, подходившие под данное описание удалялись.

Запуск дорисовки Так как DeepDepth представляет собой гетерогенную систему, когда отдельные части написаны на разных языках и осуществляют взаимодействие через файловую систему и вызовы через командную строку, то наиболее простым решением являлось запуск корневого скрипта из DeepDepth через командную строку с указанием директории, где лежат входные данные.

Для успешной работы DeepDepth требует нормализации значений в картах глубин относительно среднего значения типа *uint16*, т.е. 32768. Поэтому карты нормализовались перед запуском и переводились обратно после вычислений.

Геометрическая фильтрация в COLMAP Геометрическая фильтрация – фильтрация трехмерных точек, с учетом положений нескольких камер. Наиболее распространенная метрика ошибки при фильтрации – ошибка репроекции. Если известны положения камеры, положение 3D точки и соответствие ей на изображении, то ошибкой репроекции будет расстояние в пикселях между проекцией 3D точки и местом на изображении, ей соответствующей.

В случае достаточной точности SfM после геометрической фильтрации все явные ошибки и не консистентные значения будут удалены и не войдут в конечное облако точек.

У данного подхода есть положительная и отрицательная стороны. Положительная заключается в том, что дорисовка карт глубин не

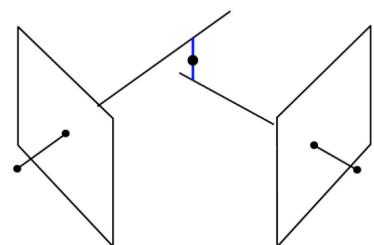


Рис. 6: Ошибка репроекции.

может привести к деградации точности. Отрицательный момент в том, что несогласованная дорисовка разных кадров приводит к удалению областей.

Вычисление карты нормалей При модификации системы PhotoScan было достаточно заменить только сами карты глубин. Для COLMAP результатом MVS являются пары и карт глубин и карт нормалей. По этой причине необходимо было дополнительно изменить исходные карты нормалей согласованно с дорисованными картами глубин.

Для вычисления карты нормалей, дорисованная карта глубин проецировалась в 3D, где у созданного облака точек вычислялись значения нормалей с помощью библиотеки Open3D [40]. Сгенерированные нормали заданы с точностью до направления. По этой причине находились все вектора, направленные в противоположную от наблюдателя сторону, и их направление менялось.

Пример результатов На Рис. 7 представлен результат для системы PhotoScan. Как можно видеть, большая часть изначально пропущенной части модели была восстановлена. В первую очередь видно восстановление стены и кресла, которые были пропущены оригинальной системой.

Дополнительно обнаружилось ещё одно свойство DeepDepth – пространственная фильтрация (spatial filtering) для имеющихся участков карт глубин: для исходно плоских объектов наблюдается “разглаживание”. Это заметно при сравнении одинаковых участков стен, но наиболее при сравнении кроватей.

Как и говорилось ранее, COLMAP осуществляет серьезную проверку согласованности перед генерацией плотного облака точек. По этой причине большая часть дорисовок не попала в конечную модель.

На Рис. 9 представлен результат вычисленных метрик качества для 3D моделей, представленных на Рис. 7 и Рис. 8. Для вычисления метрик все модели были предварительно вручную совмещены с эталоном.

От метрик изначально ожидалось, что будет стабильное повышение



Рис. 7: Photoscan.png

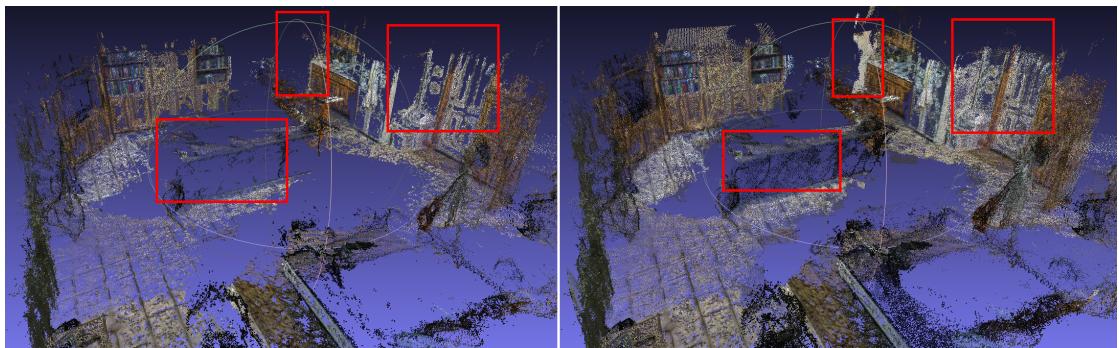


Рис. 8: ColMap.png

полноты при умеренном снижении точности. Однако результаты оказались неожиданными.

Во-первых, можно видеть, что для COLMAP точность только повысилась. Можно предположить, что это произошло вследствие того, что более плотное облако точек позволило построить более стабильную и точную полигональную сетку (mesh), хотя и нельзя утверждать что реально послужило причиной такого поведения. Важно лишь то, что модификация карт глубин для COLMAP оказалось успешной.

Во-вторых, несмотря на то, что для PhotoScan большая часть сцены была восстановлена, высокая ошибка восстановленных участков привела к тому, что метрика полноты заметно не выросла. При этом произошло существенное снижение точности. В частности, дорисованная стена оказалась слишком удалена от эталонной (около 15-20 см). Так как оба графика были построены только до 3-х см, то она не внесла вклад в метрику полноты, при этом заметно снизив точность.

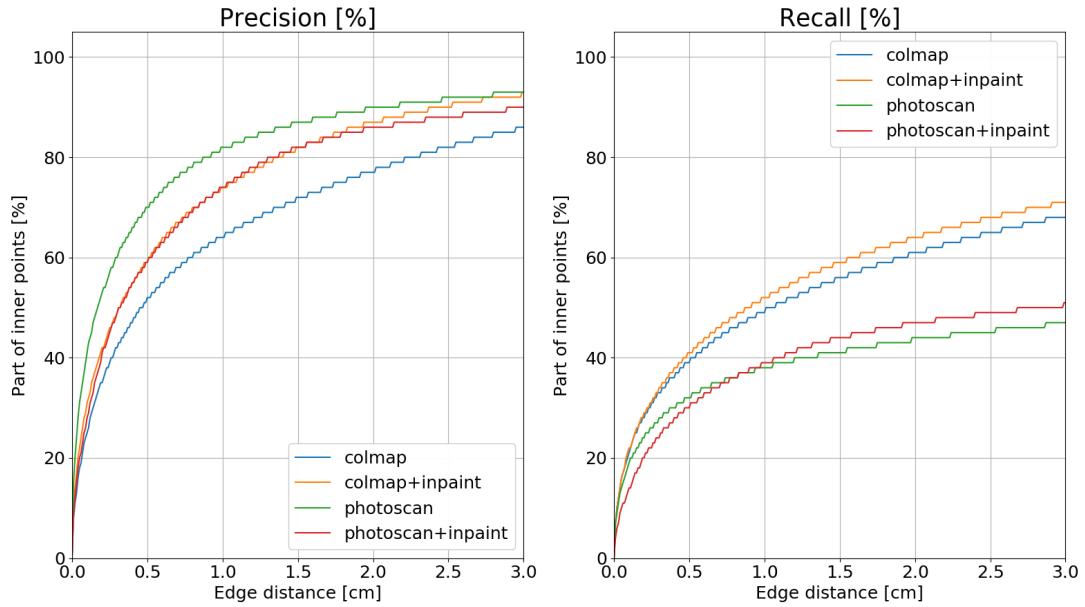


Рис. 9: Метрики качества

4. Тестирование

4.1. Выбор датасета для тестирования

Для тестирования систем реконструкции необходим датасет, содержащий фотографии реальных сцен и высококачественные 3D модели этих же сцен.

При тестировании и сравнении методов *depth inpainting* использовался датасет SceneNN [29]. Он полностью удовлетворяет заявленным выше требованиям. Единственным минусом датасета SceneNN [29] является его малый размер – всего 100 различных сканов.

При тестировании дорисовок единицей измерения являлись карты глубин. Для каждой сцены из SceneNN [29] можно сгенерировать около различных 100 карт. По этой причине общий размер тестируемого набора составлял порядка 10 000 единиц.

При вычислении метрик качества для 3D моделей единицей измерения является уже сама 3D модель. Для SceneNN размер выборки для метрик

качества тогда получается не более 100, что мало для получения принятия решения о статистической значимости результатов.

Результаты тестирования на [29], а также кросс тестирование авторами на датасетах [17] и [28] позволяют заключить, что DeepDepth обладает широкой обобщающей способностью и не заточен под конкретный датасет.

Как видно из таблицы на Рис. 10, результаты тестирования на датасете ScanNet [28] почти никак не зависят от обучающего набора. В итоге было принято решение в качестве тестового набора использовать ScanNet [28], который содержит более 1500 3D сканов, а в качестве весов взять предтренированные веса, полученные при обучении на датасете MatterPort3D [17].

Нельзя достоверно утверждать, какие именно данные использовались при тренировки моделей, выложенных в общий доступ. Однако, даже если выложенные предтренированные веса получились при обучении сразу на обоих датасетах (MatterPort3D и ScanNet), то, в случае тестирования на ScanNet, это не имеет значение, так как не может дать значимый прирост в качестве.

Train	Test	Rel	RMSE
Matterport3D	Matterport3D	0.089	0.116
ScanNet	Matterport3D	0.098	0.128
Matterport3D	Scannet	0.042	0.065
ScanNet	ScanNet	0.041	0.064

Рис. 10: Таблица результатов кросс тестирования

4.2. Описание структуры датасета и схемы тестирования

Описание датасета ScanNet Для получения датасета ScanNet использовался Robust Vision Challenge 2018 Devkits [25].

ScanNet представляет собой набор из более чем 700 различных сцен и 1500 сканов. Каждый скан представляет собой видео, полученного от RGB-D сенсора, а именно каждый кадр состоит из пары RGB фотографии и проецированной на эту фотографию карту глубины.

Имеются матрицы положения камеры для большинства кадров а также конечная 3D модель, в координатах которой и заданы данные матрицы.

Эти данные будут использоваться для сопоставления 3D моделей, полученных системами RGB реконструкции с предоставленной “эталонной” моделью.

“Эталонные” модели были получены с помощью системы RGB-D реконструкции BundleFusion [3], которая на данный момент является наиболее точной из всех существующих.

Создание подвыборки кадров В исходном виде каждый скан состоит из последовательных кадров видео. В каждый скане обычно состоит от 5000 до 10000 кадров.

COLMAP для этапа SfM имеет несколько стратегий сопоставления кадров:

- exhaustive matcher
- vocab tree matcher
- sequential matcher
- spatial matcher
- transitive matcher

Наиболее полное и точное сопоставление – *exhaustive matcher*. Оставшиеся ограничивают вычисления связей между кадрами путем тех или иных эвристик. Однако даже для наиболее быстрого метода вычисление положений камеры может занимать несколько дней до сходимости.

Для системы PhotoScan не существует выбора стратегии сопоставления и всегда вычисляется полная карта. Дополнительно оказалось, что при наличии большого количества кадров с малой разницей положений этап SfM в системе PhotoScan расходится.

В итоге для уменьшения времени вычислений и сохранения надежности алгоритма SfM в том числе и для PhotoScan (чтобы обеспечить одинаковый набор данных для обоих систем), было решено из исходной последовательности оставлять только каждый n -й кадр. n вычислялся исходя из общего количества кадров в скане. Целью было достичь количество кадров в интервале от 50 до 200. Дополнительным условием было то, чтобы n был не меньше 10, так как в противном случае PhotoScan часто расходился.

4.3. Создание системы автоматического сопоставления реконструированных и эталонных 3D моделей.

3D модели, получаемые в процессе реконструкции, определены с точностью до любых аффинных преобразований. По этой причине, для вычисления метрик качества необходимо найти преобразование между реконструированной и эталонной моделями.

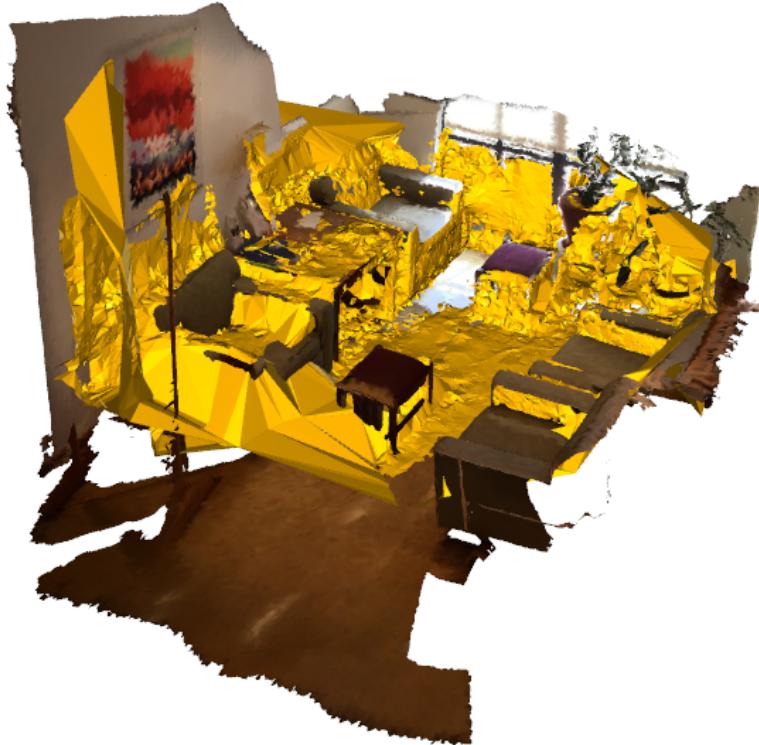


Рис. 11: Пример сопоставления 3D модели с эталоном

Нахождение преобразования между неструктурированными облаками точек является одним из главных этапов задачи регистрации – объединения нескольких разрозненных облаков точек в единую модель. По этой причине существует множество методов, решающих задачу нахождения оптимального преобразования.

Наиболее популярный метод решения задачи сопоставления на дан-

ный момент – это алгоритм Iterative Closest Point (ICP) [2, 4, 20]. Чтобы найти преобразование, ICP использует гипотезу, что для каждой точки одного облака соответствующей ей точкой из другого будет ближайшая к первой. Для большинства используемых точек данная гипотеза оказывается верной, но только в том случае, когда уже осуществлено начальное выравнивание моделей. Чем хуже начальное выравнивание, тем ICP дольше работает и тем чаще расходится.

Нахождение попарных соответствий Для начального выравнивания в случае отсутствия дополнительной информации по телеметрии наиболее распространенным методом является нахождение попарных соответствий путем вычисления (*feature detection*) и сопоставления (*feature matching*) локальных особенностей каждого облака. На данный момент имеются как традиционные [26, 35], так и нейросетевые [6, 13] методы.

Однако все методы *feature detection* обладают критическими недостатками. Во-первых, большинство из них не инвариантны по отношению к масштабу. Во-вторых, все подобные методы весьма чувствительны к шуму. Реконструированные облака точек как правило оказываются настолько шумными или неравномерно разреженными, что ни один из существующих методов не в состоянии надежно находить соответствия.

В рассматриваемом случае попарные соответствия уже имеются. Этими соответствиями являются положения камеры для каждого кадра. Данные о положениях имеют автоматическое сопоставление по номеру кадра, к тому же, среди всех точек модели, они обладают наибольшей точностью. Однако при этом существуют и

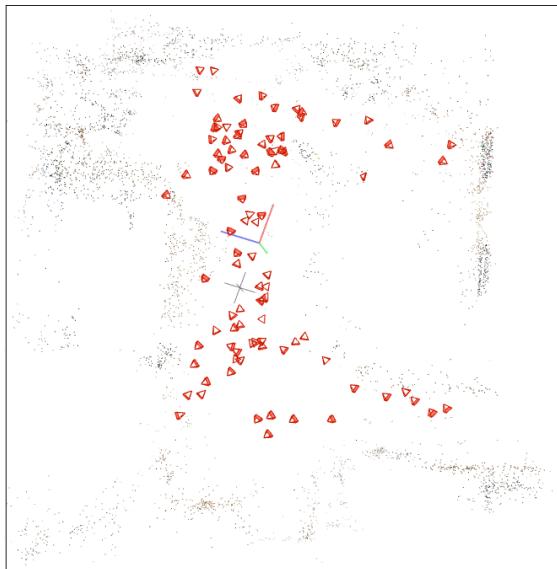


Рис. 12: Пример положения камер

отрицательные стороны. Вы-первых, малое количество кадров с известными положениями: всего 50-200 фотографий, и не для каждой из них рассматриваемые системы смогли оценить положения камер (иногда только для 7-8 фотографий). Во-вторых, довольно часто встречаются выбросы, когда для одного/двух кадров положения вычислены абсолютно не верно.

Использование случайной подвыборки Если с малым количеством ничего сделать нельзя, то отфильтровать от выбросов уже можно. Для этого бралась случайная подвыборка, для которой вычислялись начальное преобразование и нормализованная метрика ошибки для данного преобразования. Таким, образом, если для исходного множества положений камер лишь для небольшого подмножества положения были оценены с большой ошибкой, то в конечной подвыборке, данных выбросов уже не будет.

Для нахождения матрицы преобразования использовалась библиотека PCL [27]. В данной библиотеке не было возможности поиска оптимального глобального преобразования с учетом масштаба. В связи с чем масштаб находился отдельно.

Поиск масштаба Расстояние между положениями любых двух точек модели зависит только от масштаба. Более того, эта зависимость линейна: если изменить масштаб модели то все расстояния между всеми точками модели изменятся на ту же величину. Значит, если для каждой пары точек найти расстояние между ними в обоих моделях, то отношение данных расстояний и будет масштабным преобразованием между моделями.

В случае достаточной мощности выборки (больше 100) за масштаб бралась мода распределения, вычисленная как аргумент максимума гистограммы. Если же выборка была не большой, то для избежания неустойчивости бралась медиана.

После нахождения масштаба и начального выравнивания применялся ICP из библиотеки Open3D [40] для всего облака точек для уточнения.

Ограничения на совмещения В процессе работы выяснилось, что качество совмещения – наиболее значимый фактор при вычислении метрик

точности и полноты. Значит в итоге по факту меряются не системы реконструкции, а то, насколько удачно удалось вычислить совмещение для одной и для другой модели.

В итоге, чтобы преодолеть влияния данного фактора, совмещение вычислялось только для одной модели и применялась сразу для обеих. Это не приводит к расхождению, так как начальные координаты у исходной и модернизированной моделей идентичны.

Возможно, что исходные координаты немного, но отличаются. Тогда та модель, для которой совмещение оказалось наиболее удачным, имеет преимущество. По этой причине преобразование вычислялось всегда только для исходной модели.

Для двух моделей преобразование будет лучше приближать идеальное для той, для которой оно, собственно, и считалось. Тогда, если разница в качестве совмещений и влияет на конечный результат, то только в худшую сторону.

Удаления неудачных моделей из выборки Так как совмещение производилось автоматическое, то были также и неудачные случаи с ошибочным совмещением. Помимо ошибок при совмещении были также ошибки этапа SfM при вычислении положений камер.

Для детектирования данных случаев все построенные модели были вручную просмотрены и те, которые разошлись или не были построены, были удалены из выборки.

По результатам ручной фильтрации выяснилось, что для обоих систем для около 30% сцен всего набора данных этап SfM не смог выдать удовлетворительные данные по положениям камер.

Ещё около 5% оставшегося набора была удалено из рассматривания вследствие ошибок совмещения.

4.4. Вычисление метрик

Чтобы вычислить метрику точности $P(d)$ необходимо найти количество точек в реконструированной модели, для которых расстояние до эталонной модели меньше, чем d .

Для этого можно для каждой точки реконструированной модели найти расстояние до эталонной. Далее, так как не важен порядок этих точек, то полученный массив расстояний можно отсортировать, чтобы в дальнейшем за логарифмическое время вычислять значение метрики для каждого d .

Для метрики полноты все действия аналогичны. Разница лишь в том, что теперь ищутся расстояния от каждой точки эталонной модели до реконструированной.

Как видно из определения метрик, они основаны на представлении 3D модели в качестве плотного облака точек и никак не использует информацию от представления в виде полигональной сетки.

Это связано в частности с тем, что эталонные данные датасета [34] были получены с помощью высокоточного лидара и представляют собой плотные облака точек.

Если для представления в виде облака точек необходимо равномерно заполнить все пространство представляемых объектов, то меши позволяют значительно уменьшить количество используемых памяти и вычислительных ресурсов, осуществляя плотную детализацию только там, где это необходимо.

ScanNet же предоставляет эталонные 3D модели в виде мешей, поэтому имеет смысл модифицировать метрики применительно к мешам

Основное отличие мешей от облака точек заключается в дополнительной информации о связях между точками, объединяя связанные точки одним полигоном. Таким образом "точкой" 3D модели, представленной в виде меша, является любая математическая точка на любом полигоне.

Введем новые определения:

- \mathcal{G}_m – множество многоугольников (**vs полигонов**) эталонной модели
- \mathcal{R}_m – множество полигонов реконструкции

Тогда точность (**Precision**, $P_m(d)$) и полнота (**Recall**, $R_m(d)$) с учетом новых множеств будут равны будут равны:

$$P_m(d) = \frac{\text{card} \{ \mathbf{r} \in \mathcal{R} : e_{\mathbf{r} \rightarrow \mathcal{G}_m} < d \}}{\text{card} \{ \mathcal{R} \}} \quad P_m(d) = \frac{\text{card} \{ \mathbf{g} \in \mathcal{G} : e_{\mathbf{g} \rightarrow \mathcal{R}_m} < d \}}{\text{card} \{ \mathcal{G} \}} \quad (4.1)$$

Отличие от предшествующих определений в том, что для трехмерных точек

$$\mathbf{a}, \mathbf{b} \in \mathbb{R}^3 : dist(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|$$

в то время как расстояние между точкой \mathbf{a} и полигоном \mathbf{b}_m :

$$dist(\mathbf{a}, \mathbf{b}_m) = \min_{\mathbf{b}_p \in \mathbf{b}_m^+} \|\mathbf{a} - \mathbf{b}\|$$

здесь под символом \mathbf{b}_m^+ понимается все математическое множество точек, принадлежащих полигону.

В идеале, при вычислении метрик необходимо вычислить расстояния от каждой математической точки модели. Так как этих точек бесконечное количество, то сделать это невозможно. Вместо этого можно произвести повышение плотности распределения узлов в полигональной сетке.

ScanNet предоставляет 2 типа мешей: исходный и децимированный. В среднем, в децимированном меше количество точек в 50-100 раз меньше.

Для повышения плотности распределения узлов сгенерированного меша использовалась библиотека CGAL [41].

Были вычислены метрики для исходных и плотных мешей для нескольких выборочных реконструированных моделей. Разница во времени вычисления на 8-ядерном процессоре с учетом уплотнения сгенерированной модели оказалось в 20-30 раз. Разница в значениях метрик при этом находилась в пределах 0.3-0.5% в абсолютном измерении, причем для обоих вариантов (исходный и дорисованный) разница имела одинаковый знак. Отличия же в разнице между дорисованной и исходной моделью была менее 0.01% в абсолютном измерении.

Было решено не создавать равномерно распределенного по поверхности меша множества точек для реконструированных моделей, а также в качестве эталонного меша использовать децимированный. Общее время вы-

числения метрик для всего множества составила в итоге около суток. Если бы все считалось на плотных моделях, то общее время заняло бы несколько недель.

Расстояние до ближайшей точки в соседнем облаке вычислялось с помощью kd дерева [1] из библиотеки PCL [27]. Расстояния до ближайших полигонов соседнего меша вычислялись с помощью AABB дерева [42] из библиотеки CGAL [41].

4.5. Полученные результаты

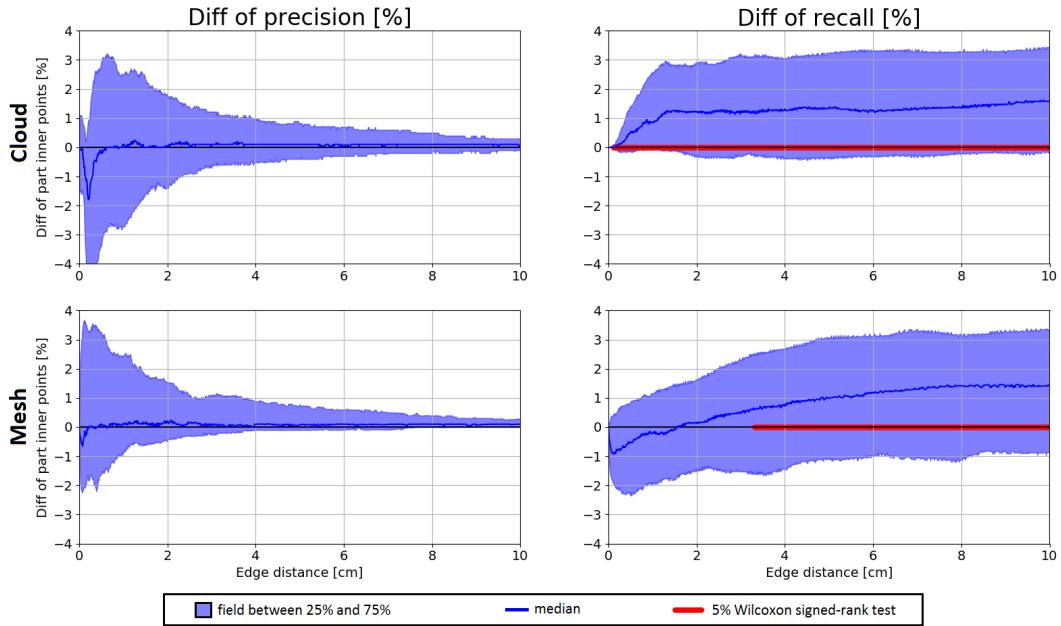


Рис. 13: Разница метрик качества модифицированной и исходной системы COLMAP

Результатом проделанной работы является повышение качества существующих систем 3D реконструкции. Поэтому основными метриками достижения данного результата должно быть значение прироста метрик качества, а не их абсолютные значения. Это особенно важно, так как выборки значений метрик для оригинальной и модифицированной системы являются парными. Если не учитывать парность, то большая часть информации может быть утеряна.

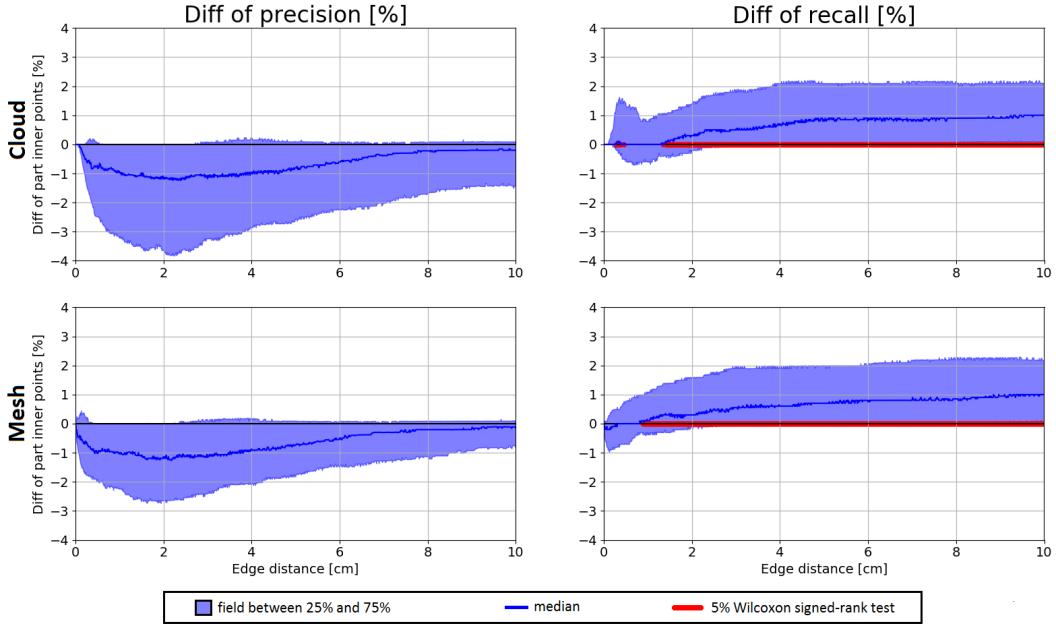


Рис. 14: Разница метрик качества модифицированной и исходной системы PhotoScan

Жирной синей линией изображена медиана выборки, полупрозрачной синей областью – граница между нижним и верхним квартилями. Красной полосой на правых графиках – область, где тест Вилкоксона с уровнем значимости 5% говорит, что распределение отлично от нуля. Так как медиана выше нуля, то и все распределение тоже. Значит, данные области – это те области, где удалось значительно улучшить существующую модель.

В качестве теста на значимость применялся тест Вилкоксона, так как выборки являются парными, а их распределение далеко от нормального.

Далее будут обсуждаться только метрики, основанные на полигональных сетках. Метрики, основанные на облаках точек, приведены лишь по той причине, что данные метрики уже используются при сравнении качества 3D моделей на датасете “Tanks and Temples” [34].

Исходя из полученных графиков можно заключить, что для системы COLMAP цель была достигнута. Для расстояний более 3 см метрика полноты значительно больше, чем у исходной системы. При этом медиана раз-

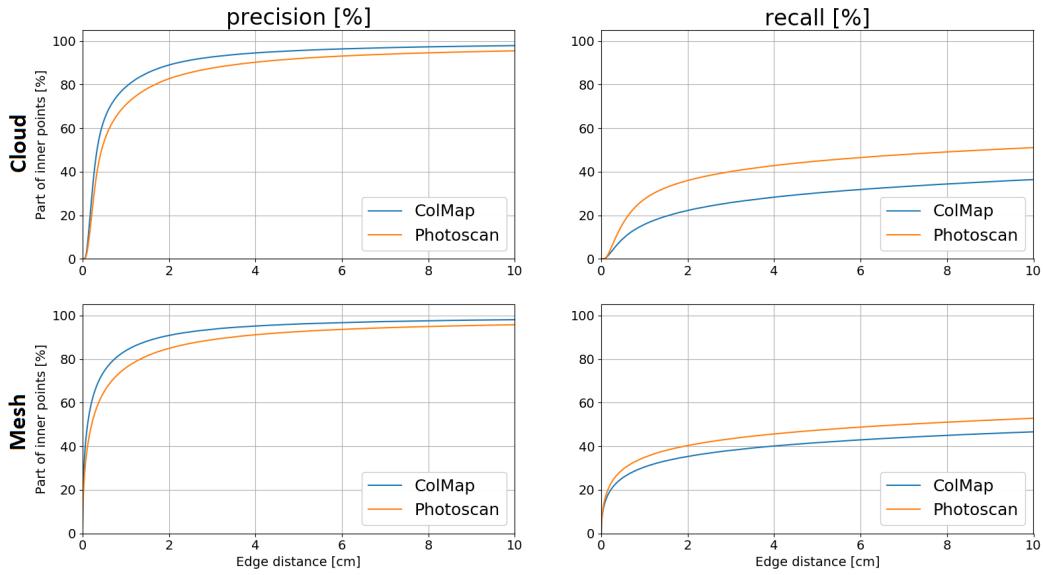


Рис. 15: Средние значения метрик на тестируемом наборе.

ницы метрик точности также выше нуля, значит точность, как минимум, не ухудшилась.

Для системы PhotoScan также удалось увеличить полноту построенных моделей, однако при этом имеется значимая деградация по точности. Возможно, что в случае наличия геометрической фильтрации при построении облака точек, точность бы не ухудшилась. Однако написание системы геометрической фильтрации выходит за рамки данной работы.

На Рис. 15 представлены средние значения метрик на тестируемом наборе. Необходимо уточнить, что данные метрики вычислялись ровно на том же наборе данных, что и при тестировании модифицированных систем с исходными. Другими словами, все те сцены, которые были исключены из тестируемого набора вследствие расхождения алгоритма SfM, или же слишком большой ошибки автоматического совмещения с эталонной моделью, также не учитывались и при построении графика на Рис. 15. Это сделано намеренно, чтобы показать именно значения метрик для моделей, которые использовались при тестировании выше, вместо того, чтобы тестировать сами системы.

Заключение

В рамках данной работы были (достигнуты следующие результаты) / (произведены следующие действия).

1. Рассмотрены методы, позволяющие повысить полноту построения трехмерных моделей.
2. Предложена идея использовать методы “depth inpaint” для коррекции карт глубин, получаемых на этапе Multi-View Stereo (MVS).
3. Создан датасет отрендеренных эталонных карт глубин и парных к ним RGB-фотографий из датасета SceneNN.
4. На сгенерированных данных, типичных для результата работы MVS систем произведено тестирование методов “depth inpaint” и выявлен лучший.
5. Осуществлено встраивание DeepDepth в лидирующие системы трехмерной RGB реконструкции: COLMAP и PhotoScan.
6. Произведено тестирование исходных и модифицированных систем реконструкций на датасете ScanNet.
7. Доказана значимость увеличения значений метрик полноты для систем COLMAP и PhotoScan и одновременное отсутствие уменьшения значения метрик точности для системы COLMAP.

Список литературы

- [1] Bentley Jon Louis. Multidimensional Binary Search Trees Used for Associative Searching // Commun. ACM. — 1975. — Sep. — Vol. 18, no. 9. — P. 509–517. — Access mode: <http://doi.acm.org/10.1145/361002.361007>.
- [2] Besl Paul J., McKay Neil D. A Method for Registration of 3-D Shapes // IEEE Trans. Pattern Anal. Mach. Intell. — 1992. — Feb. — Vol. 14, no. 2. — P. 239–256. — Access mode: <http://dx.doi.org/10.1109/34.121791>.
- [3] BundleFusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Re-integration / Angela Dai, Matthias Nießner, Michael Zollöfer et al. // ACM Transactions on Graphics 2017 (TOG). — 2017.
- [4] Chen Yang, Medioni Gérard. Object modeling by registration of multiple range images // Image and Vision Computing. — 1992. — 04. — Vol. 10. — P. 145–155.
- [5] Choi Sungjoon, Zhou Qian-Yi, Koltun Vladlen. Robust Reconstruction of Indoor Scenes // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2015.
- [6] Deng Haowen, Birdal Tolga, Ilic Slobodan. PPFNet: Global Context Aware Local Features for Robust 3D Point Matching. — 2018. — 10.
- [7] Depth from Motion for Smartphone AR / Julien Valentin, Adarsh Kowdle, Jonathan T. Barron et al. // ACM Trans. Graph. — 2018. — Dec. — Vol. 37, no. 6. — P. 193:1–193:19. — Access mode: <http://doi.acm.org/10.1145/3272127.3275041>.
- [8] Furukawa Yasutaka, Ponce Jean. Accurate, Dense, and Robust Multi-View Stereopsis // IEEE Trans. on Pattern Analysis and Machine Intelligence. — 2010. — Vol. 32, no. 8. — P. 1362–1376.

- [9] Galliani Silvano, Lasinger Katrin, Schindler Konrad. Massively Parallel Multiview Stereopsis by Surface Normal Diffusion. — 2015. — June.
- [10] Guided Depth Enhancement via a Fast Marching Method / XIAOJIN GONG, JUNYI LIU, WENHUI ZHOU, JILIN LIU // Image Vision Comput. — 2013. — Oct. — Vol. 31, no. 10. — P. 695–703. — Access mode: <http://dx.doi.org/10.1016/j.imavis.2013.07.006>.
- [11] Holynski Aleksander, Kopf Johannes. Fast Depth Densification for Occlusion-aware Augmented Reality // ACM Trans. Graph. — 2018. — Vol. 37, no. 6.
- [12] Joint Bilateral Upsampling / Johannes Kopf, Michael F. Cohen, Dani Lischinski, Matt Uyttendaele // ACM Transactions on Graphics (Proceedings of SIGGRAPH 2007). — 2007. — Vol. 26, no. 3. — P. to appear.
- [13] Khouri Marc, Zhou Qian-Yi, Koltun Vladlen. Learning Compact Geometric Features // 2017 IEEE International Conference on Computer Vision (ICCV). — 2017. — P. 153–161.
- [14] Levin Anat, Lischinski Dani, Weiss Yair. Colorization Using Optimization // ACM Trans. Graph. — 2004. — Aug. — Vol. 23, no. 3. — P. 689–694. — Access mode: <http://doi.acm.org/10.1145/1015706.1015780>.
- [15] MVSNet: Depth Inference for Unstructured Multi-view Stereo / Yao Yao, Zixin Luo, Shiwei Li et al. // European Conference on Computer Vision (ECCV). — 2018.
- [16] Ma Fangchang, Karaman Sertac. Sparse-to-Dense: Depth Prediction from Sparse Depth Samples and a Single Image. — 2018.
- [17] Matterport3D: Learning from RGB-D Data in Indoor Environments / Angel Chang, Angela Dai, Thomas Funkhouser et al. // International Conference on 3D Vision (3DV). — 2017.
- [18] Moulon Pierre, Monasse Pascal, Marlet Renaud. Adaptive Structure from Motion with a Contrario Model Estimation // Proceedings of the Asian

- Computer Vision Conference (ACCV 2012). — Springer Berlin Heidelberg, 2012. — P. 257–270.
- [19] Moulon Pierre, Monasse Pascal, Marlet Renaud, Others. OpenMVG. An Open Multiple View Geometry library. — <https://github.com/openMVG/openMVG>.
 - [20] Park Jaesik, Zhou Qian-Yi, Koltun Vladlen. Colored Point Cloud Registration Revisited // ICCV. — 2017.
 - [21] Photoscan. — Access mode: <https://www.agisoft.com> (online; accessed: 08.04.2019).
 - [22] Pix4d. — Access mode: <https://www.pix4d.com> (online; accessed: 08.04.2019).
 - [23] Pixelwise View Selection for Unstructured Multi-View Stereo / Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, Jan-Michael Frahm // European Conference on Computer Vision (ECCV). — 2016.
 - [24] 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.
 - [25] Robust Vision Challenge 2018 Devkits. — Access mode: http://cvlibs.net:3000/ageiger/rob_devkit (online; accessed: 08.04.2019).
 - [26] Rusu Radu, Blodow Nico, Beetz Michael. Fast Point Feature Histograms (FPFH) for 3D registration. — 2009. — 06. — P. 3212 – 3217.
 - [27] Rusu R. B., Cousins S. 3D is here: Point Cloud Library (PCL) // 2011 IEEE International Conference on Robotics and Automation. — 2011. — May. — P. 1–4.
 - [28] ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes / Angela Dai, Angel X. Chang, Manolis Savva et al. // Proc. Computer Vision and Pattern Recognition (CVPR), IEEE. — 2017.

- [29] SceneNN: A Scene Meshes Dataset with aNNotations / Binh-Son Hua, Quang-Hieu Pham, Duc Thanh Nguyen et al. // International Conference on 3D Vision (3DV). — 2016.
- [30] SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-training on Indoor Segmentation? / John McCormac, Ankur Handa, Stefan Leutenegger, Andrew J.Davison. — 2017.
- [31] Semantic Scene Completion from a Single Depth Image / Shuran Song, Fisher Yu, Andy Zeng et al. // IEEE Conference on Computer Vision and Pattern Recognition. — 2017.
- [32] Snavely Seitz, Szeliski. Photo Tourism: Exploring image collections in 3d // SIGGRAPH. — 2006.
- [33] Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs. / Jiejie Zhu, Liang Wang, Jizhou Gao, Ruigang Yang // IEEE Trans Pattern Anal Mach Intell. — 2010. — Vol. 32, no. 5. — P. 899–909.
- [34] Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction / Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, Vladlen Koltun // ACM Transactions on Graphics. — 2017. — Vol. 36, no. 4.
- [35] Tombari Federico, Salti Samuele, Di Stefano Luigi. Unique shape context for 3D data description. — 2010. — 01.
- [36] Wu. Changchang. Visualsfm : A visual structure from motion system. — Access mode: doi.acm.org/10.1145/1015706.1015780.
- [37] Xiao Jianxiong, Owens A., Torralba A. SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels // Computer Vision (ICCV), 2013 IEEE International Conference on. — 2013. — Dec.
- [38] Xu Qingshan, Tao Wenbing. Multi-Scale Geometric Consistency Guided Multi-View Stereo // Computer Vision and Pattern Recognition (CVPR). — 2019.

- [39] Zhang Yinda, Funkhouser Thomas. Deep Depth Completion of a Single RGB-D Image // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).—2018.
- [40] Zhou Qian-Yi, Park Jaesik, Koltun Vladlen. Open3D: A Modern Library for 3D Data Processing // arXiv:1801.09847.—2018.
- [41] Cgal, Computational Geometry Algorithms Library.—<http://www.cgal.org>.
- [42] van den Bergen Gino. Efficient Collision Detection of Complex Deformable Models using AABB Trees // journalId:00002201.—1997.—01.—Vol. 2.

Приложения

Примеры модификации 3D моделей

