Санкт-Петербургский государственный университет

АХМЕТЬЯНОВ Азат Ришатович

Выпускная квалификационная работа

Оценка глубины сцены из стереоснимков, полученных с помощью смартфонов, для 3D-реконструкции

Уровень образования: бакалавриат

Направление *09.03.04 «Программная инженерия»*

Основная образовательная программа *СВ.5080.2018 «Программная инженерия»*

Научный руководитель:
Доцент кафедры системного программирования, к.т.н. Ю. В. Литвинов

Рецензент:
Научный сотрудник, Сколковский институт науки и технологий, к.ф.-м.н. А. В. Артемов

Консультант:
Инженер-исследователь, Сколковский институт науки и технологий А. В. Корнилова

Санкт-Петербург
2022

Saint Petersburg State University

Azat Akhmetianov

Bachelor's Thesis

Depth estimation from smartphone stereo for 3D reconstruction

Education level: bachelor

Speciality *09.03.04 «Software Engineering»*

Programme *CB.5080.2018 "Software Engineering"*

Scientific supervisor:
C.Sc., docent. Y.V. Litvinov

Reviewer:
C.Sc., Researcher at Skoltech A.V. Artemov

Consultant:
Research engineer at Skoltech A.V. Kornilova

Saint Petersburg
2022

Оглавление

1. Введение **4**

2. Постановка задачи **6**

3. Обзор **7**

3.1. Синхронизация кадров 7

3.2. Получение глубины из стереоизображений 8

3.3. Монокулярные методы получения глубины на смартфонах 10

4. Сравнение методов получения глубины на смартфонах с предложенным подходом **13**

4.1. Общее описание экспериментов 13

4.2. Тестовый набор данных для сравнения 14

4.3. Выбор метрик 18

4.4. Количественное сравнение методов 20

4.5. Качественное сравнение методов 20

5. Реализация подхода к оценке глубины по стереоизображениям с двух смартфонов **23**

5.1. Архитектура приложения 23

5.2. Демонстрация работы 25

6. Результаты **27**

7. Благодарности **28**

Список литературы **29**

1. Введение

В современном мире технологии трехмерной оцифровки предметов и людей находят всё большее применение в приложениях виртуальной и дополненной реальности, онлайн-примерочных, онлайн-платформах электронной торговли (online marketplace) для демонстрации товаров. Развитие и широкое распространение смартфонов диктует спрос на портирование технологий сканирования на данные устройства — на флагманских линейках телефонов появляются датчики глубины и лидары, в магазинах мобильных приложений — всё больше приложений для трехмерного сканирования [19, 32].

Применение датчиков глубины на смартфонах увеличивает качество и реализм 3D-реконструкции за счет дополнительной информации о геометрии сцены. При этом наличие таких датчиков значительно увеличивает стоимость устройств, и, как следствие, их распространенность. Альтернативой применению датчиков глубины является использование алгоритмов Structure from Motion (SfM) [29], либо использование машинного обучения для восстановления глубины по цветному изображению [20]. В частности, SfM, основанный на анализе последовательности RGB снимков, имеет адаптацию на мобильной платформе в библиотеке ARCore [14]. Данные технологии делают возможной трехмерную оцифровку на бюджетных смартфонах с RGB камерой, однако, по качеству и скорости получения реконструкции они не могут приблизиться к подходам, использующим датчики глубины.

В качестве более доступного варианта для оценки глубины можно воспользоваться двумя бюджетными смартфонами с RGB камерами, сконструировав из них стереокамеру (рис. 1). Возможность регулировать расстояние между камерами (baseline) позволит адаптировать данную систему на широкий набор сценариев съемки, например, сканирование небольших предметов или же помещений и пространств. К достоинствам подхода с получением глубины из стереоизображения стоит отнести меньшие требования к аппаратному обеспечению устройств, по сравнению с методами, использующими глубокое обучение. Также

стереоизображение помогает избежать проблем с динамически меняющимися сценами, которые возникают в SfM методе.

Рис. 1: Пример использования системы из двух смартфонов [25]

К ограничениям данного подхода стоит отнести проблему *синхронизации съемки изображений* на двух смартфонах, которая влияет на качество полученной карты глубины во время перемещения камеры при сканировании и в случае наличия движения в снимаемой сцене.

Применение стереосистемы из двух смартфонов в задаче оценки глубины на мобильных устройствах может дать преимущества в виде гибкости к различным сценам за счет возможности изменять расстояние между камерами, устойчивости к динамическим сценам по сравнению с SfM-методом. В отличии от методов, анализирующих изображение с одной камеры с использованием глубокого обучения, анализ стереоизображений позволяет получать не относительные, а метрические данные о дистанциях до объектов. Таким образом, имеет смысл исследовать возможность применения системы из двух смартфонов для оценки глубины изображения.

2. Постановка задачи

Целью работы является исследование подхода к получению глубины на основе анализа стереоизображений, полученных с камер двух смартфонов, по сравнению с другими методами. Для её выполнения были поставлены следующие задачи:

1. выполнить обзор методов получения глубины на мобильных устройствах и методов получения синхронизированных кадров с двух смартфонов;
2. создать тестовый набор данных для сравнения методов получения глубины;
3. выбрать метрики и провести сравнение методов получения глубины на смартфонах с предлагаемым подходом;
4. реализовать подход к получению глубины из стереопары на мобильных устройствах.

3. Обзор

В данном разделе рассматриваются способы синхронизации кадров на камерах смартфонов, а также методы получения глубины на смартфонах и методы получения глубины по стереоизображению.

3.1. Синхронизация кадров

При получении глубины из стереоизображений необходимо, чтобы камеры захватывали кадры почти одновременно. В то же время, в системе из двух смартфонов каждый из них снимает свою собственную последовательность кадров. Такие последовательности могут быть смещены во времени друг от друга на значения до 15 миллисекунд — половины периода кадров при съемке с частотой 30 кадров в секунду.

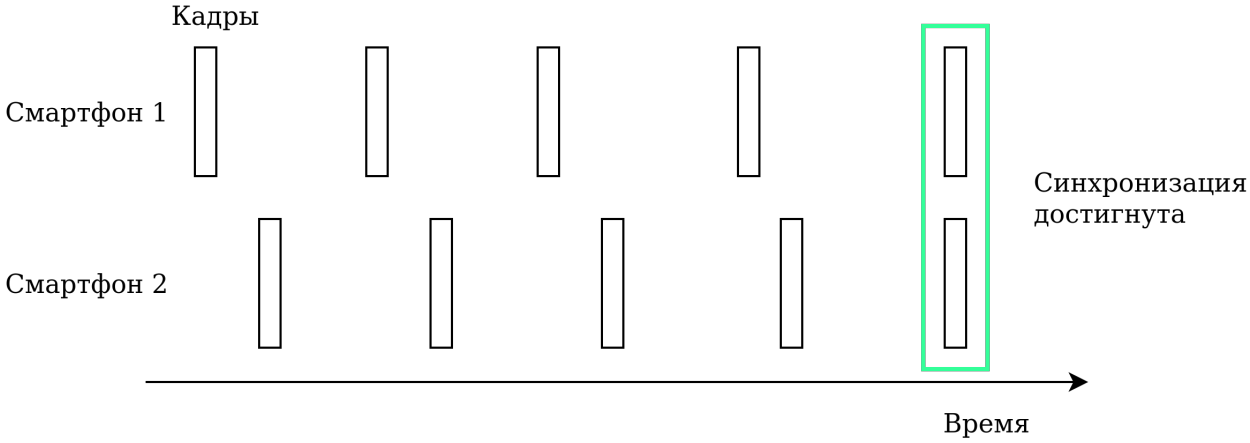


Рис. 2: Иллюстрация процесса синхронизации съемки.

На смартфонах, поддерживающих уровень Camera2API > LIMITED, в случае равенства периода кадров на устройствах, проблема синхронизации съемки отдельных изображений решается при помощи подхода смещения потока кадров камеры [22, 30], при этом среди них libsoftwaresync показывает лучшую точность синхронизации (ошибка $<$ 1 миллисекунды). Работа [23] расширяет подход из libsoftwaresync на синхронизированную запись видео в приложении RecSync на платформе Android, а также демонстрирует работоспособность предложенного метода синхронизации на 47 моделях смартфонов.

Метод	Поддержка видеозаписей	Точность синхронизации
SocialSync	+	несколько мс
libsoftwaresync	-	< 1 мс
RecSync	+	$<$ **1 мс**

Таблица 1: Сравнение методов синхронизации съемки.

3.2. Получение глубины из стереоизображений

Методы получения глубины из стереоизображений основаны на поиске похожих областей на двух изображениях и оценке их смещения (disparity). В классических методах меры «схожести» областей изображения не учитывают контекст сцены. В недавних работах исследовались методы с применением глубокого обучения [6, 2] для учета семантики изображений, например, при измерении «схожести» их областей. Такие подходы приводят к значительному улучшению качества оценки глубины, но при этом и более требовательны к аппаратным возможностям.

3.2.1. Классические методы получения глубины из стереоизображений

Методы в данной категории можно разделить на применяющие глобальную и локальную оптимизацию. В случае с локальной оптимизацией, глубина оценивается только с использованием информации из отдельных областей или блоков каждого изображения. Такие подходы [27, 21] очень чувствительны к областям изображения с резким изменением глубины или отсутствием текстуры. Для частичного преодоления этих проблем был создан метод Semi-Global Matching [16], который оценивает глубину в нескольких направлениях и минимизирует глобальную функцию энергии.

Методы для рассмотрения в этой категории выбраны по принципу наличия открытых реализаций и качества получаемой карты глубины относительно большинства других классических стереометодов [26].

- SGBM — реализованная [18] в библиотеке OpenCV (на языке C++) модификация метода SGM. Изменение по сравнению с оригинальным подходом заключается в добавленной функциональности обработки изображения по блокам заданного размера, что может использоваться с целью увеличения производительности.

- LIBELAS [11] — вероятностный метод, в котором карта глубины и связанные с ней параметры рассматриваются как случайные величины, что позволило авторам дополнительно снизить влияние отсутствия текстуры в областях изображения по сравнению с другими классическими методами (например, SGM). Данный метод реализован в виде библиотеки [1] на C++.

3.2.2. Методы получения глубины из стереоизображений с применением глубокого обучения

В недавних исследованиях в области анализа стереоизображений большое внимание уделялось методам, использующим глубокое обучение. Среди них можно выделить работу MC-CNN [33], впервые предложившую использовать глубокое обучение для измерения похожести областей изображения. В DispNet [17] глубокое обучение предлагается использовать для оптимизации на всех шагах из классических алгоритмов анализа стереоизображений. Такие работы вдохновили множество других исследований, направленных на улучшение качества в анализе стереоизображений [4, 12].

Несмотря на то, что данные методы показывают высокое качество карты глубины и занимают первые позиции в бенчмарках (например, KITTI [26]), их применимость на мобильных устройствах ограничена из-за требовательности к аппаратным ресурсам.

3.2.3. Получение глубины с использованием технологии dual-pixel

На некоторых мобильных устройствах также используется метод получения глубины с применением технологии dual-pixel, встроенной

в камерах некоторых смартфонов для улучшения качества автоматической фокусировки. В обычных камерах пиксели сенсора полностью захватывают свет, попадающий на них через диафрагму. Суть технологии dual-pixel в том, что на сенсоре камеры с этой технологией пиксели разделяются на две части, благодаря чему каждая половина пикселя захватывает свою часть изображения со слегка отличающимся обзором камеры. Это приводит к тому, что такая камера производит сразу два изображения, которые можно рассматривать как стерео с низким baseline (расстоянием между камерами) — это применяется в работах [7, 8]. Недостатком подхода является то, что малый baseline приводит к ошибкам в оценке глубины на больших расстояниях от камеры (например, для фона).

- В работе Du2Net предлагается способ совмещения информации, получаемой с помощью dual-pixel технологии, и из стерео камеры на одном смартфоне. Такой подход позволил улучшить результаты оценки глубины по сравнению с использованием каждой из данных технологий по отдельности.

3.3. Монокулярные методы получения глубины на смартфонах

Из-за низкой доступности датчиков глубины на мобильных устройствах, для большинства смартфонов актуально получение глубины с использованием изображений с одной камеры. Поэтому в данном подразделе рассматриваются доступные на смартфонах монокулярные методы получения глубины.

3.3.1. Structure from Motion

Structure from Motion (SfM) [29] — техника оценки глубины, использующая набор изображений сцены, захваченных в движении, либо с нескольких позиций в пространстве.

Получение глубины происходит за счет нескольких этапов: (1) выделение ключевых точек на изображениях, их сопоставление, (2) оценка движения камеры по сопоставленным точкам, (3) вычисление 3D структуры, используя выделенные точки и оцененное движение камеры. Такой подход подразумевает, что объекты сцены неподвижны, и в общем случае будет производить ошибки в динамичном окружении.

Для платформы Android существует библиотека ARCore [14], которая реализует подход SfM, а также распознает плоскости на сцене, благодаря чему разработчики приложений дополненной реальности могут реалистично размещать объекты на поверхностях. ARCore призван решать проблему использования дополненной реальности на смартфонах, связанную с тем, что большинство устройств имеют только одну камеру без дополнительного датчика глубины. К недостаткам библиотеки можно отнести, например, низкое разрешение получаемой карты глубины — согласно документации[1], для большинства устройств это значение близко к 160*120 пикселям.

3.3.2. Монокулярные методы с глубоким обучением

В недавних работах исследовалось также применение монокулярных подходов получения глубины с помощью глубокого обучения. В общем случае одного изображения недостаточно для того, чтобы достоверно оценить глубину сцены в произвольном окружении, однако модель обучается на наборах данных со сценами различного характера. В данных методах оценивается только относительная глубина объектов, что может приводить к проблемам со стабильностью значений глубины на последовательных кадрах с камеры.

В таких методах необходимо учитывать, что аппаратного обеспечения смартфонов (память, видеопроцессоры) не всегда достаточно для запуска методов с глубоким обучением. По этой причине в работах [20, 10] рассматриваются легковесные модели, оптимизированные для работы в условиях мобильных устройств.

[1] https://developers.google.com/ar/reference/java/com/google/ar/core/Frame#acquireRawDepthImage16Bits-

В недавнем исследовании [10] по сравнению монокулярных подходов для получения карты глубины метод, использующий модель PyDNet, показал себя одним из лучших по качеству с учетом ограничений мобильных устройств. С учетом этого факта и наличия открытой реализации на мобильной платформе рассмотрим работу [20] в данном разделе.

- Авторы метода [20] выбрали модель PyDNet по принципу требовательности к аппаратным возможностям устройств и дополнительно адаптировали ее для смартфонов. Также авторами предлагается решать проблему относительности значений глубины, комбинируя выход PyDNet с оценкой положения точек пространства из какого-либо SfM метода (например, ARCore). Данный подход реализован для платформ Android и iOS в открытом репозитории [13].

4. Сравнение методов получения глубины на смартфонах с предложенным подходом

В данном разделе приводится сравнение подхода к получению глубины, основанного на анализе изображений с двух смартфонов, с другими доступными на смартфонах подходами.

4.1. Общее описание экспериментов

К участию в сравнении были выбраны методы с открытыми реализациями из ключевых категорий, рассмотренных в обзоре.

- Стерео-методы — ELAS [11], SGBM [18].
- Монокулярные методы.
  - SfM — ARCore [14].
  - Методы с глубоким обучением, использующие одно изображение — mobilePyDNet [20].

Эксперименты предполагают два этапа: (1) количественное сравнение методов относительно референсных данных, полученных с высокоточной внешней камеры глубины, и (2) качественное сравнение методов в задаче 3D-реконструкции.

Количественное сравнение заключается в подсчете значений метрик на парах изображений: «карта глубины, полученная одним из методов — референсная карта глубины». В качественном сравнении принято решение использовать один из стандартных алгоритмов 3D реконструкции [3]. Данный алгоритм имеет открытую реализацию в библиотеке Open3D.

К набору данных для сравнения выработаны следующие требования.

1. Набор должен содержать данные о глубине изображений из библиотеки ARCore, поскольку их получение путем обработки предварительно записанных видеозаписей является нетривиальной задачей. Предоставляемая библиотекой функциональность Recording and Playback API [15] позволяет использовать предварительно записанные видео для получения глубины, но такие видеозаписи должны содержать дополнительную информацию, включенную самой библиотекой. Такой информацией, в частности, являются данные, получаемые с датчиков устройства во время записи.

2. Данные со смартфонов и референсные данные о глубине изображений должны быть синхронизированы для правильного сопоставления при сравнении.

3. Сцены, участвующие в наборе, должны соответствовать различным сценариям съемки объектов: обход небольшого объекта, комнаты, коридора.

4. Необходимо, чтобы набор данных включал в себя записи для калибровки внутренних и внешних параметров камер.

4.2. Тестовый набор данных для сравнения

Тестовый набор данных содержит четыре сцены (рис. 3) длительностью около одной минуты (1800 кадров смартфонов и 300 кадров Azure Kinect) и калибровочные записи.

Далее приводится информация о подготовке набора данных.

4.2.1. Стенд для записи тестового набора данных

Для записи тестовых данных, следуя сформулированным требованиям, был разработан стенд (рис. 4), состоящий из двух смартфонов Galaxy S20 и датчика глубины Azure Kinect. Данный датчик выбран для

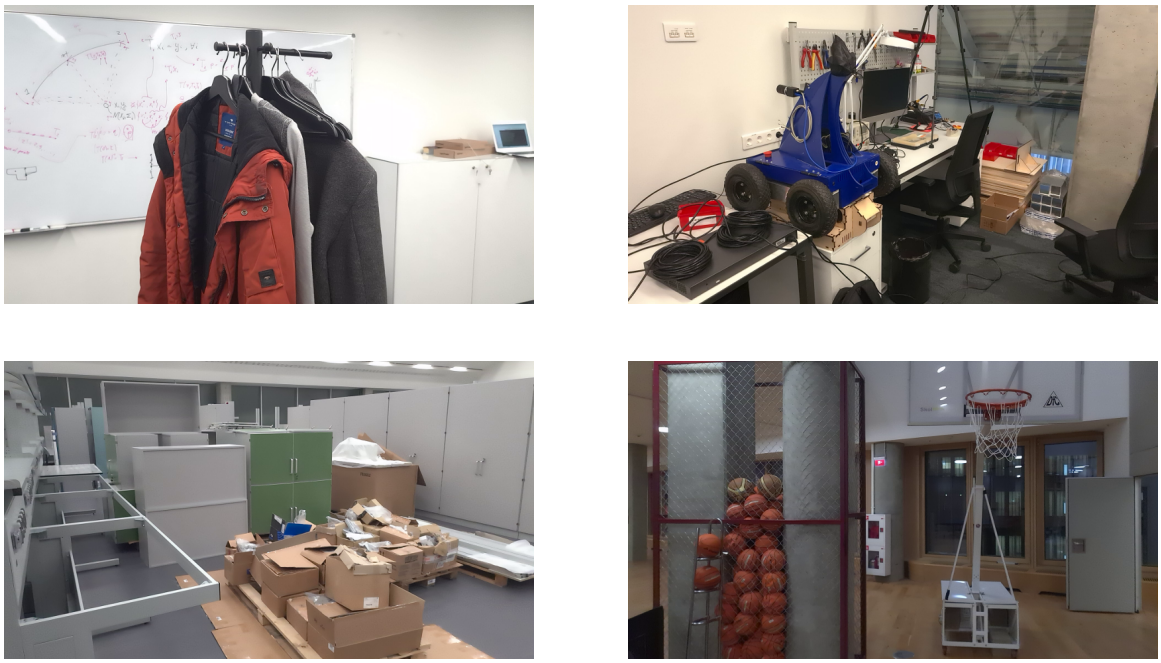


Рис. 3: Сцены из набора данных.

записи референсных значений глубины изображений ввиду его высокой точности (стандартное отклонение ошибки согласно документации — менее 17 мм[2]). Для управления записью и синхронизации устройств использовался мини-ПК NUC.

Azure Kinect использовался с параметрами для записи треков данных глубины и инфракрасных изображений. В параметрах API камер смартфонов были отключены автофокус и оптическая стабилизация, поскольку необходимо, чтобы камеры были жестко зафиксированы относительно друг друга. Смартфоны были закреплены на рейке с расстоянием между камерами приблизительно равным 10 см.

Смартфоны Galaxy S20, используемые в стенде, обладают следующими характеристиками:

- Android 10;
- стандартная камера 12 MP, f/1.8, 26mm;
- 8GB RAM.

[2] https://docs.microsoft.com/en-us/azure/kinect-dk/hardware-specification

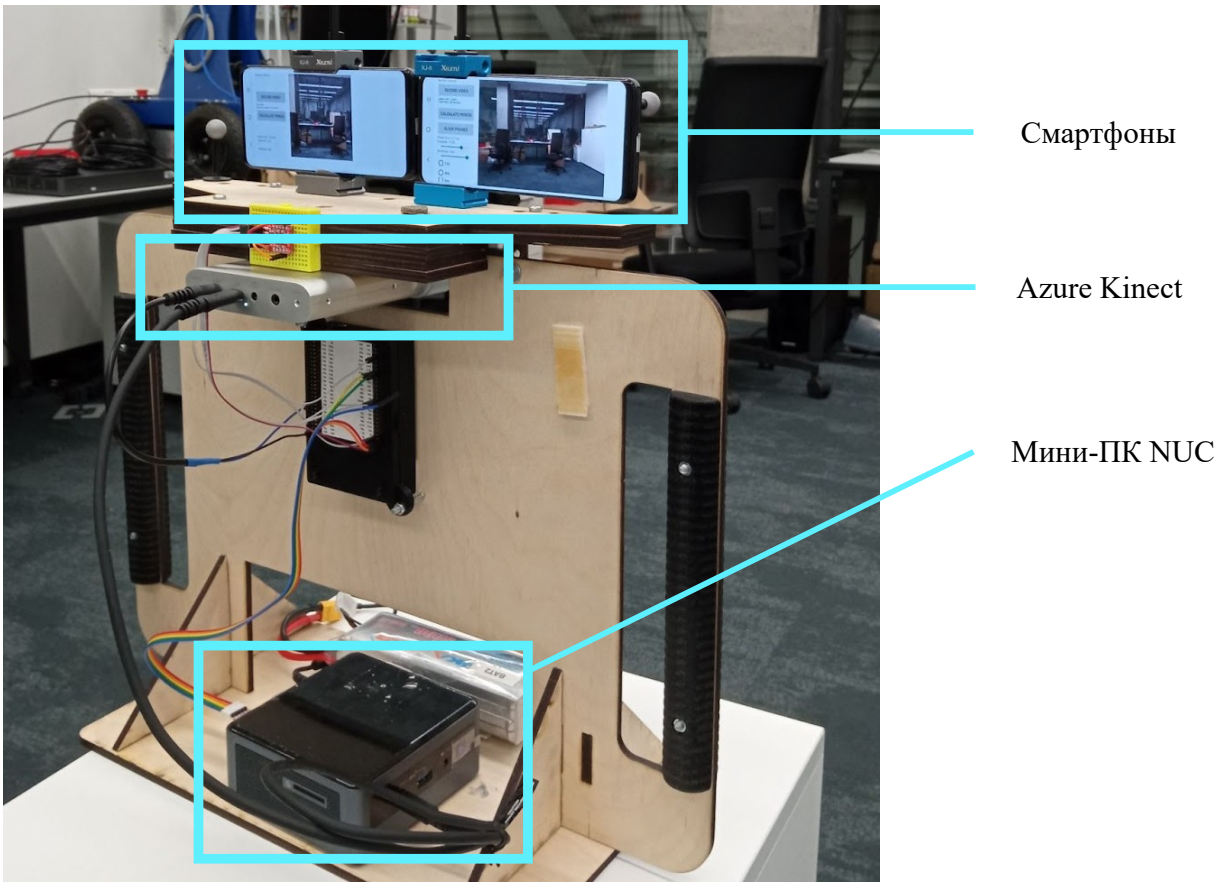


Рис. 4: Стенд для записи данных.

4.2.2. Синхронизация стенда

Поскольку стенд содержит три устройства с камерами, каждое из которых осуществляет съемку в различные моменты времени, необходимо синхронизировать съемку этих устройств. В синхронизации съемки в паре «смартфон — смартфон» применяется подход из RecSync. Для датчика Azure Kinect синхронизация съемки со смартфонами осуществляется более простым способом, так как датчик поддерживает начало съемки по сигналу, благодаря чему возможно обеспечить съемку изображений в точно определенное время. Azure Kinect получает сигнал на начало съемки в момент времени, соответствующий, с учетом периодичности, временным отметкам кадров смартфонов, что и приводит к синхронизированной съемке. Схема синхронизации в стенде проиллюстрирована на диаграмме (Рис. 5).

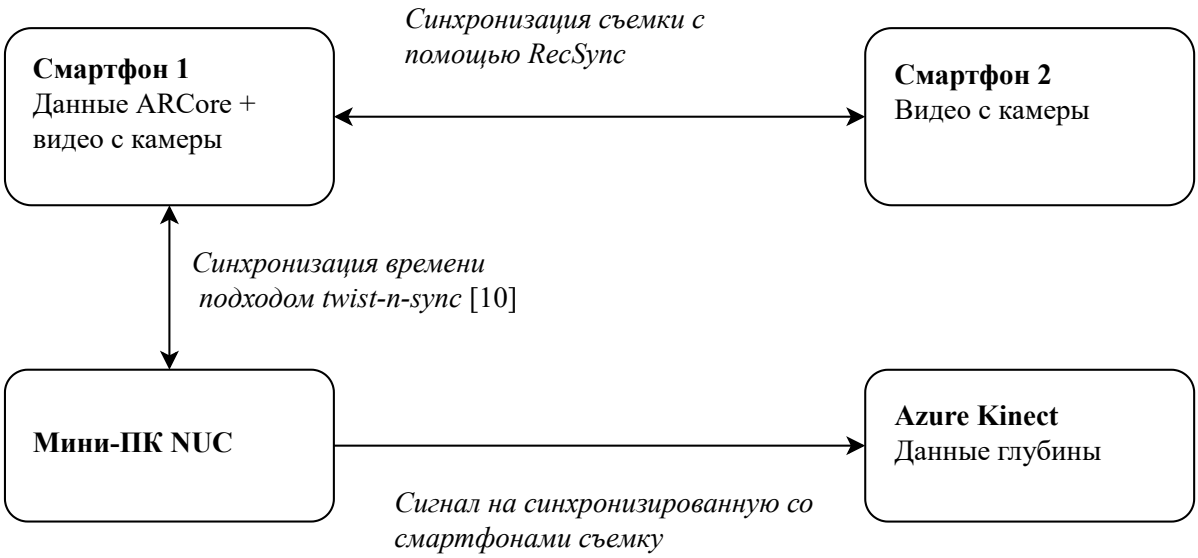


Рис. 5: Синхронизация стенда для записи данных.

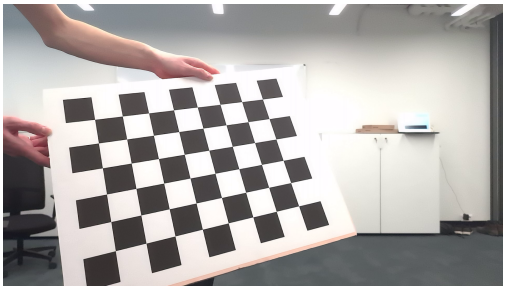


Рис. 6: Пример калибровочного паттерна.

4.2.3. Калибровка оптической системы стенда

Реализованный стенд представляет собой оптическую систему из трех камер (камеры двух смартфонов и камера глубины). Его калибровка предполагает вычисление **внутренних параметров** (матрица камеры, коэффициенты искажений) и **внешних параметров** каждой камеры (взаимного расположение камер в пространстве). Для получения данных значений для разработанного стенда использовались следующие шаги (рис. 5).

- Вычисление внутренних параметров камер смартфонов методом, предложенным Ч. Чжан [31], с использованием калибровочного паттерна (рис. 6). Для камеры Azure Kinect внутренние параметры камеры доступны в документации к устройству.

- Вычисление взаимного положения между камерами смартфонов, а также между камерой одного из смартфонов и датчиком Azure Kinect с использованием поз камер, полученных одним из методов решения проблемы Perspective-n-Point [28].

Для вычисления внутренних и внешних параметров оптической системы стенда были реализованы скрипты на языке Python, использующие реализации перечисленных методов из библиотеки (доступно в Github-репозитории[3]). Наличие перечисленных данных позволяет (1) применять методы анализа стереоизображений к парам изображений с камер смартфонов, (2) перепроецировать карту глубины, полученную с помощью Azure Kinect, на изображения смартфонов.

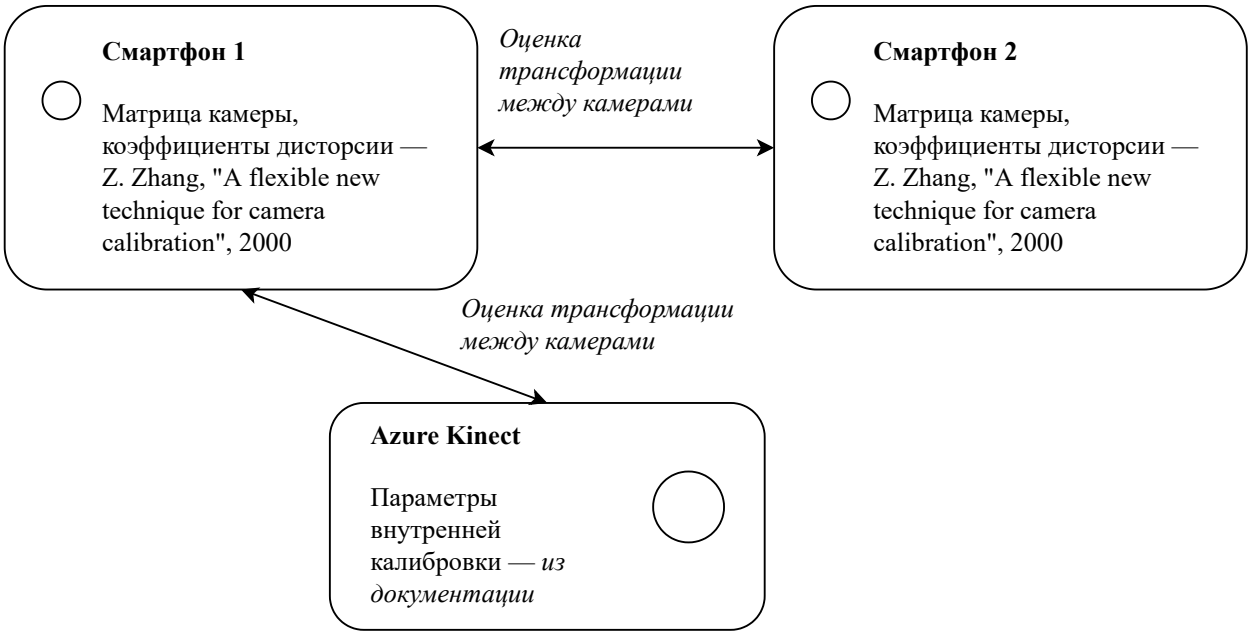


Рис. 7: Калибровка тестового набора данных.

4.3. Выбор метрик

Метрики карты глубины сравнивают близость оцененных значений глубины изображения с референсными значениями. Для этого используется статистика, основанная на попиксельном расстоянии между дву-

[3] https://github.com/azaat/sm-stereo-toolkit.

мя изображениями. Данный подход применим, когда методы возвращают глубину в абсолютных значениях. По той причине, что некоторые методы возвращают карту глубины с относительными значениями (например, методы получения глубины по одному изображению с использованием глубокого обучения), существуют метрики, не зависящие от масштаба значений дистанций — например, scale invariant. Также для таких методов возможно оптимизировать масштаб, минимизируя ошибку наименьших квадратов, а затем применять зависящие от масштаба значений метрики [24].

Для сравнения методов на тестовом наборе данных в данной работе выбраны метрики squared relative, absolute relative и scale invariant, являющиеся базовыми в работах по анализу методов получения глубины [9, 5], а также в одном из стандартных бенчмарков по оценке глубины KITTI [26].

Выбранные метрики имеют следующий вид (d_p — референсные значения глубины, $\hat{d}_p$ — предсказанные значения, $p \in 1..n$ — индексы пикселей изображения, для которых есть данные о глубине).

- Absolute relative

$$\frac{1}{n}\sum_{p}\frac{|d_p - \hat{d}_p|}{d_p}$$

- Scale invariant

$$\frac{1}{n}\sum_{p}\delta_p^2 - \frac{1}{n^2}(\sum_{p}\delta_p)^2$$
$$\delta_p = logd_p - log\hat{d}_p$$

- Squared relative

$$\frac{1}{n}\sum_{p}\frac{\delta_p^2}{d_p}$$

$$\delta_p = d_p - \hat{d}_p$$

4.4. Количественное сравнение методов

Для сравнения методов из тестового набора данных выбираются подпоследовательности кадров, не содержащие больших областей без текстуры (например, не учитываются начало и конец записи, если в них снята исключительно стена). Также не учитываются кадры, для которых библиотека ARCore не вычислила данные глубины изображения.

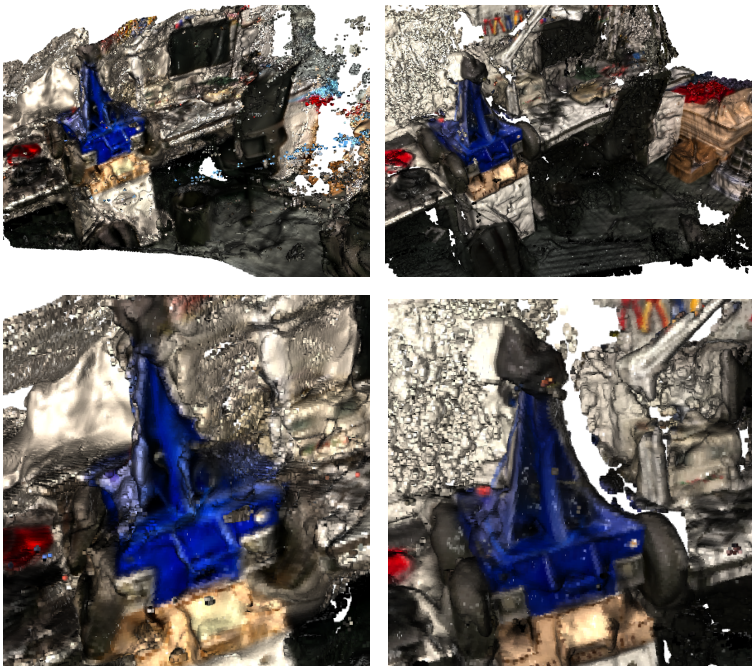
На графиках (Рис. 8) проиллюстрированы значения выбранных метрик по кадрам для каждого из методов на четырех сценах из набора данных. Сравнение показало, что применение анализа стереоизображений на базе двух смартфонов показывает меньшую ошибку, чем другие методы, доступные на смартфонах. В частности, ARCore и PyDNet, выделенные оранжевым и красным цветом на графиках, уступают результатам SGBM, выделенным синим, почти по всей длительности записей из набора данных.

4.5. Качественное сравнение методов

Качественное сравнение показало, что визуальные ошибки в пропорциях предметов также менее выражены при использовании предложенного подхода.

Приведем некоторые примеры таких ошибок. При использовании ARCore в одной из сцен набора данных на изображениях (рис. 9) видна большая по сравнению с SGBM деформация стены над столом. В другой сцене ARCore также показывает ошибки в результате реконструкции, которые выделены на изображениях (рис. 10) и отстутствуют в случае применения SGBM.

Рис. 8: Результаты количественного сравнения.



(a) ARCore (b) SGBM

Рис. 9: Результаты качественного сравнения.

(a) ARCore (b) SGBM

Рис. 10: Результаты качественного сравнения.

5. Реализация подхода к оценке глубины по стереоизображениям с двух смартфонов

Предложенный подход к получению глубины из стереопары, состоящей из двух смартфонов, был реализован в виде приложения на платформе Android. Приложение работает следующим образом: (1) пользователь устанавливает соединение между смартфонами, активируя точку доступа Wi-Fi на одном смартфоне (смартфон-лидер) и подключая к ней второй (смартфон-клиент), (2) происходит синхронизация съемки на двух устройствах, (3) изображения с камеры смартфона-клиента пересылаются на смартфон-лидер, сопоставляются в пары с изображениями смартфона-лидера и обрабатываются на нем, (4) после обработки глубина изображения визуализируется на экране смартфона в режиме мягкого реального времени. Ввиду точности синхронизации и возможности видеозаписи (Таб. 1), в качестве способа получения стереопары с двух смартфонов в приложении был выбран метод из [23].

Стоит отметить, что реализованное приложение направлено на демонстрацию предложенного подхода к оценке глубины изображения. В приложении возможно увеличение скорости передачи изображений путем применения алгоритмов трансляции видеопотока, однако это выходит за рамки данной работы.

5.1. Архитектура приложения

Приложение реализовано на языках Java, Kotlin. Язык Java применялся при модификации модулей из RecSync, Kotlin — при реализации новых модулей. Приложение содержит следующие основные компоненты, представленные на диаграмме (рис. 11).

- Компоненты из RecSync ответственны за установление соединения между устройствами, определение статуса «смартфона-лидера» и «смартфона клиента», а также синхронизацию времени и съемки.

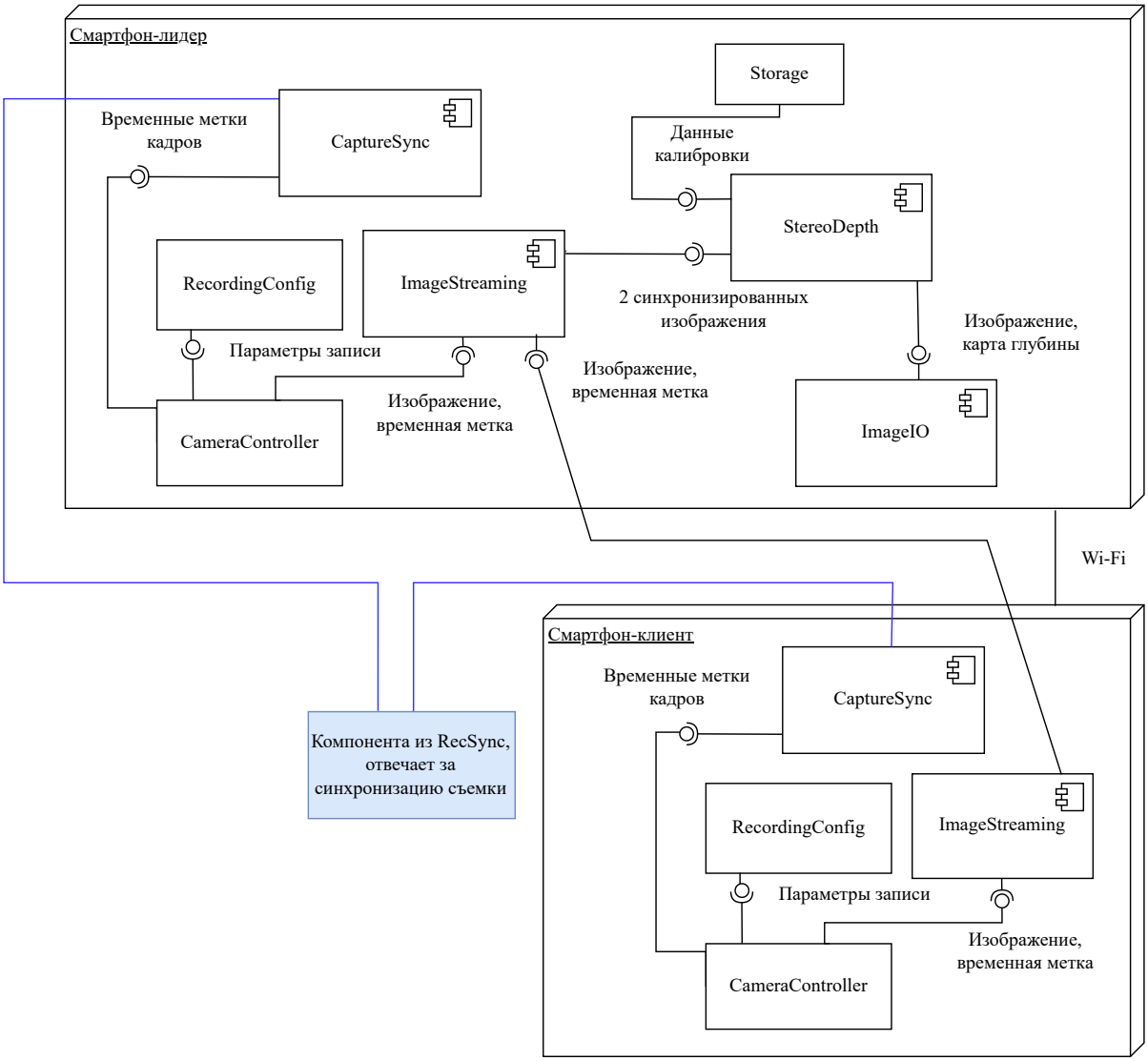


Рис. 11: Диаграмма развертывания.

- Компонента «ImageStreaming» реализует передачу изображений с одного смартфона на другой и их сопоставление в пары.
- Компонента «StereoDepth», принимая пары кадров, использует метод SGBM для получения глубины изображения. В текущей реализации калибровочные данные считываются из памяти устройства.

5.1.1. Модуль передачи изображений

Передача и сопоставление изображений для их обработки устроены следующим образом. После выполнения синхронизации съемки из

потока кадров на каждом устройстве с периодом в одну секунду выбираются кадры, временные метки которых соответствуют константному значению фазы, одинаковому для двух устройств. На смартфоне-лидере несколько последних выделенных кадров сохраняются в буфферы в оперативной памяти, а на смартфоне-клиенте выделенные кадры отправляются смартфону-лидеру с использованием TCP-сокетов. Затем смартфон-лидер, принимая отправленный кадр, сопоставляет его с кадрами из буфферов — найдя соответствующее по временным меткам изображение, мы получаем готовую к обработке стереопару.

5.1.2. Модуль получения глубины изображений

Для получения глубины изображений применяется метод SGBM из библиотеки OpenCV, портированный на платформу Android. Разработка данного модуля является частью семестровой работы студента второго курса ПИ Тимофея Пушкина.

5.2. Демонстрация работы

Сформулирован следующий сценарий демонстрации работы приложения: два смартфона, закрепленные на рейке, расположены статично и снимают поверхность, на которую ставятся различные предметы (рис. 12).

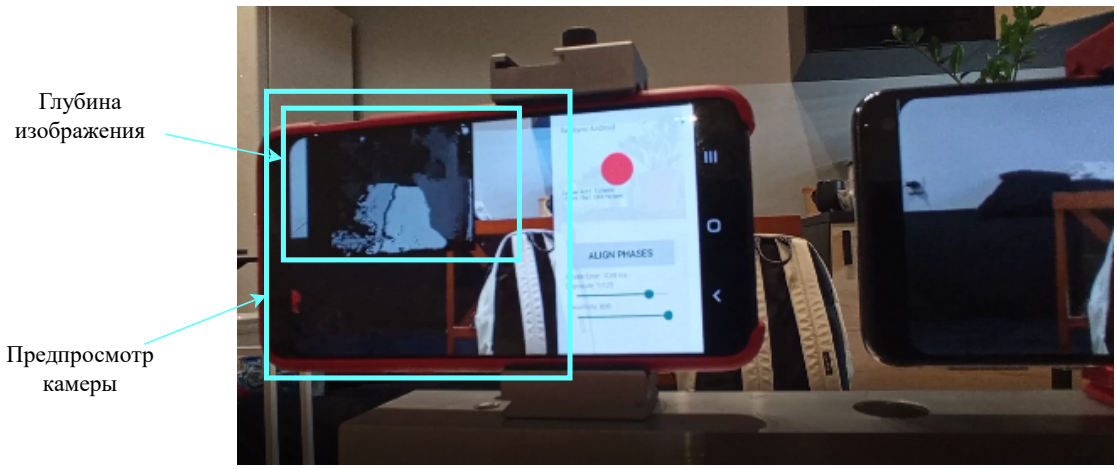


Рис. 12: Демонстрация работы приложения.

Видеозапись запуска приложения в соответствии с данным сценарием доступна по ссылке[4].

[4]Видеозапись запуска приложения https://drive.google.com/file/d/1tMm0lSjr0XL0WdBQa6VoOR12vxEg3MTL/view?usp=sharing.

6. Результаты

В результате работы над выпускной квалификационной работой были выполнены следующие задачи.

- Проведен обзор методов получения глубины на мобильных устройствах, рассмотрены их особенности и ограничения. Среди методов получения глубины изображения на смартфонах с использованием одной камеры были выбраны: ARCore, работающий по принципу Structure from Motion, а также mobilePydnet, использующий глубокое обучение. Выполнен обзор методов синхронизации кадров с камер двух смартфонов на платформе Android, выбран метод RecSync, поддерживающий съемку видео и предоставляющий синхронизацию с точностью менее 1 мс.

- Создан тестовый набор данных из синхронизированных видеозаписей с камер двух смартфонов и изображений с датчика глубины Azure Kinect, выполнена калибровка набора. В наборе содержатся записи 4-х сцен различного характера длительностью около 1 мин.

- Проведено количественное сравнение методов получения карты глубины на смартфонах и предлагаемого подхода на тестовом наборе данных, в результате сравнения показано улучшение значений метрик при использовании стереоизображений с камер двух смартфонов по сравнению с использованием ARCore и mobilePydnet. Проведено качественное сравнение методов, показавшее визуальные улучшения в пропорциях объектов при 3D реконструкции с использованием глубины изображений полученных путем анализа стереоизображений в сравнении с методом ARCore.

- Реализован подход к получению карты глубины из стереоизображений с двух смартфонов в виде приложения на платформе Android (доступно в Github-репозитории[5]).

[5] https://github.com/azaat/RecSync-android/tree/depth

7. Благодарности

Автор выражает благодарность помогавшим в данной работе коллегам:

- Корниловой Анастасии Валерьевне — за консультации на всех этапах работы, помощь в редактировании текста и презентации;
- Ярошу Дмитрию Сергеевичу — за комментарии по презентации работы и созданию демо-приложения на Android;
- Файзуллину Марселю Фарисовичу — за помощь в реализации тестового стенда для сравнения;
- Пушкину Тимофею Дмитриевичу — за портирование модуля получения глубины изображений на платформу Android.

Список литературы

[1] Andreas Geiger. LIBELAS: Library for Efficient Large-scale Stereo Matching. — 2021. — [Online; accessed 16-December-2021]. URL: http://www.cvlibs.net/software/libelas/.

[2] Chang Jia-Ren, Chen Yong-Sheng. Pyramid Stereo Matching Network. — 2018. — 06. — P. 5410–5418.

[3] Choi Sungjoon, Zhou Qian-Yi, Koltun Vladlen. Robust reconstruction of indoor scenes // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2015. — P. 5556–5565.

[4] Continuous 3D Label Stereo Matching Using Local Expansion Moves / Tatsunori Taniai, Yasuyuki Matsushita, Yoichi Sato, Takeshi Naemura // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2018. — 11. — Vol. 40. — P. 2725–2739.

[5] Deep Two-View Structure-from-Motion Revisited / Jianyuan Wang, Yiran Zhong, Yuchao Dai et al. // CVPR. — 2021.

[6] Duggal Shivam, Wang Shenlong, Ma Wei-Chiu et al. DeepPruner: Learning Efficient Stereo Matching via Differentiable PatchMatch. — 2019. — 09.

[7] Zhang Yinda, Wadhwa Neal, Orts-Escolano Sergio et al. Du^{2}Net: Learning Depth Estimation from Dual-Cameras and Dual-Pixels. — 2020. — 2003.14299.

[8] Pan Liyuan, Chowdhury Shah, Hartley Richard et al. Dual Pixel Exploration: Simultaneous Depth Estimation and Image Restoration. — 2020. — 2012.00301.

[9] Eigen David, Puhrsch Christian, Fergus Rob. Depth Map Prediction from a Single Image Using a Multi-Scale Deep Network // Proceedings of the 27th International Conference on Neural Information Processing

Systems - Volume 2. — NIPS'14. — Cambridge, MA, USA : MIT Press, 2014. — P. 2366–2374.

[10] Fast and Accurate Single-Image Depth Estimation on Mobile Devices, Mobile AI 2021 Challenge: Report / Andrey Ignatov, Grigory Malivenko, David Plowman et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. — 2021. — June. — P. 2545–2557.

[11] Geiger Andreas, Roser Martin, Urtasun Raquel. Efficient Large-Scale Stereo Matching. — 2010. — 12. — P. 25–38.

[12] Gidaris Spyros, Komodakis Nikos. Detect, Replace, Refine: Deep Structured Prediction For Pixel Wise Labeling. — URL: https://arxiv.org/abs/1612.04770.

[13] GitHub. PyDNet on mobile devices v2.0. — 2021. — [Online; accessed 16-December-2021]. URL: https://github.com/FilippoAleotti/mobilePydnet.

[14] Google. ARCore | Google Developers. — 2021. — [Online; accessed 16-December-2021]. URL: https://developers.google.com/ar.

[15] Google. Recording and playback introduction. — 2022. — [Online; accessed 26-April-2022]. URL: https://developers.google.com/ar/develop/recording-and-playback.

[16] Hirschmüller Heiko, Buder Maximilian, Ernst Ines. Memory Efficient Semi-Global Matching. — Vol. I-3. — 2012. — 08.

[17] A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation / Nikolaus Mayer, Eddy Ilg, Philip Hausser et al. — 2016. — 06. — P. 4040–4048.

[18] OpenCV. cv::stereo::StereoBinarySGBM Class Reference. — 2021. — [Online; accessed 16-December-2021]. URL: https://docs.opencv.org/3.4/d1/d9f/classcv_1_1stereo_1_1StereoBinarySGBM.html.

[19] Polycam. Polycam - LiDAR 3D Scanner. — 2022. — [Online; accessed 26-April-2022]. URL: https://poly.cam/.

[20] Real-time single image depth perception in the wild with handheld devices / Filippo Aleotti, Giulio Zaccaroni, Luca Bartolomei et al. // Sensors. — 2021. — Vol. 21.

[21] Scharstein Daniel, Szeliski Richard, Zabih Ramin. A taxonomy and evaluation of dense two-frame stereo correspondence algorithm. — Vol. 47. — 2001. — 02. — P. 131–140.

[22] Socialsync: Sub-frame synchronization in a smartphone camera network / Richard Latimer, Jason Holloway, Ashok Veeraraghavan, Ashutosh Sabharwal // European Conference on Computer Vision / Springer. — 2014. — P. 561–575.

[23] Akhmetyanov Azat, Kornilova Anastasiia, Faizullin Marsel et al. Sub-millisecond Video Synchronization of Multiple Android Smartphones. — 2021. — 2107.00987.

[24] Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer / Rene Ranftl, Katrin Lasinger, David Hafner, Vladlen Koltun // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2020. — 08. — Vol. PP. — P. 1–1.

[25] Tridimesional. You can use 2 smartphones as a powerful 3D camera with Camarada. — 2022. — [Online; accessed 27-April-2022]. URL: https://www.tridimensional.info/2018/01/you-can-use-2-smartphones-as-a-powerful-3d-camera-with-camarada

[26] Vision meets Robotics: The KITTI Dataset / Andreas Geiger, Philip Lenz, Christoph Stiller, Raquel Urtasun // International Journal of Robotics Research (IJRR). — 2013.

[27] Weber Michael, Humenberger Martin, Kubinger Wilfried. A very fast census-based stereo matching implementation on a graphics processing unit. — 2009. — 11. — P. 786 – 793.

[28] Wikipedia. Perspective-n-Point. — 2022. — [Online; accessed 26-April-2022]. URL: https://en.wikipedia.org/wiki/Perspective-n-Point.

[29] Wikipedia contributors. Structure from motion — Wikipedia, The Free Encyclopedia. — 2021. — [Online; accessed 16-December-2021]. URL: https://en.wikipedia.org/w/index.php?title=Structure_from_motion&oldid=1057307916.

[30] Wireless software synchronization of multiple distributed cameras / Sameer Ansari, Neal Wadhwa, Rahul Garg, Jiawen Chen // 2019 IEEE International Conference on Computational Photography (ICCP) / IEEE. — 2019. — P. 1–9.

[31] Zhang Z. A flexible new technique for camera calibration // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2000. — Vol. 22, no. 11. — P. 1330–1334.

[32] scandy. Scandy Pro - A full-color 3D scanner on your iPhone X. — 2022. — [Online; accessed 26-April-2022]. URL: https://www.scandy.co/apps/scandy-pro.

[33] Žbontar Jure, Lecun Yann. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. — 2015. — 10. — Vol. 17.