

Санкт-Петербургский политехнический университет Петра Великого
Институт машиностроения, материалов и транспорта
Высшая школа автоматизации и робототехники

ОТЧЁТ О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

**Исследование алгоритмов SLAM для создания карты помещения
с использованием наземного мобильного робота**

Выполнил
студент гр.3331506/60401

Е.А. Цмугунова

Руководитель
доцент, к.т.н.

В.Н. Уланов

Научный консультант
доцент, к.т.н.

А.В. Бахшиев

«___» _____ 201__ г.

Санкт-Петербург

2019

Санкт-Петербургский политехнический университет Петра Великого

Институт машиностроения, материалов и транспорта
Высшая школа автоматизации и робототехники

ЗАДАНИЕ

на выполнение научно-исследовательской работы

Цмугуновой Елизавете Андреевне

студенту гр. 3331506/60401

1 Тема работы

«Исследование алгоритмов SLAM для создания карты помещения
с использованием наземного мобильного робота»

2 Срок сдачи студентом законченной работы

Декабрь 2019

3 Исходные данные к работе

4 Содержание расчетно-пояснительной записки

Обзор и сравнение типов датчиков, применяемых для измерения
расстояния до объектов в решении задачи SLAM; обзор и сравнение
существующих алгоритмов RGBD-SLAM, применимых для работы наземных
мобильных роботов внутри помещений.

5 Перечень графического материала (с точным указанием обязательных
чертежей)

нет

6 Консультанты по работе

Бахшиев Александр Валерьевич, доцент, к.т.н.

7 Дата выдачи задания

9 сентября 2019

Руководитель

Уланов Владимир Николаевич, доцент, к.т.н.

(ФИО, должность, подпись руководителя)

Задание принял к исполнению

(дата)

(подпись студента)

РЕФЕРАТ

53 с., 15 табл., 60 источников.

SLAM, ЛОКАЛИЗАЦИЯ, ПОСТРОЕНИЕ КАРТЫ, МОБИЛЬНЫЙ НАЗЕМНЫЙ РОБОТ, RGBD-КАМЕРА, КАМЕРА ГЛУБИНЫ, RGBD-SLAM

Целью данной работы является исследование алгоритмов метода SLAM для создания карты помещения с использованием выбранного датчика для наземного мобильного робота.

Проведен обзор и сравнение типов датчиков, применяемых для измерения расстояния до объектов в решении задачи SLAM. Проведен обзор и анализ существующих алгоритмов RGBD-SLAM, применимых для работы наземных мобильных роботов внутри помещений. Представлены результаты сравнения датчиков и алгоритмов по подобранным критериям.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	6
1 Аналитический обзор датчиков, используемых для получения значений расстояний в алгоритмах SLAM	8
1.1 Ультразвуковые дальномеры	8
1.2 Инфракрасные дальномеры	9
1.3 Лидары	10
1.4 Радары	10
1.5 Стереокамеры	11
1.6 RGB-D-камеры	12
1.7 Выводы по разделу	15
2 Аналитический обзор алгоритмов RGBD-SLAM для наземных мобильных роботов	19
2.1 RGBDSLAMv2	20
2.2 ORB-SLAM2	23
2.3 RTAB-map	28
2.4 RGBiD-SLAM	29
2.5 OpenVSLAM	31

2.6 SlamDunk	32
2.7 DVO SLAM	34
2.8 BAD SLAM	35
2.9 CVO SLAM	37
2.10 ElasticFusion	39
2.11 BundleFusion	41
2.12 Использование RGBD-камер для лазерных методов SLAM	42
2.13 Выводы по разделу	43
Направление дальнейших исследований	46
ЗАКЛЮЧЕНИЕ	47
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	48

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ИК – инфракрасный

УЗ - ультразвуковой

ATE – absolute trajectory error

BA – Bundle Adjustment

EMM – Environment Measurement Model

IMU – Inertial measurement unit

RGBD – Red, Green, Blue, Depth

RMSE – root-mean-square error

ROS – Robot Operating System

SLAM – Simultaneous Localization and Mapping

ToF – Time-of-Flight

ВВЕДЕНИЕ

Важной задачей в разработке системы управления автономного мобильного робота является обеспечение способности ориентироваться в

исследуемой среде. Одним из способов решения данной проблемы является разработка алгоритмов метода одновременной локализации и построения карты (SLAM).

Метод SLAM (Simultaneous localization and mapping) заключается в построении карты окружения мобильным роботом и использовании данной карты для определения своего расположения в исследуемом пространстве [1]. При этом предполагается, что для оценки своей траектории и положения ориентиров робот либо не имеет априорных знаний о среде и строит новую карту, либо в процессе работы обновляет уже строящуюся карту.

Первые методы для решения задачи SLAM основывались на фильтрах Калмана, фильтре частиц, информационных фильтрах [2]. Со временем появились и получили широкое распространение методы, основанные на оптимизации графов. Также появилась возможность разделять процесс на несколько потоков [3].

SLAM используется для навигации управляемых и автономных роботов в помещении, на улице, в воздухе и под водой. Применяемый алгоритм при этом зависит от конкретных условий среды и используемых датчиков. Основными составляющими алгоритмов SLAM являются:

- определение текущего положения робота по данным датчиков;
- поиск ориентиров в окружающем пространстве;
- сопоставление новых ориентиров с уже исследованными;
- обновление карты окружения.

Построенная в результате решения задачи SLAM карта имеет важное значение для последующего планирования движения и автономной навигации мобильного робота, поскольку по точным данным об окружающей среде, включающим положение ориентиров и препятствий, робот имеет возможность оценить свое положение и спланировать траекторию движения без столкновений [4].

Целью данной работы является исследование алгоритмов метода SLAM для создания карты помещения с использованием выбранного сенсора для наземного мобильного робота.

Для выполнения работы были поставлены следующие задачи:

- 1) Сделать обзор датчиков для измерения расстояний, используемых для решения задачи метода SLAM.
- 2) Сравнить рассмотренные датчики и выбрать наиболее подходящий для создания карты помещения наземным мобильным роботом.
- 3) Сделать обзор существующих алгоритмов SLAM для выбранного датчика.
- 4) Выполнить сравнение алгоритмов и выбрать наиболее подходящие для решения задачи построения карты помещения с использованием наземного мобильного робота.

Для выбора подходящего решения принимаются следующие условия:

- считается, что среда статична;
- окружающие объекты имеют произвольную форму и размер более 2 см;
- минимальный размер проемов составляет 1 м;
- скорость движения робота составляет примерно 0,2 м/с;
- выполнение задачи роботом проводится в автономном режиме, поэтому необходима работа алгоритма в режиме реального времени.

1 Аналитический обзор датчиков, используемых для получения значений расстояний в алгоритмах SLAM

Для получения пространственной информации о среде, в которой находится мобильный робот, и следовательно локализации и картографирования необходимо использовать датчики, по измерениям которых можно напрямую или косвенно получить значения расстояний до наблюдаемых объектов. Каждый тип датчиков имеет свои недостатки и преимущества, способные повлиять на результаты работы алгоритма. Для решения задачи SLAM обычно используются следующие датчики измерения дальности:

- ультразвуковые дальномеры;
- инфракрасные дальномеры;
- лидары;
- радары;
- стереокамеры;
- RGB-D-камеры.

Выбор датчика зависит от условий среды, в которой находится робот, а также от требований, предъявляемых к точности и форме получаемых измерений. Кроме того, используемые датчики влияют на выбор подходящего алгоритма.

В данном разделе приведено описание и сравнение наиболее распространенных датчиков, используемых для измерения расстояний до объектов при решении задачи SLAM.

1.1 Ультразвуковые дальномеры

Для измерения расстояния до объекта ультразвуковой дальномер испускает ультразвуковые волны (частота 20-40 кГц), которые отражаются от объекта и возвращаются в приемник, и использует время с момента импульса до приема эха для расчета [5].

Главным недостатком ультразвукового дальномера является сложность измерения расстояния до конкретной точки. Поскольку звуковые волны распространяются конусом, целесообразно использовать ультразвуковой дальномер для измерения расстояния до сплошных стен, но при работе с небольшими объектами погрешность измерений увеличивается, что делает практически невозможным определение границ предметов. Вследствие принципа работы точность измерений относительна и составляет 0,5 – 1 % от измеренного расстояния, что делает измерения недостоверными при работе с расстояниями больше диапазона 0,01 – 10 м.

Основным достоинством является совместимость с большинством типов поверхностей (металлическими и неметаллическими, прозрачными и

непрозрачными), если они обладают достаточной отражательной способностью [6]. Ультразвуковые дальномеры компактны (размер исчисляется несколькими сантиметрами), энергопотребление составляет порядка 1 Вт, имеют невысокую стоимость (порядка 10-100 \$), поэтому при поломке для них легко найти замену [7].

1.2 Инфракрасные дальномеры

Инфракрасные дальномеры являются распространенными оптическими датчиками, использующими метод триангуляции в измерении расстояний. Излучатель испускает световой луч, который отражается от объекта и возвращается в приемник, попадая в точку светочувствительной матрицы. По положению данной точки вычисляется угол отражения и расстояние до наблюдаемого объекта [8].

В основном ИК-датчики имеют низкую стоимость (порядка 10-100 \$), небольшие размеры. Энергопотребление составляет порядка 1 Вт. Диапазон измерений инфракрасных датчиков составляет 0,01 – 10 м с точностью 0,01 м. На их работу при измерении расстояний влияет отражательная способность поверхностей в исследуемой среде. Поскольку такие датчики обладают низкой точностью, они чаще применяются в качестве датчиков приближения. Для повышения точности измерений (дискретность измерений примерно 5 см для дистанции до 10 м) в более дорогих ИК-датчиках дополнительно используется измерение сдвига фазы [9].

1.3 Лидары

Распространенным в решении задач SLAM является лидар (LIDAR – Light Detection and Ranging) – технология, которая использует явления отражения и рассеивания света в прозрачных и полупрозрачных средах для получения и обработки информации об удаленном объекте [10]. По количеству испускаемых лучей бывают 2D и 3D лидары. Также по технологии

производства лидары подразделяются на механические, гибридные твердотельные и твердотельные [3].

Лидары имеют высокий уровень детализации (точность измерений порядка 0,01 м и дискретность угла измерений 0,2 градуса), диапазон измерений в пределах 1-200 м, вертикальный угол обзора может составлять 10-40 градусов, горизонтальный 120-360 градусов [11]. Энергопотребление лидаров составляет 10-20 Вт. Измерения, полученные с помощью лидаров, имеют достаточно высокую точность как на улице, так и в помещении [10], лидары устойчивы к изменению освещения, но подвержены влиянию погодных условий [11]. По сравнению с радаром лидар имеет более сфокусированный лазерный луч, большее число сканируемых слоев в вертикальном направлении и большую плотность измерений [12]. Одновременно может проводиться от 32 до 128 измерений. Основным недостатком таких систем является высокая стоимость (порядка 1000-10000\$), особенно для систем высокого разрешения [10]. По сравнению с другими датчиками, лидары в основном достаточно крупные и тяжелые, из-за наличия движущихся частей требуют более сложного обслуживания [7].

1.4 Радары

Радар – это устройство, которое по отраженным радиоволнам (частота 24/77/79 ГГц) обнаруживает и определяет местоположение объектов. Основу устройства составляют передатчик, антенна и приемник. Для измерения скорости объектов используется время пролета волны и эффект Доплера. С помощью радара можно обнаружить сразу несколько объектов, измерить расстояние до них, направление движения и скорость [5].

Радары из-за длины радиоволн не позволяют обнаруживать маленькие объекты, а также точную форму поверхности и границы объектов. Энергопотребление радаров составляет 4 – 12 Вт. Диапазон измерения радаров составляет 0,5 – 250 м с точностью 0,05 м, горизонтальный угол обзора варьируется в пределах 15 – 90 градусов, вертикальный – 4 – 20 градусов.

Стоимость радаров варьируется в пределах 500 – 5000 \$. Скорость обработки данных для радаров ниже, чем для лидаров и камер, но по сравнению с ними радар обладает меньшим разрешением [12]. Радары устойчивы к изменению погодных условий и освещения, могут быть скрыты защитными панелями, поскольку сигнал способен проникать сквозь них [11].

1.5 Стереокамеры

Стереокамеры имеют два объектива, расположенные на расстоянии друг от друга. Это расстояние называется стереобазой. Стереокамеры не выдают напрямую информацию о глубине, но по двум изображениям методом триангуляции можно вычислить расстояние до объектов, присутствующих на двух кадрах [13]. Чем больше стереобаза, тем лучше оценка глубины. Размер стереокамеры зависит от стереобазы и исчисляется несколькими сантиметрами (в основном до 15 см в длину) [7].

Поскольку принцип работы стереокамер не основан на активном излучении, они могут использоваться при солнечном свете и не подвержены влиянию со стороны аналогичных устройств [14]. Стереокамеры потребляют порядка 2 Вт. Разрешение таких камер обычно составляет 1280×720 пикселей. Стоимость стереокамер варьируется в пределах 100-200 \$.

Качество и плотность карты глубины, полученной при расчетах по двум изображениям, зависит от освещенности, наличия текстур и различных цветов в обозреваемой сцене, а также от расстояния до объектов. Погрешность измерений пропорциональна квадрату расстояния до объекта. При наличии однотонных стен в наблюдаемой среде появляются проблемы с определением расстояния до них, поскольку два изображения практически не отличаются. Повторяющаяся структура и симметрия исследуемой сцены так же может привести к ошибочным измерениям вследствие нахождения ложных соответствий [14]. Поскольку данный подход к измерению глубины зависит от освещенности, в случае недостаточного количества света в помещении также могут наблюдаться ошибки вычислений [6].

1.6 RGB-D-камеры

С развитием технологий появилась возможность измерять трехмерную геометрию наблюдаемой сцены. Одним из распространенных средств для выполнения данной задачи стали камеры глубины, или RGBD-камеры. Такие камеры помимо традиционного цветного изображения предоставляют данные о глубине каждого пикселя, то есть о расстоянии до точек наблюдаемых объектов. RGBD-камеры имеют относительно небольшой размер, приемлемую стоимость и благодаря серийному производству легко заменяемы [7].

Камеры, позволяющие получить данные о глубине изображения, работают по следующим принципам:

- использование времени пролета;
- использование структурированного света;
- активное стерео.

Далее рассмотрены более подробно данные принципы работы, а также их преимущества и недостатки.

1.6.1 ToF-камеры

ToF (Time-of-Flight) камеры для измерения глубины используют время пролета или сдвиг фазы излучаемого светового импульса от излучателя до объекта и обратно до приемника. Несмотря на то, что данный принцип работы позволяет производить измерения при тусклом свете и независимо от структуры и цветов объектов сцены, он накладывает ряд особенностей и ограничений.

Данный тип камер не предназначен для работы при солнечном свете, а также не может работать совместно с подобными устройствами, если они имеют такой же сдвиг фазы. Более того, ошибки измерений наблюдаются при взаимодействии с зеркальными и прозрачными поверхностями, так как в приемник может прийти несколько лучей вместо одного. Из-за темных неотражающих поверхностей наблюдаются провалы в измерении глубины. Также для времяпролетных камер характерно наличие «пикселей в воздухе» в

местах сильного перепада расстояний, то есть данные пиксели принимают промежуточные значения глубины между передним и фоновым объектом. Быстрое движение камеры в свою очередь приводит к размытию карты глубины [14].

ToF-камеры производят измерения в диапазоне 0,1 – 13 м, значение горизонтального угла обзора находится в пределах 40-74 градусов, а вертикального – 40-60 градусов, разрешение камеры может достигать 512×424 пикселей [15]. Точность измерений составляет 0,01-0,02 м [16]. Энергопотребление ToF-камер находится в пределах 4-10 Вт. По сравнению с камерами структурированного света ToF-камеры менее чувствительны к солнечному свету, однако обладают меньшим разрешением, выдают больше шумов при измерении, имеют более высокую стоимость (порядка 150-400\$) и требуют более сложной калибровки [17].

1.6.2 Камеры структурированного света

Принцип работы камер структурированного света заключается в проецировании известного узора (например, множества точек или полосок) на сцену инфракрасным излучателем и последующем вычислении глубины по искаженному поверхностью объектов узору, принимаемого ИК-камерой [18].

Добавление в сцену проецируемой структуры позволяет избежать проблем определения глубины в сценах с недостаточным освещением и с малым количеством характерных особенностей текстуры и цвета. Такие камеры способны проводить измерения в диапазоне 0,1 – 9 м, значение горизонтального угла обзора находится в пределах 57-72 градусов, а вертикального – 43-55 градусов, разрешение камеры обычно 640×480 пикселей [15]. Точность измерений составляет 0,01-0,04 м [16]. Энергопотребление составляет примерно 2 Вт. Стоимость камер структурированного света составляет 80-200 \$ в зависимости от конкретной модели.

Как и другие датчики, использующие инфракрасный диапазон, камеры структурированного света подвержены негативному влиянию на результаты

измерений при наличии солнечных лучей или аналогичных источников ИК-излучения (схожих устройств) в наблюдаемой сцене. В местах сильных перепадов глубины в сцене расчет расстояния может быть ненадежен, поскольку при этом проецируемый рисунок частично не попадает в область видимости ИК-камеры. Как следствие, на карте глубины образуются области с отсутствующими измерениями вокруг объекта. Также существует проблема определения глубины для тонких и маленьких объектов, размеры которых меньше расстояния между элементами проецируемого узора, так как данные объекты просто не определяются. Кроме того, излучаемый структурированный свет может поглощаться очень темными поверхностями, отражаться от зеркальных и преломляться прозрачными поверхностями [14].

1.6.3 Активное стерео

Активное стерео проецирует рисунок в наблюдаемую сцену и использует его для последующего стереосравнения. Как и в случае пассивного стерео, точность вычисления глубины зависит от расстояния до наблюдаемого объекта.

Камеры с технологией активного стерео производят измерения в диапазоне 0,2-10 м, значение горизонтального угла обзора находится в пределах 59-85 градусов, а вертикального – 40-58 градусов, разрешение камеры может достигать 1280×720 пикселей [15]. Энергопотребление составляет примерно 1-2,7 Вт. Стоимость таких камер в основном составляет 150-200 \$.

В частности, данный подход используется камерами Intel Realsense серии D400, которые совмещают активное и пассивное стерео. Основными компонентами данной камеры являются ИК-излучатель, две ИК-камеры и RGB-камера. Интенсивность ИК излучения настраивается в зависимости от освещения. Если есть влияние постороннего излучения (солнечного света или излучения от других устройств), то происходит традиционное стереосравнение между двумя изображениями без использования ИК рисунка [14].

Данный подход прежде всего подходит для помещений, но работает и при солнечном свете, используя не только проецируемую структуру, но и

естественную текстуру сцены на изображении [6]. Излучаемый узор позволяет корректно проводить измерения в случае недостаточного освещения или отсутствия явных элементов структуры и цвета.

1.7 Выводы по разделу

В данном разделе были рассмотрены следующие датчики измерения расстояния, применяемые при решении задачи SLAM: инфракрасный дальномер, ультразвуковой дальномер, лидар, радар, стереокамера и RGBD-камеры (камера с технологией активного стерео, камера структурированного света и ToF-камера). Сравнение рассмотренных датчиков измерения расстояния, представлено в таблицах 1.1-1.3. При сравнении учитывались характеристики датчиков, а также критерии, влияющие на их использование и обслуживание и на качество измерений.

При выборе датчика немаловажными критериями являются стоимость, сложность обслуживания, а также размер датчика и энергопотребление, которые могут повлиять на габариты мобильного робота. При хорошем соотношении цены-качества для решения задачи предпочтение отдается датчикам невысокой стоимости. Наиболее дорогими среди рассмотренных датчиков являются лидар и радар. Также лидар, радар и ToF-камера являются наиболее энергозатратными.

Таблица 1.1 – Сравнение датчиков измерения расстояния

	Стоимость	Сложность обслуживания	Размер
Лидар	1000 - 10 000 \$	Высокая	Средний
Стереопара	100 - 200 \$	Средняя	Малый
Активное стерео	150 - 200 \$	Низкая	Малый
Камера структур. света	80 - 200 \$	Низкая	Малый
ToF-камера	150 - 400 \$	Средняя	Малый
ИК дальномер	10 - 100 \$	Низкая	Малый
УЗ дальномер	10 - 100 \$	Низкая	Малый
Радар	500 - 5000 \$	Низкая	Малый

Таблица 1.2 – Сравнение датчиков измерения расстояния. Характеристики

	Диапазон, м	Горизонт. угол обзора	Вертик. угол обзора	Точность измерения, м	Количество измерений	Энерго- потребление, Вт
Лидар	1 - 200	120° - 360°	10° - 40°	0,01 (0.2°)	32 - 128	10 - 20
Стереопара	1 - 100	30° - 120°	5° - 45°	0,01	1280 × 720	2
Активное стерео	0,2 - 10	59° - 85°	40° - 58°	0,01 - 0,02	1280 × 720	1 - 2,7
Камера структур. света	0,1 - 9	57° - 72°	43° - 55°	0,01 - 0,04	640 × 480	2
ToF-камера	0,1 - 13	40° - 74°	40° - 60°	0,01 - 0,02	512 × 424	4 - 10
ИК дальномер	0,01 - 10	-	-	0,01	1	1
УЗ дальномер	0,01 - 10	-	-	0,5 - 1 %	1	1
Радар	0,5 - 250	15° - 90°	4° - 20°	0,05 (0.5°)	32 - 256	4 - 12

Для работы мобильного робота в помещении достаточно дальности измерения до 10 м, при этом необходима возможность измерения близких расстояний до 1 м. Верхняя граница измерений лидара, стереокамеры и радара является избыточной для данной задачи, более того они менее пригодны для измерения близких расстояний. Для ориентации в пространстве и построения карты целесообразнее получать в каждом положении робота измерения сразу до нескольких объектов в области видимости, а не проводить единичные измерения при каждой смене положения, как в случае с ультразвуковым и инфракрасным дальномером.

Таблица 1.3 – Сравнение датчиков измерения расстояния. Особенности использования

	RGB-изобр.	Возможность определения границ предмета	Влияние посторонних излучений	Влияние тусклого освещения	Влияние текстур	Влияние зеркальных/прозрачных поверхностей
--	------------	---	-------------------------------	----------------------------	-----------------	--

Лидар	Нет	Да	Да	Нет	Нет	Низкое
Стереопара	Да	Да	Нет	Да	Да	Высокое
Активное стерео	Да	Да	Да	Нет	Нет	Высокое
Камера структ. света	Да	Да	Да	Нет	Нет	Высокое
ToF-камера	Да	Да	Да	Нет	Нет	Высокое
ИК дальномер	Нет	Да	Да	Нет	Нет	Среднее
УЗ дальномер	Нет	Нет	Нет	Нет	Нет	Низкое
Радар	Нет	Нет	Нет	Нет	Нет	Низкое

Для построения карты и ориентации в помещении необходимо определять очертания объектов в исследуемой среде, поэтому принцип работы датчика должен позволять определять границы предметов, что затруднительно при использовании радара и ультразвукового дальномера. Под влиянием посторонних излучений подразумевается возникновение ошибок измерений при наличии солнечных лучей или ИК-излучений от приборов в зоне работы датчика. Очевидно, что данному влиянию, а также трудностям при наличии стеклянных и зеркальных поверхностей, подвержены датчики, использующие инфракрасный диапазон излучений для измерений. Влияние тусклого освещения понимается как возникновение сложности при работе в среде с недостаточной освещенностью. Такие сложности появляются при работе со стереокамерой, для которой необходимо хорошее освещение. Для остальных рассматриваемых датчиков данная проблема не возникает, поскольку в них используются дополнительные источники излучения. Под влиянием текстуры имеется в виду необходимость наличия видимых отличий (особенностей) поверхности, например, узоров, изменений цветов, визуальных границ предметов. Влиянию данного критерия также подвержены измерения с использованием стереокамеры, поскольку выполняется визуальное сравнение

двух изображений, поэтому практически невозможно определить с помощью стереокамеры расстояние до однотонных стен, характерных для помещений.

По результатам сравнения наиболее подходящими датчиками для построения карты помещения с использованием наземного мобильного робота являются RGBD-камеры. Они обеспечивают необходимый диапазон и точность измерений, соответствуют условиям работы в среде данного типа и имеют приемлемую стоимость. При этом RGBD-камера с технологией активное стерео имеет более высокую точность измерений и разрешение по сравнению с камерой структурированного света, а также более высокое разрешение и более низкое энергопотребление по сравнению с ToF-камерой. Таким образом, для дальнейших исследований выбрана RGBD-камера с технологией активное стерео.

2 Аналитический обзор алгоритмов RGBD-SLAM для наземных мобильных роботов

С развитием средств обработки графики, а также удешевлением и облегчением камер, широкое распространение получил визуальный SLAM. Для получения данных чаще всего используются одиночные камеры, стереокамеры, RGBD-камеры [3].

Методы визуального SLAM разделяются на прямые методы (direct) и методы, основанные на извлечении особых точек (feature-based) [19]. В первом случае используются различия в интенсивности пикселей кадров и соответственно минимизация фотометрической ошибки. Во втором случае – расстояния между особыми точками и минимизация ошибки проецирования, такой подход имеет устойчивость к поворотам изображения, изменению масштаба и освещения.

Важным моментом является возможность получить в результате выполнения алгоритма карту занятости, пригодную для дальнейшего

использования. В общем случае двумерная карта занятости представляет собой равномерную сетку, в которой каждой ячейке соответствует двоичная переменная, определяющая свободна ячейка или нет [2].

В различных методах на выходе алгоритма получается разреженная (sparse), полу-плотная (semi-dense) или плотная (dense) карта. Разреженная карта представляет собой небольшое подмножество точек, описывающих сцену, т.е. содержит только особые точки. Плотная карта, напротив, использует большинство или все точки кадра, поэтому алгоритмы, обрабатывающие такую карту, зачастую требуют использования графического процессора из-за большой вычислительной стоимости. Методы, работающие с полу-плотной картой, используют только пиксели с большими перепадами яркости, т.е. границы объектов [20].

Основными составляющими алгоритмов для решения задачи SLAM являются [21]:

- определение текущего положения робота по данным датчиков;
- поиск ориентиров в окружающем пространстве;
- ассоциация данных, т.е. сопоставление новых и ранее исследованных ориентиров;
- обновление карты окружения.

Первым шагом алгоритма всегда является инициализация карты, которая заключается в инициализации глобальной системы координат начальным положением робота.

Главным отличием визуального SLAM от методов визуальной одометрии является возможность замыкания петель и оптимизации глобальной карты [19]. Замыкание петель (или циклов) заключается в определении повторного посещения уже исследованной локации и последующей корректировки карты с учетом новых и накопленных данных. Под оптимизацией понимается нахождение наиболее вероятной траектории движения робота и/или положения ориентиров с учетом минимизации ошибки.

В данном разделе приведен обзор и сравнение известных алгоритмов SLAM, использующих в качестве датчиков RGBD-камеры.

2.1 RGBDSLAMv2

Метод RGBDSLAMv2 [17] разработан в 2014 году и основан на графах. Систему можно представить в виде следующих основных составных частей:

- модуль обработки данных, поступающих с датчика;
- модуль оптимизации графа;
- окончательное представление карты.

В модуль обработки данных входит извлечение особых точек из RGB-изображения и сохранение их дескрипторов вместе с положением относительно позиции камеры по измерениям с карты глубины. Геометрические связи позволяют оценить положение робота между позициями камеры. Для сравнения дескрипторов вычисляется расстояние между ними в пространстве дескрипторов. Для решения проблемы ошибочных совпадений точек при вычислении преобразований подобия кадров используется метод оценки параметров модели на основе случайных выборок RANSAC [22]. В вычислениях также используются особые точки, не имеющие соответствующих измерений глубины, чтобы повысить устойчивость системы к отсутствующим измерениям. Для вычисления преобразования между кадрами выбираются несколько последних кадров, кандидаты на замыкание петли среди соседних на графе с предыдущим кадром, а также несколько кадров с остова дерева, полученного из графа положений (при этом последние кадры исключаются из дерева). Ключевые кадры добавляются, когда текущий кадр не соответствует предыдущему ключевому.

Модуль оптимизации графа отвечает за создание графа положения, который несет информацию о геометрических связях и их неопределенности, а также за оптимизацию этого графа для нахождения наиболее вероятной траектории движения камеры. Для оптимизации графа используется нелинейная оптимизация и библиотека g2o [23].

По информации о траектории и относительном положении особых точек строится вероятностная трехмерная карта занятости с использованием фреймворка OctoMap [24]. Древовидная структура представления вокселей (элементов объемного изображения) позволяет снизить объем памяти для хранения карты. Для использования карты «на ходу» необходимо понизить ее разрешение (в экспериментах авторов RGBDSLAMv2 до 10 см).

Одним из распространенных методов оценки алгоритмов SLAM является метрика ATE RMSE – среднеквадратичное значение абсолютной ошибки траектории [25]. Абсолютное расстояние между истинной (эталонной) траекторией и траекторией, найденной алгоритмом, в момент времени i вычисляется по формуле

$$ATE_i = T_i^{-1} S X_i, \quad (1)$$

где T – истинная траектория движения, X – траектория движения, найденная алгоритмом, S – трансформация твердого тела, найденная методом Хорна [26] (используется, поскольку траектории могут быть в произвольных системах координат). Затем оценивается среднеквадратичная ошибка по всем интервалам времени по формуле

$$RMSE(ATE) = \sqrt{\frac{1}{n} \sum_{i=1}^n ATE_i^2} \quad (2)$$

где n – количество точек траектории.

В таблице 2.1 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Каждый набор представляет собой видеоряд, снятый с помощью RGBD-камеры, и истинную траекторию движения камеры. Наборы f1 сняты в офисной среде, а наборы f2 – в промышленном помещении. Набор данных fr1/desk представляет собой видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета для проверки замыкания петель, fr2/desk – видеозапись вокруг двух офисных столов с компьютером.

Таблица 2.1 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом RGBDSLAMv2

Набор данных TUM (Мюнхенский технический университет)	ATE RMSE (м)
fr1/desk	0,026
fr1/room	0,087
fr2/desk	0,057

Использование дескриптора ORB [28] (Oriented FAST and Rotated BRIEF) или комбинации Shi-Tomasi [29] с SURF [30] (Speeded Up Robust Features) больше всего подходит для работы в режиме реального времени, но из-за накапливаемой ошибки лучше использовать дополнительную одометрию. Если есть возможность использования графического процессора, лучший результат достигается с использованием дескриптора SIFT [31] (Scale-Invariant Feature Transform) [17].

2.2 ORB-SLAM2

Метод ORB-SLAM2 [32] предложен в 2016 году как усовершенствование ORB-SLAM (2015) [33]. Дополнением является возможность использовать не только одиночные камеры, но и стерео- и RGBD-камеры. Алгоритм подходит для локализации и картографирования в помещении и на больших территориях снаружи и работает в режиме реального времени без использования графического процессора.

ORB-SLAM2 строит разреженную карту окружения. Для обнаружения и описания точек данный метод использует дескриптор ORB [28], который содержит усовершенствованный детектор FAST [34] (Features from Accelerated Test) и дескриптор BRIEF [35] (Binary Robust Independent Elementary Features). «На ходу» можно восстановить по ключевым кадрам и картам глубины локальную карту. Для восстановления 3D-модели всей исследованной сцены из последовательности ключевых кадров и разреженной карты используется метод полного уравнивания связок (full Bundle Adjustment). В общем случае

уравнивание связок (BA) заключается в минимизации ошибки перепроецирования между положением наблюдаемых и прогнозируемых точек. При этом уточняются трехмерные координаты, описывающие геометрию сцены, параметры относительного движения и оптические характеристики камеры.

На примере метода ORB-SLAM2 можно рассмотреть подробно структуру методов SLAM, основанных на графах и извлечении особых точек.

Карта, которая обновляется в процессе выполнения алгоритма, содержит:

- точки карты;
- ключевые кадры;
- граф общей видимости (Convisibility Graph);
- минимальное остовное дерево графа общей видимости.

Каждая точка карты содержит трехмерное положение точки в пространстве, среднее направление обзора (лучей, соединяющих точку и оптические центры кадров), характерный ORB-дескриптор, максимальное и минимальное расстояния, с которых может наблюдаться точка. Каждый ключевой кадр содержит позицию камеры, характеристики камеры (например, фокусное расстояние и оптический центр), все найденные на кадре ORB-особенности. В вершинах графа общей видимости находятся ключевые кадры, а ребра между ними существуют, если ключевые кадры наблюдают как минимум 15 общих точек карты. Основной граф содержит остовное дерево графа общей видимости, ребра графа общей видимости со связностью минимум 100 общих точек и ребра замыкания петель.

Алгоритм также содержит блок распознавания места, основанный на библиотеке DBoW2 [36] (модель bag-of-words – «мешок слов») и используемый для релокализации, повторной инициализации и обнаружения петель.

Для совместимости метода как со стерео, так и с RGBD-камерами, входные кадры предварительно обрабатываются. В случае стереокамеры из двух изображений извлекаются ORB-особенности, затем в результате стереосравнения обнаруживаются парные (stereo) и одиночные (mono) ключевые

точки. При использовании RGBD-камеры ORB-особенности извлекаются из одного изображения, а с помощью карты глубины для них создаются стерео-координаты, из которых получаются парные и одиночные ключевые точки. Далее полученные ключевые точки используются алгоритмом.

Парная ключевая точка имеет пару на правом кадре и описывается трехмерной координатой (u_L, v_L, u_R) – координатами на левом кадре и горизонтальной координатой на правом кадре. Для RGBD-камеры вычисляется виртуальная правая координата

$$u_R = u_L - \frac{f_x b}{d}, \quad (3)$$

где f_x – горизонтальная составляющая фокусного расстояния, b – стереобаза, т.е. расстояние между ИК-излучателем структурированного света и ИК-камерой, d – значение глубины. Если глубина точки как минимум в сорок раз больше стереобазы, точка считается дальней (предоставляет точную информацию о повороте), а если меньше этого значения – ближней (предоставляет информацию о масштабе, повороте и сдвиге).

Ключевые точки, которые не имеют пары на правом изображении или которые имеют недействительное значение глубины, становятся одиночными ключевыми точками и описывается двумерной координатой на левом изображении. Они не несут информации о масштабе, но используются в вычислении оценки поворота и сдвига.

Рассмотрим основные шаги алгоритма ORB-SLAM2.

При инициализации из первого кадра создается ключевой, его положение становится началом координат, а все парные точки наносятся на карту.

Система разделена на три главных потока:

- отслеживание (локализация);
- построение локальной карты;
- замыкание петель.

При отслеживании из кадра извлекаются особые точки, как было описано выше, затем по предыдущему кадру прогнозируется первоначальная оценка

положения (модель движения – с постоянной скоростью) или выполняется глобальная релокализация, если отслеживание было прервано. Далее отслеживается локальная карта, которая содержит ключевые кадры, имеющие общие точки карты с текущим, их соседние кадры на графе общей видимости и опорный ключевой кадр, наиболее соответствующий текущему. Для этого на текущем кадре выполняется поиск точек локальной карты, а затем выполняется оптимизация положения камеры с использованием всех найденных на кадре точек. Обработка нового кадра имеет приоритет над локальным картографированием и может остановить уравнивание связок (Bundle Adjustment – BA). Для выбора нового ключевого кадра должны в обязательном порядке выполняться определенные условия (заданное количество кадров после глобальной оптимизации, после добавления ключевого кадра, ожидание входных данных потоком локального картографирования, количество особых точек и их принадлежность к опорному кадру, количество ближних точек). В потоке отслеживания также выполняется уравнивание связок для оптимизации положения камеры (motion-only BA).

Поток локального картографирования добавляет ключевой кадр, обновляя при этом граф общей видимости и основное дерево, а также вычисляя представление кадра в форме bag-of-words. Затем с карты удаляются недавние точки, если при отслеживании точка была найдена в менее 25% случаев, когда ее видимость была предсказана, или если она не видна как минимум на трех ключевых кадрах, когда с ее создания прошло больше одного кадра. Создание новых точек карты происходит путем триангуляции найденных с помощью дескриптора ORB особых точек с точками ключевых кадров графа. Для оптимизации ключевых кадров локальной карты и их точек выполняется локальное уравнивание связок (local BA), посторонние точки удаляются в середине и конце оптимизации. Далее удаляются лишние локальные ключевые кадры, т.е. те, у которых 90% точек были видны на других ключевых кадрах такого же или меньшего масштаба.

Поток замыкания петель определяет кандидатов на замыкание путем вычисления минимального совпадения описания текущего кадра и соседних с ним в форме bag-of-words и последующего запроса в базу данных распознавания для поиска кандидатов с равным или большим совпадением. При этом исключаются кандидаты, напрямую связанные на графе с текущим кадром. В результате вычислений замыкание принимается, если имеется достаточное количество совпадающих точек, и происходит слияние петли – корректируется положение текущего и соседних ключевых кадров, сливаются совпадающие точки, обновляется граф общей видимости.

После оптимизации основного графа вызывается четвертый поток – оптимизация всех точек карты и всех ключевых кадров, кроме начального, с помощью полного уравнивания связок (full BA) и обновление карты. Замыкание петель имеет приоритет над четвертым потоком и останавливает его в случае обнаружения петли.

В систему добавлен режим локализации, при котором отключаются потоки картографирования и замыкания петель, и положение камеры отслеживается с помощью релокализации или поиска совпадений между текущим кадром и точками карты, а также точками, созданными из предыдущего кадра.

При сравнении ORB-SLAM2 [32] с методами PUT SLAM [37], CCNY_RGBD [38], RGBDSLAMv2 [17] рассматриваемый метод показал наиболее точные результаты оценки траектории движения камеры, также было выявлено, что он теряет положение камеры при резких поворотах, хотя и восстанавливает его при возвращении в уже посещенную локацию [39]. В таблице 2.2 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Наборы данных fr1/desk и fr1/desk2 представляет собой видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета, fr2/desk – видеозапись вокруг двух офисных столов с компьютером, fr2/xyz – видеозапись офисного стола по

пространственным осям, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности).

Таблица 2.2 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом ORB-SLAM2

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/desk	0,016
fr1/desk2	0,022
fr1/room	0,047
fr2/desk	0,009
fr2/xyz	0,004
fr3/office	0,019

2.3 RTAB-map

RTAB-Map (Real Time Appearance Based Mapping) [40] относится к визуальным методам 3D-SLAM для стерео- и RGBD-камер и может дополнительно использовать одометрию, IMU (inertial measurement unit), лидар. При запуске можно настраивать такие параметры, как размеры областей памяти, разрешение карты занятости, используемые детекторы и дескрипторы, способ оценки соответствий и скорость обнаружения [41].

Алгоритм позволяет получить на выходе плотное цветное облако точек, а также 2D и 3D карту занятости. Карта хранится в виде графа, вершинами которого являются положения робота и соответствующие им кадры, а ребрами являются преобразования одометрии и замыкания петель [42].

Алгоритм данного метода основан на извлечении особых точек и оптимизации графа положений. По соответствию особых точек на двух кадрах вычисляются их трехмерные координаты, которые подаются на вход алгоритма PnP (Perspective-n-Point) для оценки положения камеры. Новый ключевой кадр

добавляется при недостаточном количестве соответствий на опорном ключевом и текущем кадре [41].

Для достижения работы в режиме реального времени используемая память подразделяется на три области фиксированного размера: краткосрочную, рабочую и долгосрочную (STM, WM, LTM соответственно). Краткосрочная память хранит последние ключевые кадры и их представление в форме bag-of-words. Недавние ключевые кадры, за исключением кадров из краткосрочной памяти, хранятся в рабочей памяти. Долгосрочная память содержит ключевые кадры «давно» посещенных локаций. В процессе выполнения алгоритма кадры добавляются и удаляются из определенных разделов памяти [41].

Оптимизация положения камеры выполняется путем уравнивания связей (Bundle Adjustment), причем при вычислениях используются предыдущие положения камеры и координаты соответствующих точек карты. Для обнаружения замыкания петель текущий кадр, представленный в форме bag-of-words, сравнивается с кадрами в рабочей памяти. Если замыкание найдено, выполняется оптимизация графа положений [41].

В таблице 2.3 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Наборы данных fr1/desk, fr1/desk2 представляет собой видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета, fr2/desk – видеозапись вокруг двух офисных столов с компьютером, fr2/xyz – видеозапись офисного стола по пространственным осям, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности), fr3/nst – видеозапись объекта с четкой изгибающейся формой и явным рисунком. При использовании визуальной одометрии Fovis [43] достигается режим работы в реальном времени [40].

Таблица 2.3 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом RTAB-map с одометрией Fovis

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/desk	0,048
fr1/desk2	0,088
fr1/room	0,119
fr2/desk	0,047
fr2/xyz	0,007
fr3/office	0,051
fr3/nst	0,106

2.4 RGBiD-SLAM

RGBiD-SLAM (2018) [44] – это прямой визуальный алгоритм SLAM, основанный на плотной визуальной одометрии для RGBD-камер с использованием обратной глубины (2015) [45]. Положение камеры между кадрами оценивается попиксельной минимизацией фотометрической, а также геометрической ошибки, выраженной через обратную глубину текущего кадра.

Основу системы составляют два одновременных потока центрального процессора (CPU), которые для выполнения определенных вычислений задействуют потоки графического процессора (GPU).

Первый поток CPU отвечает за отслеживание положения камеры и слияние измерений обратной глубины на близких кадрах в один ключевой кадр. При этом задействуется поток GPU, предназначенный для нахождения плотного пространственного соответствия кадров и слияния обратной глубины. Для оценки движения камеры используется модель постоянной скорости. Кадры из буфера ключевых кадров удаляются после их слияния с последним обработанным кадром. Слияние происходит, если пиксели на обоих кадрах соответствуют одной точке наблюдаемой сцены.

Из буфера ключевых кадров кадры подаются на вход второго потока CPU, который отвечает за сегментацию ключевых кадров, представление их в форме bag-of-words (DBoW2 [36]) и сохранение в базе, а также за обнаружение

замыкания петель и оптимизацию графа положений. Вершины графа являются трехмерными точками, определенными по расположению пикселей, а ребра добавляются для точек со смежными расположениями пикселей на кадре. Для ребер рассчитываются весовые коэффициенты, и вследствие сегментации группируются точки, соединенные ребрами малого веса. После вычисления гистограммы нормалей групп точек, для них вычисляется энтропия. Для замыкания петель используется модель bag-of-words и дескрипторы ORB [28]. Обнаружение замыкания петель выполняется путем сравнения последнего обработанного кадра с кадрами, сохраненными в базе. При этом для плотного пространственного соответствия ключевых кадров при обнаружении петли используется второй поток GPU. В процессе оптимизации графа траектория вычисляется путем минимизации функции стоимости, использующей ограничения одометрии и петель.

В таблице 2.4 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Наборы данных fr1/desk и fr1/desk2 представляет собой видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета, fr2/desk – видеозапись вокруг двух офисных столов с компьютером, fr2/xyz – видеозапись офисного стола по пространственным осям, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности), fr3/nst – видеозапись объекта с четкой изгибающейся формой и явным рисунком. Алгоритм дает сравнимые (на некоторых наборах данных лучшие) с современными подходами результаты (например, ElasticFusion [46]) [44].

Таблица 2.4 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом RGBiD-SLAM

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/desk	0,018

fr1/desk2	0,036
fr1/room	0,040
fr2/desk	0,017
fr3/office	0,025
fr3/nst	0,015

2.5 OpenVSLAM

OpenVSLAM [47] был предложен как фреймворк визуального SLAM в 2019 году. Главным отличием данного метода является возможность использовать различные камеры: не только одиночные, стерео- и RGBD-камеры, которые относятся к камерам с традиционной перспективой, но и «рыбий глаз» и камеры кругового обзора. При этом пользователь может добавлять новые классы для других моделей камер. Система также позволяет сохранять и загружать карты и использовать их для локализации. Алгоритм строит разреженную карту исследуемого пространства. Системой используются библиотеки Eigen [48], OpenCV [49] и g2o [23] соответственно для операций с матрицами, обработки изображений и оптимизации карты.

OpenVSLAM использует алгоритм, основанный на графах и извлечении особых точек из обрабатываемых кадров. Для извлечения особых точек используется дескриптор ORB [28]. Как и в любом алгоритме SLAM, в OpenVSLAM основными компонентами являются модули отслеживания, картографирования и глобальной оптимизации. Модуль отслеживания принимает на вход текущий кадр, извлекает из него особые точки, находит соответствия с локальной картой и после оптимизации положения принимает решение о добавлении нового ключевого кадра. Модуль картографирования добавляет новый ключевой кадр, создает новые точки карты и выполняет локальное уравнивание связок (local BA). Модуль глобальной оптимизации отвечает за обнаружение замыкания петель, оптимизацию графа положений для решения проблемы дрейфа масштаба (т.е. накопления ошибки

масштабирования, что характерно для монокулярных алгоритмов SLAM вследствие отсутствия возможности напрямую измерить расстояния до объектов) и дрейфа траектории, а также глобальное уравнивание связок (global BA).

2.6 SlamDunk

SlamDunk [50] предложен как метод, работающий в режиме реального времени без использования графического процессора. Алгоритм разделен на три главных модуля: локальное картографирование, отслеживание камеры и локальная оптимизация. Траектория представлена в виде последовательности ключевых кадров, положения которых хранятся в дереве квадрантов.

Для локального картографирования и отслеживания не задействуется вся информация о траектории, а используются только последние ключевые кадры, т.е. особенные точки текущего кадра сопоставляются с особыми точками кадров, принадлежащих активному окну. Для фильтрации ошибочных совпадений, как и во многих реализациях, используется RANSAC [22]. После оценки положения камеры и выбора нового ключевого кадра, выполняется локальная оптимизация положений. Локальная оптимизация также запускается при обнаружении замыкания петли при сравнении текущего кадра и кадров, соответствующих локальной карте. Сдвиг окна выполняется, когда смещение текущего положения относительно центра активного окна преодолевает установленный порог, тогда в дереве квадрантов запрашивается новое квадратное окно определенного размера с центром в новом положении. Алгоритм выполняет реконструкцию исследуемой среды, из которой можно получить плотную трехмерную карту занятости.

В таблице 2.5 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Набор данных fr1/desk представляет собой видеозапись четырех офисных столов, fr1/xyz – видеозапись рабочего стола в офисной среде по пространственным осям,

fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета, рисунок поверхности), fr3/nst – видеозапись объекта с четкой изгибающейся формой и явным рисунком. По опубликованным результатам метод SlamDunk показал лучшую точности траектории по сравнению с методом RGBDSLAMv2 [17], который для работы в режиме реального времени задействует графический процессор [50].

Таблица 2.5 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом SlamDunk

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/xyz	0,017
fr1/desk	0,022
fr3/office	0,023
fr3/nst	0,012

2.7 DVO SLAM

Метод DVO SLAM [51] использует плотную визуальную одометрию, а также граф положений. Карта представляет собой граф с положениями ключевых кадров в вершинах и преобразованием положений в ребрах. Схожесть кадров при выборе нового ключевого кадра и подтверждении замыкания петель определяется с использованием энтропии.

Плотная визуальная одометрия для оценки движения камеры использует минимизацию фотометрической (разница интенсивности пикселей) и геометрической (разница измерений глубины) ошибки. Общим ошибкам присваивается вес – существенным ошибкам соответствует меньший вес, чем малым ошибкам. При этом для присвоения малого веса не играет роли, какая составляющая ошибки – фотометрическая, геометрическая или обе – имеет большое значение. Ошибки с малым весом соответствуют посторонним точкам.

Для уменьшения локального дрейфа траектории текущий кадр сравнивается с ключевым. Рассчитывается коэффициент, равный отношению энтропии оценки перемещения от последнего ключевого кадра до текущего кадра к энтропии перемещения от последнего ключевого кадра до кадра, следующего сразу после него. Если для текущего кадра данный коэффициент ниже установленного порога, предыдущий кадр становится ключевым и добавляется на карту.

Поскольку метод предназначен для использования в помещениях небольшой площади, в которых визуальная одометрия показывает достаточно точные результаты, поиск кандидатов на замыкание петли выполняется среди метрически близких соседних кадров. В частности, поиск происходит в области сферы с центром в текущем ключевом кадре и заданным радиусом. Для подтверждения замыкания вычисляется коэффициент, равный отношению энтропии оценки перемещения от последнего ключевого кадра до текущего кадра к средней энтропии, рассчитанной для промежуточных кадров. При обнаружении замыкания данный коэффициент превышает установленный порог. После подтверждения замыкание добавляется на граф в виде нового ребра. При компенсации ошибки сильнее изменяются ребра графа с наименьшей точностью. Для оптимизации графа используется библиотека g2o [23].

В таблице 2.6 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Набор данных fr1/xyz представляет собой видеозапись рабочего стола в офисной среде по пространственным осям, fr1/desk и fr1/desk2 – видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета, fr2/desk – видеозапись вокруг двух офисных столов с компьютером, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности). Метод DVO SLAM показал меньшую

ошибку траектории, чем реализация алгоритма KinectFusion [52] и метод RGBDSLAM [17] (на большинстве проверяемых наборов данных) [51].

Таблица 2.6 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом DVO SLAM

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/xyz	0,011
fr1/desk	0,021
fr1/desk2	0,046
fr1/room	0,053
fr2/desk	0,017
fr3/office	0,023

2.8 BAD SLAM

Метод BAD SLAM [53] использует уравнивание связок (BA), в отличие от других прямых методов, в которых применяется аппроксимация (оптимизация графа положений или графа деформаций). Для работы плотного уравнивания связок на небольших сценах в режиме реального времени используется графический процессор.

Траектория движения представлена положениями ключевых кадров, а геометрия сцены – плотными элементами поверхности (surfel – surface element). Элемент поверхности описывается центром в трехмерной системе координат, вектором нормали к поверхности, радиусом и дескриптором. Такие элементы могут изменять положение только вдоль вектора нормали, а для их оптимизации используются только измерения со схожим вектором нормали.

Отслеживание камеры заключается в начальной оценке ее положения, а также геометрии наблюдаемой сцены. При поступлении нового кадра в систему, выполняется оценка его положения по фотометрическому и геометрическому сравнению с последним ключевым кадром. При этом используется не интенсивность пикселей, а градиенты интенсивностей.

Ключевым кадром становится каждый десятый обрабатываемый кадр. Для поддержания работы в режиме реального времени итерации уравнивания связок (ВА) прерываются в случае добавления нового ключевого кадра.

Для обнаружения замыкания петель используется модель bag-of-words и дескриптор BRIEF [35]. Сначала относительное положение последнего ключевого кадра и кандидата на замыкание оценивается по совпадению особых точек, затем оно обновляется с помощью прямого сопоставления, также сопоставляется текущий кадр с соседними кадрами кандидата. Замыкание подтверждается, если соблюдаются пороговые условия перемещения и поворота. Затем выполняется оптимизация графа положений и уравнивание связок (ВА).

При оптимизации параметров элементов поверхности, траектории и параметров камеры элементы поверхности проецируются на соответствующие ключевые кадры для вычисления фотометрической и геометрической ошибки. Ложным соответствием считается совпадение центра элемента поверхности с пикселем без значения глубины, с сильно отличающимся значением глубины или нормали (нормаль должна быть направлена к камере). В предложенном алгоритме прямого уравнивания связок (ВА) для всех ключевых кадров добавляются недостающие элементы поверхности, т.е. если в квадратной ячейке со стороной равной 4 пикселям ни одному пикселю не соответствует какой-либо элемент поверхности, то создается новый элемент поверхности со значением глубины одного из пикселей ячейки. Затем несколькими итерациями обновляются нормали элементов, оптимизируются их положения и дескрипторы, положения ключевых кадров и параметры камеры. В первой и последней итерации выполняется слияние похожих элементов. В последнем шаге алгоритма удаляются посторонние элементы и устанавливаются в соответствии с измерениями минимальные радиусы элементов.

В таблице 2.7 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на общедоступных наборах данных Мюнхенского технического университета (TUM) [27]. Набор данных

fr1/desk представляет собой видеозапись четырех офисных столов, fr2/xyz – видеозапись офисного стола по пространственным осям, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности). BAD SLAM [53] показал точность траектории, схожую с методом BundleFusion [54].

Таблица 2.7 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом BAD SLAM

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/desk	0,017
fr2/xyz	0,011
fr3/office	0,017

2.9 CVO SLAM

Метод CVO SLAM [55] основан на непрерывной визуальной одометрии [56], использующей ключевые кадры. Текущая реализация не способна работать в режиме реального времени и поддерживает только граф положений.

Для нового кадра в рамках визуальной одометрии вычисляются преобразования подобия по сравнению с предыдущим кадром и с последним ключевым, данные преобразования добавляются на локальный граф положений. После заполнения локального графа положений, он оптимизируется и добавляется в глобальный граф.

Поиск замыканий между текущим и остальными ключевыми кадрами выполняется при добавлении локального графа в глобальный. Для обнаружения замыкания петель используется дескриптор ORB [28]. Извлечение особенностей и сравнение ключевых кадров выполняется аналогично ORB-SLAM2 [32]. Для уточнения оценки положения камеры используется полуплотная прямая непрерывная визуальная одометрия. После подтверждения замыкания оно добавляется в глобальный граф положений, затем для уменьшения накопленной ошибки граф оптимизируется следующим образом:

сначала выполняется оптимизация текущего локального графа с фиксированным положением текущего ключевого кадра, после этого выполняется оптимизация всех ключевых кадров и ребер.

В таблице 2.8 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Набор данных fr1/xyz представляет собой видеозапись рабочего стола в офисной среде по пространственным осям, fr1/desk и fr1/desk2 – видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета. CVO SLAM [55] имеет точность траектории сравнимую с DVO SLAM [51] и лучшую, чем BAD SLAM [53]. В экспериментах на данных с отсутствующей явной текстурой и структурой лучших результатов достигает CVO SLAM [55], ORB-SLAM2 [32] не способен обрабатывать данные без явной текстуры. По результатам сравнения накопительной ошибки траектории после выполнения нескольких последовательностей данных лучшие результаты показывает метод BAD SLAM [53], затем DVO SLAM [51], ORB-SLAM2 [32] и CVO SLAM [55], худшие результаты по сравнению с данными методами у ElasticFusion [46] и BundleFusion [54].

Таблица 2.8 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом CVO SLAM

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/desk	0,025
fr1/desk2	0,034
fr1/room	0,111
fr1/xyz	0,017

2.10 ElasticFusion

Метод ElasticFusion [46] относится к прямым методам SLAM и способен в режиме реального времени строить плотную карту наблюдаемой сцены,

имеющий размер комнаты. Вместо использования графа положений и оптимизации траектории движения, данный метод использует граф деформации и оптимизирует карту. Важно отметить, что метод использует графический процессор.

Геометрия сцены представлена элементами поверхности (surfel – surface element), которые описываются положением в пространстве, вектором нормали к поверхности, цветом, весом, радиусом, временем инициализации и временем последнего обновления. Элементы поверхности делятся на активные и неактивные. Для оценки положения камеры вычисляется геометрическая и фотометрическая разность текущего кадра и соответствующего прогноза по активным элементам поверхности, полученным при предыдущей оценке, а затем минимизируется функция стоимости, учитывающая геометрическую и фотометрическую ошибки. Поверхности ставится в соответствие граф деформации, в узлах которого записано положение в пространстве, время инициализации и аффинное преобразование. Узлы графа деформации связаны направленными ребрами с соседними узлами (число соседних узлов фиксировано). Такой граф создается для элементов поверхности каждого нового кадра, при этом положению и времени инициализации узлов присваивается таковое соответствующих элементов поверхности. Для замыкания петель в соответствии с ограничениями, наложенными на участки поверхности, и обновления карты после оптимизации деформируются активные и неактивные элементы поверхности с помощью графа. Неактивные элементы становятся активными после обнаружения и выполнения локального замыкания.

В таблице 2.9 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Набор данных fr1/hyz представляет собой видеозапись рабочего стола в офисной среде по пространственным осям, fr1/desk и fr1/desk2 – видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета, fr2/desk – видеозапись

вокруг двух офисных столов с компьютером, fr2/xyz – видеозапись офисного стола по пространственным осям, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности), fr3/nst – видеозапись объекта с четкой изгибающейся формой и явным рисунком.

Таблица 2.9 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом ElasticFusion

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/desk	0,020
fr1/desk2	0,048
fr1/xyz	0,011
fr1/room	0,068
fr2/xyz	0,011
fr2/desk	0,071
fr3/office	0,017
fr3/nst	0,016

2.11 BundleFusion

Метод BundleFusion [54] может применяться для построения карты больших помещений. Оптимизация положения и обновление строящейся карты происходит с каждым новым кадром. Для первоначальной оценки положения, релокализации и обнаружения замыкания петель используются разреженные особенности изображения (применяется SIFT [31]), при этом для сопоставления выполняется графическим процессором. Для исключения ложных соответствий (в том числе ложных замыканий петель) используется фильтр соответствия особенностей, фильтр площади поверхности и подтверждение соответствия с помощью плотного сопоставления с учетом геометрических и фотометрических изменений.

Первоначальная оценка положения уточняется путем оптимизации с использованием фотометрической и геометрической ошибки. Оптимизация выполняется на двух уровнях – локальном и глобальном. На локальном уровне для оптимизации положения определяется пространственное соответствие нескольких последовательных кадров (набора кадров), соответствующих текущему временному окну. На глобальном уровне определяется пространственное соответствие наборов кадров при обнаружении соответствий особенностей ключевых кадров (первых кадров в наборах).

Для плотной реконструкции наблюдаемая среда описывается вокселями (элементами объемного изображения), т.е. используется разреженное объемное представление и слияние. Реконструкция изменяется при каждом обновлении положений, при этом кадры удаляются из своего предыдущего положения и снова интегрируются с учетом оптимизированного положения.

В таблице 2.10 представлено среднеквадратичное значение абсолютной ошибки траектории при тестировании алгоритма на наборах данных Мюнхенского технического университета (TUM) [27]. Набор данных fr1/desk представляет собой видеозапись четырех офисных столов, fr2/xyz – видеозапись офисного стола по пространственным осям, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности), fr3/nst – видеозапись объекта с четкой изгибающейся формой и явным рисунком.

Таблица 2.10 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методом BundleFusion

Набор данных TUM (Мюнхенский технический университет)	RMSE of ATE (м)
fr1/desk	0,016
fr2/xyz	0,011
fr3/office	0,022
fr3/nst	0,012

2.12 Использование RGBD-камер для лазерных методов SLAM

В некоторых исследованиях предпринимаются попытки применения RGBD-камер в качестве датчиков для методов SLAM, изначально разработанных для использования с дальномерами. Методы Gmapping [57] и HectorSLAM [58] могут принимать на вход данные с RGBD-камеры, имитирующей лазерный сканнер, но при этом снижается точность, возникают проблемы с замыканием петель и параллельностью стен [41].

Gmapping [57] относится к методам 2D-SLAM, которые используют данные одометрии и измерения дальномеров и в процессе выполнения строят карту занятости. Данный метод использует фильтр частиц Рао-Блэквелла [41].

HectorSLAM [58] – это метод, обрабатывающий двумерные данные с дальномеров, для повышения точности можно дополнительно использовать инерционно-измерительные устройства (IMU), которые дают возможность строить трехмерную карту. Для соединения данных с разных датчиков используется фильтр Калмана [42].

По результатам экспериментов [42] HectorSLAM [58] показывает неудовлетворительную работу с использованием RGBD-камер, которая выражается в построении карт с большим количеством артефактов и неточностей, в целом не пригодных для дальнейшего использования.

Применение Gmapping [57] с RGBD-камерами имеет место быть, но проигрывает в точности алгоритмам, разработанным специально для камер глубины. Оба алгоритма оказывают меньшую нагрузку на центральный процессор по сравнению с ORB-SLAM2 [32] и RTAB-map [40], поскольку в отличие от них являются двумерными.

2.13 Выводы по разделу

Сравнение рассмотренных алгоритмов, предназначенных для обработки входных данных с RGBD-камер, приведено в таблице 2.11. Некоторые из представленных алгоритмов не способны отслеживать положение камеры и строить карту в режиме реального времени, часть из них способна выполнять

данное условие только при использовании графического процессора. Под режимом реального времени подразумевается способность алгоритма обрабатывать один кадр за менее чем 0,033 с при частоте кадров примерно 30 Гц.

Для получения качественной карты занятости желательно, чтобы алгоритм строил плотную карту окружения. Поскольку Gmapping [57] и HectorSLAM [58] показывают результаты с низкой точностью и большим количеством артефактов, они исключены из сравнения.

Поскольку отличием фреймворка OpenVSLAM [47] от метода ORB-SLAM2 [32] в основном является адаптация под различные модели камер, он так же не рассматривается в сравнении.

Распространенным программным обеспечением для решения задач робототехники является универсальный кроссплатформенный программный пакет для роботов ROS (Robot Operating System) [59]. ROS имеет открытый программный код, обеспечивает удобную разработку и использование за счет модульной структуры, поддерживает повторное использование программного кода, а также обладает набором готовых решений для типовых задач [60]. При этом ROS является бесплатным для использования в исследованиях и коммерции.

Таблица 2.11 – Сравнение методов RGBD-SLAM

Метод RGBD-SLAM	Работа в режиме реального времени	Тип карты	Процессор	Реализация в ROS
ORB-SLAM2	+	Разреженная	CPU	+
RTAB-map	+	Плотная	CPU	+
RGBDSLAMv2	+	Плотная	CPU (+GPU)	+
RGBiD-SLAM	+(GPU)	Плотная	CPU (+GPU)	+
DVO-SLAM	-	Плотная	CPU	+
SlamDunk	+	Плотная	CPU	+
CVO SLAM	-	Разреженная	CPU	+

BadSLAM	-	Плотная	CPU (+GPU)	-
ElasticFusion	+ (GPU)	Плотная	CPU (+GPU)	-
BundleFusion	+ (GPU)	Плотная	CPU (+GPU)	-

В таблице 2.12 представлено среднеквадратичное значение абсолютной ошибки траектории, т.е. отклонения вычисленной траектории от эталонной, при тестировании на наборах данных Мюнхенского технического университета (TUM) [27]. Каждый набор представляет собой видеоряд, снятый с помощью RGBD-камеры, и истинную траекторию движения камеры. Набор данных fr1/xyz представляет собой видеозапись рабочего стола в офисной среде по пространственным осям, fr1/desk и fr1/desk2 – видеозапись четырех офисных столов, fr1/room – видеозапись офисного кабинета, fr2/desk – видеозапись вокруг двух офисных столов с компьютером, fr2/xyz – видеозапись офисного стола по пространственным осям, fr3/office – видеозапись офисной сцены с сильно выделяющейся структурой (формы и изгибы объектов) и текстурой (цвета и рисунок поверхности), fr3/nst – видеозапись объекта с четкой изгибающейся формой и явным рисунком.

Таблица 2.12 – Среднеквадратичное значение абсолютной ошибки траектории, полученной методами RGBD-SLAM на наборах данных TUM

Метод RGBD-SLAM	Процессор	RMSE of ATE (м)							
		fr1				fr2		fr3	
		desk	desk2	room	xyz	desk	xyz	office	nst
ORB-SLAM2	Intel Core i7-4790 (16 Гб)	0,016	0,022	0,047	-	0,009	0,004	0,010	0,019
RTAB-map	Intel Core i7-3770 (6 Гб)	0,048	0,088	0,119	-	0,047	0,007	0,051	0,106
RGBDSLAMv2	Intel Core i7 (3,40 ГГц), nVidia GeForce GTX 570	0,026	-	0,087	-	0,057	-	-	-
RGBiD-SLAM	Intel Core i7-4710HQ	0,018	0,036	0,040	-	0,017	-	0,025	0,015

	(2,50 ГГц, 16 Гб), nVidia GeForce GTX 850M (4 Гб)								
DVO-SLAM	Intel Core i7-2600 (3,40 ГГц, 16 Гб)	0,02 1	0,04 6	0,05 3	0,01 1	0,01 7	-	0,03 5	-
SlamDunk	CPU	0,02 2	-	-	0,01 7	-	-	0,02 3	0,01 2
CVO SLAM	Intel Core i7- 6700HQ (2,60 ГГц, 16 Гб)	0,02 5	0,03 4	0,11 1	0,01 7	-	-	-	-
BadSLAM	Intel Core i7- 6700K, MSI GeForce GTX 1080 Gaming X (8 Гб)	0,01 7	-	-	-	-	0,01 1	0,01 7	-
ElasticFusion	Intel Core i7- 4930K (3,4 ГГц, 32 Гб), nVidia GeForce GTX 780 Ti (3 Гб)	0,02 0	0,04 8	0,06 8	0,01 1	0,07 1	0,01 1	0,01 7	0,01 6
BundleFusion	Intel Core i7 (3.4GHz, 32 Гб), nVidia GeForce GTX Titan X, GTX Titan Black	0,01 6	-	-		-	0,01 1	0,02 2	0,01 2

В таблице так же указаны центральные и графические процессоры, на которых проводилось тестирование алгоритмов. Легко заметить, что следствием способности к работе в режиме реального времени при меньших затратах вычислительных ресурсов является снижение точности работы алгоритма. Однако повышение точности за счет увеличения вычислительной мощности сказывается на увеличении габаритов мобильного робота.

В результате сравнения алгоритмов RGBD-SLAM для дальнейших исследований выбраны методы RTAB-map [40], RGBDSLAMv2 [17], SlamDunk [50], поскольку они способны строить плотную карту окружения в режиме реального времени без использования графического процессора.

ЗАКЛЮЧЕНИЕ

В рамках данной работы проведено сравнение датчиков для измерения расстояний, используемых при решении задачи метода SLAM. В качестве датчика, наиболее подходящего для создания карты помещения с использованием наземного мобильного робота, была выбрана RGBD-камера с технологией активного стерео. Данная камера при относительно невысокой стоимости и сравнительно невысоком энергопотреблении обеспечивает диапазон и точность измерений, достаточный для работы в помещении, а также соответствует условиям работы в среде данного типа.

Сделан обзор существующих алгоритмов SLAM для использования с RGBD-камерой и выполнено сравнение рассмотренных алгоритмов. Основными критериями выбора алгоритмов для поставленной задачи являются способность отслеживать положение камеры и строить карту исследуемой среды в реальном времени, задействование минимальных вычислительных ресурсов (т.е. предпочтительны алгоритмы, способные выполнять задачу без использования графического процессора), построение алгоритмом плотной карты окружения для получения качественной карты занятости. В результате сравнения алгоритмов RGBD-SLAM для дальнейших исследований выбраны алгоритмы RTAB-map, RGBDSLAMv2 и SlamDunk, поскольку они способны строить плотную карту окружения в реальном времени без использования графического процессора.

В план дальнейших исследований входит изучение математического описания выбранных алгоритмов, подбор методов и критериев сравнения алгоритмов для решения поставленной задачи, а также проверка пригодности данных алгоритмов для решения задачи SLAM с заданными условиями в среде ROS. По результатам тестирования необходимо сравнить алгоритмы и рассмотреть возможности улучшения работы выбранного алгоритма.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Durrant-Whyte H., Bailey T. Simultaneous localization and mapping: Part I. // IEEE Robotics and Automation Magazine. - 2006. - Vol. 13, no. 2. - P. 99-110.
2. Thrun S., Burgard W., Fox D. Probabilistic Robotics (Intelligent Robotics and Autonomous Agents). - The MIT Press, 2005.
3. Huang B., Zhao J., Liu J. A Survey of Simultaneous Localization and Mapping. - 2019.
4. Hadji S.E., Kazi S., Hing T.H., Ali M.S. A Review: Simultaneous Localization and Mapping Algorithms. // Jurnal Teknologi. - Vol. 73, no. 2. - 2015.
5. Rosique F., Navarro Lorente P., Fernandez C., Padilla A. A Systematic Review of Perception System and Simulators for Autonomous Vehicles Research. // Sensors. - 2019. - Vol. 19. - P. 648.
6. Chong T.J., Tang X.J., Leng C.H., Yogeswaran M., Ng O.E., Chong Y.Z. Sensor Technologies and Simultaneous Localization and Mapping (SLAM). // Procedia Computer Science. - 2015. - Vol. 76. - P. 174-179.
7. Zaffar M., Ehsan S., Stolkin R., Maier K. Sensors, SLAM and Long-term Autonomy: A Review. // 2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS). - 2018. - P. 285-290.
8. Infrared distance sensor [Электронный ресурс]. - 2018. - URL: https://home.roboticlab.eu/en/examples/sensor/ir_distance (Дата обращения: 30.10.2019).
9. Benet Gilabert G., Blanes F., Simo J., Perez P. Using Infrared sensors for distance measurement in mobile robots. // Robotics and Autonomous Systems. - 2002. - No. 40. - P. 255-266.
10. Лидары [Электронный ресурс]. - URL: <https://learnnc.info/blog/lidars.html/> (Дата обращения: 02.11.2019).

11. Schoettle B. Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles. - Ann Arbor, MI.: University of Michigan, Sustainable Worldwide Transportation, 2017.
12. Kocic J., Jovičić N., Drndarevic V. Sensors and Sensor Fusion in Autonomous Vehicles. // 26th Telecommunications Forum (TELFOR). - Belgrade, 2018. - P. 420–425.
13. Zanuttigh P., Marin G., Dal Mutto C., Dominio F., Minto L., Cortelazzo G.M. Time-of-Flight and Structured Light Depth Cameras: Technology and Applications. - Cham: Springer, 2016. - P. 8-17.
14. Zollhöfer M. Commodity RGB-D Sensors: Data Acquisition // Rosin P.L., Lai Y.K., Shao L., Liu Y. RGB-D Image Analysis and Processing. - Cham: Springer, 2019. - P. 3-13.
15. Giancola S., Valenti M., Sala R. State-of-the-Art Devices Comparison. // A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies. - Cham: Springer, 2018. - P. 29-39.
16. Wasenmüller O., Stricker D. (2017) Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision. // Chen CS., Lu J., Ma KK. Computer Vision - ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science. - Cham: Springer, 2017. - Vol. 10117. - P. 34-45.
17. Endres F., Hess J., Sturm J., Cremers D., Burgard W. 3-D mapping with an RGB-D camera. // IEEE Transactions on Robotics. - 2014. - Vol. 30, no. 1. - P. 177-187.
18. Myers B. Capturing depth: structured light, time of flight, and the future of 3D imaging [Электронный ресурс]. - 2018. - URL: <https://www.androidauthority.com/structured-light-3d-imaging-870016/> (Дата обращения: 24.10.2019).
19. Taketomi T., Uchiyama H., Ikeda S. Visual SLAM algorithms: a survey from 2010 to 2016. // IPSJ Transactions on Computer Vision and Applications. - Singapore: Springer, 2017. - Vol. 9, no. 1.

20. Cadena C., Carlone L., Carrillo H., Latif Y., Scaramuzza D., Neira J., Reid I., Leonard J.J. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. // IEEE Transactions on Robotics. - 2016. - Vol. 32, no. 6. - P. 1309-1332.
21. Собченко М. И., Ухандеев В. И. Алгоритмы SLAM: обзор существующих решений. // Электронные информационные системы. - 2014. - N 1 (1). - С. 69-78.
22. Fischler M., Bolles R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. // Commun. ACM. - 1981. - Vol. 26, no. 6. - P. 381-395.
23. Kummerle R., Grisetti G., Strasdat H., Konolige K., Burgard W. g2o: A general framework for graph optimization. // Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). - 2011.
24. Hornung A., Wurm K. M., Bennewitz M., Stachniss C., Burgard W. OctoMap: An efficient probabilistic 3D mapping framework based on octrees. // Autonomous Robots. - 2013.
25. Sturm J., Engelhard N., Endres F., Burgard W., Cremers D. A benchmark for the evaluation of RGB-D SLAM systems. // Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). - 2012.
26. Horn B. Closed-form solution of absolute orientation using unit quaternions. // Journal of the Optical Society of America A. - 1987. - Vol. 4. - P. 629-642.
27. Sturm J. RGB-D SLAM Dataset and Benchmark [Электронный ресурс]. - 2012. - URL: <https://vision.in.tum.de/data/datasets/rgbd-dataset/> (Дата обращения: 25.11.2019).
28. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: an efficient alternative to SIFT or SURF. // Proceedings of the IEEE/RSJ International Conference on Computer Vision (ICCV). - 2011. - Vol. 13.

29. Shi J., Tomasi C. Good Features to Track. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). - 1994. - P. 593-600.
30. Bay H., Ess A., Tuytelaars T., VanGool L. Speeded-up robust features (SURF) // Computer Vision and Image Understanding. - 2008. - Vol. 110, no.3. - P. 346-359.
31. Lowe D. Distinctive image features from scale-invariant keypoints. // International Journal of Computer Vision. - 2004. - Vol. 60, no. 2. - P. 91-110.
32. Mur-Artal R., Tardos J. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. // IEEE Transactions on Robotics. - 2017. - Vol. 33, no. 5. - P. 1255-1262.
33. Mur-Artal R., Montiel J.M.M, Tardos J.D. ORB-SLAM: a versatile and accurate monocular SLAM system. // IEEE Transactions on Robotics. - 2015. - Vol. 31, no. 5. - P. 1147-1163.
34. Rosten E., Drummond T. Fusing points and lines for high performance tracking. // Proceedings of the IEEE International Conference on Computer Vision. - 2005. - Vol. 2. - P. 1508-1515.
35. Calonder M., Lepetit V., Strecha C., Fua P. BRIEF: Binary Robust Independent Elementary Features. // 11th European Conference on Computer Vision (ECCV). - 2010. - P. 778-792.
36. Galvez-Lopez D., Tardos J.D. Bags of binary words for fast place recognition in image sequences. // IEEE Transactions on Robotics. - 2012. - Vol. 28, no. 5. - P. 1188-1197.
37. Belter D., Nowicki M., Skrzypczyński P. Accurate map-based RGB-D SLAM for mobile robots. // Reis L., Moreira A., Lima P., Montano L., Muñoz-Martinez V. Robot 2015: Second Iberian Robotics Conference. Advances in Intelligent Systems and Computing. - Cham: Springer, 2016. - Vol. 418. - P. 533-545.
38. Dryanovski I., Valenti R., Xiao J. Fast visual odometry and mapping from RGB-D data. // Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). - 2013. - P. 2305-2310.

39. Kostusiak A., Nowicki M., Skrzypczyński P. On the Application of RGB-D SLAM Systems for Practical Localization of Mobile Robots. // Journal of Automation, Mobile Robotics and Intelligent Systems. - 2017. - Vol. 11. - P. 57-66.
40. Labbé M., Michaud F. RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. // Journal of Field Robotics. - 2018. - Vol. 36.
41. Silva B., Xavier R., Gonçalves L. Mapping and Navigation for Indoor Robots under ROS: An Experimental Analysis. - 2019.
42. Silva B., Xavier R., Nascimento T., Gonçalves L. Experimental evaluation of ROS compatible SLAM algorithms for RGB-D sensors. // 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR). - 2017. - P. 1-6.
43. Huang, A. S., Bachrach, A., Henry, P., Krainin, M., Maturana, D., Fox, D., Roy, N. Visual odometry and mapping for autonomous flight using an RGB-D camera. // Proceedings International Symposium on Robotics Research. - 2011.
44. Gutierrez-Gomez D., Guerrero J. RGBiD-SLAM for Accurate Real-time Localisation and 3D Mapping. - 2018.
45. Gutierrez-Gomez D., Mayol-Cuevas W., Guerrero J. Dense RGB-D visual odometry using inverse depth. // Robotics and Autonomous Systems. - 2016. - Vol. 75. - P. 571-583.
46. Whelan T., Salas-Moreno R.F., Glocker B., Davison A. J., Leutenegger S. ElasticFusion: Real-time dense SLAM and light source estimation. // The International Journal of Robotics Research. - 2016. - Vol. 35, no. 14. - P. 1697-1716.
47. Sumikura S., Shibuya M., Sakurada K. OpenVSLAM: A Versatile Visual SLAM Framework. // Proceedings of the 27th ACM International Conference on Multimedia. - 2019. - P. 2292-2295.
48. Eigen. A C++ template library for linear algebra [Электронный ресурс]. - URL.: <http://eigen.tuxfamily.org/> (Дата обращения: 29.11.2019)
49. Open Source Computer Vision Library [Электронный ресурс]. - URL.: <http://opencv.org/opencv/> (Дата обращения: 29.11.2019)

50. Fioraio N., Di Stefano L. SlamDunk: Affordable Real-Time RGB-D SLAM. // Agapito L., Bronstein M., Rother C. Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science. - Cham: Springer, 2015. - Vol. 8925. - P. 401-414.
51. Kerl C., Sturm J., Cremers D. Dense visual SLAM for RGB-D cameras. // Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. - 2013. - P. 2100-2106.
52. Newcombe R. A., Izadi S., Hilliges O., Molyneaux D., Kim D., Davison A. J., Kohli P., Shotton J., Hodges S., Fitzgibbon A. KinectFusion: Real-time dense surface mapping and tracking. // IEEE International Symposium on Mixed and Augmented Reality (ISMAR). - 2011.
53. Schöps T.M., Sattler T., Pollefeys M. BAD SLAM: Bundle adjusted direct RGB-D SLAM. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. - 2019. - P. 134-144.
54. Dai A., Nießner M., Zollhöfer M., Izadi S., Theobalt C. BundleFusion: Real-Time Globally Consistent 3D Reconstruction Using On-the-Fly Surface Reintegration. // ACM Transactions on Graphics. - 2017. - Vol. 36.
55. Lin X., Sun D., Lin T.Y., Eustice R., Ghaffari M. A Keyframe-based Continuous Visual SLAM for RGB-D Cameras via Nonparametric Joint Geometric and Appearance Representation. - 2019.
56. Ghaffari M., Clark W., Bloch A., Eustice R.M., Grizzle J.W. Continuous direct sparse visual odometry from RGB-D images. // Proceedings of the Robotics: Science and Systems Conference. - Freiburg, 2019.
57. Grisetti G., Stachniss C., Burgard W. Improved techniques for grid mapping with Rao-Blackwellized particle filters. // IEEE Transactions on Robotics. - 2007. - Vol. 23. - P. 34-46.
58. Kohlbrecher S., Stryk O., Meyer J., Klingauf U. A flexible and scalable SLAM system with full 3D motion estimation. // IEEE International Symposium on Safety, Security, and Rescue Robotics. - 2011. - P. 155-160.

59. Программный пакет ROS [Электронный ресурс]. – URL.: <http://www.ros.org/> (Дата обращения: 11.10.2019).

60. Михайлова У.В., Михайлов Е.А., Сарваров А.С. Использование фреймворка ROS для разработки архитектуры системы управления роботом // Электротехнические системы и комплексы. - 2013. - N 21. - С. 117-121.