

**УТВЕРЖДАЮ**

Генеральный директор  
АО «НИИМА «Прогресс»

\_\_\_\_\_ З.К. Кондрашов  
«\_\_» \_\_\_\_\_ 2021 г.  
М.П.

**УТВЕРЖДАЮ**

Генеральный директор  
ООО «Навигационные решения»

\_\_\_\_\_ А.А. Панев  
«\_\_» \_\_\_\_\_ 2021 г.  
М.П.

Разработка и производство аппаратно-программного комплекса средств программно-аппаратной платформы для системы интеллектуальной навигации с использованием технологии искусственного интеллекта

### **ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

ТП в части обоснования технических решений

Использование алгоритма ORB-SLAM

Главный конструктор  
СЧ НИОКР «Интеллект-Н»  
\_\_\_\_\_ В.И. Катречко  
«\_\_» \_\_\_\_\_ 2021 г.

Технический директор  
ООО «Навигационные решения»  
\_\_\_\_\_ А.А. Панев  
«\_\_» \_\_\_\_\_ 2021 г.

Москва 2021 г

# Оглавление

Оглавление .....	2
Введение .....	5
Работы в области SLAM .....	6
Метод распознавания места .....	6
Инициализация карты .....	7
Принцип работы алгоритма.....	11
Общая архитектура системы .....	14
ORB.....	15
Bundle adjustment.....	16
Автоматическая инициализация карты .....	18
Навигация.....	18
Замыкание графа карты .....	18
Использование стереокамер и камер глубины .....	18
ORB SLAM 3.....	19
Общая архитектура системы .....	21
Использование инерциальных данных в модели ORB-SLAM 3[2].....	23
Общая характеристика системы .....	24
Производительность алгоритма .....	24
Список литературы.....	25

Документ представлен в следующем порядке:

- разбор основных методов и задач в области применения алгоритмов SLAM
- обзор работы алгоритма ORB-SLAM, его ключевые элементы
- подробный разбор отдельных ключевых блоков алгоритма

Описание алгоритма ведется в историческом порядке разработки системы.

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

СПО АТБ-С	Специальное программное обеспечение абонентского терминала
СПО МОБ	Специальное программное обеспечение мобильного терминала на базе смартфона или планшета с ОС Аврора или ОС Android
НУ	Навигационное устройство – Абонентский терминал либо смартфон
СПО КСО ИИ	Специальное программное обеспечение комплекта серверного оборудования комплекса интеллектуальной платформы навигации
ИНС	Инерциальная навигационная система
IMU	Инерциально-измерительный блок
ЛСН	Локальная система навигации
СЧ ПО	Специальное программное обеспечение Составная часть
Wi-Fi	Технология беспроводной локальной сети с устройствами на основе стандартов IEEE 802.11
BLE	Беспроводная технология связи Bluetooth с низким энергопотреблением
ИИ, AI	Искусственный интеллект, Artificial Intelligence
ML	Методы машинного обучения
SLAM	Simultaneous localization and mapping, одновременное позиционирование и картографирование
Sensor fusion	Алгоритмы комплексирования различной навигационной информации
ORB-SLAM	метод локализации и навигации
ORB	ORB (Oriented FAST and Rotated BRIEF) дескриптор изображения, метод для получения характерных точек по изображению
BO	Bundle Adjustment - алгоритм добавления связей в граф карты
loop closing	замыкание цикла, объединение вершин графа карты для при возвращении в старую точку
RGBD, RGB+D	камера (rgb) с датчиком глубины (depth sensor)
RMSE	корень среднеквадратичной ошибки

RPE	Relative Pose Error - относительная ошибка позиционирования, средняя относительная ошибка на участках заданной длины
ATE	Absolute Trajectory Error - абсолютная позиционная ошибка (Евклидово расстояние или Расстояние Махаланобиса)

# Введение

Одной из основных задач, связанных с проблемой навигации, является построение локальной карты. Данная задача как правило решается алгоритмами SLAM - одновременной локализации и навигации. Среди современных алгоритмов SLAM можно выбрать несколько основных методов которые будут решать задачу в выбранном контексте максимально эффективно.

В этом документе представлен алгоритм ORB-SLAM[1], [2], функциональная система SLAM, работающая в реальном времени в малых и больших помещениях и на открытом воздухе. Система устойчива к серьезным помехам от движения, позволяет работать с картами больших масштабов, обладает возможностью записывать большие участки маршрута в карту. Алгоритм также включает в себя полную автоматическую инициализацию. Система использует одни и те же функции для всех задач SLAM: отслеживание, отображение, перемещение и замыкание цикла. Стратегия алгоритма обеспечивает надежную работу построения карты и навигации. Алгоритм создает компактную и отслеживаемую карту, которая увеличивается только при изменении содержимого сцены, что позволяет не пересоздавать карту заново, а динамически дополнять существующее представление карты.

Одним из наиболее надёжных, проверенных и развитых алгоритмов является ORB-SLAM. Алгоритм соответствует основным критериям простоты, надёжности и производительности для работы в системах реального времени. Существующие аналоги незначительно отличаются по производительности, при необходимости, интеграция алгоритма ORB-SLAM позволяет без значительных усилий заменить элементы данного алгоритма на аналогичные решения. По тестам производительности, ORB-SLAM равен или превосходит аналогичные подходы. Код алгоритма полностью находится в открытом доступе.

Обновлённая версия ORB-SLAM 3 [2] — это система SLAM, которая поддерживает визуально-инерциальную навигацию и гибридные карты. Метод может работать на одиночной камере, стерео и RGB-D камерах, а также с использованием широкоугольных камер типа «рыбий глаз».

Средняя точность ORB-SLAM3 - 3,6 см для набора данных EuRoC[3] и 9 мм для набора данных TUM-VI (сцена AR / VR), использует портативные устройства для быстрого перемещения.

Подробнее о свойствах и работе самого алгоритма ORB-SLAM далее.

# Работы в области SLAM

Алгоритм ORB-SLAM является многолетней разработкой, более поздние версии этого алгоритма были доработками предыдущих. Существуют и другие алгоритмы SLAM с похожими характеристиками.

В данной части представлено позиционирование алгоритма ORB-SLAM относительно аналогов и сравнение с существующими методами.

## Метод распознавания места

Первоначальным элементом системы SLAM является модуль определения места, метод сопоставляет изображения и карту, определяет попадание в предыдущую точку карты.

В обзоре Williams et al. [4] сравнили несколько подходов к технике распознавания места и пришли к выводу, что методы, основанные на сопоставлении изображения с изображением, лучше масштабируются в больших средах, чем методы «карта-карта» или «изображение-карта». Среди методов, основанных на прямом сопоставлении изображений, на передний план выходят методы набора слов [5], такие как вероятностный подход FAB-MAP [6], из-за их высокой эффективности.

DBoW2 [5] впервые использовал пакеты двоичных слов, полученные из дескрипторов BRIEF [7] вместе с очень эффективным детектором признаков FAST [5]. Это сократило более чем на порядок время, необходимое для извлечения признаков, по сравнению с функциями SURF [8] и SIFT [9], которые до сих пор использовались в подходах к пакетам слов.

Прямое сопоставление с использованием BRIEF, не зависит от вращения или масштаба изображений и ограничивает систему траекториями в плоскости и обнаружением петель с аналогичных точек зрения. Хотя система продемонстрировала свою очень эффективную и надежную работу, для лучшего распознавания характерных точек следует использовать дескрипторы инвариантные к вращению и масштабированию.

В работе [10] представлен метод распознавания места, построенный на DBoW2 с ORB[9]. ORB — это двоичные функции, инвариантные к вращению и масштабированию (в определенном диапазоне), что приводит к очень быстрому распознаванию с хорошей инвариантностью к точке обзора. Время работы распознавателя на четырех различных наборах данных - менее 39 мс (включая извлечение признаков) для извлечения кандидата в цикл из базы данных изображений размером 10K. В работе ORB-SLAM[1] используется улучшенная версия распознавателя мест, использующая информацию о взаимной видимости характерных точек из положения камеры и возвращающую несколько гипотез при запросе к базе данных, а не только наилучшее совпадение.

## Инициализация карты

Для монокулярного SLAM чтобы определить положение характерных точек на карте в трехмерном пространстве требуется восстановить глубину на изображении. Существует три известных метода для восстановления глубины в изображении: использование специальной камеры с RGBD сенсором которая напрямую выдает данные глубины в изображении, реконструкция трехмерной сцены по двум жестко связанным камерам (работа в режиме стереосъемки), реконструкция глубины точек по серии кадров с разных направлений. Таким образом монокулярный SLAM требует особой процедуры для создания начальной карты, потому что глубина не может быть восстановлена из одного изображения.

При разборе методов SLAM необходимо в первую очередь рассматривать монокулярный SLAM.

Задача монокулярного SLAM изначально была решена с помощью методов фильтрации [11]–[14].

При таком подходе каждый кадр обрабатывается фильтром для совместной оценки местоположения объектов на карте и положения камеры. Он имеет недостатки, заключающиеся в бесполезной трате вычислений при обработке последовательных кадров с небольшим количеством новой информации и накоплении ошибок линеаризации.

С другой стороны, подходы на основе ключевых кадров [4], [15] оценивают карту, используя только выбранные кадры (ключевые кадры), что позволяет выполнять более дорогостоящую, но точную оптимизацию настройки пакетов, поскольку отображение не привязано к частоте кадров. Strasdat et. al. [14] продемонстрировал, что методы, основанные на ключевых кадрах, более точны, чем фильтрация, при тех же вычислительных затратах.

Наиболее представительной системой SLAM на основе ключевых кадров, вероятно, является PTAM [4], алгоритм разделяет отслеживание и отображение камеры в параллельных потоках, и успешно работает для приложений дополненной реальности в реальном времени в небольших средах. Точки карты PTAM соответствуют углам FAST, сопоставленным с помощью корреляции фрагментов. Это делает точки полезными только для отслеживания, но не для распознавания места. Фактически, PTAM не обнаруживает большие циклы, и определение перемещения основано только на корреляции сжатых ключевых кадров с низким разрешением, что снижает точность алгоритма поскольку учитывает положение точки обзора менее точно.

Монокулярный SLAM без определения глубины изображения обладает проблемой масштаба. По исходному видеопотоку невозможно определить масштаб изображения и расстояние до объектов, следовательно, приходится корректировать карту после добавления объектов.

Strasdat et. [14] представили крупномасштабную монокулярную систему SLAM реализующим замыкание петель с помощью оптимизации графа положений камеры и характерных точек с ограничениями 7DoF. Использование 7DoF ограничений позволяет скорректировать ошибку масштаба в монокулярном SLAM. ORB-SLAM заимствует тип ограничений модели из этой работы, и реализует замыкание цикла с помощью оптимизации графа позы 7DoF, применяя оптимизацию к выбранной части графа, называемой в алгоритме Essential Graph [9].

Ранние монокулярные методы применяют локальную оптимизацию графа. Strasdat et. al. [14] использовал интерфейс PTAM (параллельное отслеживание и построение карты), но выполнял отслеживание только на локальной карте, полученной из графа совместной видимости. В PTAM реализована серверная модель для непрерывной оптимизации двойного окна, выполняя обновление и оптимизацию графа (BA) во внутреннем окне и обновляя граф во внешнем окне ограниченного размера. Однако такой алгоритм замыкания цикла эффективен только в том случае, если размер внешнего окна достаточно велик, чтобы охватить весь цикл. В ORB-SLAM используется идея локальной карты для построения и дополнения графа поз на основе графа взаимной видимости без внешнего сервера. Другое отличие ORB-SLAM от PTAM состоит в том, что вместо использования определенных характерных точек для обнаружения петель (SURF) выполняется распознавание места в графе для совмещения тех же отслеживаемых и сопоставленных характерных точек, получая необходимую производительность метода локализации и обнаружения петель.

CD-SLAM [16] - полноценная система SLAM, включающая замыкание циклов, релокализацию при обновлении карты, поддержку работы с крупномасштабными картами и динамическими средами. Однако отсутствует общедоступная реализация и нет большого количества данных о точности и надежности данного метода.

Визуальная одометрия Song et al. [17] использует функции ORB для отслеживания и серверную часть BA с временным скользящим окном. Система ORB-SLAM является более универсальной, поскольку содержит методы глобальной релокализации, замыкания циклов и переиспользования карты. Song et al. также используют ограничение на фиксированное расстояние от камеры до земли.

Lim et. al [18] также использует те же функции для отслеживания, отображения и обнаружения замыканий что и ORB-SLAM. Однако выбор BRIEF ограничивает систему траекториями в плоскости.

Если система отслеживает только точки из последнего ключевого кадра аналогично работе Lim et. al., то карта не используется повторно при повторном



посещении (аналогично визуальной одометрии) и количество точек на карте будет линейно неограниченно возрастать. ORB-SLAM не имеет такой проблемы, при повторном добавлении точек в карту рост размера карты не происходит, это позволяет использовать метод на больших локациях.

Работа Engel et. [19], известная как LSD-SLAM, способна строить крупномасштабные полу-плотные карты (в карте содержатся графы характерных точек и часть информации с самих изображений), используя прямые методы (т. е. оптимизацию непосредственно по интенсивности пикселей изображения) вместо решения неплотной системы, то есть совместной оптимизации графа характерных точек. LSD-SLAM может работать в реальном времени, без использования графического процессора, создавая полу-плотную карту с большим количеством потенциальных приложений для робототехники. Тем не менее, качество обнаружения петель и суммарная точность определения местоположения камеры значительно ниже, чем в ORB-SLAM и PTAM.

Между прямым LSD-SLAM и дескрипторными методами (ORB-SLAM, PTAM) находится визуальная одометрия SVO Forster et al. [20]. Не требуя извлечения элементов из каждого кадра, метод может работать с высокой частотой кадров, получая впечатляющие результаты в квадрокоптерах. Обнаружение петель не выполняется, и реализация в основном предназначена для камер, смотрящих вниз.

Стратегия выбора ключевых кадров — это отдельная задача в SLAM по данным камеры, никакие существующие решения задачи SLAM не выполняют обновление и оптимизацию графа со всеми точками и всеми кадрами. Работа Strasdat et al. [14] показывает, что наиболее экономичный подход - сохранить как можно больше точек, сохраняя только не избыточные ключевые кадры. Подход PTAM основан на очень осторожной вставке ключевых кадров в набор для оптимизации, чтобы избежать чрезмерного роста вычислительной сложности. Эта ограничительная политика вставки ключевых кадров приводит к сбою отслеживания в сложных условиях исследования. У ORB-SLAM применена стратегия выживания наиболее соответствующих остальному набору кадров, что обеспечивает надежность алгоритма в сложных сценариях. Алгоритм умеет добавлять и удалять кадры из задачи оптимизации, что дает избежание дополнительных затрат.

	SLAM or VO	Pixels used	Data association	Estimation	Relocation	Loop closing	Multi Maps	Mono	Stereo	Mono IMU	Stereo IMU	Fisheye	Accuracy	Robustness	Open source
Mono-SLAM [13], [14]	SLAM	Shi Tomasi	Correlation	EKF	-	-	-	✓	-	-	-	-	Fair	Fair	[15] <sup>1</sup>
PTAM [16]–[18]	SLAM	FAST	Pyramid SSD	BA	Thumbnail	-	-	✓	-	-	-	-	Very Good	Fair	[19]
LSD-SLAM [20], [21]	SLAM	Edgelets	Direct	PG	-	FABMAP PG	-	✓	✓	-	-	-	Good	Good	[22]
SVO [23], [24]	VO	FAST+ Hi.grad.	Direct	Local BA	-	-	-	✓	✓	-	-	✓	Very Good	Very Good	[25] <sup>2</sup>
ORB-SLAM2 [2], [3]	SLAM	ORB	Descriptor	Local BA	DBoW2	DBoW2 PG+BA	-	✓	✓	-	-	-	Exc.	Very Good	[26]
DSO [27]–[29]	VO	High grad.	Direct	Local BA	-	-	-	✓	✓	-	-	✓	Good	Very Good	[30]
DSM [31]	SLAM	High grad.	Direct	Local BA	-	-	-	✓	-	-	-	-	Very Good	Very Good	[32]
MSCKF [33]–[36]	VO	Shi Tomasi	Cross correlation	EKF	-	-	-	✓	-	✓	✓	-	Fair	Very Good	[37] <sup>3</sup>
OKVIS [38], [39]	VO	BRISK	Descriptor	Local BA	-	-	-	-	-	✓	✓	-	Good	Very Good	[40]
ROVIO [41], [42]	VO	Shi Tomasi	Direct	EKF	-	-	-	-	-	✓	-	-	Good	Good	[43]
ORB-SLAM-VI [4]	SLAM	ORB	Descriptor	Local BA	DBoW2	DBoW2 PG+BA	-	✓	✓	✓	-	-	Very Good	Very Good	-
VINS-Fusion [7], [44]	VO	Shi Tomasi	KLT	Local BA	DBoW2	DBoW2 PG	✓	-	✓	✓	✓	✓	Very Good	Exc.	[45]
VI-DSO [46]	VO	High grad.	Direct	Local BA	-	-	-	-	-	✓	-	-	Very Good	Exc.	-
BASALT [47]	VO	FAST	KLT (LSSD)	Local BA	-	ORB BA	-	-	-	-	✓	-	Very Good	Exc.	[48]
Kimera [8]	VO	Shi Tomasi	KLT	Local BA	-	DBoW2 PG	-	-	-	-	✓	-	Good	Exc.	[49]
ORB-SLAM3 (ours)	SLAM	ORB	Descriptor	Local BA	DBoW2	DBoW2 PG+BA	✓	✓	✓	✓	✓	✓	Exc.	Exc.	[5]

Рисунок 0-1 Сводка наиболее репрезентативных визуальных (вверху) и визуально-инерциальных (внизу) систем SLAM в хронологическом порядке [2]

Алгоритм ORB-SLAM позволяет осуществлять одновременную локализацию и навигацию, обладает рядом преимуществ перед другими решениями SLAM, интегрирован во множество решений и проверен историей использования.

Алгоритм строит карту и ведёт навигацию по данным с внешних устройств. Алгоритм поддерживает различные возможные конфигурации физического оборудования. Изначально была предложена версия, работающая с единственной камерой. Последняя версия ORB-SLAM поддерживает совместное или раздельное использование стереокамер, камер глубины, лидаров и моно-камеры. Выбор подходящей для конкретной задачи конфигурации для данного алгоритма может основываться на уже существующие исследования.

Основой алгоритма ORB-SLAM является техника работы с картой, называемая редактирование связей (bundle adjustment, BA). Алгоритм предлагает реализацию данной техники с доступной вычислительной сложностью, что позволяет обеспечить производительность в режиме реального времени на современном оборудовании.

Обновлённая версия ORB-SLAM 3 [2] - это система SLAM, которая поддерживает визуально-инерциальную навигацию и гибридные карты. Метод может работать на одиночной камере, стерео и RGB-D камерах, а также с использованием широкоугольных камер типа «рыбий глаз».

Средняя точность ORB-SLAM 3 (3,6) см для набора данных EuRoC [3] и (9) мм для набора данных TUM-VI (сцена AR / VR), использует портативные устройства для быстрого перемещения.

Это первая плотно-связанная система визуально-инерциальной одометрии (VIO), основанная на извлечении характерных точек, которая полагается только на максимальную апостериорную оценку (включая IMU во время инициализации). Эффект от такой системы заключается в следующем: она может надёжно работать в реальном времени, а её точность улучшена в 2–5 раз по сравнению с предыдущей версией. Использование максимальной апостериорной оценки гарантирует получение лучшего решения для любого набора данных.

Второе нововведение в этой статье - создание гибридной карты на основе нового модуля локализации, которое уменьшает ошибку позиционирования.

Новый модуль локализации позволяет использовать ORB-SLAM 3 в сценах с плохими характеристиками карты.

Долгосрочная работа: при выходе из строя модуля навигации и невозможности решения для данной карты, система будет перестраивать карту и выравнивать её с исходной. ORB-SLAM 3 - первая система, которая может повторно использовать всю предыдущую информацию на всех этапах алгоритма.

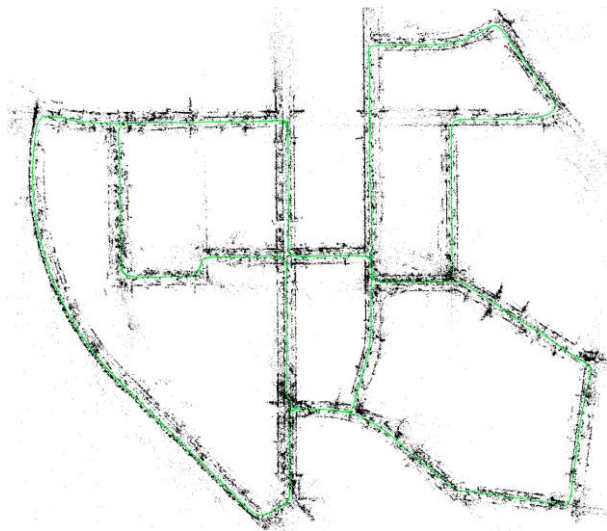
Обновленный механизм сопоставления кадров может использовать ключевые кадры, которые были собраны независимо и находятся далеко на графе связей в карте, даже если два кадра были сняты в разное время.

## **Принцип работы алгоритма**

Алгоритм SLAM строит карту в неизвестном окружении и позиционируют объект на карте по данным сенсоров.

Основным требованием к алгоритму SLAM является его работа в реальном времени. Это требование в основном существенно для автономного движения

робота в неизвестном статическом окружении или для автономного движения автомобиля в динамическом окружении. Если карту и окружение можно рассматривать как статические, то возможно значительное упрощение архитектуры алгоритма и снижения требований по производительности.



*Рисунок 0-1 Работа ORB-SLAM 2 [21] в городской среде с использованием стереокамеры, траектория и разреженная реконструкция карты с закрытием нескольких петель*

Среди различных типов датчиков камеры дешевы и предоставляют обширную информацию об окружающей среде, что позволяет надёжно и точно осуществлять позиционирование и навигацию.

Алгоритмы Visual SLAM, в которых основным датчиком является камера являются наиболее точными и универсальными решениями в локализации и навигации, при этом они зачастую бывают требовательны к производительности и оборудованию.

Распознавание места/определение места по изображению — это ключевой модуль системы SLAM для замыкания контуров (т.е. обнаружения, когда датчик возвращается в нанесённую на карту область и исправления накопленной ошибки в исследовании) и для перемещения камеры после сбоя отслеживания из-за окклюзии или агрессивного движения, или при повторной инициализации системы.

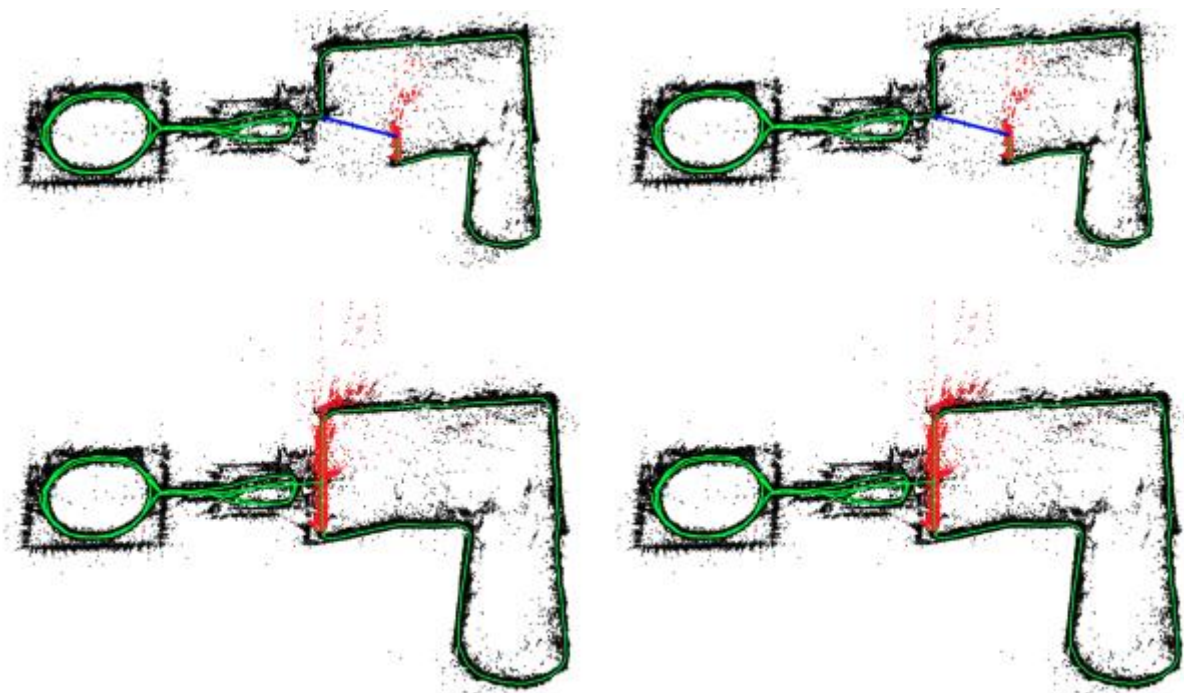


Рисунок 0-2 Работа ORB-SLAM 2, реконструкция карты до и после замыкания петли, ORB-SLAM 2 [21]

Самым большим преимуществом отображения SLAM является то, что оно позволяет сопоставить и использовать предыдущие наблюдения, которые выполняют три ассоциации данных в bundle adjustment:

- Краткосрочная ассоциация данных: сопоставление элементов карты за последние несколько секунд. Так же, как это делается в визуальной одометрии, отбрасывание кадров, которые больше не видны, приведёт к накоплению дрейфа.
- Среднесрочная ассоциация данных: сопоставление карты с небольшой совокупной ошибкой камеры. Это также можно использовать в bundle adjustment. Когда система работает на уже построенной карте, она может достичь нулевого дрейфа.
- Долгосрочная ассоциация данных: многократное распознавание сцены для сопоставления текущих наблюдений и предыдущих наблюдений, независимо от накопленной ошибки. Позволяет создать ассоциацию между кадрами даже в случае сбоя отслеживания. Долгосрочное сопоставление может использовать оптимизацию позиционного графа чтобы минимизировать ошибку каждого отдельного перемещения. Bundle

adjustment является ключевым методом для получения точности в SLAM на больших картах.

## Общая архитектура системы

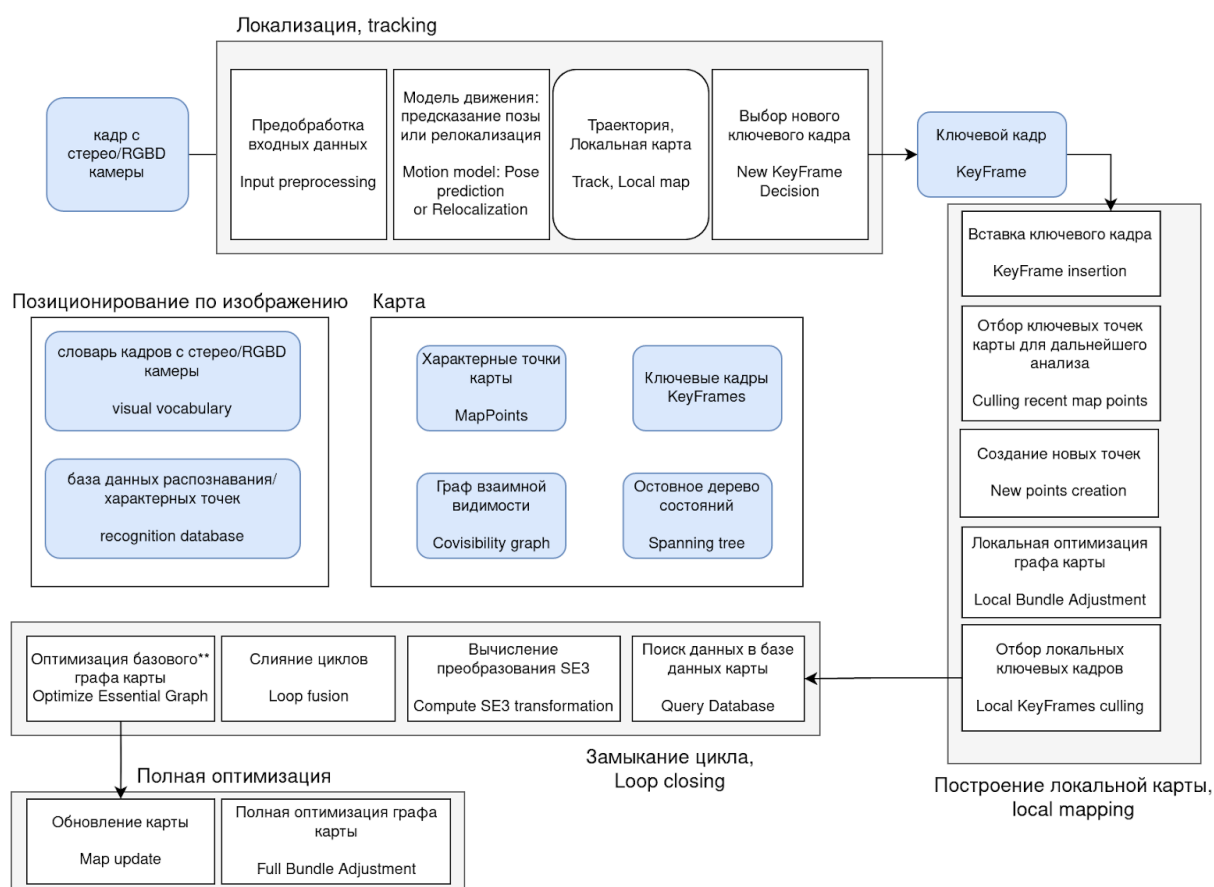


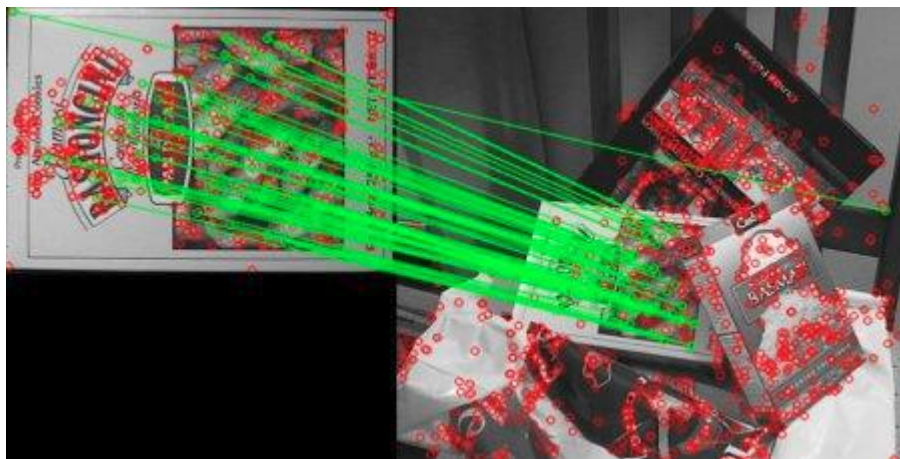
Рисунок 0-3 ORB-SLAM 2 [21], Структурная схема, потоки и модули

ORB-SLAM 2 [21] является одним из самых популярных решений для SLAM среди всех применений. Недавняя новая версия ORB-SLAM 3 ещё не получила широкого распространения, но обладает рядом изменений в архитектуре. Обзор изменений в системе будет представлен после разбора базовой версии алгоритма.

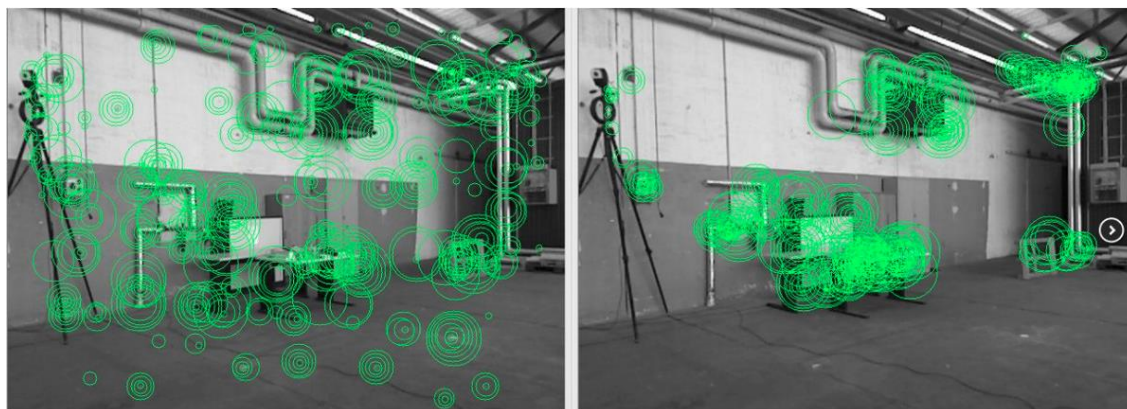
Общая архитектура системы представлена тремя параллельными процессами. Четвёртый процесс глобальной оптимизации выполняется в фоне и независимо. Процессы локализации, построения карты и замыкания карты образуют конвейер обработки данных.



## ORB



*Рисунок 0-4 ORB дескрипторы, Определение ассоциаций между характерными точками изображений*



*Рисунок 0-5 Характерные точки ORB [9] зафиксированные на изображении (слева) и используемые в радиокarte (справа)*

Алгоритм получения характерных точек по изображению — это детерминированный метод, который использует один определённый вид кодирования. Метод ORB - дескриптор, учитывающий расстояние между точками и их взаимную ориентацию. После сопоставления положения соответствующих точек между кадрами вычисляется перемещение камеры.

Из-за непостоянства освещённости и других артефактов съёмки, повторяемость определения характерных точек будет не абсолютная. Мы видим, что повторяемость характерных точек возникает для участков изображения с определённой текстурой - объект в кадре. Если объекты в кадре отсутствуют (пустые коридоры, большое пустое пространство), то навигация с данным методом будет неточной.

# Bundle adjustment

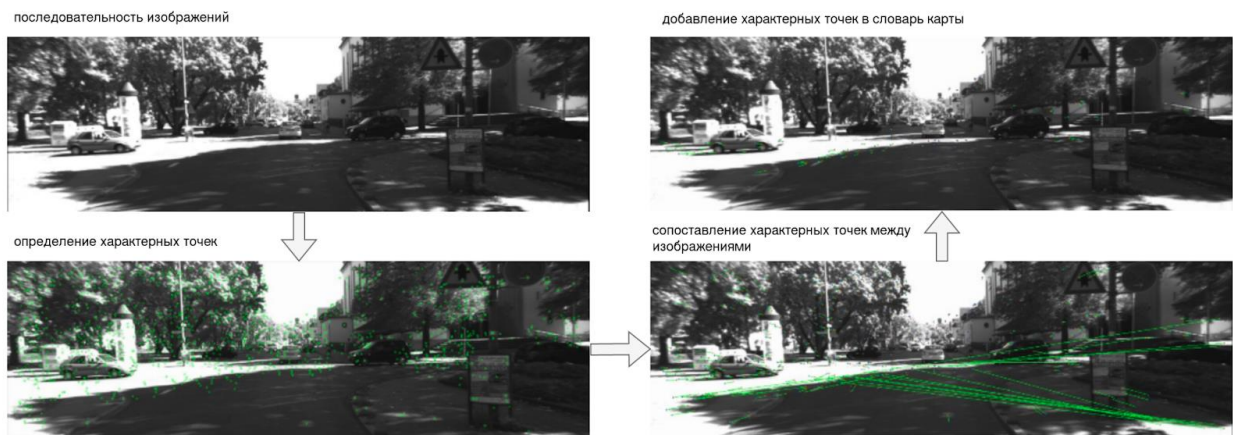


Рисунок 0-6 ORB-SLAM, последовательность определения характерных точек

Алгоритм ORB-SLAM [1], [2], [21] основан на работе с характерными точками (дескрипторами), их кодировании в связный граф и дальнейшей работе с этим связным графом.

Определение движения 2D-2D: основная матрица преобразования координат, плоская проекция

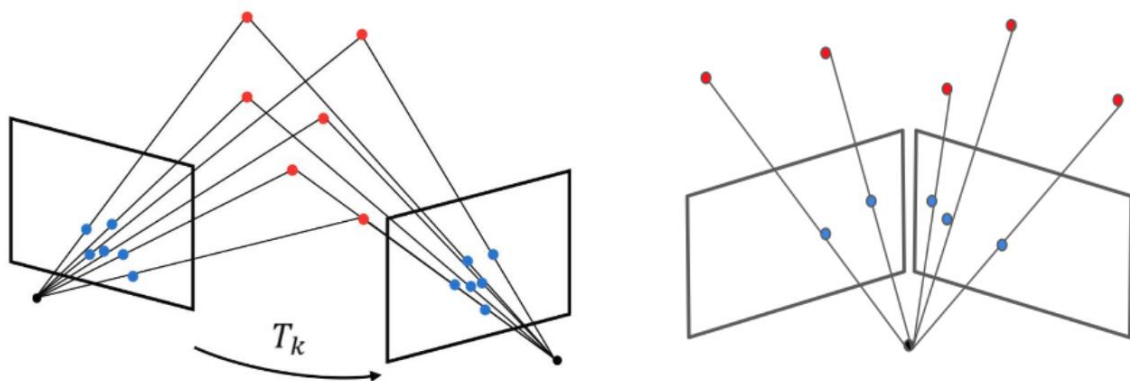


Рисунок 0-7 Реконструкция матрицы преобразования по изображениям

На изображении выше представлено возможное преобразование геометрии точек. По двум плоским изображениям (синие точки) необходимо получить положение характерных точек в трёхмерном пространстве. Не для всех точек в кадре возможно определить их перемещения.



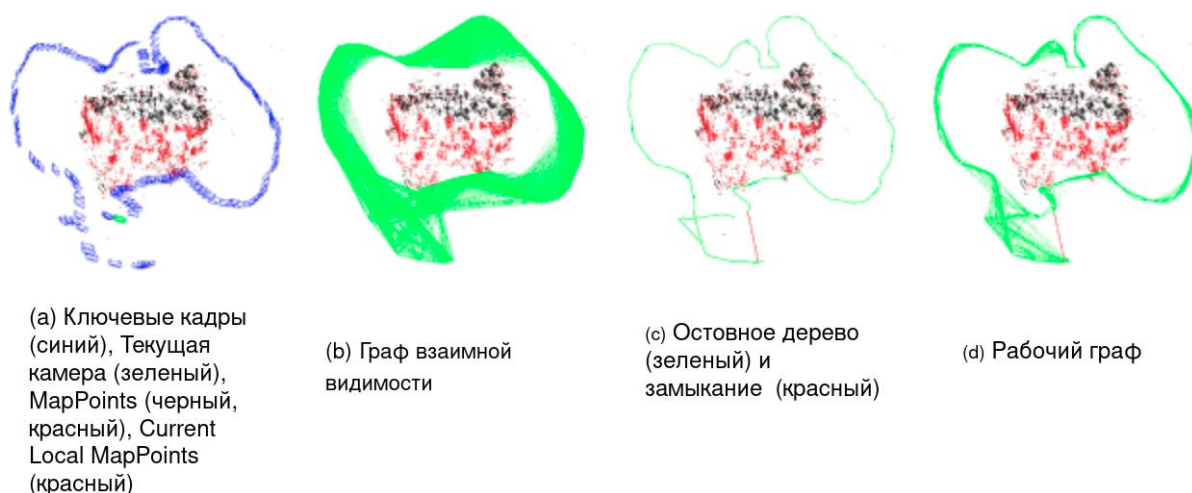


Рисунок 0-8 ORB-SLAM 2, Реконструкция представлений графов из теста TUM RGB-D [22]

На рисунке представлены основные сущности в процессе обработки ключевых кадров. После построения основных связей между характерными точками и положениями данный граф поступает на следующую стадию алгоритма.

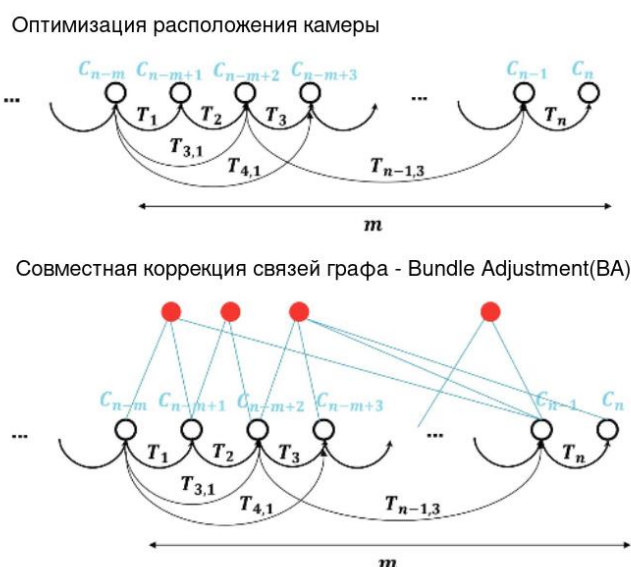


Рисунок 0-9 Факторный граф, представление модели для оптимизации. Сравнение графа позиций и совместного графа с учётом характерных точек.

В отличие от оптимизации только графа положений камеры, оптимизируется совместно граф положений камеры и характерных точек карты. Характерные точки динамически добавляются в граф, определяется их видимость из каждого положения камеры, строится граф связности, производится его локальная или глобальная оптимизация. Для методов с плотной картой, в графе дополнительно хранятся также все данные изображений с камеры (плотное облако точек).

# Автоматическая инициализация карты

## Навигация

### Замыкание графа карты

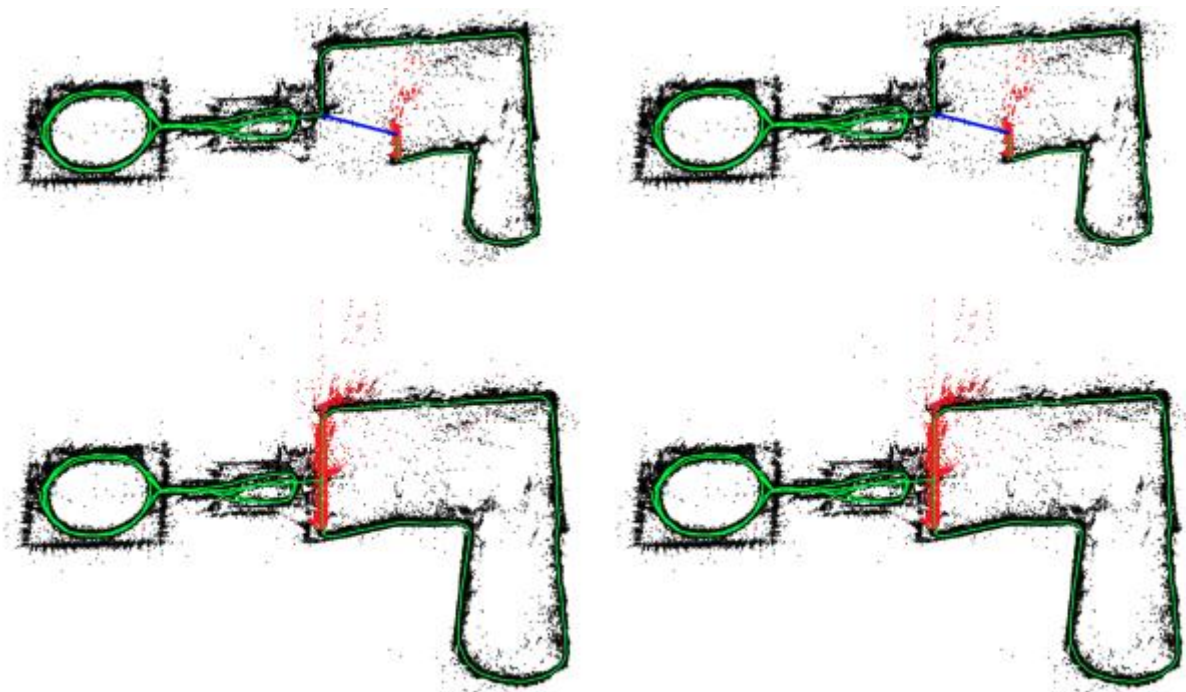


Рисунок 0-1 Карта до и после замыкания, ORB-SLAM [23]

### Использование стереокамер и камер глубины

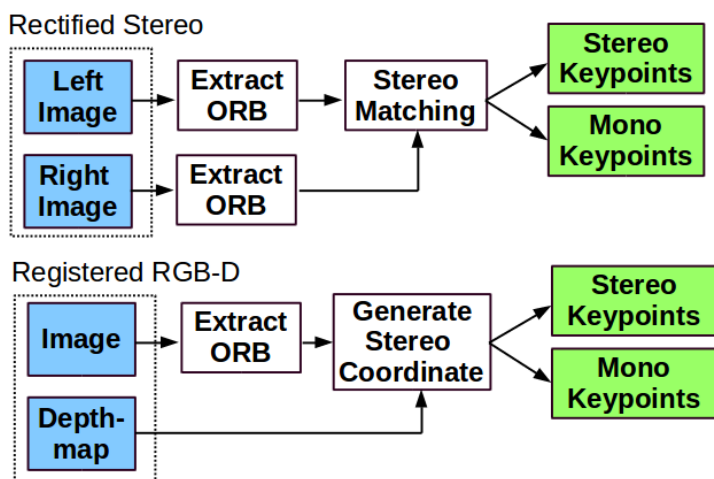


Рисунок 0-2 ORB-SLAM 2[21], Предобработка входных данных (Input preprocessing), сравнение совмещения стерео реконструкции (верх) и использования камеры глубины (низ)

Структура работы алгоритма ORB-SLAM2 представлена на Рисунок 0-3 ORB-SLAM 2 [21], Структурная схема, потоки и модули.

ORB-SLAM2 состоит из трех основных параллельных потоков:

позиционирования, локального построения карты и механизма замыкания циклов в графе карты. Отдельно выделяется четвертый поток для выполнения полного ВА после закрытия цикла, который выполняется независимо и только по требованию, при необходимости пересчета всей карты.

Поток отслеживания предварительно обрабатывает входной стереозвук или RGB-D, так что остальная часть системы работает независимо от входного датчика. Хотя это не показано на этом рисунке, ORB-SLAM2 также работает с монокулярным входом, как в [1].

## ORB SLAM 3

Алгоритм ORB-SLAM3 основывается на работах [21], [22], это полнофункциональная библиотека с открытым исходным кодом для запуска визуального, визуально-инерционного SLAM с поддержкой множества карт и сессий одновременно, с моно, стерео или RGB-D камерами, поддержкой стандартных и широкоугольных камер. Основное улучшение по сравнению с предыдущими версиями состоит в точной интеграции данных с IMU сенсоров, новом методе инициализации и слияние карт и одновременной поддержке множества сессий и карт, улучшенной отзывчивости и точности метода.

Эксперименты показывают, что ORB-SLAM3 способна эффективно использовать краткосрочные, среднесрочные, долгосрочные и мульти-картографические данные ассоциаций, достигая уровня точности, выходящего за рамки охвата существующих систем. По результатам сравнения точности, возможность использования всех типов ассоциации данных делает этот метод преобладающим над другими вариантами, такими как использование прямые методы сравнения изображений вместо извлечения характерных элементов или выполнение усреднения ключевых кадров для локальной оптимизации графа карты вместо использования внешнего набора статических ключевых кадров, как в ORB-SLAM3.

ORB-SLAM3 невозможно применять в случае низкого количества текстур в среде/окружении/сцене (гладкие стены, отсутствие объектов в кадре).

Прямые методы [19] более устойчивы к низкому количеству текстур,

но ограничены краткосрочными [23] и среднесрочными [14] ассоциациями между данными. С другой стороны, соответствующие дескрипторы функций успешно решает долговременную ассоциацию данных с несколькими картами, но кажется менее надежным для отслеживания, чем LSD-SLAM [19], который использует фотометрическую информацию полностью.

В сравнении четырех различных конфигурациях датчиков нет неоднозначностей, стерео SLAM с использованием инерциальных датчиков обеспечивает наибольшую точность и стабильность решения. Кроме того, инерционный датчик позволяет оценить позу со скоростью обновления IMU, которая на порядки выше, чем частота кадров решения SLAM.

Для приложений, где стереокамера нежелательна из-за более высокого объема данных, стоимости устройства или требований к обработке, возможно использовать монокулярно-инерциальную конфигурацию, не теряя многого в надежности и точности.

В приложениях с медленными движениями или поворотов снимающего устройства, например, движение автомобиля на ровной местности – систему может быть сложно инициализировать. В тех случаях по возможности необходимо использовать стерео SLAM или камеру глубины (RGBD SLAM).

## Общая архитектура системы

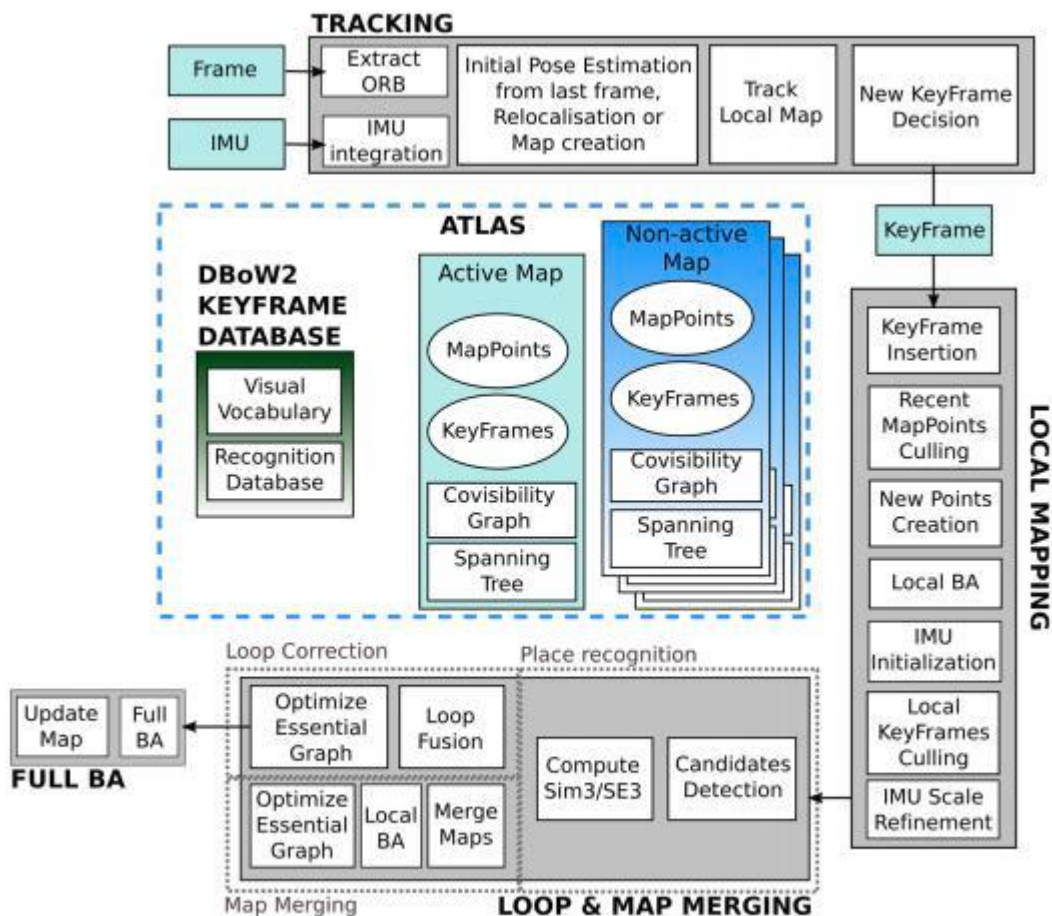


Рисунок 0-1 Основные компоненты системы ORB-SLAM 3

ORB-SLAM3 построен на ORB-SLAM2 [21] и ORB-SLAM-VI [22].

Алгоритм поддерживает работу с множеством карт и сессий. Алгоритм способен работать в чисто визуальном или визуально-инерционном режимах, с моно, стерео или RGB-D камерами, используя модель камеры «точечное отверстие» или широкоугольную «рыбий глаз».

На рисунке выше показана основная система. Большинство компонентов соответствуют компонентам ORB-SLAM2

Основные нововведения:

- База данных карт Atlas — это представление с несколькими картами, состоящее из набора независимых неактивных карт. Есть активная карта, где поток отслеживания локализует входящие кадры и постоянно оптимизируется, и расширяется за счёт новых ключевых кадров локальным потоком отображения. Система строит базу данных ключевых

кадров DBoW2, которая используется для релокализации, определения циклов и слияния карт.

- изменение архитектуры потоков обработки информации
  - Поток локализации
  - Поток локального построения карты
  - Поток определения циклов и замыканий
- Абстрактная модель камеры
  - поддержка стереорежима для разных камер
  - работа с широкоугольными объективами
- Улучшенная поддержка инерциальных сенсоров

1. Поток локализации обрабатывает информацию с датчиков и передаёт информацию. Поток определяет расположение текущего кадра относительно активной карты в реальном времени, что сводит к минимуму ошибку перепроецирования при сопоставлении различных участков карты. Он также решает, будет ли текущий кадр становится ключевым. Когда локализация не может быть решена, поток локализации пытается переместить текущий кадр на всех картах Атласа. Если релокализация произведена успешно, отслеживание возобновляется возможно при этом переключая активную карту. В противном случае через определённое время активная карта определяется как неактивная, и инициализируется новая активная карта.

3. Поток локального построения карты добавляет новые ключевые кадры и характерные точки привязки на активную карту, удаляет лишние и уточняет карту с помощью визуальной или визуально-инерционной процедуры определения связностей в графе (data association, bundle adjustment). Работа ведётся в локальном окне ключевых кадров, близких к текущему кадру.

Кроме того, параметры с датчиков IMU если они присутствуют в модели тоже инициализируются и уточняются, используя технику MAP-оценки.

4. Поток слияния петель и карт обнаруживает общие области между активной картой и всем набором карт, с частотой соответствующей обновлению ключевого кадра (если невозможно выполнить обновление в реальном времени, то следует уменьшить размер карты, не должно быть проблемой для реальных

сценариев.). Коррекция циклов и слияние карт производится на общем массиве карт. После коррекции петли полная оптимизация графа (ВА) запускается в независимом потоке для дальнейшего уточнения карты, не влияя на производительность в реальном времени.

Алгоритм работает в параллели. За счёт возможности использовать множество версий карты можно организовать серверную модель построения карты, которая будет максимально приближена к реальному времени:

1. сбор данных на клиенте, построение карты
2. передача данных на сервер, построение полной карты на сервере
3. оптимизация полной карты, передача решения пользователю

В общем случае на сервере можно запустить любой метод (продублировать запуски), но ORB-SLAM 3 в целом упрощает распределенную интеграцию системы.

## Использование инерциальных данных в модели ORB-SLAM 3[2]

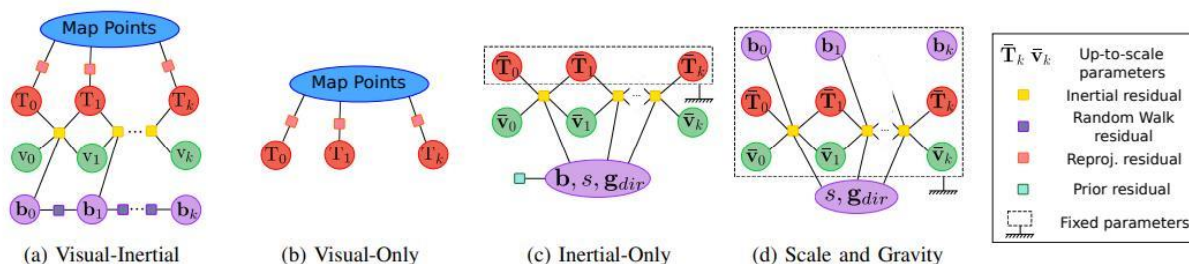


Рисунок 0-2 Конфигурация факторного графа для разных вариантов систем: (а) визуально-инерциальный, (b) визуальный, (c) инерциальный, (d) ограничения на масштаб и гравитацию

В визуально-инерционном режиме скорость тела и отклонения ускорений оцениваются как включение инерционных невязок в оптимизацию.

По инерционным невязкам строится факторный граф, потом производится его полная оптимизация. Факторный граф — это представление всех ограничений между состояниями системы в виде единого графа, как правило граф



представляется в двухдольном виде, при учете дополнительных ограничений, например инерциальных данных – граф будет представлен в трехдольном виде.

Если не рассматривать теорию факторных графов более подробно, то логика такого решения заключается в введении большего количества ограничений в модель. Увеличение количества информации уменьшает неопределённость системы, и можно проводить более точную оптимизацию. Если инерциальные измерения шумные, то жёсткость ограничений не позволит получить решение оптимизации на графе, и алгоритм либо ослабит ограничения, либо удалит их (ORB SLAM bundle adjustment). Использование внешних сенсоров помимо камеры (инерциальных и иных датчиков) позволяет улучшить точность решения, особенно для ситуации, когда камера не может опознать точки привязки на изображении (нет близких объектов, белая стена, засветка объектива, движение в толпе людей, отсутствие света вообще).

## **Общая характеристика системы**

## **Производительность алгоритма**



# Список литературы

- [1] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System,” 2021.
- [2] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM,” *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021, doi: 10.1109/TRO.2021.3075644.
- [3] M. Burri *et al.*, “The EuRoC micro aerial vehicle datasets,” *Int. J. Robot. Res.*, vol. 35, no. 10, pp. 1157–1163, 2016, doi: 10.1177/0278364915620033.
- [4] G. Klein and D. Murray, “Parallel Tracking and Mapping for Small AR Workspaces,” in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nara, Japan, Nov. 2007, pp. 1–10. doi: 10.1109/ISMAR.2007.4538852.
- [5] D. Galvez-López and J. D. Tardos, “Bags of Binary Words for Fast Place Recognition in Image Sequences,” *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1188–1197, Oct. 2012, doi: 10.1109/TRO.2012.2197158.
- [6] M. Cummins and P. Newman, “Appearance-only SLAM at large scale with FAB-MAP 2.0,” *Int. J. Robot. Res.*, vol. 30, no. 9, pp. 1100–1123, Aug. 2011, doi: 10.1177/0278364910385483.
- [7] D. Hutchison *et al.*, “BRIEF: Binary Robust Independent Elementary Features,” in *Computer Vision – ECCV 2010*, vol. 6314, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 778–792. doi: 10.1007/978-3-642-15561-1\_56.
- [8] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded Up Robust Features,” in *Computer Vision – ECCV 2006*, vol. 3951, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417. doi: 10.1007/11744023\_32.
- [9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in *2011 International Conference on Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 2564–2571. doi: 10.1109/ICCV.2011.6126544.
- [10] R. Mur-Artal and J. D. Tardos, “Fast relocalisation and loop closing in keyframe-based SLAM,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, May 2014, pp. 846–853. doi: 10.1109/ICRA.2014.6906953.
- [11] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, “MonoSLAM: Real-Time Single Camera SLAM,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007, doi: 10.1109/TPAMI.2007.1049.
- [12] J. Civera, A. J. Davison, and J. M. M. Montiel, “Inverse Depth Parametrization for Monocular SLAM,” *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 932–945, 2008, doi: 10.1109/TRO.2008.2003276.

- [13] E. Eade and T. Drummond, “Scalable Monocular SLAM,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, 2006, vol. 1, pp. 469–476. doi: 10.1109/CVPR.2006.263.
- [14] H. Strasdat, J. M. M. Montiel, and A. J. Davison, “Visual SLAM: Why filter?,” *Image Vis. Comput.*, vol. 30, no. 2, pp. 65–77, 2012, doi: <https://doi.org/10.1016/j.imavis.2012.02.009>.
- [15] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, “Real Time Localization and 3D Reconstruction,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, New York, NY, USA, 2006, pp. 363–370. doi: 10.1109/CVPR.2006.236.
- [16] K. Pirker, M. Rüther, and H. Bischof, “CD SLAM - continuous localization and mapping in a dynamic world,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 3990–3997. doi: 10.1109/IROS.2011.6094588.
- [17] S. Song, M. Chandraker, and C. C. Guest, “Parallel, real-time monocular visual odometry,” in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 4698–4705. doi: 10.1109/ICRA.2013.6631246.
- [18] H. Lim, J. Lim, and H. J. Kim, “Real-time 6-DOF monocular visual SLAM in a large-scale environment,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 1532–1539. doi: 10.1109/ICRA.2014.6907055.
- [19] J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: Large-Scale Direct Monocular SLAM,” in *Computer Vision – ECCV 2014*, vol. 8690, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 834–849. doi: 10.1007/978-3-319-10605-2\_54.
- [20] C. Forster, M. Pizzoli, and D. Scaramuzza, “SVO: Fast semi-direct monocular visual odometry,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 15–22. doi: 10.1109/ICRA.2014.6906584.
- [21] R.-D. Camas, R. ur-Artal, Juan, and D. Tardos, “ORB SLAM 2 : an Open-Source SLAM System for Monocular, Stereo and.”
- [22] R. Elvira, J. Tardós, and J. Montiel, “ORB SLAM-Atlas: a robust and accurate multi-map system,” Aug. 2019.
- [23] “The reconstruction of a plane surface from two perspective projections,” *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 227, no. 1249, pp. 399–410, May 1986, doi: 10.1098/rspb.1986.0030.