

Подпишитесь на DeepL Pro для редактирования данного документа.  
Дополнительную информацию можно найти на странице [www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document).

Совместный визуальный SLAM с использованием сжатого обмена характеристиками

Доминик Ван Опденбошанд Эккехард Штайнбах

***Аннотация В* области робототехники совместная одновременная ло-кализация и картографирование (SLAM) все еще остается сложной проблемой. Исследование неизвестных крупномасштабных сред выигрывает от разделения работы между несколькими агентами, возможно, обладающими различными способностями, такими как воздушные или наземные транспортные средства. В этом письме мы специально рассматриваем вопрос эффективности обмена ви- суальной информацией в совместной визуальной системе SLAM. Для эффективного обмена данными мы расширяем схему сжатия локальных бинарных признаков двумя дополнительными режимами, обеспечивающими поддержку локальных признаков с дополнительной информацией о глубине и режим межвидового кодирования, использующий пространственные отношения между видами стереокамеры. Для демонстрации схемы кодирования мы используем централизованную архитектуру системы, основанную на ORB-SLAM2, где агенты с ограниченным энергопотреблением извлекают локальные бинарные признаки и отправляют сжатую версию по сети более мощному агенту, который способен параллельно выполнять несколько визуальных экземпляров SLAM. Мы используем информацию от других агентов, обнаруживая перекрытие между уже нанесенными на карту областями и последующее объединение карт. Таким образом, участники вносят свой вклад в совместное представление и пользуются общей информацией о карте. Мы демонстрируем снижение скорости передачи данных на 70,8% при использовании ком- прессии признаков и снижение абсолютной ошибки траектории на 53,7% при использовании стратегии совместного картирования с тремя агентами на известном наборе данных KITTI. В интересах сообщества мы предоставляем публичную версию исходного кода.**

***Индексные термины - многороботные* системы, SLAM, локализация, картографирование, навигация на основе визуальных данных.**

1. ВВЕДЕНИЕ

**M**

Любые роботизированные задачи выполнимы только при наличии глубоких знаний об окружающей среде, текущем положении

робота и расположение возможных препятствий. С появлением доступных, универсальных и компактных визуальных датчиков визуальный SLAM стал очень практичным подходом к решению проблемы SLAM. Хотя сегодня доступны небольшие системы SLAM, фокус исследований смещается в сторону крупномасштабного картографирования. Использование не только одного, но и нескольких агентов в таких сценариях повторно

Рукопись получена 24 июля 2018 г.; принята 19 октября 2018 г. Дата

публикация 31 октября 2018 года; дата текущей версии 15 ноября 2018 года. Данное письмо было рекомендовано к публикации младшим редактором Х. Алонсо-Мора и редактором Н. Я. Чонг после оценки замечаний рецензентов. Работа выполнена при поддержке космического агентства Германского аэрокосмического центра на средства Федерального министерства экономики и технологий (FKZ: 50NA1515). *(Автор-корреспондент: Доминик Ван Опденбош).*

Авторы работают на кафедре медиатехнологий Мюнхенского технического университета, Мюнхен 80333, Германия (e-mail: dominik.van-opdenbosch@tum.de; eckehard.steinbach@tum.de).

Это письмо содержит дополнительный материал, доступный для загрузки по адресу http:// ieeexplore.ieee.org, предоставленный авторами. Дополнительный материал содержит видео, демонстрирующее результаты предложенного подхода к сжатию локальных признаков. Размер этого материала составляет 47,5 МБ.

Цифровой идентификатор объекта 10.1109/LRA.2018.2878920

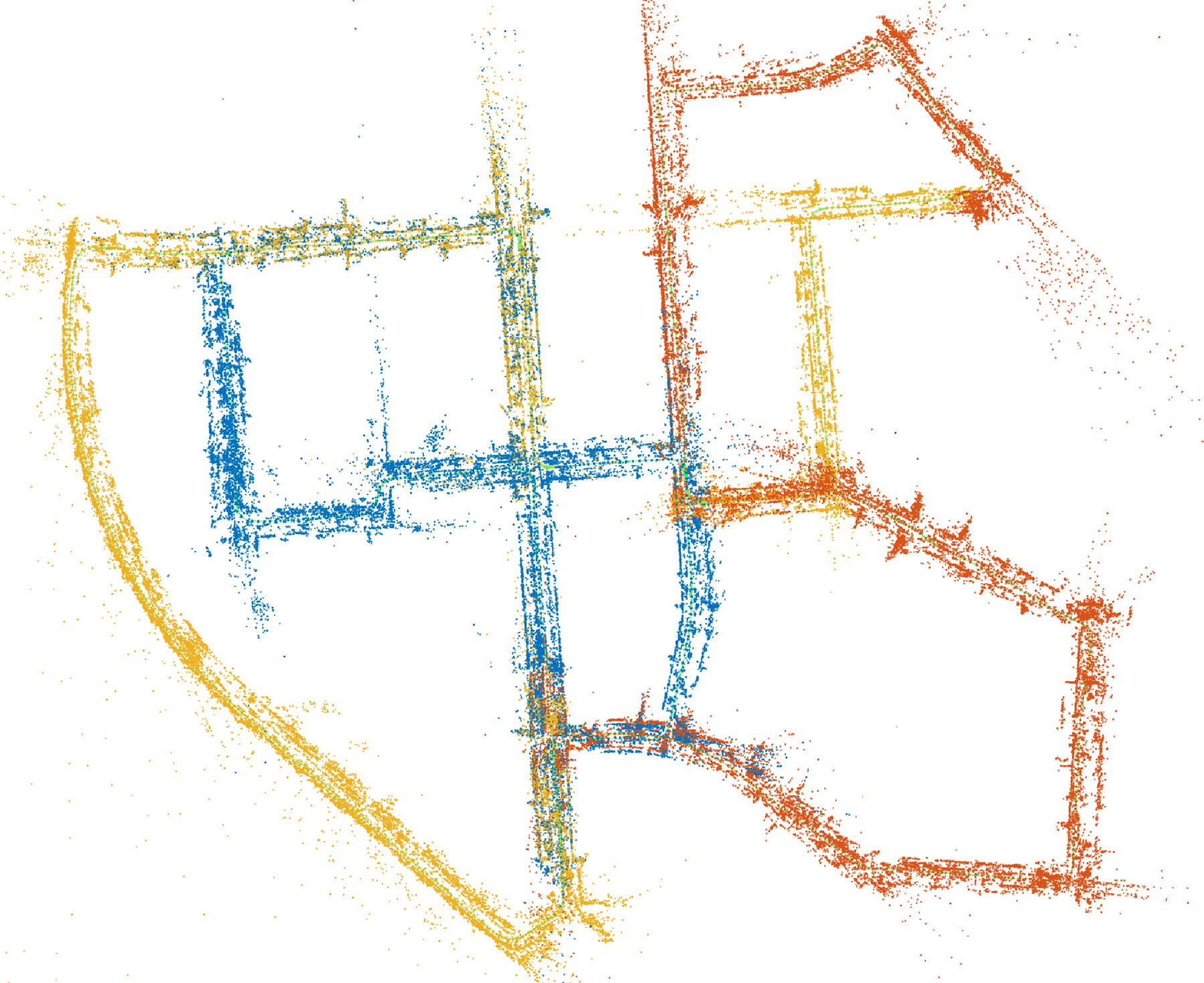


Рис. 1. Глобально согласованная карта после объединения индивидуальных карт, полученных от трех клиентов, работающих с последовательностью KITTI 00. Цвет указывает, какой агент создал точки карты.

сокращает требуемое время отображения и добавляет отказоустойчивость в случае отказа агента. Поэтому совместная визуальная SLAM недавно привлекла внимание научного сообщества. В этом письме мы рассматриваем сценарий, в котором небольшая команда совместно исследует окружающую среду. Требования к энергоэффективности зависят от типа агента. В то время как для микроаэромобилей (MAV) энергоэффективность является ключевым вопросом для долгосрочной работы, роверы обычно не так ограничены в потреблении энергии и могут выполнять более сложные вычислительные задачи за счет дополнительных вычислительных возможностей и необходимых батарей. Чтобы использовать это различие, мы используем централизованную архитектуру системы, в которой только извлечение визуальных признаков выполняется на клиентах с ограниченным энергопотреблением, а сжатые визуальные сигналы, необходимые для выполнения визуального SLAM метрического масштаба, передаются на центральный сервер, переносимый мощным агентом. Этот сервер запускает отдельные экземпляры визуального SLAM для каждого клиента, выполняет построение карты и при необходимости передает информацию о положении и карте обратно соответствующему клиенту. Кроме того, сервер способен обнаруживать перекрытия между картами визуального SLAM и выполнять слияние карт. После слияния подключенные агенты вносят вклад в общую карту, как показано на рис. 1. В качестве примера, эта глобальная карта может быть использована для централизованного управления группой исследователей. Для этого в данной работе представлены следующие материалы:

1. Мы расширили существующую схему кодирования локальных бинарных признаков [1], включив в нее дополнительную информацию о глубине.

2377-3766 © 2018 IEEE. Разрешается личное использование, но для переиздания/распространения требуется разрешение IEEE.

Дополнительную информацию см. на сайте <http://www.ieee.org/publications>standards/publications/rights/index.html.

с RGB-D камер или стереофонического сопоставления признаков для обеспечения визуального SLAM в метрическом масштабе.

1. Мы дополняем систему кодирования режимом стереокодирования признаков, способным передавать признаки с обеих точек зрения, используя пространственные корреляции.
2. Для демонстрации кодирования в сценарии совместной работы мы реализовали централизованную визуальную систему SLAM.

Остальная часть письма организована следующим образом: В разделе II мы обсуждаем смежные работы по визуальному SLAM, совместному ви- суальному SLAM и локальному сжатию признаков. В разделе III мы представляем архитектуру нашей системы. В разделе IV подробно описана схема кодирования, включая предложенное кодирование глубины и стерео. В разделе V представлен подход совместного отображения. Мы демонстрируем эффективность системы в разделе VI и делаем вывод в разделе VII.

1. СООТВЕТСТВУЮЩАЯ РАБОТА
2. *Визуальный SLAM*

Большинство визуальных SLAM-систем можно разделить на две категории по двум ключевым свойствам [2]. Они могут быть прямыми или косвенными, плотными или разреженными методами. В то время как косвенные методы используют промежуточное представление, такое как локальные особенности или шаблоны, основанные на участках изображения, прямые методы оперируют интенсивностью пикселей. Плотные методы пытаются восстановить полную 3D-модель окружающей среды на основе всех пикселей, в то время как разреженные методы используют только подмножество точек изображения для создания 3D-точек. Представителями прямых и плотных методов являются плотное слежение и картирование (DTAM) [3] и крупномасштабный прямой SLAM (LSD SLAM) [4]. Оба метода пытаются минимизировать фотометрическую ошибку между кадрами на основе интенсивности пикселей для оценки движения, что требует больших вычислительных затрат. Поэтому некоторые методы сочетают прямой и разреженный подход, например, прямая разреженная одометрия (DSO) [2]. Он обеспечивает только визуальную одометрию без замыкания контура, что очень важно для коррекции дрейфа в крупномасштабных сценариях картографирования. Для замыкания контура приходится полагаться на методы, основанные на признаках [5]. Другие подходы, такие как быстрая полупрямая монокулярная визуальная одометрия (SVO) [6], используют прямые методы для оценки движения и косвенные методы для построения карты. Прямые методы обычно требуют наличия исходной интенсивности пикселей, что требует большого количества обмена данными в сценариях совместной работы. Косвенные и разреженные подходы привлекли внимание к совместным приложениям благодаря абстракции изображений и разреженному представлению карты. Примерами косвенных и разреженных систем являются MonoSLAM [7] и PTAM [8]. Недавним представителем является ORB- SLAM2 [9], [10], которая представляет собой комплексную систему SLAM, способную в реальном времени выполнять параллельное слежение, отображение, замыкание цикла и повторную локализацию.

1. *Совместный визуальный SLAM*

Имея под рукой точные системы визуального SLAM, способные работать в реальном времени, подходы совместного картирования вызывают все больший интерес. Zou *и др.* [11] предложили систему CoSLAM, использующую несколько камер одновременно. Они группируют камеры в соответствии с их перекрытием обзора и предлагают совместную работу.

картографировать окружающую среду, используя как внутрикамерный, так и межкамерный картографический пинг. Форстер *и др.* [12] предложили монокулярную SLAM-систему для совместного картирования с использованием БПЛА. Риазуэло *и др.* [13] предложили облачный подход для совместного слежения и отображения. Они передают дорогостоящую задачу оптимизации более мощным вычислительным узлам в облаке и сохраняют только легкое отслеживание на локальном устройстве. Шмук *и др.* [14] предложили совместную систему SLAM для нескольких MAV на основе ORB- SLAM2. В их концепции каждый MAV запускает легкую ви- суальную систему SLAM, а информация собирается на центральном сервере, где выполняются задачи, требующие больших вычислительных затрат, например, оптимизация карты. Позже авторы расширили свою систему до визуально-инерциальной совместной системы SLAM [15]. Что касается обмена картами, то в [16], [17] были предложены подходы для синхронизации карт. Чтобы ограничить неограниченный рост карт и уменьшить объем данных при обмене статической картографической информацией, в нескольких работах рассматривался вопрос о разрежении карты, сохраняя только ограниченное количество полезных точек карты [18]-[20] или оптимизируя целевой размер карты с точки зрения количества необходимых битов [21]. Полностью децентрализованные системы были предложены Cieslewski *и др.* [22] на основе предыдущих работ по распределенному обнаружению перекрытий [23], [24]. Их система позволяет обмениваться квантованными локальными признаками в виде визуальных индексов слов, как это было предложено в [25]. Однако большинство распределенных систем игнорируют аспект эффективного сжатия "на лету" для обмена визуальной информацией.

1. *Сжатие характеристик*

Занимаясь сжатием признаков, Бароффио *и др.* представили концепции гибридного кодирования видео как для реально-значных [26], так и для бинарных дескрипторов признаков [27]. Они предложили различные режимы кодирования, такие как внутри- и межкадровое кодирование, чтобы использовать как зависимости между элементами дескриптора признаков, так и временную корреляцию между последовательными кадрами. Кроме того, Бонди *и др.* [28] предложили межвидовое кодирование. Редонди *и др.* [29] представили анализ производительности сети визуальных датчиков с использованием принципов межвидового кодирования. В предыдущей работе мы рассмотрели совместное кодирование локальных и глобальных характеристик изображения, используя зависимости между де-скрипторами и их соответствующим представлением Bag-of-Words [30]. Некоторые из вышеупомянутых парадигм легли в основу стандарта MPEG CDVS (Compact Descriptors for Visual Search) [31], который нацелен на эффективные дескрипторы для задач визуального поиска и извлечения изображений, а также был использован в кон- тексте визуального SLAM [32]. Совсем недавно были предприняты некоторые усилия по созданию эффективных дескрипторов для анализа видео для стандарта MPEG CDVA (Compact Descriptors for Video Analysis) [33], включая возможность использования временных зависимостей и дескрипторов с глубоким обучением признаков. Визуальное сжатие дескрипторов с использованием квантования произведений было применено к визуальным картам в [34]. Обзор различных компактных представлений признаков приведен в [35].

1. ОБЗОР СИСТЕМЫ

Наша система основана на нашей предыдущей работе [1], в которой мы объединили концепцию кодирования бинарных признаков с монокулярностью.

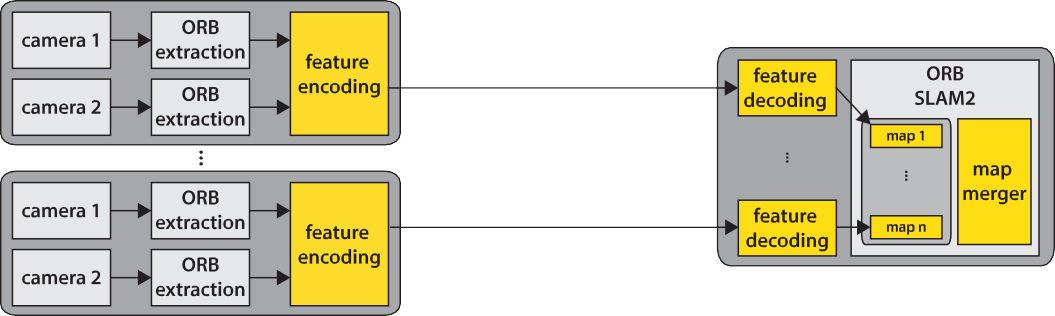
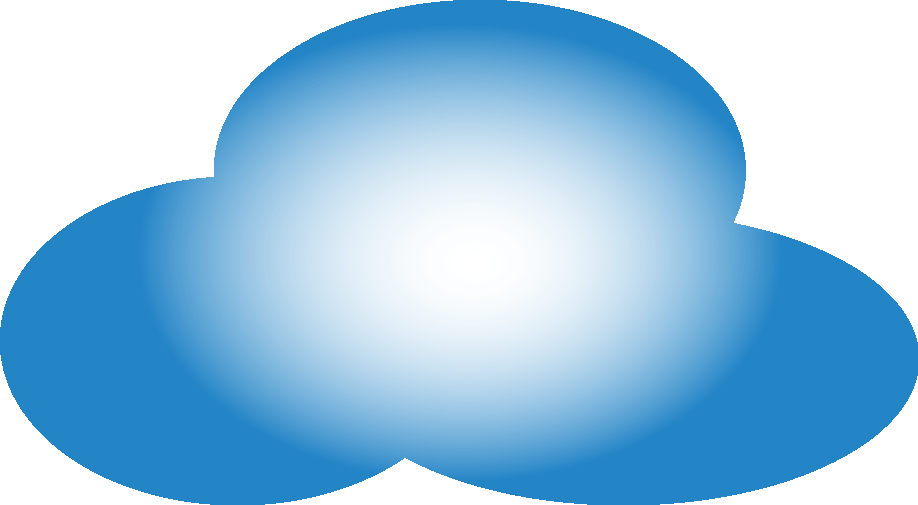


Рис. 2. Обзор системы: Клиентские агенты выполняют только извлечение визуальных признаков и кодирование признаков. Сервер декодирует потоки признаков и создает индивидуальные карты для всех клиентов. Модуль объединения карт срабатывает при создании новых ключевых кадров и объединяет перекрывающиеся карты в единое представление.

версия ORB-SLAM2. В этом письме мы предлагаем три дополнения к этому подходу, чтобы получить совместную визуальную SLAM-систему метрического масштаба на основе стереоинформации, как показано на рис. 2. Во-первых, мы используем централизованную архитектуру, в которой клиенты получают изображения, извлекают и кодируют признаки ORB [36] и отправляют их по сети на сервер, на котором работает несколько экземпляров декодирования признаков, обрабатывающих все входящие потоки данных. После декодирования признаки передаются соответствующему экземпляру визуального SLAM. Каждый экземпляр визуального SLAM состоит из модуля отслеживания, отображения и замыкания цикла, работающего на карте, назначенной данному агенту. Для того чтобы использовать аспект совместного отображения, мы реализовали модуль объединения карт. Подобно подходу к объединению карт, предложенному Шмуком *и другими* [14], мы постоянно проверяем новые ключевые кадры, вставленные в любую карту, на соответствие в любой другой карте в отдельном фоновом потоке. В случае обнаружения перекрытия карты объединяются, и отныне подключенные агенты вносят вклад в общую согласованную карту.

Для SLAM, использующего только визуальные подсказки, датчик глубины или калиброванная многокамерная установка необходимы для обеспечения метрической информации о масштабе. Следовательно, следующий вклад - это расширение монокулярного кодирования признаков для включения значений глубины, полученных, например, непосредственно с RGB-D камеры или путем сопоставления стерео признаков. Вместо передачи представления глубины с плавающей запятой мы используем неравномерное скалярное квантование, адаптированное к типичным значениям глубины, используемым в визуальном SLAM.

В качестве последнего вклада мы добавляем схему кодирования стереовида, использующую пространственную корреляцию между обоими видами в стереокамере. Для эффективного кодирования визуальной информации мы используем основную концепцию, известную из кодирования признаков, извлеченных из общих сетей визуальных датчиков [28], где мы передаем только различия между сопоставленными признаками в обоих ракурсах. Благодаря предложенному вкладу мы решаем проблему эффективного сжатия в системах совместного визуального SLAM.

1. КОДИРОВАНИЕ СЕМЬИ
2. *Рамки кодирования*

Мы кратко перескажем основную схему двоичного кодирования признаков, представленную в нашей предыдущей работе [1], а затем подробно расскажем о ее расширении для кодирования глубины и стерео. Входными данными используемой визуальной SLAM-системы являются признаки ORB, которые состоят из двух элементов

части: Первая часть описывает местоположение, ориентацию и масштаб, где была обнаружена локальная особенность, обычно называемая ключевой точкой. Мы следуем обозначениям из [1] и обозначаем набор свойств ключевой точки для признака *i* из изображения *n* как **k** *n,i*= [*x, y, σ, θ*], где *x* и *y* - положение ключевой точки в пиксельных координатах, σ обозначает уровень масштаба, а θ - ориентацию ключевой точки. Вторая часть локального признака - дескриптор, который является результатом набора парных тестов пикселей, содержащихся в двоичной строке **d** [0*,* 1]*D* длиной *D* = 256.

В нашей предыдущей работе [30] мы ввели режим внутрикадрового кодирования, который использует зависимости между дескрипторами и их соответствующим представлением Bag-of-Words. Мы используем визуальный словарь в качестве общего знания, вычисляем вектор разности между дескриптором и ближайшим визуальным словом и передаем только этот остаточный вектор, содержащий бинарные разности вместе с соответствующим индексом визуального слова. Вектор различий содержит в основном нули, что в сочетании с энтропийным кодированием (в нашем случае арифметическим кодированием) приводит к снижению битрейта. Помимо преимущества использования визуального словаря для минимизации энтропии передаваемых данных, ORB-SLAM2 может напрямую использовать визуальные слова в представлении Bag-of-Words для эффективного сопоставления признаков, обнаружения замыкания цикла и повторной локализации. Для того чтобы использовать временную корреляцию, мы добавили режим межкадрового кодирования [1], аналогичный подходу, предложенному Бароффио *и др.* [27], где передаются только различия ключевой точки и дескрипторов по отношению к признаку, содержащемуся в последнем кадре, наряду с идентификатором. Кроме того, мы добавили режим пропуска, который способен копировать опорные признаки из предыдущих кадров без дополнительного кодирования, тем самым снижая битрейт при отсутствии движения в видео. За подробностями мы отсылаем читателя к [1].

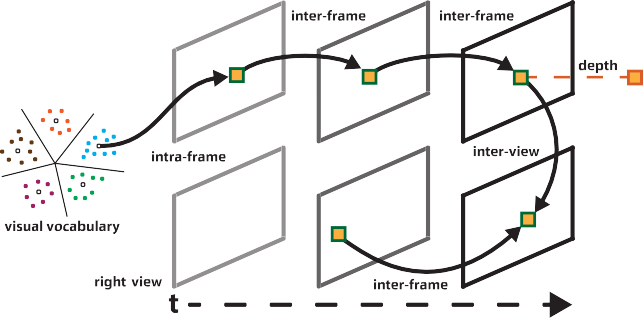
∈

Начиная с этих режимов кодирования, мы завершаем работу над каркасом, добавляя режим кодирования дополнительной информации о глубине и режим кодирования стереохарактеристик. На протяжении всего письма мы обозначаем режимы кодирования как *m I, P, S, M* , где I - внутрикадровое кодирование, P - прогнозируемое межкадровое кодирование, S - режим пропуска, а M - новый многоракурсный режим стереокодирования. Мы используем

∈ {}

+D для обозначения наличия дополнительных данных о глубине. Различные концепции показаны на рис. 3. Для принятия решения о режиме мы рассчитываем скорость для всех режимов кодирования для каждого признака (I, P, S для вида слева и дополнительно M для вида справа) и выбираем режим с наименьшим битрейтом.

эртили *R* *M,kpt*даны как



*n,i*

*RM* = *RM,ref* + RM*,des* (*j*) + *RM,kpt,* (1)

*n,i*

*n,i*

*n,i*

*n,i*

где *j* обозначает эталонный признак. После того, как мы определили набор признаков-кандидатов с левого вида вблизи эпиполярной линии, мы проверяем признак *j*∗, который имеет минимальное расстояние Хэмминга *H* между текущим дескриптором **d** *n,i*и эталонным признаком.

дескриптор **d** *M*

*j*

Рис. 3. Режимы прогнозирования, использующие либо визуальный словарь (внутрикадровый), временные (межкадровые) или пространственные корреляции (межвидовые) между локальными признаками.

*j*∗ = arg min *j*(*H*(**d** *n,i,* **d** *M*))*.* (2)

* 1. *Кодирование эталона:* В предположении равномерной вероятности, энтропия для кодирования эталонного признака равна

*j*

В качестве альтернативы можно передавать квантованную информацию о глубине.

*R M,ref*= *log* (*N*

)*,* (3)

*n,i*

2l *,n*

1. *Глубинное кодирование*

Кодирование глубины может быть использовано для эффективного добавления информации о глубине к каждой особенности левого вида. Глубина определяется либо непосредственно RGB-D камерой, либо с помощью стереофонического сопоставления. В последнем случае мы предполагаем наличие предварительно откалиброванной системы камер с

где *Nl,n* обозначает количество признаков в левом представлении.

1. *Кодирование дескриптора:* Для кодирования остатка мы вычисляем вектор разности между текущим и опорным дескриптором в терминах двоичной операции XOR **r** *M*=

*n,i*

**d***n,i* **d** *M*. При этом *hM* (*j*∗ ) обозначает расстояние Хэмминга между текущим и эталонным признаком, задаваемое следующим образом

*j ∗*

*n,i*

⊕

известная внутренняя и внешняя калибровка камеры. Мы определяем

*hMj* ∗) = *H*(**d**

*,* **d** *M*)*,* (4)

*n,i*

*n,ij ∗*

совпадения признаков, расположенных в том же масштабе, путем поиска вдоль эпиполярной линии. Для учета несовершенной калибровки камеры мы допускаем отклонение в два пикселя от эпиполярной линии в пространственно-масштабном представлении. Вместо того чтобы напрямую передавать сигнал в представление с плавающей запятой, требующее 32 бита на значение глубины,

мы предлагаем использовать *неравномерную* схему *квантования*, обученную

на траектории образца, чтобы минимизировать ошибку реконструкции. Из-за окклюзии не каждый объект сопровождается значением глубины, поэтому для каждого объекта мы сначала посылаем один бит, указывающий на наличие значения глубины. Если глубина доступна, мы сигнализируем об уровне реконструкции с помощью фиксированного числа битов *ND*. В этой схеме оценка глубины выполняется на стороне клиента, и все признаки с правого ракурса могут быть опущены.

1. *Межракурсное кодирование*

В общем случае, при использовании визуального SLAM-приложения, которое использует визуальную информацию обоих видов, необходимо передавать все признаки с обеих камер. Для этого мы предлагаем использовать *пространственную корреляцию* между обоими видами, находя соответствия между характеристиками, и передавать только различия между дескрипторами вместе с идентификатором опорной характеристики и недостающей информацией для восстановления информации о ключевой точке. Из-за окклюзий это не всегда возможно, а визуальные объекты, расположенные близко к системе камер, могут привести к значительно отличающимся дескрипторам признаков, поскольку угол обзора существенно отличается. Мы разработали нашу систему таким образом, что для каждого объекта в правом ракурсе можно выбрать режимы кодирования, введенные для кодирования признаков в одном ракурсе. Для установления соответствия мы используем то же стереосоответствие, что и для глубины.

минимальное количество битов может быть рассчитано с помощью функции двоичной энтропии следующим образом

*RM,des* (*j*∗) = - (*D* - *hM (j* ∗)) - *log*2(*p M*)

*n,i*

*n,i*

0

*- hM (j*∗) - *log (*1 - *p M*)*,* (5)

*n,i*

2

0

где *D* - длина дескриптора, а *pM* - вероятность того, что любая запись остаточного вектора равна нулю для режима межпросмотрового кодирования. Эта нижняя граница достигается с помощью арифметических действий

0

кодирование, применяемое к каждой записи остаточного вектора. В этой схеме выбранный признак *j*∗ является не только наилучшим стереосовпадением, но и приводит к минимальным затратам на кодирование.

1. *Кодирование ключевых точек:* Для ключевых точек мы передаем квантованную ориентацию и положение ключевой точки. Мы ограничиваем сопоставление объектов одним и тем же масштабом σ*n,i*, что позволяет напрямую использовать информацию о масштабе эталонного объекта. Мы кодируем положение *x* аналогично подходу внутреннего кодирования [1], где мы масштабируем положение ключевой точки на уровень масштабного пространства, в котором был извлечен признак. В таком представлении координаты ключевой точки располагаются в целочисленных позициях, что позволяет кодировать положение ключевой точки без потерь. Для *координаты y* мы просто передаем разницу по отношению к положению опорной ключевой точки в общем масштабном пространстве. Мы допускаем отклонение в 2 пикселя от эпиполярной линии в соответствующем масштабно-пространственном представлении, требующем log 2(5) бит для передачи сигнала. Ориентация квантуется в *Nθˆ* бинов. Предполагая равномерное распределение, общие затраты на ключевые точки можно записать как

±

*R M,kpt*= log 2(*width*(σ *n,i*)) + log 2(5) + log 2(*N* ˆ)*,* (6)

*n,iθ*

оценка стоимости. Затраты на кодирование *R M*

*n,i*

в кусочках для интер-вью

где *width*(σ*n,i* ) обозначает ширину изображения в соответствующем...

кодирование признака *i* из изображения *n* состоит из отдельных затрат на кодирование индекса эталонного признака *RM,ref*, затрат на остаточный вектор, содержащий различия между сопоставленными дескрипторами *RM,des*, и затрат на кодирование ключевой точки.

*n,i*

*n,i*

σ *n,i*. Квантованные ориентация и *x-компонента* также могут быть закодированы дифференцированно, а отдельные вероятности могут быть оценены, но это дает лишь незначительный выигрыш по сравнению с вычислительными затратами.

1. СОВМЕСТНОЕ КАРТИРОВАНИЕ
2. *Слияние карт*

Объединение карт облегчается путем адаптации существующих методов закрытия петель в ORB-SLAM2 для работы с несколькими картами. Сначала выявляются возможные кандидаты на замыкание цикла путем оценки визуального сходства между текущим ключевым кадром и возможными целевыми ключевыми кадрами, содержащимися во всех доступных картах, используя представление Bag- of-Words. После получения начального набора кандидатов рассчитывается преобразование сходства между точками карты, содержащимися в совпадающих ключевых кадрах, по методу Хорна [37]. Этот шаг служит для геометрической проверки совпадений, чтобы избежать ложного слияния карт. После успешной проверки две карты объединяются путем добавления всех ключевых кадров и точек карты из исходной карты в целевую карту. Следующим шагом является выравнивание ключевых кадров и точек карты в местах наложения с использованием оцененного преобразования сходства. Это позволяет выявить больше дублирующих точек карты и установить связи между точками карты из исходной карты и ключевыми кадрами в целевой карте и наоборот. После этого шага выполняется оптимизация эссенциального графа, который представляет собой охватывающее дерево между ключевыми кадрами. Для этого мы соединяем графики сущностей обеих карт. Затем в фоновом режиме асинхронно запускается глобальная настройка связки для оптимизации общей карты.

1. *Совместное картирование*

Во время слияния карт поток отслеживания и поток отображения исходной карты присоединяется к целевой карте. Нить замыкания контура исходной карты останавливается. Отныне поток закрытия циклов целевой карты отвечает только за обнаружение циклов в объединенной карте. Чтобы избежать одновременного доступа и сохранить согласованность, мы используем схему общей блокировки, при которой только один поток отображения может добавлять новые точки карты к определенному ключевому кадру. Кроме того, только один поток может одновременно обновлять ключевой кадр и положение точки карты.

1. ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА

Для согласованности и подтверждения обобщения мы использовали те же параметры, что и в нашей монокулярной оценке [1], и обучали вероятности *pM* для стереокодирования на тех же учебных данных, что и в предыдущей версии, а именно на наборе данных EuRoC [38] Machine Hall sequences. Он состоит из изображений, полученных с борта микроаэромобиля с разрешением 752 480 пикселей при частоте 20 Гц в промышленной среде. Мы оценили нашу схему на хорошо известном наборе данных KITTI [39]. В нем представлены различные последовательности, собранные автомобилем в городской среде с помощью калиброванной стереокамеры. Неискаженные изображения имеют разрешение 1241 376 пикселей, снятые с частотой 10 Гц. Набор данных KITTI представляет собой сложный сценарий как для кодирования признаков из-за больших расстояний между кадрами, так и для картографического аспекта из-за пространственной протяженности охваченной области. Чтобы получить характеристики квантования для значений глубины, мы использовали се-квенты Машинного зала и KITTI для обучения кодовой книги для значений глубины, чтобы учесть различные диапазоны глубины в сценариях внутри и снаружи помещений.

0

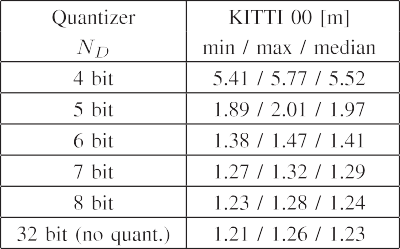
×

×

ТАБЛИЦА I

СРАВНЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ РАЗЛИЧНЫХ КВАНТОВАТЕЛЕЙ ГЛУБИНЫ НА ПОСЛЕДОВАТЕЛЬНОСТИ KITTI 00 С ТОЧКИ ЗРЕНИЯ АБСОЛЮТНОЙ ОШИБКИ ТРАЕКТОРИИ

ПОЛУЧЕННЫЕ В РЕЗУЛЬТАТЕ ПЯТИ ПРОГОНОВ



1. *Кодирование характеристик*

Во-первых, мы даем некоторое представление о кодировании значений глубины. Мы представляем результаты отображения в терминах *абсолютной ошибки траектории* (ATE) [40] в зависимости от количества битов для квантования значений глубины для последовательности KITTI 00 в таблице I, используя настройки по умолчанию ORB-SLAM2. При *N* *D*= 8 бит на значение глубины и с дополнительным битом, сигнализирующим о том, было ли значение глубины оценено, мы достигли производительности наравне с исходными значениями глубины. Поскольку ORB-SLAM2 использует информацию о глубине для воспроизведения оригинальных стереокоординат [10], результат неквантованной глубины и стереокодирования аналогичен.

Во втором эксперименте мы представляем результаты предложенной схемы сжатия признаков для последовательности KITTI 00 на рис. 4. Для внутреннего кодирования мы тратим около 225,4 бита для левого и правого вида без информации о глубине. Интеркодирование ограничено использованием только прошлого кадра для предсказания и требует

150,3 бита и 145,8 бита для левого и правого вида соответственно. Добавление большего количества опорных кадров уменьшит битрейт за счет увеличения сложности. Значения глубины для вида слева могут быть оценены для 47,9% признаков с помощью внутреннего кодирования и 64,5% признаков с помощью интеррежима. В среднем, это добавляет 4,8 бита к режиму внутреннего кодирования и 6,2 бита к режиму интеркодирования в левом ракурсе. Интуитивно понятно, что признаки, закодированные в режиме внутреннего кодирования, считаются более стабильными и не располагаются на границе изображения, чтобы их можно было отследить на протяжении нескольких кадров или наблюдать с соседних ракурсов, что и обуславливает разное процентное соотношение. Режим пропуска, который кодирует только индекс опорного признака в прошлом кадре, занимает 13,3 бита, но используется только для менее чем 1% признаков в этой последовательности. Все режимы кодирования включают накладные расходы в два бита на сигнализацию режима. Предлагаемый режим стереокодирования использует 11,3 бита для передачи сигнала, 144,6 бита для остаточной информации и 16,9 бита для разницы ключевых точек. Включая затраты на передачу сигнала, на стереокодированный признак уходит в среднем около 174,8 бита. Мы также показываем долю признаков, закодированных с помощью определенного режима кодирования. Из-за большого межкадрового расстояния многие признаки кодируются с использованием режима внутрикадрового кодирования. Доля внутрикадрового кодирования для такого набора данных с высокой частотой кадров, как EuRoC, значительно выше, что делает кодирование более эффективным.

Мы измерили время кодирования на двух различных системах. Во-первых, мы показываем медианные результаты измерения времени на Intel Core i7-7700 с

3,6 ГГц, измеренные на последовательности KITTI 00 для монокуляра

+ глубинное кодирование для двух профилей кодирования в таблице II. Хотя,

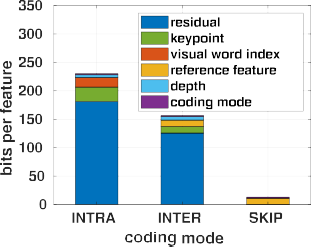
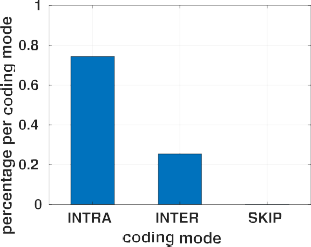
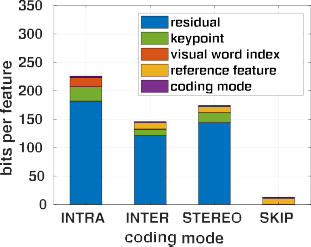
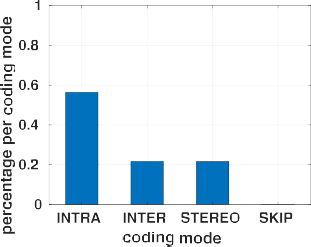
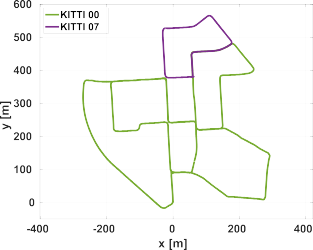
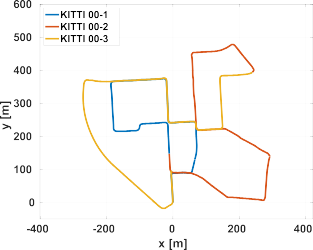
   



Рис. 4. Сравнение количества битов, необходимых для внутривидового кодирования, временного предсказания, режима пропуска и межвидового стерео предсказания, как средние значения для каждого признака для последовательности KITTI 00. Левый вид включает биты, необходимые для кодирования значения глубины. Мы показываем долю признаков, использующих тот или иной режим кодирования для левого (b) и правого (d) вида соответственно.

ТАБЛИЦА II

СРЕДНЕЕ ВРЕМЯ И КОЛИЧЕСТВО БИТ НА ПРИЗНАК ДЛЯ МОНОКУЛЯРНОГО + ГЛУБИННОГО КОДИРОВАНИЯ И ДЕКОДИРОВАНИЯ, ИЗМЕРЕННОЕ НА ПОСЛЕДОВАТЕЛЬНОСТИ KITTI 00 С

РАЗЛИЧНЫЕ КОНФИГУРАЦИИ РЕЖИМОВ

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |





Рис. 5. Обзор двух последовательностей из набора данных KITTI, использованных для экспериментальной оценки, показанных в общей системе координат.



ТАБЛИЦА III

МЕДИАННЫЕ ТАЙМИНГИ И СРЕДНИЕ БИТЫ НА ФУНКЦИЮ ДЛЯ СТЕРЕО КОДИРОВАНИЯ И

ДЕКОДИРОВАНИЕ ИЗМЕРЕНО НА ПОСЛЕДОВАТЕЛЬНОСТИ KITTI 00 С

РАЗЛИЧНЫЕ КОНФИГУРАЦИИ РЕЖИМОВ

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  | | |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  | | |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

в этой схеме мы передаем только 2k признаков из левого вида, для оценки глубины нужны признаки, извлеченные из обоих видов. В первой колонке мы разрешаем только внутрирежимный режим плюс информация о глубине (I+D). Этот режим не использует никаких временных корреляций в видеопотоке. Это позволяет нам начать декодирование с любого кадра без каких-либо предварительных знаний о предыдущих кадрах, что также известно как *случайный доступ*. Более того, это самый быстрый режим кодирования и может быть использован в сценариях с *малой задержкой,* но с недостатком увеличения среднего битрейта. Далее разрешается временное предсказание и пропуск признаков (I+P+S+D), что приводит к увеличению времени кодирования, но также к уменьшению среднего количества бит на признак. В таблице III мы оценили режим стереокодирования для трех различных профилей кодирования. В первой колонке мы

разрешить только внутривидовой режим (I) для обоих видов. Затем мы добавляем предложенный нами режим межвидового кодирования (I+M), что приводит к небольшому увеличению времени кодирования, но за счет уменьшения количества битов на функцию. Далее мы добавляем временное предсказание и режим пропуска (I+P+S+M), что добавляет значительное время обработки, но также обеспечивает значительное снижение битрейта.

Мы также провели несколько тестов на NVIDIA Jetson TX2 в режиме MAX-P ARM, используя 7,2 Вт при частоте 2,0 ГГц, включенных в таблицы II и III. Мы не проводили никакой специальной оптимизации для встроенной платформы. Видно, что параллельное извлечение признаков ORB для двух изображений создает препятствие для работы ORB-SLAM2 или нашей системы на встроенной платформе. Однако система способна кодировать 2 тыс. изображений, включая оценку глубины, примерно за 26,5 мс. Для многих приложений компьютерного зрения или сценариев визуального SLAM может быть достаточно меньшего количества функций, что снижает время кодирования. Кроме того, существуют оптимизированные локальные двоичные признаки для встроенных устройств [41], но их применение выходит за рамки данного письма.

1. *Совместное картирование*

Чтобы продемонстрировать аспект совместного отображения, мы провели эксперимент с двумя агентами, работающими с последовательностями KITTI 00 и 07, как показано на рис. 5(a). Мы симультанно запустили обе последовательности и оценили результат с точки зрения кодирования признаков и точности SLAM. Следующие эксперименты проводились на виртуальной машине, работающей в облаке с 16 vCPU на базе Intel Xeon Platinum 8124M с частотой 3,00 ГГц. Для получения базовой оценки точности, достижимой с помощью

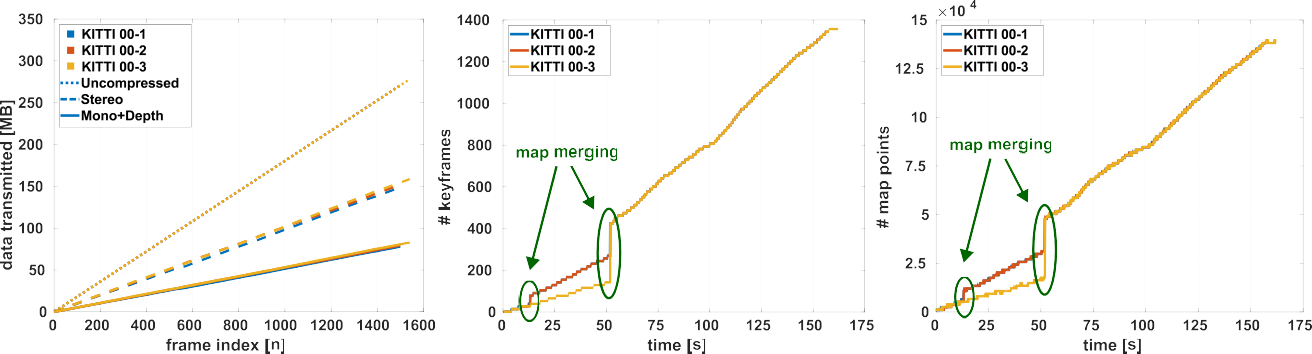




Рис. 6. Анализ кумулятивной скорости передачи данных, измеренной на кодере на одного агента, где стерео использует режимы I+P+S+M, а моно+глубина - режимы I+P+S+D. Рядом показана эволюция ключевых кадров и точек карты системы SLAM с использованием трех агентов на последовательности KITTI 00.

ТАБЛИЦА IV

СРАВНЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ ПОСЛЕДОВАТЕЛЬНОСТЕЙ KITTI 00 И 07 ПРИ ИСПОЛЬЗОВАНИИ ПРЕДЛОЖЕННОГО СОВМЕСТНОГО ПОДХОДА В ТЕРМИНАХ

АБСОЛЮТНАЯ ОШИБКА ТРАЕКТОРИИ, ПОЛУЧЕННАЯ В РЕЗУЛЬТАТЕ ПЯТИ ПРОГОНОВ

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

ТАБЛИЦА V

СРАВНЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ НА ПОСЛЕДОВАТЕЛЬНОСТИ KITTI 00 ПРИ ИСПОЛЬЗОВАНИИ ПРЕДЛОЖЕННОГО СОВМЕСТНОГО ПОДХОДА С ТОЧКИ ЗРЕНИЯ АБСОЛЮТНОЙ ОШИБКИ ТРАЕКТОРИИ, ПОЛУЧЕННОЙ В РЕЗУЛЬТАТЕ ПЯТИ ЗАПУСКОВ

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

В нашей установке мы сначала провели эксперименты для обеих последовательностей с деактивированным объединением карт, которое обозначается как автономное. После этого мы активировали модуль объединения карт, позволяющий учитывать картографическую информацию от обоих агентов. В таблице IV мы приводим результаты в виде оценки ATE для каждой траектории в отдельности, полученные в пяти отдельных прогонах. Результаты показывают улучшение на 5,7% для KITTI 00 и на 16,9% для последовательности KITTI 07. В конце процесса отображения получается полностью связанная карта, охватывающая обе последовательности.

В нашем следующем эксперименте мы использовали три агента совместно на последовательности KITTI 00. Для этого мы разделили последовательность на три части, как показано на рис. 5(b). Мы одновременно запустили все последовательности и в определенный момент обнаружили значительное перекрытие между картами, после чего карты были последовательно объединены в одну карту. Результаты по ATE, полученные в результате пяти отдельных запусков, приведены в таблице V. Первый агент достигает немного меньшей ошибки при работе в автономном режиме, в то время как результаты других последовательностей явно улучшаются при использовании функции объединения карт. Например, ошибка третьего агента может быть уменьшена с 3,11 м

до 1,44 м, что на 53,7% меньше. Улучшение происходит частично за счет повторного использования картографированных областей, а также за счет дополнительных контуров, обнаруженных в объединенных картах, что позволяет исправить смещение. На рис. 6(a) показан объем данных, которыми обменивается каждый клиентский агент с сервером. Мы сравнили суммарную скорость передачи данных, необходимую для обмена характеристиками при использовании передачи без сжатия, стереокодирования характеристик и кодирования глубины для всех агентов. Достижимое снижение битрейта составляет около 44,1% при использовании стереокодирования и 70,8% при использовании моно+глубинного кодирования при условии 360 бит на несжатый признак, что является суммой 256 бит для дескриптора,

3 32 бита для передачи сигналов *x*, *y* и ориентации признака и дополнительные 8 бит для информации о масштабе. Мы показываем эволюцию карт во времени в терминах ключевых кадров на рис. 6(b) и в терминах точек карты на рис. 6(c). Мы выделяем случаи объединения карт, когда существующие ключевые кадры и точки карты объединяются в общее представление. Используя этот подход, мы сократили время построения полной и последовательной карты последовательности KITTI 00, как показано на рис. 1, до трети длины исходной последовательности при одновременном использовании трех агентов. Кроме того, мы измерили среднее время, необходимое для отслеживания отдельных кадров - 28,8 мс для агента, работающего на KITTI 00-3. Это соответствует времени, о котором сообщают авторы ORB-SLAM2 [10] без выделения признаков ORB, которое передается клиенту. В нашей централизованной системе количество клиентов ограничено вычислительной мощностью центрального сервера. Для больших команд требуется расширение системы для поддержки нескольких серверов, способных обмениваться информацией.

×

1. ВЫВОДЫ

В этом письме мы представляем комплексную схему для коллаборативного визуального SLAM для небольшой команды. Для этого мы расширяем нашу предыдущую систему двоичного кодирования признаков методом глубинного и межракурсного кодирования. Эта система может использоваться автономно во многих сценариях применения и не ограничивается визуальным SLAM. Кроме того, мы реализовали схему совместного отображения на основе ORB-SLAM2 , в которой несколько карт SLAM строятся параллельно и могут быть объединены при обнаружении перекрытия между картами. Мы объединили оба подхода в системную архитектуру, где требовательный к вычислениям визуальный SLAM

Система работает на центральном вычислительном узле, и только ви- суальные сигналы извлекаются и сжимаются на клиентских агентах. Мы оценили наш подход с точки зрения эффективности кодирования, времени и абсолютной ошибки траектории на наборе данных KITTI, показав значительное снижение требуемой скорости передачи данных до 70,8% и снижение ATE на 53,7% при использовании трех агентов и совместного отображения по сравнению с автономным отображением. Таким образом, наша система закрывает пробел, удовлетворяя потребность в экономии данных в совместных визуальных системах SLAM. Сжатие характеристик также может быть использовано при обмене ключевыми кадрами из локального визуального одома с центральным сервером. Интеграция такой облегченной одоме- этрии в клиент для локального контура управления оставлена для будущей работы. Дальнейшая оценка и исходный код доступны в Интернете. 1

ССЫЛКИ

1. D. Ван Опденбош, М. Оелш, А. Гарсеа и Э. Стейнбах, "Выбор и сжатие локальных бинарных признаков для удаленного визуального SLAM", в *Proc. IEEE Int. Conf. Robot. Automat.* , 2018, pp. 7270-7277.
2. J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Anal. Mach. Intell.* , vol. 40, no. 3, pp. 611-625, Mar. 2018.
3. R. А. Ньюкомб, С. Дж. Лавгроув и А. Дж. Дэвисон, "DTAM: плотное отслеживание и отображение в реальном времени", в *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 2320-2327.
4. J. Engel, T. Scho¨ps, and D. Cremers, "LSD-SLAM: Direct monocular SLAM," in *Proc. Eur. Conf. Comput. Vision*, vol. 8690, 2014, pp. 834- 849.
5. X. Гао, Р. Ванг, Н. Деммель и Д. Кремерс, "LDSO: Прямая разреженная одометрия с замыканием цикла", в *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* , 2018, pp. 2198-2204.
6. C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monoc- ular visual odometry," in *Proc. IEEE Int. Conf. Robot. Automat.* , 2014, pp. 15-22.
7. A. J. Davison, I. D. Reid, N. D. Molton и O. Stasse, "MonoSLAM: однокамерный SLAM в реальном времени", *IEEE Trans. Pattern Anal. Mach. Intell.* , том 29, № 6, стр. 1052-1067, июнь 2007.
8. Г. Клейн и Д. Мюррей, "Параллельное отслеживание и отображение для небольших рабочих пространств AR", в *Proc. IEEE ACM Int. Symp. Mixed Augmented Reality*, 2007, pp. 225-234.
9. R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: versa- tile and accurate monocular SLAM system," *IEEE Trans. Robot.* , vol. 31, no. 5, pp. 1147-1163, Oct. 2015.
10. R. Мур-Артал и Дж. Д. Тардос, "ORB-SLAM2: Система SLAM с открытым исходным кодом для монокулярных, стерео и RGB-D камер", *IEEE Trans. Robot.* , vol. 33, no. 5, pp. 1255-1262, Oct. 2017.
11. D. Зоу и П. Тан, "CoSLAM: совместный визуальный SLAM в динамических средах", *IEEE Trans. Pattern Anal. Mach. Intell.* , том 35, № 2, стр. 354-366, февраль 2013 г.
12. C. Forster, S. Lynen, L. Kneip, and D. Scaramuzza, "Collaborative monoc- ular SLAM with multiple micro aerial vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* , 2013, pp. 3963-3970.
13. L. Riazuelo, J. Civera, and J. M. Montiel, "C2TAM: облачная структура для совместного отслеживания и отображения", *Robot. Auton. Syst.* , vol. 62, no. 4, pp. 401-413, 2014.
14. P. Шмук и М. Чли, "Совместная монокулярная SLAM для нескольких БПЛА", в *Proc. IEEE Int. Conf. Robot. Automat.* , 2017, pp. 3863-3870.
15. M. Karrer, P. Schmuck, and M. Chli, "CVI-SLAM-collaborative visual- inertial SLAM," *IEEE Robot. Autom. Lett.* , vol. 3, no. 4, pp. 2762-2769, Oct. 2018.
16. T. Cieslewski, S. Lynen, M. Dymczyk, S. Magnenat, and R. Siegwart, "Map API-Scalable decentralized map building for robots," in *Proc. IEEE Int. Conf. Robot. Automat.* , 2015, pp. 6241-6247.
17. M. Гэдд и П. Ньюман, "Проверьте мою карту: Контроль версий для визуальной локализации всего флота", в *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* , 2016, pp. 5729-5736.
18. H. S. Park, Y. Wang, E. Nurvitadhi, J. C. Hoe, Y. Sheikh, and M. Chen, "3D point cloud reduction using mixed-integer quadratic programming," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2013, pp. 229-236.

1https://rebrand.ly/ral18

1. M. Dymczyk, S. Lynen, T. Cieslewski, M. Bosse, R. Siegwart, and P. Furgale, "The gist of maps-Summarizing experience for lifelong localiza- tion," in *Proc. IEEE Int. Conf. Robot. Automat.* , 2015, pp. 2767-2773.
2. M. Dymczyk, S. Lynen, M. Bosse, and R. Siegwart, "Keep it brief: Масштабируемое создание сжатых карт локализации", в *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* , 2015, pp. 2536-2542.
3. D. Ван Опденбош, Т. Айкут, М. Оэльш, Н. Альт и Э. Штайнбах, "Эффективное сжатие карты для совместного визуального SLAM", в *Proc. IEEE Winter Conf. Appl. Comput. Vision*, 2018, pp. 992-1000.
4. T. Cieslewski, S. Choudhary, and D. Scaramuzza, "Data-efficient decen- tralized visual SLAM," in *Proc. IEEE Int. Conf. Robot. Autom.* , 2018, pp. 2466-2473.
5. T. Cieslewski и D. Scaramuzza, "Efficient decentralized visual place recognition using a distributed inverted index," *IEEE Robot. Autom. Lett.* , vol. 2, no. 2, pp. 640-647, Apr. 2017.
6. T. Cieslewski и D. Scaramuzza, "Эффективное децентрализованное визуальное распознавание места по дескрипторам полного изображения", в *Proc. IEEE Int. Symp. Multi- Robot Multi-Agent Syst.* , 2017, pp. 78-82.
7. D. Tardioli, E. Montijano, and A. R. Mosteo, "Visual data association in narrow-bandwidth networks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* , 2015, pp. 2572-2577.
8. L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi, and S. Tubaro, "Coding visual features extracted from video sequences," *IEEE Trans. Image Process.* , vol. 23, no. 5, pp. 2262-2276, May 2014.
9. L. Baroffio, A. Canclini, M. Cesana, A. Redondi, M. Tagliasacchi, and S. Tubaro, "Coding local and global binary visual features extracted from video sequences," *IEEE Trans. Image Process.* , vol. 24, no. 11, pp. 3546- 3560, Nov. 2015.
10. L. Bondi, L. Baroffio, M. Cesana, A. Redondi, and M. Tagliasac- chi, "Multi-view coding of local features in visual sensor net- works," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2015, pp. 1-6.
11. A. E. Redondi, L. Baroffio, M. Cesana, and M. Tagliasacchi, "Multi-view coding and routing of local features in visual sensor networks," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.* , 2016, pp. 1-9.
12. D. Ван Опденбош, М. Оелш, А. Гарсеа и Э. Стейнбах, "Совместная схема сжатия для локальных бинарных дескрипторов признаков и их соответствующего представления в виде мешка слов", в *Proc. IEEE Conf. Visual Com- mun. Image Process.* , 2017, pp. 1-4.
13. L.-Y. Дуан *и др.*, "Обзор стандарта MPEG-CDVS", *IEEE Trans. Image Process.* , vol. 25, no. 1, pp. 179-194, Jan. 2016.
14. P. P. De Gusmao, S. Rosa, E. Magli, S. Lepsoy, and G. Francini, "Loop detection in robotic navigation using MPEG CDVS," in *Proc. IEEE 17th Int. Workshop Multimedia Signal Process.* , 2015, pp. 1-6.
15. L.-Y. Duan *et al.* , "Компактные дескрипторы для анализа видео: Появляющийся стандарт MPEG", 2017, arXiv:1704.08141.
16. S. Линен, Т. Саттлер, М. Боссе, Дж. Хеш, М. Поллефейс и Р. Зиг- варт, "Убирайтесь из моей лаборатории: Крупномасштабная визуально-инерциальная ло-кализация в реальном времени", в *Proc. Robot: Sci. Syst. XI*, 2015. [Online]. Доступно: <http://www.roboticsproceedings.org/rss11/p37.html>
17. L. Baroffio, A. Redondi, M. Tagliasacchi, and S. Tubaro, "A survey on compact features for visual content analysis," *APSIPA Trans. Signal Inf. Process.* , vol. 5, 2016. [Online]. Доступно: [https:](http://www.cambridge.org/)//www.cambridge.org/ core/journals/apsipa-transactions-on-signal-and-information-processing/ article/survey-on-compact-features-for-visual-content-analysis/A80C3E 52CE3D92443DEF40EA44B82DE7
18. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: эффективная альтернатива SIFT или SURF," in *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 2564-2571.
19. B. К. П. Хорн, "Закрытое решение абсолютной ориентации с использованием единичных кватернионов", *J. Opt. Soc. Amer. A*, vol. 4, no. 4, p. 629, 1987.
20. M. Burri *et al.* , "The EuRoC micro aerial vehicle datasets," *Int. J. Robot. Res.* , vol. 35, no. 10, pp. 1157-1163, 2016.
21. A. Гейгер, П. Ленц и Р. Уртасун, "Готовы ли мы к автономному вождению? Набор эталонных тестов для технического зрения KITTI", в *Proc. IEEE Conf. Comput. Vision Pattern Recognit.* , 2012, pp. 3354-3361.
22. J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A bench- mark for the evaluation of RGB-D SLAM systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* , 2012, pp. 573-580.
23. L. Baroffio, A. Canclini, M. Cesana, A. Redondi, and M. Tagliasacchi, "Briskola: BRISK optimized for low-power ARM architectures," in *Proc. IEEE Int. Conf. Image Process.* , 2014, pp. 5691-5695.