# Событийная стерео-визуальная одометрия

Система событийного стереозрения построена по аналогии с биологическими системами зрения у животных, где каждая часть изображения обрабатывается параллельно и независимо с двух глаз. Все участки независимо и асинхронно реагируют на изменение яркости с микросекундной разрешающей способностью. Эти преимущества позволяют решать сложные задачи в робототехнике, такие как высокоскоростные сцены и сцены с высоким динамическим диапазоном. Представляется система визуальной одометрии на основе данных, полученных стереокамерой, основанной на событиях. Система использует параллельный подход отслеживания и отображения, где новые решения для каждой подпроблемы (трехмерное (3-D) построение карты пространства и оценка положения камеры) разработаны с целями быть принципиальными и эффективными для работы в реальном времени с аппаратным обеспечением. Это достигается за счет максимизации пространственно-временной согласованности данных, основанных на стереособытиях, используя при этом простое и эффективное представление. В частности, модуль отображения строит полуплотную трехмерную карту сцены, объединяя оценки глубины с нескольких точек обзора (полученные благодаря пространственно-временной согласованности) в вероятностном режиме. Модуль отслеживания восстанавливает позу стереосистемы, решая проблему регистрации, которая естественно возникает из-за выбранной карты и представления данных о событиях. Эксперименты на общедоступных наборах данных демонстрируют универсальность предложенного метода в естественных сценах с общим 6-DoF движением. Система успешно использует преимущества камер, основанных на событиях, для выполнения визуальной одометрии в сложных условиях освещенности, таких как низкая освещенность и высокий динамический диапазон, работая при этом в реальном времени на стандартном процессоре.

Событийный метод работы с камерами, как называются его авторы () позволяет отслеживать ситуации, когда происходит пиксельные скачкообразные изменения интенсивности, так называемые "события".

Событийные камеры отдают кадры асинхронно только в момент фиксации события. Этот асинхронный и дифференциальный принцип работы подавляет временную избыточность (необходимость фильтровать кадры на видео для определения ключевых кадров в обычных SLAM методах) и, следовательно, снижает энергопотребление и вычислительные затраты. Фактически основное отличие заключается в том, что фильтрация ключевых кадров осуществлена аппаратно. Обладая микросекундным разрешением, событийные камеры способны фиксировать высокоскоростные движения, которые на стандартных камерах вызвали бы сильное смазывание движения, что также актуально для работы в динамических системах. Кроме того, событийные камеры имеют очень высокий динамический диапазон (HDR) (например, 140 дБ по сравнению с 60 дБ у стандартных камер), что позволяет использовать их в широких условиях освещенности. Таким образом, событийные камеры открывают возможности для решения сложных задач в робототехнике, таких как высокоскоростное и/или HDR отслеживание объектов [3]-[5], отслеживание камеры [6]-[9], управление [10]-[12], одновременная локализация и отображение (SLAM) [13]-[16]. Основная проблема восприятия с помощью этих датчиков заключается в разработке новых алгоритмов, которые обрабатывают незнакомый поток изменений интенсивности ("событий") и способны раскрыть потенциал камеры [2]. В некоторых работах эта проблема решалась путем объединения событийных камер с дополнительными датчиками, такими как датчики глубины [17] или стандартные камеры [18], [19], чтобы упростить задачу восприятия. Однако это создает сложности в синхронизации таких систем и добавляет узкие места в работу системы, поскольку комбинированная система всегда ограничена скоростью самого медленного датчика и его динамическим диапазоном.

В работе () решается проблема стерео визуальной одометрии (ВО) с помощью камер событий в естественных сценах и произвольном 6-DoF движении. Для этого разработана система, которая обрабатывает стереопоток событий в реальном времени и выдает эго-движение стереосистемы и карту трехмерной (3-D) сцены (см. рис. 1). Предлагаемая система в основном следует философии параллельного отслеживания и картографирования [20], где основные модули работают в чередующемся режиме, оценивая движение камеры и трехмерную структуру сцены соответственно (более подробный обзор системы приведен на рис. 2).

Система содержит новый метод отображения стереокадров в пространственные координаты, основанный на оптимизации целевой функции, предназначенной для измерения пространственно-временной согласованности потоков стереособытий. Выработана стратегия слияния на основе вероятностных характеристик оцененной обратной глубины для улучшения плотности и точности восстановленной трехмерной структуры. Представлен новый метод отслеживания камеры на основе 3D.2D регистрации, который использует соответствующую полю глубины природу компактного и эффективного представления событий. Система также эффективна с вычислительной точки зрения и работает в реальном времени, что позволяет рассматривать возможность использования такой или аналогичной системы для задач быстрой локализации и навигации.

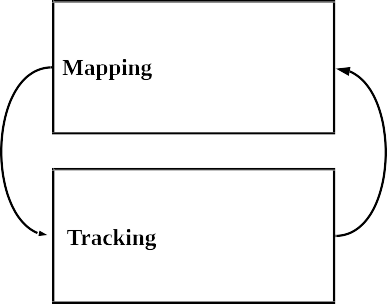
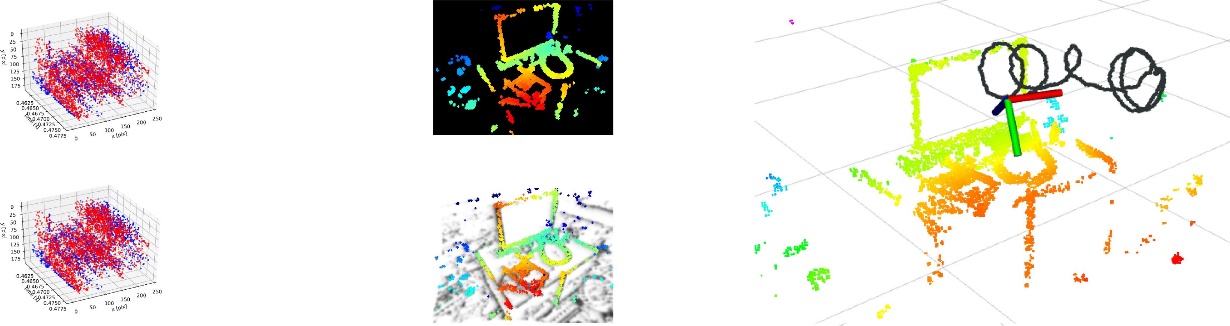


Рис. 1. Система принимает на вход асинхронные данные, полученные парой камер в стереоконфигурации (слева), и восстанавливает движение камер, а также полуплотную карту сцены (справа). Система использует пространственно-временную согласованность событий в плоскостях изображения камер для решения подпроблемы локализации (т.е. отслеживания 6-DoF) и отображения (т.е. оценки глубины) визуальной одометрии (средняя). Система работает в реальном времени на стандартном процессоре.

Работа () значительно расширяет стандартные методы решения проблемы стерео отображения при работе в реальном времени.

Работы в области событийной стерео визуальной одометрии

Событийная стерео визуальная одометрия (ВО) связана с несколькими проблемами оценки структуры и движения с помощью событийных камер, таких как динамический датчик зрения (DVS).

Оценка глубины на основе событий (3-D реконструкция)

Мгновенное стерео: в литературе по оценке глубины стерео на основе событий преобладают методы, которые решают проблему трехмерной реконструкции с использованием данных от пары синхронизированных и жестко закрепленных событийных камер в течение очень короткого времени (в идеале - на основе каждого события). Цель состоит в том, чтобы использовать преимущества событийных камер для реконструкции динамических сцен на очень высокой скорости и с низким энергопотреблением. Эти работы [22]-[24] обычно следуют классическому двухэтапному подходу к задаче поиска эпиполярных совпадений и последующей триангуляции 3-D точки [25]. Сопоставление событий часто решается путем наложения нескольких ограничений, включая временную когерентность (например, одновременность) событий на обеих камерах. Например, [26] объединили эпиполярные ограничения, временную несогласованность, несогласованность движения и фотометрическую ошибку (доступную только для полутоновых событий, полученных с помощью камер ATIS [27]) в объективную функцию для вычисления наилучших совпадений.

Другие работы, такие как [28]-[30], расширяют кооперативное стерео [31] на случай событийных камер [32]. Эти методы хорошо работают со статичными камерами в не загромождённых сценах, так что совпадения событий легко найти среди небольшого количества движущихся объектов.

Монокулярные методы: Оценка глубины с помощью камеры с одним событием была показана в [13], [33] и [34]. Поскольку мгновенная оценка глубины в монокулярных установках нестабильна, эти методы решают проблему оценки глубины для VO или SLAM: следовательно, они требуют знания движения камеры, чтобы интегрировать информацию от событий за более длительный промежуток времени и быть в состоянии произвести полуплотную трехмерную реконструкцию сцены. Одновременность событий не применяется, следовательно, временную когерентность гораздо сложнее использовать для сопоставления событий во времени, поэтому разрабатываются другие методы.

## Оценка позы камеры на основе событий

Исследования локализации камеры на основе событий развивались путем тестирования сценариев все большей сложности. С точки зрения типа движения, ограниченные движения, такие как чистое вращение [8], [35]-[37] или плоское движение [38], [39] были изучены до исследования наиболее общего случая произвольного 6-DoF движения. Что касается типа сцен, то решения для искусственных моделей, таких как высококонтрастные текстуры и/или структуры (линейные или плоскостные карты) [6], [38], [40], были предложены до решения более сложных случаев: естественных сцен с произвольной трехмерной структурой и фотометрическими вариациями [7], [9], [36].

С точки зрения методологии, вероятностные фильтры [7], [36], [38] обеспечивают обновление навигации по событиям, достигая таким образом минимальной задержки (микросекунды), в то время как методы, основанные на кадрах (часто нелинейная оптимизация), дают компромисс между задержкой и более стабильными и точными результатами [8], [9].

### Событийная визуальная одометрия (ВО) и SLAM, основанные на событиях

Монокулярный SLAM: Два метода выделяются как решение проблемы монокулярного событийного VO для 6-DoF движений в естественных трехмерных сценах. В подходе [13] одновременно работают три чередующихся байесовских фильтра, которые оценивают интенсивность изображения, глубину и положение камеры. Восстановление информации об интенсивности и регуляризация глубины делают метод вычислительно трудоемким, поэтому для работы в реальном времени требуется специальное оборудование (GPU). В отличие от этого, [14] предложили геометрический подход, основанный на технике полуплотного отображения из [33] (фокусировка событий [41]) и трекере выравнивания изображений, который работает на изображениях событий. Он не требует восстановления абсолютной интенсивности и работает в реальном времени на центральном процессоре. До сих пор ни один из этих методов не опубликован и не может быть протестирован.

Стерео: метод итеративно-ближайшей точки (ICP) лежит в основе технологии работы большинства реконструкторов стерео в трехмерную сцену. Метод () основан на нашей предыдущей работе [21], переработанной, и новом устройстве слежения за камерой, которое повторно использует структуры данных, используемые для отображения. Для отображения не используется классическая модель сопоставления событий плюс триангуляция, а используется подход прямой проекции, который позволяет оценивать глубину без явного установления соответствия событий. Для отслеживания система использует нелинейную оптимизация на временных поверхностях, что напоминает парадигму, основанную на кадрах, которая дает компромисс между задержкой и эффективностью и точностью. Как и в [14], система () не нуждается в восстановлении абсолютной интенсивности и является эффективной, способной работать в реальном времени без специального оборудования (GPU); достаточно стандартного оборудования, такого как CPU ноутбука.

## ОБЗОР СИСТЕМЫ

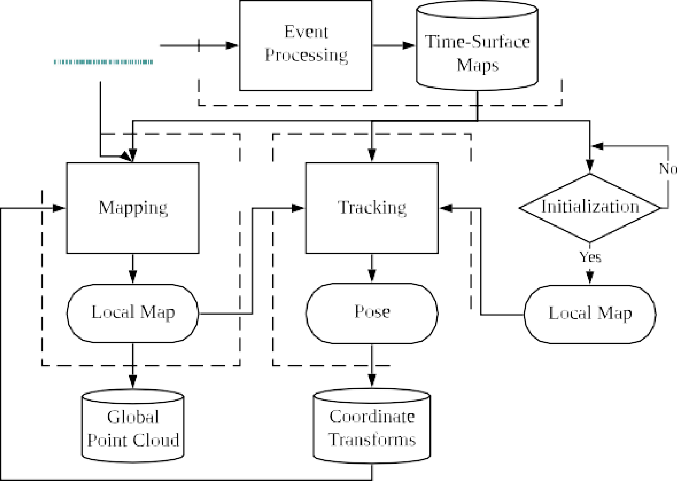


Рис. 2. Блок-схема системы. Основные модули системы, включая Предварительная обработка событий (см. раздел III-A), картирование (см. раздел IV) и отслеживание (см. раздел V) отмечены пунктирными прямоугольниками. Единственным входным сигналом для системы являются необработанные стереоизображения с калиброванных камер, а выходной сигнал состоит из поз камер и облака точек трехмерных краев сцены.

Предлагаемая стереосистема VO принимает на вход только необработанные события от калиброванных камер и одновременно определяет позу стереокамеры и реконструирует окружение с помощью полуплотных карт глубины. Обзор системы приведен на рис. 2, где основные модули выделены пунктирными линиями. Аналогично классическим SLAM конвейерам [20], ядро системы () состоит из двух чередующихся модулей: локализация (отображение) и картирование. Кроме того, существует третий ключевой модуль: предварительная обработка событий.

Модуль обработки событий генерирует представление событий, называемое картами временных поверхностей (или просто "временными поверхностями"), которые используются другими модулями. Теоретически, эти карты времени обновляются асинхронно, с каждым входящим событием (наносекундное время обновления). Однако, учитывая, что одно событие не приносит много информации для обновления состояния системы VO, временные поверхности стерео обновляются с более практичной скоростью: например, при наступлении определенного количества событий или с фиксированной частотой (например, 100 Гц в нашей реализации).

Краткая история временных поверхностей хранится в базе данных (см. правую верхнюю часть рис. 2) для доступа других модулей. Во-вторых, после фазы инициализации модуль отслеживания непрерывно оценивает положение левой камеры относительно локальной карты. Полученные оценки позы хранятся в базе данных преобразований координат (например, TF в ROS [44]), которая способна вернуть позу в любой момент времени путем интерполяции в SE(3). Наконец, модуль сопоставления берет события, временные поверхности и оценки позы и обновляет локальную карту (представленную в виде вероятностной полуплотной карты глубины), которая используется модулем отслеживания. Локальные карты хранятся в базе данных глобального облака точек для визуализации.

Инициализация: для начальной загрузки системы применен стерео метод (модифицированный метод SGM [45], о котором говорится в разделе VI-C), который обеспечивает грубую начальную карту. Это позволяет модулю отслеживания начать работу, в то время как модуль отображения также запущен и создает лучшую полуплотную обратную карту глубины (более точную и плотную).

## Представление событий

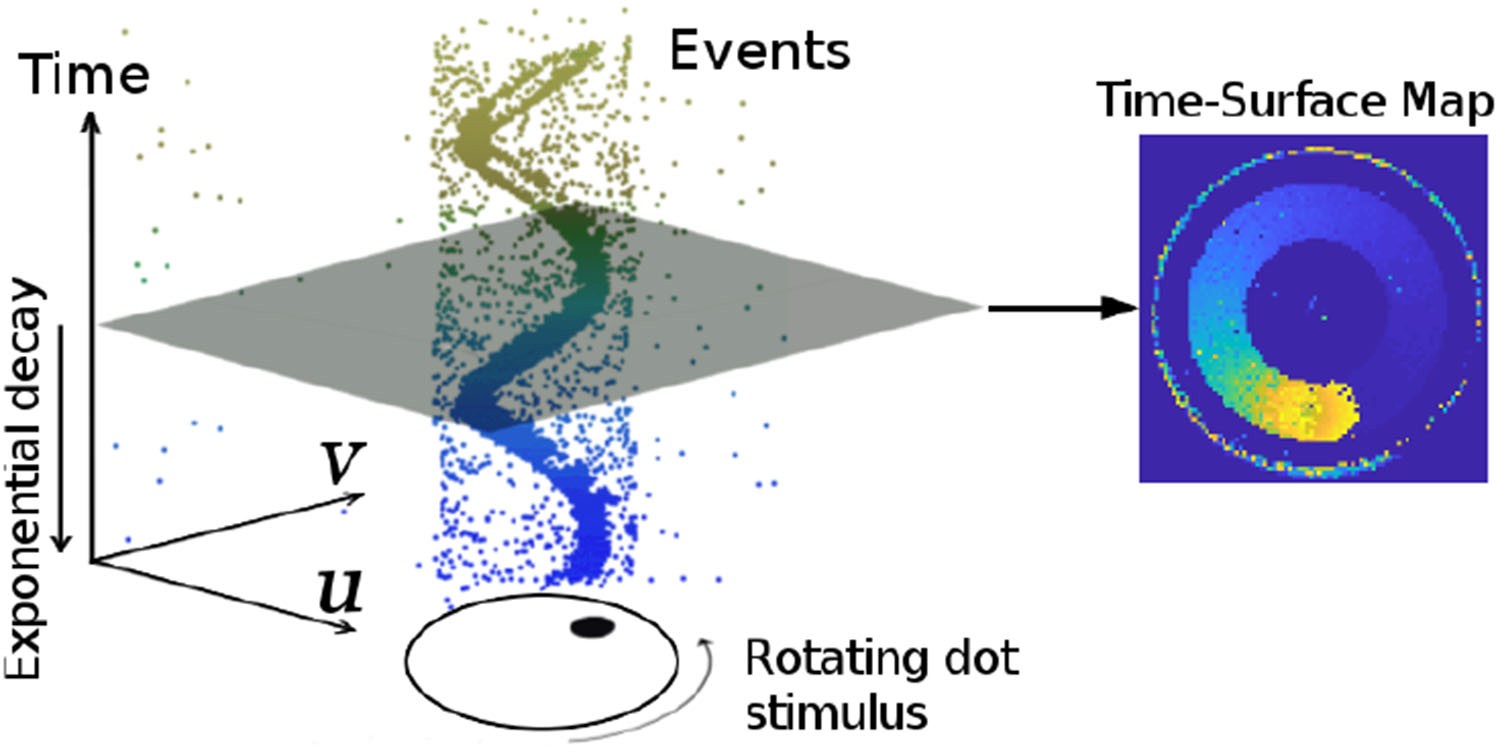


Рис. 3. Представление событий. Слева: выход камеры событий при просмотре вращающейся точки. Справа: Карта временной поверхности в момент времени, которая измеряет, насколько далеко во времени (относительно текущего) произошло последнее событие в каждом пикселе. Чем ярче цвет, тем более позднее событие было вызвано. [46].

Как показано на рис. 3 слева, выходной сигнал камеры событий представляет собой поток асинхронных событий. Каждое событие состоит из пространственно-временных координат, в которых произошло изменение интенсивности заданного размера, и знака этого изменения. Предлагаемая система (см. рис. 2) использует как отдельные события, так и альтернативное представление, называемое временной поверхностью (см. рис. 3- справа). Временная поверхность (ВП) — это двухмерная карта, где каждый пиксель хранит одно значение времени, например, временную метку последнего события в этом пикселе [47]. Используя ядро экспоненциального распада [43], которое увеличивает вес последних событий по сравнению с старыми.

Как показано на рис. 3-справа, значение времени пикселя представляют недавнюю историю движущихся краев в компактном виде (с использованием 2-мерной сетки). Обсуждение нескольких представлений событий (воксельные сетки, рамки событий и т.д.) можно найти в [2] и [48]. Такой подход эффективен с точки зрения памяти и вычислений, представления информативны (края - наиболее описательные области сцены для SLAM), интерпретируемы и потому что они доказали свою успешность для движения (оптический поток) [47], [49], [50] и оценки глубины [21]. В частности, для отображения (см. раздел IV) применяется попиксельное сравнение стереопары TS [21] в качестве замены критерия фотосогласованности стандартных камер [51]. Поскольку значение времени пикселя кодирует временную информацию, сравнение пятен на этом представлении равносильно измерению пространственно-временной согласованности на небольших объемах данных на плоскостях изображения. Для сопоставления изображения в локализации (см. раздел V) применен тот факт, что значение времени пикселя действует как анизотропное поле расстояний [52], определяемое последними местоположениями краев, для регистрации событий относительно трехмерной карты. Для удобства визуализации и обработки (1) масштабируется из [0,1] в диапазон [0, 255].

## Картографирование: оценка глубины стереоизображения по пространственно-временной согласованности и слияние

Рис. 4. Картирование. Геометрия (обратной) оценки глубины. Трехмерные точки, совместимые с событием e = (x,t E, p) на левой камере, параметризуются обратной глубиной ρ на луче обзора через пиксель x в момент времени t E. Истинное местоположение 3D-точки, вызвавшей событие, соответствует значению ρ\*, которое максимизирует временную согласованность по всему стереонаблюдению. left( , t), right( , t). Интервал поиска [ρmin , ρ max] определяется для ограничения оптимизации вдоль луча обзора.

Модуль отображения состоит из двух этапов: 1) вычисление оценок глубины событий и 2) объединение таких оценок глубины в точную и наполненную карту глубины. Обзор модуля отображения представлен на рис. 7(a).

Основополагающими принципами, часто используемыми для оценки глубины стерео на основе событий, являются совместное существование событий и эпиполярное ограничение, которые просто означают, что трехмерная грань должна быть зафиксирована на двух одновременных событиях на соответствующих эпиполярных линиях обеих камер. Однако, как показано в [28] и [53], стерео-временное совпадение не является строго верным на уровне пикселей из-за задержек, дрожания и несоответствия пикселей (например, различия в частоте срабатывания событий). Следовательно, критерий временной согласованности стереоизображения можно определить по пространственно-временным окрестностям событий, а не путем сравнения временных меток событий в двух отдельных пикселях. В работе () представлены такие окрестности с помощью временных поверхностей (благодаря их свойствам и естественной интерпретации как информации, связанной во времени) и преобразуют проблему реконструкции трехмерной сцены в задачу минимизации по выбранному критерию.

В работе [21] был использован вышеупомянутый двухэтапный процесс и принцип. Основные различия обусловлены целью работы в реальном времени:

* Целевая функция построена только по временной несогласованности для одного события карты времени-поверхности, а не на более длительных промежутках времени (таким образом, предлагаемый подход становится ближе к стратегии из [51], чем из [54]).
* Система может быть модифицирована предложенным новым алгоритмом глубинного слияния, который построен на основе исследования вероятностных характеристик временных остатков и обратной параметризации глубины, чтобы обеспечить точную оценку глубины на более длительных временных интервалах по сравнению с одной стереокартой временной поверхности.
* Начальное предположение для минимизации цели определяется с помощью метода блочного сопоставления, который более эффективен, чем перебор [21].
* Неотрицательные невязки по блокам изображений [21] заменяются на подписанные попиксельные невязки, что гарантирует ненулевые якобианы для достоверного распространения неопределенности в динамической модели движения и обновления карты.

## Обратная репараметризация глубины для события



Рис. 4 иллюстрирует геометрию предлагаемого подхода. Без потери общности применена обратная репараметризация глубины, используя левую камеру. Обозначен интервал сопоставления [min, max] глубины, в котором проводится оптимизация.

Для оценки обратной глубины событий, произошедших до стереонаблюдения в момент времени применена процедура оптимизации энергии модели.

Стереонаблюдение в момент времени t относится к паре временных поверхностей левой и правой (left( , t), right( , t)), созданных с помощью (1) [см. также рис. 5(c) и (d)]. Система решает задачу стереосопоставления кадров как задачу минимизацию выбранного критерия. Предполагается, что калибровка камеры (внутренние и внешние параметры) известна и не меняется.

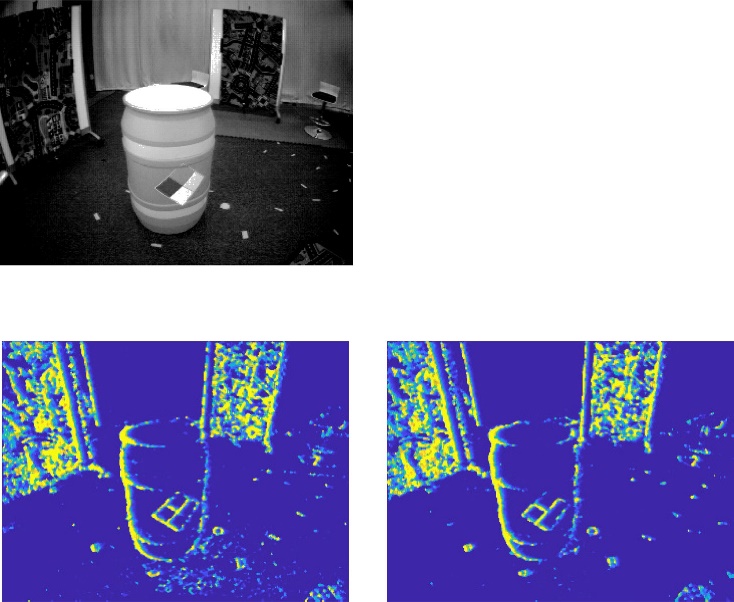


Рис. 5. Сопоставление. Пространственно-временная согласованность. (a) Кадр интенсивности показывает визуальный вид сцены. Наш метод не использует кадры интенсивности; только события. (b) Объективная функция измеряет несоответствие между содержанием истории движения [временные поверхности (c) и (d)] на левой и правой сетчатках, заменяя таким образом фотометрическую ошибку в стерео на основе кадров. В частности, (b) изображает изменение C(x, ρ, left( , t), right( , t), T tδt:t) с обратной глубиной ρ. Вертикальная пунктирная линия (черная) указывает на обратную глубину. (c), (d) показывают временные поверхности стереокамеры событий в момент наблюдения, где пиксели для измерения временного остатка в (b) выделены красным цветом. (a) Сцена в наборе данных [55]. (b) Целевая функция (3) (красным цветом). (c) Временная поверхность (левый DVS). (d) Временная поверхность (правая DVS).

Как показано на рис. 4, предположение об обратной глубине ρ дает трехмерную точку (x, ρ), проекция которой на обе плоскости стереоизображения в момент времени t дает точки x 1(ρ) и x 2(ρ), окрестности которых сравниваются в объективной функции. Следовательно, обратная гипотеза глубины ρ устанавливает кандидата на совпадение стереособытий, а наилучшее совпадение обеспечивается тем ρ, который минимизирует цель. Нелинейный решатель для оценки глубины: Объективная функция оптимизируется с помощью нелинейных методов наименьших квадратов, таких как метод Гаусса-Ньютона, которые итеративно находят корень из необходимого условия оптимальности.

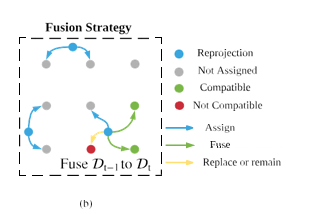
Рис. 6. Распределение вероятностей временных остатков r i: эмпирическое (зеленая гистограмма) и t-образное распределение Стьюдента (синяя кривая). (a) simulation\_3planes [58]. (b) upenn\_flying1 [55].

Каждое событие трансформируется с использованием положения камеры в момент его временной метки. Функция отображения проецирует трехмерную точку на плоскость изображения камеры, а ее обратная функция обратно проецирует пиксель в трехмерное пространство с учетом обратной глубины ρ. right2left преобразование обозначает преобразование от левой камеры события к правой, которое является постоянным. Все координаты события x не искажаются и стереокорректируются с помощью известной функции калибровки камер.

На рис. 5 показан пример объективной функции из реальной стереопоследовательности события с камерой [55], которая имеет истинную глубину. Это подтверждает, что предложенная объективная функция (3) действительно приводит к оптимальной глубине для общего события. Она визуализирует профиль Аналитические производные используются для ускорения вычислений.

Инициализация нелинейного решателя: Успешное решение обратной оценки глубины (7) зависит от хорошего начального значения ρ0 . Для этого, вместо того, чтобы проводить интенсивный поиск по сетке инверсной глубины [21], мы применяем более эффективную стратегию, использующую каноническую стерео конфигурацию: сопоставление блоков по эпиполярным линиям стерео наблюдения left( , t), right( , t) с использованием целопиксельной сетки диспаратности. То есть, мы максимизируем нулевую нормированную кросскорреляцию (ZNCC), используя центры патчей (пиксельные координаты события).

Сопоставление изображений: что данный подход отличается от классических двухэтапных методов обработки событий [22]-[24], [26], которые сначала решают задачу стереосопоставления, а затем триангулируют трехмерную точку. Такие двухэтапные подходы работают по принципу "обратной проекции", отображая двухмерные измерения событий в трехмерное пространство. В отличие от них, данный подход объединяет сопоставление и t E, не компенсируется в xt 1, x t2. Тем не менее, эта аппроксимация обеспечивает повторное и эффективное начальное предположение ρ (используя d), временное согласованность уточняется в ходе последующей нелинейной оптимизации.



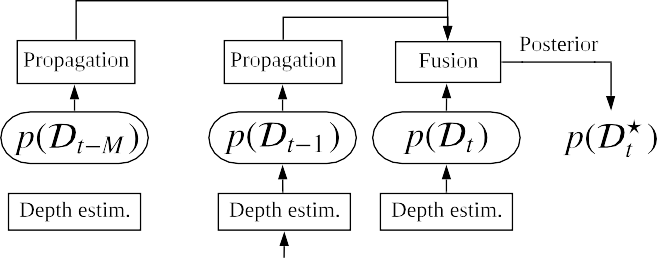


Рис. 7. Модуль картирования: (a) Стереонаблюдения (временные поверхности) создаются в выбранные моменты времени t, …​,t - M (например, 20 Гц) и подаются в модуль отображения вместе с событиями и положением камеры. Обратные оценки глубины, представленные вероятностными распределениями p(D t−k), распространяются на общее время t и объединяются для получения обратной карты глубины p(D\*). Для создания p(D\*) мы объединяем оценки из 20 стереонаблюдений (т.е. M = 19). (b) Принимая слияние от t - 1 до t в качестве примера, правила слияния указаны в пунктирном прямоугольнике, который представляет собой область 3 × 3 плоскости изображения (пиксели отмечены сеткой серых точек). Трехмерная точка, соответствующая средней глубине p(D t−1), проецируется на плоскость изображения в момент времени t в виде синей точки. Такая синяя точка и p(D t−1) влияют (т.е. присваивают, объединяют или заменяют) распределения p(D\*), оцененные по четырем ближайшим пикселям. (a) Блок-схема модуля отображения. (b) Правила слияния глубины в местах на сетке 3 × 3 пикселя.

### Алгоритм 1: Обратная оценка глубины.

Для каждого события, выбрать соответствующий патч, провести оптимизацию с учетом Якобиана.

* пока не найдено решение
  + Вычислите остатки r(ρ) в (4).
  + Вычислить якобиан J(ρ) (аналитические производные).
  + Обновление: ρρ + Δρ, используя (7).
    - возврат результата конвергированной обратной глубины ρ (т.е. ρ \*на рис. 4).

Такая реконструкция приближается к полуплотному уровню, создавая значения глубины для большинства краевых пикселей.

Вероятностная модель оценки обратной глубины: моделируется обратная глубина в пикселе на опорном виде не числом ρ, а фактическим распределением вероятности. Алгоритм 1 обеспечивает "среднее" значение ρ. Теперь известно, как распространяется неопределенность (т.е. разброс вокруг среднего) поступательное движение стереоустановки между периодами событий и можно провести эмпирическое исследование для определения распределения события и стереонаблюдения, а также оценить постоянные внешние параметры между обеими камерами.

Обратная глубина каждого рассматриваемого события оценивается независимо, поэтому вычисления можно распараллелить.

## Полуплотная реконструкция

Метод трехмерной реконструкции (алгоритм 1), производит обратные оценки глубины для отдельных событий, и в соответствии с параметризацией (см. рис. 4) каждая оценка имеет свою временную метку.

Система представляет вероятностный подход для слияния обратных оценок глубины для создания полуплотной карты глубины в текущий момент времени (см. рис. 7), которая в дальнейшем используется для отслеживания. Слияние глубин очень важно, поскольку позволяет отнести все оценки глубины к единому времени, уменьшает неопределенность оцениваемой трехмерной структуры и улучшает плотность реконструкции. Далее используются вероятностные характеристики обратных оценок глубины. На основе этих характеристик представлена стратегия слияния, которая применяется постепенно по мере получения оценок глубины по новым стереонаблюдениям.

Гистограммы остатков функции глубины представлены на рис. 6 и кажутся хорошо центрированными на нуле (по сравнению с их разбросом и диапазоном абсцисс), и поэтому мы можем установить μ r0. Параметры подогнанных t-распределений Стьюдента приведены в таблице I для двух использованных последовательностей из двух различных наборов данных.

Соответствующее место в плоскости целевого изображения обычно имеет нецелую координату x float. Следовательно, распространяемая обратная глубина будет влиять на распределения в четырех ближайших точках пикселя x [int4см. рис. 7(b)]. Используя x качестве примера, слияние выполняется на основе следующих правил. Если в точке xint не существует предыдущего распределения, оно инициализируйтся с помощью степени свободы ν) Если уже существует обратное распределение глубины в точке и оно примерно соответствует распределению t Стьюдента то проверяется совместимость двух гипотез ρ a, Если две гипотезы совместимы, они объединяются в одну инверсную глубину распределение, в противном случае остается распределение с наименьшей дисперсией.

Стратегия слияния показана пунктирным прямоугольником на рисунке 7.

Надежная оценка: Полученная вероятностная модель может быть использована для робастной оценки обратной глубины в присутствии шума и выбросов, поскольку тяжелые хвосты t-распределения Стьюдента учитывают их. Для этого каждый квадратичный остаток в (3) перевзвешивается на коэффициент ω(ri ), который является функцией вероятностной модели p(r). Полученная оптимизационная задача решается с помощью итеративно перевзвешенного метода наименьших квадратов (IRLS), заменяющего решатель Гаусса-Ньютона в алгоритме 1. Подробности о выводе весовой функции приведены в [52] и [57].

Инверсные фильтры глубины: Объединение обратных оценок глубины, полученных от нескольких стереопар, выполняется в два этапа. Во-первых, обратные оценки глубины распространяются от времени каждого события до времени стереонаблюдения (т.е. текущего времени). Это делается аналогично операции распространения неопределенности в (9) и (10). Во-вторых, распространяемая обратная оценка глубины сливается (обновляется) с предыдущими оценками на данном пикселе. Шаг обновления выполняется с использованием робастного байесовского фильтра для распределения Стьюдента t.

Вероятностные оценки в различные моменты времени распространяются и объединяются с распределением карты обратной глубины в самый последний момент времени t, p( \*). Предложенное слияние приводит к полуплотной обратной карте глубины \*с достаточно хорошим соотношением сигнал-шум, что необходимо для метода слежения, обсуждаемого в следующем разделе.

Примечания: Все события участвуют в создании временных поверхностей, которые используются для отслеживания и отображения. Однако глубина не оценивается для каждого события, поскольку это дорого, а мы стремимся достичь работы в реальном времени при ограниченных вычислительных ресурсах (см. раздел VI-F).

Количество слитых стереонаблюдений, M+1 = 20 на рис. 7, было определено эмпирически как разумный выбор для получения хорошей плотности полуплотной карты глубины в большинстве тестируемых последовательностей. Более теоретическим подходом было бы иметь адаптивное число, основанное на статистических критериях, таких как кажущаяся плотность точек или уменьшение неопределенности в слитой глубине, но это остается будущей работой.

# Слежение за камерой

Модуль слежения в алгоритме, представленный на рис. 2 принимает события и локальную карту в качестве входных данных и вычисляет положение стереосистемы относительно карты. В принципе, каждое событие имеет свою временную метку и, следовательно, разную координату камеры при перемещении стереосистемы. Но, как правило, нет необходимости вычислять позы с микросекундным разрешением, следовательно, можно применять более точные подходы в ущерб более высокочастотным.

Решение вероятностного слияния обратной глубины предложенное в () заключается в использовании всех преимуществ временных поверхностей, уже определенных для отображения. Для этого реализован метод отслеживания, основанный на глобальной регистрации изображений с использованием "негативов" временных поверхностей. Он вдохновлен методом выравнивания краев для RGB-D камер с использованием полей смещения [52]. Далее формально определяется задачу слежения и решается с помощью прямого композиционного метода Лукаса-Канаде [60].

## Использование временных поверхностей как полей расстояний

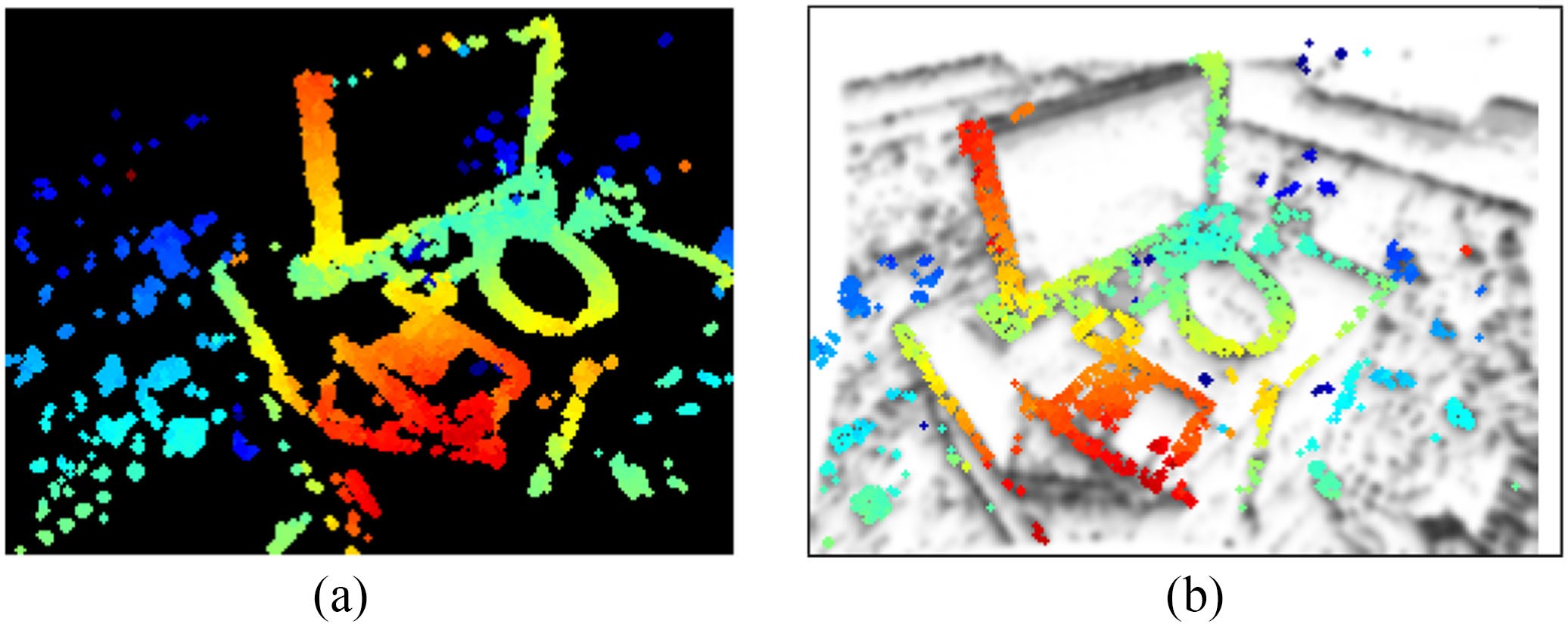


Рис. 8. Отслеживание. Облако точек, восстановленное по инверсной карте глубины (а), деформируется в негатив временной поверхности в текущий момент времени (б) с использованием оцененной относительной позы. Результат (b) - хорошее выравнивание между проекцией облака точек и минимумами (темными областями) отрицательной временной поверхности. (a) Карта глубины в опорной точке обзора с известной позой. (b) Искаженная карта глубины, наложенная на негатив временной поверхности в текущий момент времени.

TS (см. раздел III-A) кодирует историю движения краев в сцене. Большие значения TS (1) соответствуют недавно произошедшим событиям, т.е. текущему расположению края. Обычно эти большие значения имеют пандус с одной стороны (сигнализирующий о предыдущих местоположениях края) и "обрыв" с другой. Это можно интерпретировать как анизотропное поле расстояний: следуя по рампе, можно плавно достичь текущего местоположения края.

Метод выравнивании темных областей негатива TS и опоры инверсной карты глубины при деформации кадра TS на позу кандидата представляется как метод выравнивания по изображениям, при этом изображения представляют информацию о времени, а края сцены - "нулевое время". На рис. 8(b) приведен пример успешного отслеживания, демонстрирующий выравнивание по карте краев.

Композиционный алгоритм

Мы переформулируем задачу определения положения камеры с помощью прямого композиционного метода Лукаса-Канаде [60], который итеративно уточняет инкрементные параметры позы.

Он минимизирует включение правильного TS не приводит к значительному увеличению точности, в то время как затраты на вычисления удваиваются.



Рис. 9. Отслеживание. Срезы объективной функции (15). Графики (a)-(c) и (d)- (f) показаны изменения объективной функции относительно каждого DoF в ориентации и переводе, соответственно. Вертикальная черная пунктирная линия обозначает истинную позу, а зеленая - минимум функции. (a-f) Соответствующие компоненты целевой функции.

Композиционный подход более эффективен, чем аддитивный метод (15), поскольку некоторые части якобиана остаются постоянными на протяжении всей итерации и могут быть предварительно вычислены. Это связано с тем, что линеаризация всегда выполняется в позиции нулевого приращения. В качестве примера на рис. 9 показаны срезы объективной функции относительно каждой степени свободы θ, оцениваемой вокруг наземной истинной относительной позы Δθ = 0. Очевидно, что объективная функция, сформулированная с использованием композиционного метода является гладкой, дифференцируемой и имеет уникальный локальный оптимум вблизи реального значения. Чтобы увеличить область сходимости, применяется гауссово размытие (размер ядра 5 пикселей).

## Надежная и эффективная оценка движения

Насколько мы заметили, нелинейный решатель методом наименьших квадратов уже достаточно точен. Однако для повышения устойчивости при наличии шума и выбросов в обратной карте глубины рассматривается робастная норма. Для эффективности применяется норма Хубера, а для решения полученной задачи используется метод IRLS. Для ускорения оптимизации мы решаем задачу методом Левенберга-Марквардта (LM) со стохастической стратегией выборки (как в [14]). На каждой итерации случайным образом выбирается только партия из N трёхмерных точек в системе отсчета, которые используются для оценки объективной функции (обычно Np = 300). Метод LM может справиться с неотрицательностью остатка и выполняется только одна итерация на партию. Пять итераций часто достаточно для успешного сближения, поскольку начальная поза обычно близка к оптимальной.

# Эксперименты

Теперь давайте оценим предложенную систему стерео VO на основе событий. Сначала мы представим наборы данных и стереокамеру, используемую в качестве источника данных о событиях (см. раздел VI-A). Затем мы оценим эффективность метода с помощью двух наборов экспериментов. В первом наборе мы показываем эффективность одного только модуля отображения, используя исходные позы, предоставленные внешней системой захвата движения.

# Экспериментальная установка и используемые наборы данных



Рис. 10. Пользовательская стереокамера, состоящая из двух камер DAVIS346 с горизонтальной базовой линией 7,5 см.

Оценка алгоритма стерео VO на данных из общедоступных наборов данных и симуляторов [21], [55], [58]. Данные, представленные в [21], были собраны с помощью ручной стереокамеры в помещении. Последовательности, использованные в [55], были собраны с помощью стереокамеры, установленной на беспилотнике, летающем в просторном помещении. Симулятор [58] предоставляет синтетические последовательности с простой структурой (например, плоскопараллельные структуры, геометрические примитивы и т.д.) и "идеальной" моделью камеры событий.

Кроме вышеупомянутых наборов данных, использованы несколько последовательностей, используя стереосистему событийной камеры, показанную на рис. 10. Стереокамера состоит из двух динамических и активных пиксельных датчиков зрения (DAVIS 346). 346 260 пикселей, которые калибруются внутренне и внешне. DAVIS состоит из кадровой камеры и датчика событий (DVS) на одном пиксельном массиве, поэтому калибровка может быть выполнена стандартными методами по кадрам интенсивности и применяется к событиям. Алгоритм () работает с неискаженными и стерео-правильными координатами, которые предварительно вычисляются с учетом калибровки камеры.

## Сравнение критериев оптимизации картирования:

IRLS против LS: Используя синтетические данные из [58], рис. 11 показывают, что предложенный вероятностный подход приводит к более точным трехмерным реконструкциям, чем метод, используемый в [58]. стандартный объективный критерий наименьших квадратов (LS). Синтетическая сцена на рис. 11 состоит из трех плоскостей, параллельных плоскостям изображения камер на разной глубине. Результаты реконструкции на рис. 11(b) показывают более точные плоские структуры, чем на рис. 11(a). Стандартное отклонение ошибки глубины для задачи на основе распределения Стьюдента t в 2-3 раза меньше, чем для стандартной задачи LS, что объясняет более компактные плоские реконструкции на рис. 11(b) по сравнению с (a).

## Сравнение методов стерео 3-D реконструкции

Производительность системы оценивается в сравнении с тремя стереометодами. Базовые методы сокращенно называются GTS [26], SGM [45] и CopNet [62].

Описание базовых методов: Метод в [26] предлагает сопоставить события, используя критерий соответствия, основанный на времени каждого события, который также работает для полутоновых событий с камеры ATIS [27]; после этого классическая триангуляция обеспечивает 3-D местоположение точки.

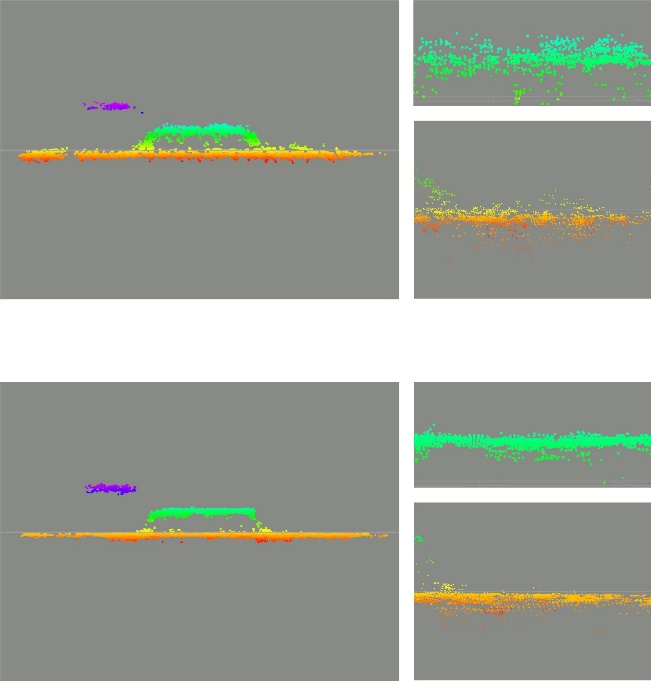


Рис. 11. Сопоставление. Качественное сравнение между стандартным LS-решателем и IRLS-решателем на основе распределения Стьюдента t. Области, выделенные пунктиром, увеличены для лучшей визуализации деталей. (a) Стандартный LS-решатель. (b) IRLS-решатель на основе распределения t Стьюдента.

Алгоритм semiglobal matching (SGM) [45], доступный в OpenCV, изначально разработан для решения задачи стереосовмещения в плотном режиме на основе кадров. Он адаптирован к данной задаче, с помощью запуска его на временных стереоповерхностях и маскирования полученной карты глубины так, чтобы оценки глубины давались только для пикселей, в которых произошли последние события.

Метод в [62] (CopNet) применяет стратегию кооперативного стерео [31] в асинхронном режиме. Использована реализация [63], где применяются идентичные параметры.

Для справедливого сравнения с предложенным методом (), который последовательно объединяет последовательные оценки глубины, также распространяются оценки глубины, полученные GTS и SGM.

Поскольку данные методы не предоставляют оценок неопределенности, просто искажаются оценки глубины от прошлого к настоящему времени (т.е. к моменту, когда слияние запускается в нашем методе). Все методы запускаются и завершаются в одно и то же время и используют наземные истинные позы для распространения оценок глубины во времени, чтобы оценка не зависела от модуля слежения.

Результаты: на рис. 12 сравниваются обратные карты глубины, полученные вышеупомянутыми стереометодами. В первом столбце показаны необработанные кадры в градациях серого из программы DAVIS [64], которые только иллюстрируют внешний вид сцен, поскольку методы не используют информацию об интенсивности. Со второго по последний столбцы показаны обратные карты глубины, созданные GTS, SGM, CopNet и нашим методом, соответственно.

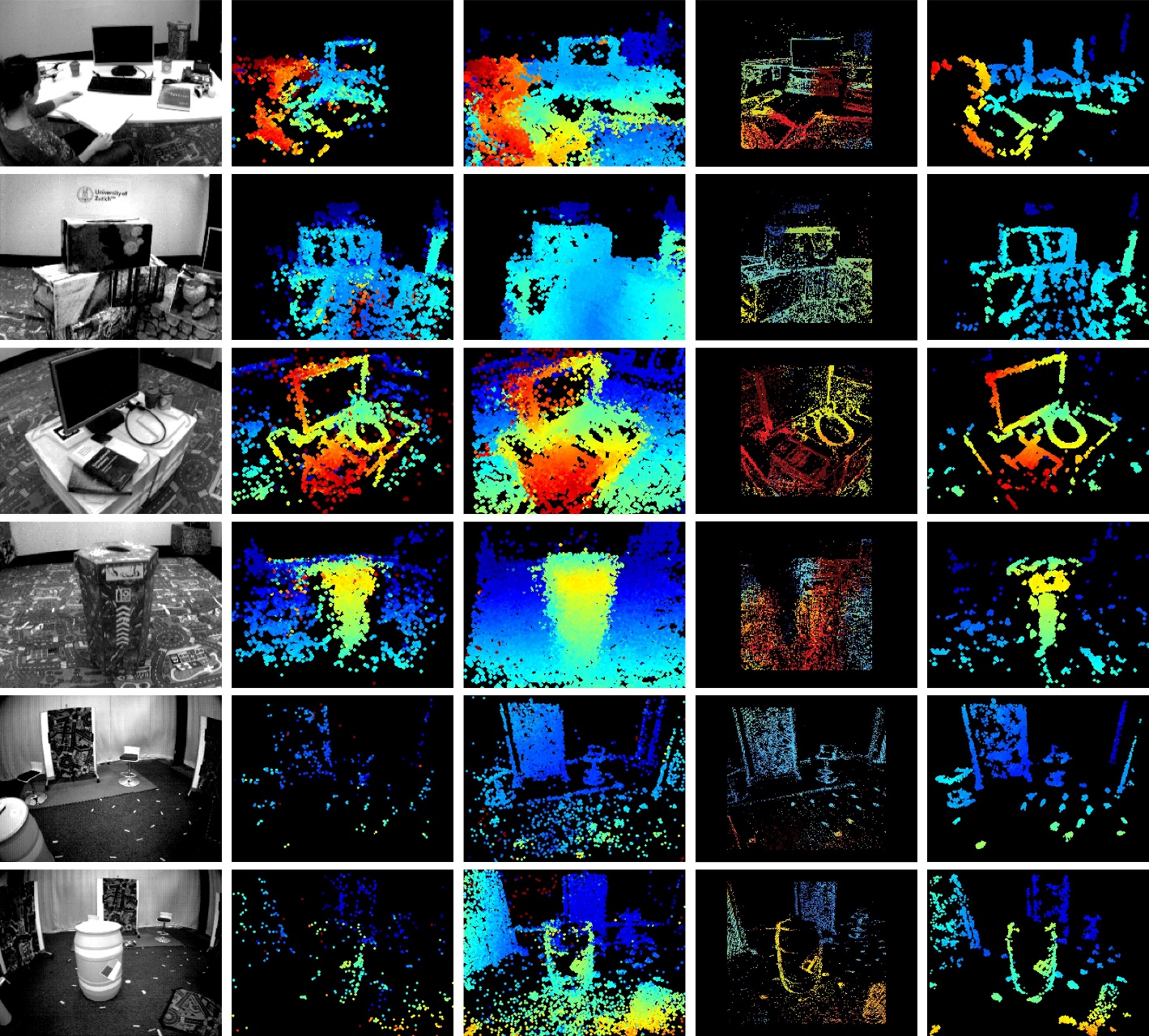


Рис. 12. Картирование. Качественное сравнение результатов картирования (оценки глубины) на нескольких последовательностях с использованием различных стереоалгоритмов. В первом столбце показаны кадры интенсивности с камеры DAVIS (не используется, только для визуализации). В колонках со 2 по 5 показаны результаты обратной оценки глубины GTS [26], SGM [45], CopNet [62] и нашего метода, соответственно. Карты глубины имеют цветовую кодировку от красного (близко) до синего (далеко) на черном фоне в диапазоне 0,55-6,25 м для четырех верхних строк (последовательности из [21]) и в диапазоне 1-6,25 м для двух нижних строк (последовательности из [55]).

Как и ожидалось, поскольку камеры реагируют на видимое движение краев, методы создают полуплотные карты глубины, которые представляют трехмерные края сцены.

Это более очевидно в GTS, CopNet и нашем методе, чем в SGM, потому что регуляризатор в SGM помогает сгладить оценки глубины в областях, где пространственно-временная согласованность неоднозначна, что приводит к наиболее плотным картам глубины.

Хотя CopNet дает удовлетворительные результаты по плотности, он работает хуже предложенного метода с точки зрения точности глубины. Это может быть связано с тем, что диспаратность на выходе CopNet квантуется с точностью до пикселя. Кроме того, используемый относительно большой размер окрестности (предложенный создателями [62]) вносит эффект пересглаживания.

Наконец, можно заметить, что предложенный метод дает лучшие результаты с точки зрения компактности и отношения сигнал/шум. Это связано с тем, что мы моделируем как (обратную) глубину, так и ее неопределенность, что обеспечивает принципиальное слияние глубины в нескольких ракурсах и отсечение ненадежных оценок. Поскольку метод инкрементально объединяет последовательные оценки глубины, плотность результирующих карт глубины остается стабильной, даже если скорость потока событий может меняться.

По количественной оценке, ошибок глубины для последовательностей рис. 12, где доступна истинная глубина (полученная с помощью LiDAR [55]), рассмотренный метод превосходит базовые методы по всем критериям: среднее значение, медиана и относительная ошибка (относительно диапазона глубин).

Оценка производительности полной системы VO выполнена на шести последовательностях с истинными траекториями, а результаты оценки приведены в таблицах V и VI. Лучшие результаты по каждой последовательности выделены жирным шрифтом. Очевидно, что наш метод превосходит базовое решение на основе событий во всех последовательностях. Чтобы сравнение с ORB-SLAM2 было справедливым, глобальная корректировка пучков (BA) была отключена; тем не менее, результаты с включенной глобальной BA также представлены в таблицах для справки. Наша система немного менее точна, чем ORB-SLAM2 на наборе данных rpg, но при этом показывает лучшую производительность на наборе данных upenn\_indoor\_flying. Это связано с эффектом мерцания в наборе данных rpg, вызванным системой захвата движения, который немного ухудшает производительность нашего метода, но не проявляется на полутоновых кадрах, используемых ORB-SLAM2.

Сравнение траекторий, полученных методами, основанными на событиях, представлено на рис. 13. Наш метод значительно превосходит базовый метод SGM+ICP, основанный на событиях. Оценка полной системы VO с помощью рис. 13 оценивает, остаются ли отображение и отслеживание согласованными друг с другом. Это требует, чтобы модуль отображения был устойчив к ошибкам, вызванным модулем отслеживания, и наоборот. Наша система отлично справляется с этой задачей. В результате вышеупомянутых явлений мерцания в наборах данных rpg пространственно-временная согласованность между стереокартами "время-поверхность" может не всегда сохраняться. Мы обнаружили, что наша система устойчиво работает в этом сложном сценарии, если это не происходит во время инициализации. Читатели могут получить лучшее представление о феномене мерцания, посмотрев сопровождающее видео.

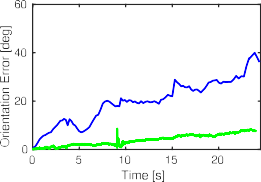
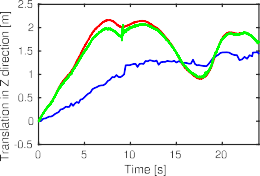
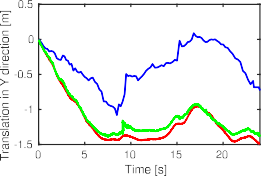
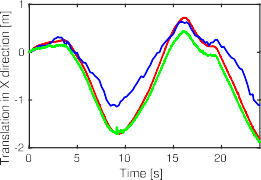
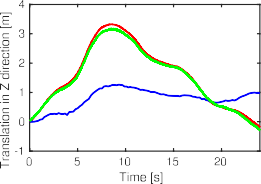
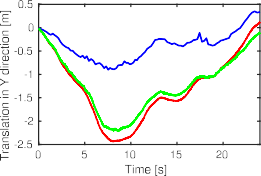
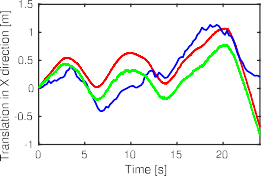
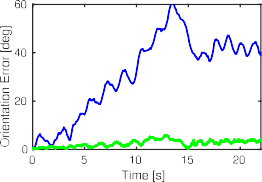
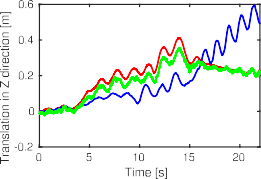
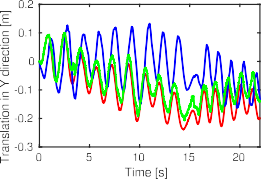
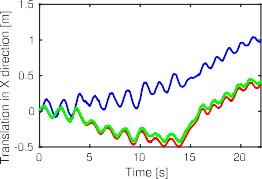
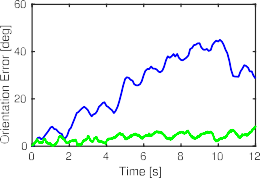
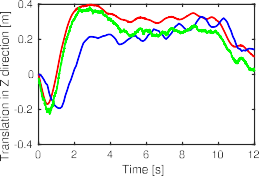
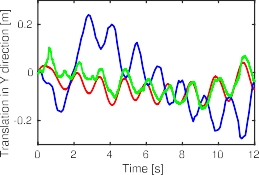
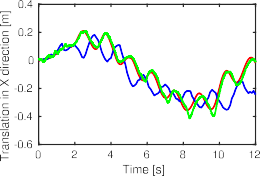
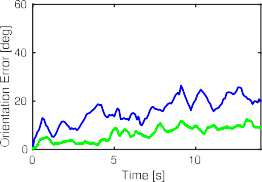
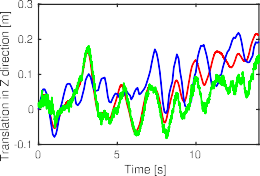
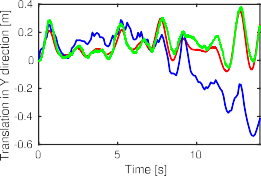
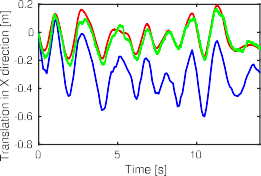
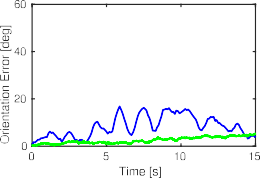
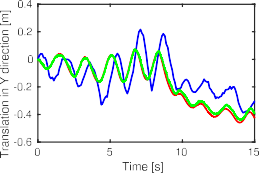
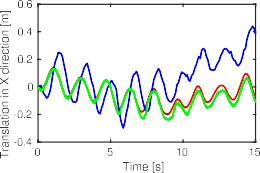


Рис. 13. Отслеживание - графики DoF. Сравнение двух методов отслеживания с истинной траекторией камеры, предоставленной системой захвата движения. Столбцы 1-3 показывают трансляционные степени свободы (в метрах). Последний столбец показывает вращательную ошибку в терминах геодезического расстояния в SO(3) (угол относительного вращения между истинным и расчетным вращением). Каждая строка соответствует отдельной последовательности: rpg\_bin, rpg\_box, rpg\_desk, rpg\_monitor, upenn\_flying1 и upenn\_flying3, соответственно. Красным цветом () изображена грунтовая истина, синим () - метод "SGM+ICP", зеленым () - наш метод. На графиках ошибок "грунтовая истина" соответствует эталону, т.е. нулю. Последовательности rpg [21] сняты с помощью ручной стерео установки, движущейся в условиях локально зацикленного поведения (верхние четыре строки). В отличие от них, последовательности upenn\_flying [55] получены с помощью стереоустановки, установленной на беспилотнике, который переключается между зависанием и преимущественно поступательным движением (два нижних ряда).

Результаты VO на наборе данных upenn показывают худшую точность по сравнению с результатами на наборе данных rpg. Это может быть объяснено следующими двумя причинами. Точность системы (слежения и отображения) ограничена относительно небольшим пространственным разрешением датчика. Использование камер событий с более высоким разрешением (например, VGA [68]) повысило бы точность системы. Также, что когда дрон останавливается и замирает, генерируется мало событий, и поэтому временные поверхности, срабатывающие с постоянной скоростью, становятся ненадежными. Это может привести к тому, что описанная система () будет переинициализирована. Это поведение можно смягчить, используя более сложные стратегии для сигнализации о создании временных поверхностей, например, постоянное или адаптивное количество событий [69].

Таким образом, на динамическом участке набора данных система показывает хорошую производительность.

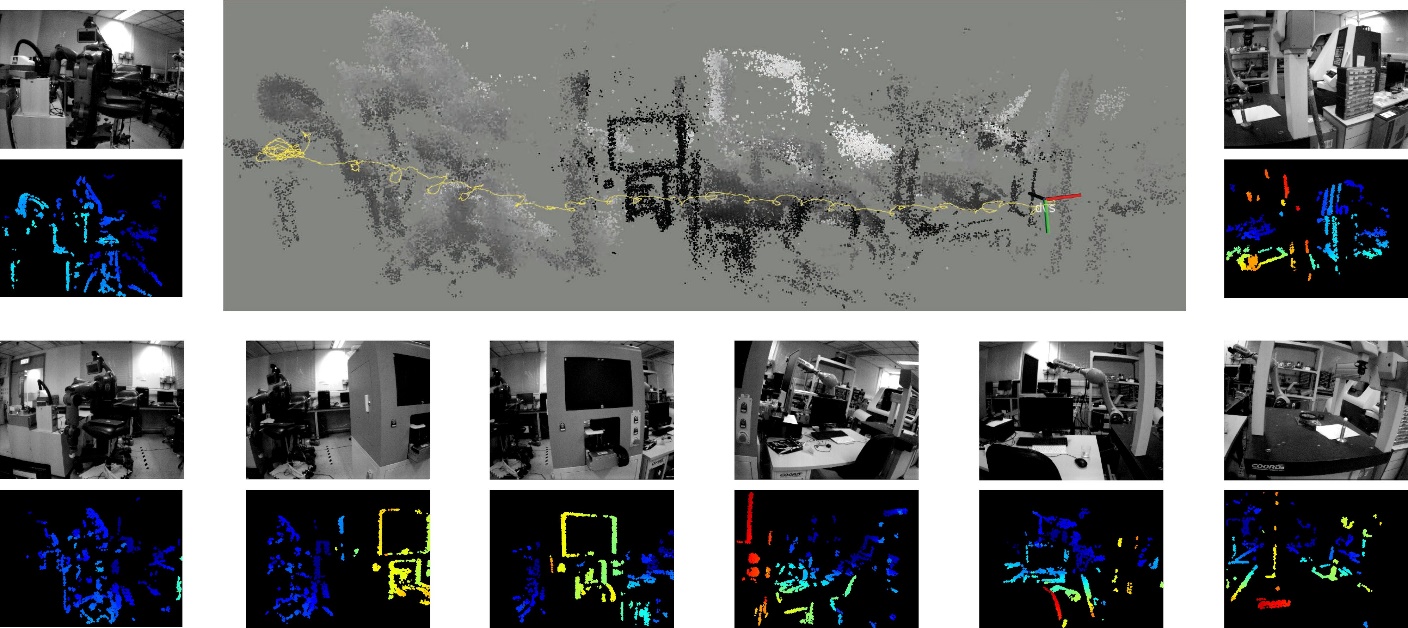


Рис. 14. Расчетная траектория движения камеры и трехмерная реконструкция последовательности hkust\_lab. Вычисленные обратные карты глубины в выбранных точках обзора визуализируются последовательно, слева направо. Кадры интенсивности показаны только для визуализации.

Приведена оценка предложенной системы на примере сцены hkust\_lab, собранной с помощью стереокамеры. Сцена представляет собой загроможденную среду, состоящую из различных машинных установок. Стереокамера была ручной и перемещалась слева направо при локально зацикленном поведении. Трехмерное облако точек вместе с траекторией движения датчика отображаются на экране. на рис. 14. Кроме того, показаны оцененные обратные карты глубины на выбранных ракурсах.

# Эксперименты в условиях низкой освещенности и HDR-среды

Помимо оценки в условиях нормальной освещенности, мы тестируем систему VO в сложных условиях для камер, основанных на кадрах. Для этого мы запустили алгоритм на двух сериях, собранных в темной комнате. Одна из них освещена лампой

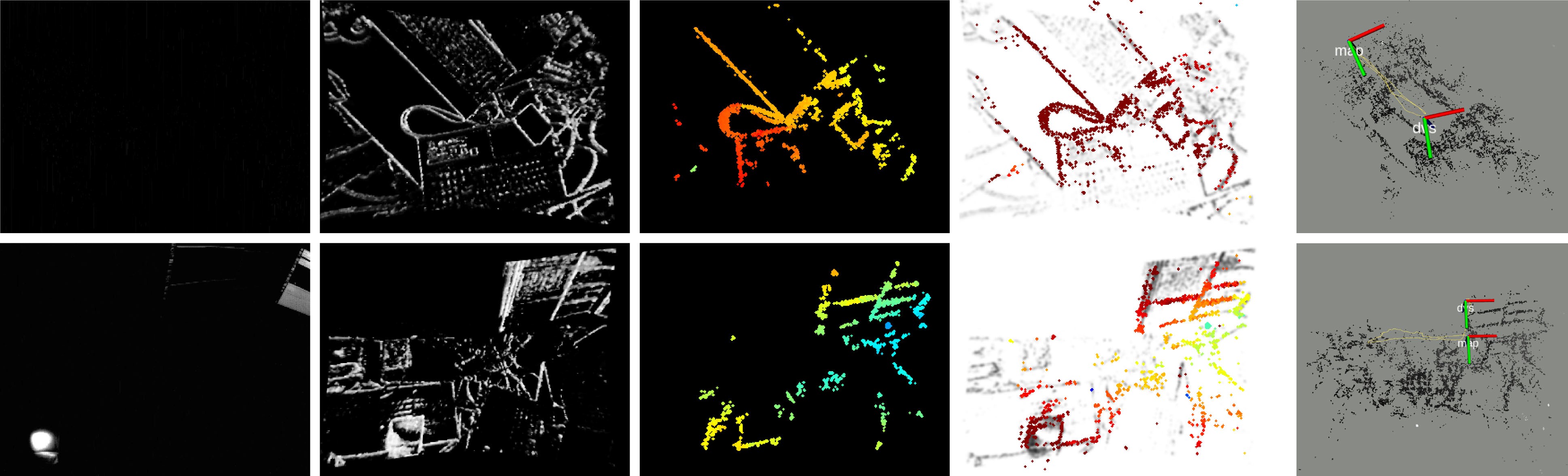


Рис. 15. Сцены с низкой освещенностью и HDR. Верхний ряд: результаты в темной комнате; нижний ряд: результаты в темной комнате с направленной лампой. Слева направо: полутоновые кадры (только для визуализации), временные поверхности, расчетные карты глубины, репроецированные карты на негативы временных поверхностей (отслеживание) и 3-D реконструкция с наложенными оценками траектории движения камеры, соответственно.

для увеличения диапазона изменения яркости сцены, создавая условия высокого динамического диапазона. Результаты показаны на рис. 15. В таких условиях датчик DAVIS на основе кадров (с динамическим диапазоном 55 дБ) практически ничего не видит в темных областях, используя встроенную автоэкспозицию, что привело бы к сбою в работе конвейеров VO, работающих с этой визуальной модальностью. В отличие от этого, представленный метод, основанный на событиях, способен устойчиво работать в таких сложных условиях освещенности благодаря чувствительности и высокого разрешения событийных камер (диапазон 120 дБ).

# Вычислительная производительность

Вышеописанная система стерео визуальной одометрии реализована на языке C++ в ROS и работает в реальном времени на ноутбуке с процессором Intel Core i7-8750H. Для ускорения обработки некоторые узлы (отображения и отслеживания) реализованы с использованием технологии гиперпоточности.



Рис. 16. Влияние количества событий, используемых для (обратной) оценки глубины, на плотность слитой карты глубины. (a) 500 событий. (b) 1000 событий. (в) 2000 событий.

Создание карт "время-поверхность" занимает около 5-10 мс, в зависимости от разрешения датчика. Узел инициализации, активный только во время загрузки, занимает 12-17 мс (в зависимости от разрешения датчика) для создания первой локальной карты (карта глубины, заданная методом SGM и замаскированная картой событий). Узел сопоставления использует 4 потока и занимает около 41 мс, расходуясь на три основные функции. 1) Функция сопоставления занимает 6 мс для поиска 1000 соответствующих пятен на паре временных поверхностей. Успешность совпадения составляет 40-50%, в зависимости от того, насколько хорошо сохраняется пространственно-временная согласованность в данных. 2) Функция уточнения глубины возвращает 500 обратных оценок глубины за 15 мс. 3) Функция слияния (шаги распространения и обновления) выполняет 60 000 операций за 20 мс. Таким образом, узел отображения работает с типичной частотой 20 Гц. Что касается выбора количества событий, учитываемых при обратной оценке глубины (т.е. 1000, как упоминалось ранее), мы обосновываем его, показывая его влияние на плотность структуры оценочных карт глубины. На рис. 16 показаны результаты картографирования с использованием 500, 1000 и 2000 событий для обратной оценки глубины. Выбираются эти события случайным образом из последних 10 000 событий. Для корректного сравнения количество шагов слияния остается постоянным. Как видно, чем больше событий используется, тем более плотной становится обратная карта глубины.

Карта, полученная при использовании 500 событий, является самой ненасыщенной. Также заметно, что использование 1000 или 2000 событий дает практически одинаковую плотность реконструкции, последний вариант (2000 событий) более затратный с вычислительной точки зрения (время вычислений примерно пропорционально количеству событий), поэтому для работы в реальном времени лучше использовать 1000 событий. Узел слежения использует 2 потока и требует 10 мс для решения задачи оценки позы с помощью IRLS-решателя (партия из 300 точек случайным образом выбирается на каждой итерации и выполняется не более пяти итераций). Следовательно, алгоритм может работать с частотой до 100 Гц (0.36 метра смещения на кадр при движении автомобиля со скоростью 100 км/ч).

# Общий анализ системы

Представленная система визуальной стерео одометрии Event-Based Stereo Visual Odometry на основе событий для пары калиброванных и синхронизированных событийных камер в стереоконфигурации показывает хорошую производительность, точность и позволяет осуществлять работу с новым типом камер с большим разрешением и частотой, не поддерживаемых в стандартных ORB-SLAM методах. Это первая опубликованная работа, в которой решается данная проблема.

Описанный метод отображения основан на оптимизации объективной функции, предназначенной для измерения пространственно-временной согласованности потоков стереособытий. Для повышения плотности и точности восстановленной трехмерной структуры использована стратегия слияния, основанная на изученных вероятностных характеристиках оцененной обратной глубины. Метод отслеживания основан на 3d-2d регистрации, которая использует природу поля расстояний, присущую компактному и эффективному представлению событий (временные поверхности).

Приведены данные оценки работы системы на общедоступных наборах данных, которые демонстрируют универсальность данной системы. Производительность сопоставима со зрелыми, современными методами визуальной локализации и навигации для работы стереокамер в нормальных условиях.