

Санкт-Петербургский государственный университет

Громова Арабелла Михайловна

Выпускная квалификационная работа

Исследование применимости тепловизионных камер в задачах SLAM

Уровень образования: бакалавриат

Направление *02.03.03 “Математическое обеспечение и администрирование
информационных систем”*

Основная образовательная программа *СВ.5162.2021 “Технологии программирования”*

Научный руководитель:
старший преподаватель, к.т.н. Ю. В. Литвинов

Консультант:
руководитель исследовательской группы, лаборатория «ИИ для Автономных Систем»,
ААНО ВО «Сколковский Институт Науки и Технологий», А. В. Корнилова

Рецензент:
ведущий научный сотрудник, МФТИ, д.ф.-м.н. С. Э. Парсегов

Санкт-Петербург
2025

Saint Petersburg State University

Arabella Gromova

Bachelor's Thesis

Thermal cameras in SLAM algorithms

Education level: bachelor

Speciality *02.03.03 "Software and Administration of Information Systems"*

Programme *CB.5162.2021 "Programming Technologies"*

Scientific supervisor:
C.Sc., senior lecturer Y.V. Litvinov

Consultant:
head of research group at "AI for Autonomous Systems" Lab, Skoltech A.V. Kornilova

Reviewer:
Sc.D., senior research scientist at MIPT S.E. Parsegov

Saint Petersburg
2025

Оглавление

1. Введение	4
2. Постановка цели и задач	6
3. Обзор	7
3.1. SLAM-алгоритмы, использующие тепловизионные снимки	7
3.2. Предобработка изображений	10
3.3. Датасеты	13
3.4. Метрики	15
4. Обзор визуальной SLAM-системы	16
4.1. Детектирование и ассоциация ключевых точек	16
4.2. Инициализация карты	17
4.3. Модуль отслеживания	17
4.4. Локальное картографирование	19
5. Реализация	21
6. Экспериментальное сравнение	23
6.1. Цель и вопросы эксперимента	23
6.2. Условия эксперимента	24
6.3. Результаты	24
6.4. Выводы	34
7. Заключение	35
Список литературы	36

1. Введение

Задачи, связанные с определением местоположения и оценкой геометрии окружающей среды, играют ключевую роль в функционировании автономных систем. Для решения такого вида задач используются SLAM-алгоритмы (Simultaneous Localization And Mapping), которые одновременно решают две взаимосвязанные задачи: локализацию автономного устройства и создание карты окружающей среды. SLAM-системы в основном состоят из двух ключевых компонентов: фронтенда, который отвечает за извлечение ориентиров из данных сенсоров и установление ассоциаций между ними, и бэкенда, который вычисляет текущее местоположение автономной системы в пространстве, а также оценивает и обновляет карту.

Для эффективной работы SLAM-систем в различных условиях необходимо применять разнообразные сенсоры. Одним из таких сенсоров является тепловизор — прибор для наблюдения за распределением температуры исследуемых объектов. Тепловизоры играют важную роль в системах SLAM при работе в сложных условиях, таких как темнота, туман, задымление и запыленность. Их использование особенно актуально в тех случаях, когда нужно точно определить местоположение, например, при пожарах [19, 75], в шахтах [35], туннелях [30] или в темное время суток [65]. Кроме того, тепловизоры могут использоваться для локализации на ледяных спутниках [26]. Пример разницы видимости между RGB и тепловизионными снимками представлен на рисунке 1.

Однако следует учитывать, что тепловизионные снимки в отличие от RGB могут обладать высоким уровнем шума, низкой контрастностью и слабой выраженностью контуров, что может снижать их эффективность в определенных условиях. Эти аспекты усложняют точное восприятие и интерпретацию данных системами SLAM. Поэтому тепловизионные изображения требуют предварительной обработки, включая фильтрацию и нормализацию [68, 63, 34].

Таким образом, возникает вопрос о поиске и сравнении различных алгоритмов детекции и ассоциации ключевых точек и оценке их каче-

ства в рамках работы SLAM-алгоритмов. Стоит отметить, что систематического обзора и сравнения на близком к реальным задачам датасете на данный момент не проводилось. В лаборатории «ИИ для автономных систем» разрабатывается собственная SLAM-система для работы с визуальными данными, которая проектируется с возможностью масштабирования на новые модальности, а также новые виды ориентиров и детекторов для них. В рамках данной работы предлагается реализовать тепловизионную модальность для системы, разрабатываемой в лаборатории, определить наилучшие для задач SLAM алгоритмы детекции и ассоциации ключевых точек, методы предобработки тепловизионных снимков, а также сравнить полученную SLAM-систему с существующими аналогами.

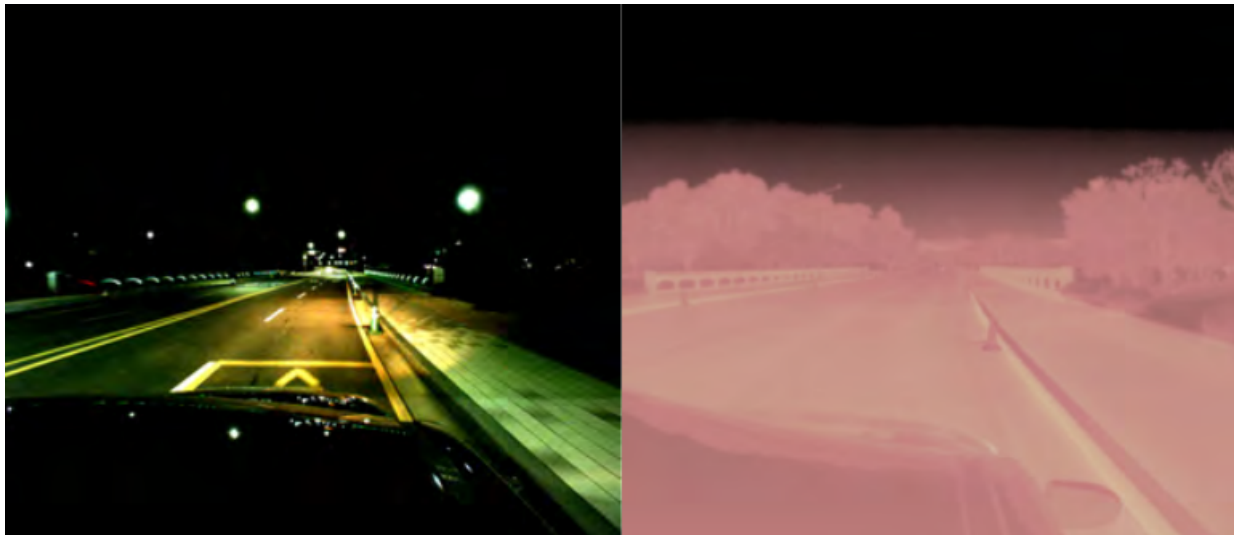


Рис. 1: RGB и тепловизионный снимки соответственно

2. Постановка цели и задач

Целью работы является исследование применимости тепловизионных снимков в рамках визуальной SLAM-системы.

Для достижения данной цели были поставлены следующие задачи.

1. Выполнить обзор существующих решений SLAM-задачи с использованием тепловизионных снимков, включая наборы данных для экспериментов, методы предобработки, а также алгоритмы детекции и ассоциации ключевых точек.
2. Выполнить обзор визуальной SLAM-системы, разрабатываемой в лаборатории «ИИ для автономных систем».
3. Реализовать в рамках SLAM-системы различные алгоритмы предобработки тепловизионных снимков, а также устойчивые к ним детекторы и алгоритмы ассоциации ключевых точек.
4. Провести экспериментальное сравнение алгоритмов предобработки, детекции и ассоциации ключевых точек.
5. Выбрать наиболее оптимальные алгоритмы предобработки снимков, детекции, ассоциации особенностей и провести экспериментальное сравнение полученной системы с другими SLAM-алгоритмами.

3. Обзор

Использование тепловизионных изображений в системах визуально-инерциальной одометрии (VIO) и SLAM-алгоритмах сопряжено с рядом проблем, включая низкую контрастность, высокий уровень шумов и слабую выраженность контуров. Для решения этих проблем современные системы часто дополняются другими типами сенсоров, такими как лидары [57, 12], радары [1], инерциальные измерительные блоки (IMU) [52, 64], камеры видимого спектра [76, 36], стереопары тепловизоров [65].

В данном разделе будет проведён обзор существующих систем визуально-инерциальной одометрии и SLAM-алгоритмов, способы предобработки данных, таких как нормализация, фильтрация шумов. Кроме того, будут описаны существующие наборы данных тепловизионных снимков и метрики для оценки качества работы SLAM-алгоритмов.

3.1. SLAM-алгоритмы, использующие тепловизионные снимки

Для решения SLAM-задачи с использованием тепловизионных камер существуют три основных подхода: прямой (direct), косвенный (indirect) и полупрямой (semi-direct).

3.1.1. Прямой подход

Прямой подход в SLAM-системах использует непосредственно значения яркости пикселей для оценки движения камеры и построения карты, минимизируя фотометрическую ошибку между соответствующими участками последовательных кадров. В отличие от косвенных методов, этот подход учитывает все пиксели изображения, что позволяет более точно восстанавливать геометрию карты. Одним из таких алгоритмов является ROVIO [53] — система, которая объединяет инерциальные и визуальные данные в рамках расширенного фильтра Калмана (Extended Kalman Filter, EKF), используя прямое фотометриче-

ское выравнивание участков изображения. Авторы исследования [52] адаптировали систему ROVIO для обработки тепловизионных изображений, объединяя данные длинноволнового инфракрасного диапазона с инерциальными данными. Это позволило достичь трекинга в реальном времени в условиях плохой видимости, таких как задымленные помещения. Авторы работы [20] предложили алгоритм DSO, основанный на оптимизации фотометрической ошибки. Одним из основных преимуществ прямого подхода является его эффективность в работе с изображениями, обладающими низкой текстурой, но в то же время данный подход неустойчив к шуму. Однако прямой метод ограничивается оценкой положения камеры только на основе снимков с одной камеры и не допускает использование других модальностей.

3.1.2. Косвенный подход

В основе этих методов лежит выделение и отслеживание визуальных ориентиров, например, ключевых точек, линий. Данные SLAM-алгоритмы применяют особенности изображений и их сопоставление для построения карты и определения положения автономного устройства. Такие алгоритмы включают несколько этапов: извлечение особенностей, установление ассоциаций между ними, оценка положения, построение карты и оптимизация как положения, так и карты.

Авторы работы [75] представили SLAM-систему на основе ORB SLAM3 [51], в которой метод извлечения признаков заменен на SURF [59], а в качестве дескриптора используется BRIEF [8]. Для сопоставления особенностей и для замыкания петли (loop closure) используется «мешок визуальных слов» (bag of visual words). Авторы статьи [66] представили SLAM-систему на основе MCSLAM [15], где для выделения особенностей используется нейросетевой алгоритм SuperPoint [14], а для сопоставления — GlueStick [28]. Также для увеличения контрастности тепловизионных снимков применяется CLANE [72]. Алгоритм позволяет сохранять карту в дневное время и точно оценить местоположение ночью. Авторы исследования [3] предложили систему на основе ORB SLAM2 [48], которая объединяет тепловизионную и RGB. Эта си-

система обладает возможностью переключения между монокулярным и стереорежимом, что повышает её точность и надежность. Также формируется три типа ключевых точек: стерео-точки, полученные с обеих камер, монокулярные точки с тепловизоров и с RGB камер. Для улучшения процесса сопоставления между двумя камерами был представлен дескриптор bi-directional BRIEF [8] (BDBRIEF). Авторы статьи [68] представили визуально-инерциальный SLAM, в котором для детекции ключевых точек использовался алгоритм Shi-Tomasi [56], а для их отслеживания — алгоритм Kanade-Lukas-Tomasi (KLT) [4]. Для оценки инлайеров применялся метод RANSAC [22]. Кроме того, в качестве ориентиров были выделены линии с использованием метода LSD [39], а для их сопоставления использовались алгоритмы LBD [71] и KNN [23], при этом короткие линии отбрасывались. Оригинальные изображения с тепловизора, обладающие шумом, обрабатывались с использованием двустороннего фильтра (bilateral filter) [6] для уменьшения шума, полосового фильтра в частотной области (frequency-domain bandpass filter) [29] для устранения полосового шума и метода CLAHE [?] для повышения контрастности. В исследованиях [62, 64] в качестве детектора особенностей используется FAST [54], а для их трекинга применяется алгоритм KLT [4]. Кроме того, авторы статьи [62] дополнительно используют детектор Кэнни вместе с FAST [54]. Авторы исследований [26, 19] используют EKF-SLAM [11] в качестве основы системы, а для обнаружения признаков используют SIFT [44]. В статьях [18, 17] для выделения углов использовался алгоритм Difference of Gaussian [45]. Авторы исследования [1] разработали SLAM-систему с тепловизором и LiDAR, в которой для детекции признаков на снимках применяется LoFTR [43]. В статье [16] был предложен визуальный SLAM, использующий две камеры — RGB и тепловизионную, алгоритмом выделения особенностей является ORB [49].

Можно отметить, что в некоторых работах классические методы детекции ключевых точек (ORB [49], SIFT [44], SURF [59]) заменяются на нейросетевые. Например, нейросетью SuperPoint [14]. Также недавно появились нейросетевые алгоритмы XFeat [70], DeDoDe [13], которые обе-

щают повышенную точность. Аналогичный процесс можно наблюдать и в методах сопоставления — нейросетевые алгоритмы (GlueStick [28], LightGlue [42]) обеспечивают более надежное сопоставление даже при наличии значительных шумов. Авторы исследования [67] представили алгоритм ThermalPoint, схожий с SuperPoint [14], но в отличие от него специально адаптированный для работы с тепловизионными изображениями. Однако SuperPoint [14] демонстрирует более высокую обобщающую способность, что позволяет ему устойчиво работать с различными типами тепловизоров и в разнообразных условиях съемки.

3.1.3. Полупрямой подход

Данный подход интегрирует принципы косвенного и прямого подходов. Авторы работы [25] предложили алгоритм SVO (Semi-direct Visual Odometry), комбинирующий разреженную инициализацию (инициализация с минимумом ключевых точек) с прямым трекингом (отслеживание по пиксельным интенсивностям без дескрипторов). Khattak и др. [37] предложили визуально-инерциальную систему KTIO, основанную на минимизации ошибки репроекции 3D точек и ошибок инерциальных измерений. Авторы статьи [40] комбинируют DSO [20] для трекинга в реальном времени и модифицированный ORB-SLAM [47] для построения глобально-согласованной карты. В работе N.Krombach и др. [38] также был предложен гибридный подход, сочетающий прямой метод (LSD-SLAM [21]) с косвенным (LIBVISO2 [27]). При этом LIBVISO2 [27] используется для начальной инициализации положения камеры в рамках прямого метода. Стоит отметить, что методы данного подхода затруднительно адаптировать к другим сенсорам, как и алгоритмы прямого подхода.

3.2. Предобработка изображений

Предобработка тепловизионных снимков включает в себя нормализацию изображения и уменьшение шума. Шум в тепловизионных изображениях может возникать по нескольким причинам, включая ограни-

ченную чувствительность сенсоров, колебания температуры окружающей среды, а также влияние других источников тепла. В данном разделе рассматриваются современные алгоритмы предварительной обработки тепловизионных изображений, используемые в системах SLAM и VIO, а также широко применяемые в компьютерном зрении методы обработки снимков.

3.2.1. Методы удаления шума

В данном разделе будут рассмотрены фильтры для удаления шума.

Для удаления полосового шума (периодические полосы на изображении из-за помех или неравномерного считывания сигнала) авторы исследования [68] использовали полосовой фильтр в частотной области (frequency-domain bandpass filter) [29]. Полосовой фильтр в частотной области используется для удаления нежелательного шумового сигнала, который находится вне заданного диапазона частот. Математически фильтр $H(u, v)$ описывается следующим образом:

$$H(u, v) = \begin{cases} 1, & \text{если } \text{lower_border} \leq D(u, v) \leq \text{upper_border} \\ 0, & \text{иначе} \end{cases}$$

где $D(u, v)$ – расстояние от точки (u, v) в частотной области до центра изображения, lower_border и upper_border обозначают соответственно минимальную и максимальную частоты. Сначала к изображению применяется двумерное преобразование Фурье. Затем применяется сам фильтр, что приводит к следующему выражению:

$$G(u, v) = F(u, v) \cdot H(u, v)$$

где $F(u, v)$ – преобразование Фурье в точке (u, v) в частотной области. Для получения отфильтрованного изображения необходимо применить обратное преобразование Фурье для $G(u, v)$.

Также в статье [68] был применен двусторонний фильтр (bilateral filter) [6] для удаления шума. Он работает, учитывая расстояние между пикселями и различия в их яркости, что позволяет ему сглаживать

однородные участки изображения, не размывая при этом края объектов. Формула для отфильтрованного значения пикселя в позиции p :

$$BF[I]_p = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) \cdot G_{\sigma_r}(|I(p) - I(q)|) \cdot I(q)$$

где $BF[I]_p$ – отфильтрованное значение пикселя в позиции p , $I(q)$ – значение интенсивности пикселя в позиции q , S – окрестность пикселя p , σ_s – параметр, определяющей, насколько сильно расстояние влияет на вес пикселя, σ_r – параметр, определяющий, насколько сильно различия в значениях интенсивности между пикселями влияет на вес, W_p – нормализующий коэффициент, который обеспечивает, что сумма весов равна 1:

$$W_p = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) \cdot G_{\sigma_r}(|I(p) - I(q)|)$$

$G(x)$ — одномерная функция Гаусса:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

В статьях [17, 18] для обнаружения краёв используется разница Гауссиан (difference of Gaussians). Математически ее можно описать следующим образом:

$$K = G(x, y)_{\sigma_1} - G(x, y)_{\sigma_2}$$

где $\sigma_1 < \sigma_2$, $G(x, y)_{\sigma}$ – двумерная функция Гаусса:

$$G(x, y)_{\sigma} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Авторы исследования [63] предложили предобработку на основе сингулярного разложения для подавления полосового шума.

В качестве методов шумоподавления широко применяются медианная фильтрация и алгоритм тотальной вариации. Медианный фильтр [32] демонстрирует высокую эффективность против импульсного шума (случайные выбросы яркости пикселей) за счёт замены значений пикселей медианой их локального окружения, что обеспечивает устойчивость к выбросам интенсивности. Для более сложных случаев,

требующих сохранения чётких границ при устранении шума без чрезмерного сглаживания, используется метод тотальной вариации (TV), предложенный Chambolle [10] и основанный на минимизации глобального энергетического функционала.

Для устранения шумов в изображениях также существуют различные нейросетевые подходы, включая U-Net-подобную архитектуру NAFNet [58], трансформерную модель SwinIR [61] и генеративную сеть ImageDenoisingGAN [33].

3.2.2. Нормализация

Для нормализации тепловизионных изображений авторы исследований [68, 66] использовали CLANE (Contrast Limited Adaptive Histogram Equalization) [72]. Данный алгоритм применяется для повышения контрастности изображений. CLANE работает следующим образом: изображение делится на небольшие блоки, для каждого блока вычисляется гистограмма и выполняется выравнивание гистограммы. Затем вводится ограничение на максимальное значение гистограммы, чтобы избежать чрезмерного усиления контраста. Далее результаты объединяются с помощью билинейной интерполяции, что сглаживает переходы между блоками. В результате получается изображение с улучшенным контрастом и более четкими деталями. В OpenCV [7] реализация CLANE позволяет настраивать ключевые параметры: `clipLimit`, определяющий порог ограничения контраста, и `tileGridSize`, задающий размер блоков обработки. Авторы исследования [17] применили стандартное выравнивание гистограммы для улучшения контраста.

3.3. Датасеты

В данном разделе будут рассмотрены датасеты, которые находятся в открытом доступе и упоминаются в работах, посвященных SLAM с тепловизорами.

- **ViViD++** [69] — набор данных, записанный в помещениях и на открытом воздухе. Датасет включает тепловизионные, RGB сним-

ки, LiDAR и инерциальные данные. Референсные траектории получены с помощью системы захвата движений Vicon и лидарной одометрии.

- **Датасет из статьи [24]** — набор данных, записанный в темных помещениях. Данные содержат RGB, тепловизионные снимки, облака точек, IMU. Точная траектория была получена при помощи системы слежения Vicon.
- **Multi-spectral Dataset [74]** — набор данных, записанный внутри помещений и снаружи. Данные содержат тепловизионные, RGB снимки, карты глубины, измерения акселерометра и гироскопа. Точная траектория была получена при помощи системы захвата движений.
- **OdomBeyondVision [50]** — набор данных, записанный в помещениях вручную и с помощью беспилотного летательного аппарата (Unmanned Aerial Vehicle, UAV), беспилотного наземного транспортного средства (Unmanned Ground Vehicle, UGV). Датасет содержит тепловизионные, RGB-D снимки, IMU, облака точек. Точная траектория для UGV была определена с помощью колесной одометрии, в то время как для UAV траектория была получена с использованием системы отслеживания Vicon. Для датасета, записанного вручную, траектория была получена с помощью одометрии на основе LiDAR.
- **Brno-Urban-Dataset [41]** — набор данных, записанный в открытых пространствах с использованием установленной на автомобиле системы камер. Датасет включает в себя RGB, тепловизинные снимки, облака точек, IMU, GNSS измерения.
- **Caltech Aerial RGB-Thermal [9]** — набор данных, снятый в открытой местности. Датасет содержит RGB, тепловизионные аэрофотоснимки, инерциальные данные, семантические маски, GNSS данные.

3.4. Метрики

В этом разделе будут рассмотрены метрики для оценки алгоритмов детекции и ассоциации различных типов ориентиров, а также работы SLAM-алгоритмов. Поскольку для алгоритмов детекции и ассоциации ориентиров нет эталонных данных, их оценка будет осуществляться визуально и на основе количества детекций, количества ассоциаций.

Поскольку в рассматриваемых наборах данных имеется эталонная траектория, для оценки SLAM-алгоритмов будет применяться референсный подход. В этом случае используются следующие метрики для оценки SLAM-алгоритмов:

- Absolute Trajectory Error (абсолютная ошибка траектории) — метрика, показывающая абсолютную разницу между референсной и оцененной траекториями [46]:

$$E_i := P_{est,i}^{-1} P_{ref,i}$$

$$ATE = \left(\frac{1}{n} \sum_{i=1}^n \|\text{trans}(E_i)\|^2 \right)^{\frac{1}{2}}$$

где $P_{est,i}, P_{ref,i} \in SE(3)$ — оцененная и референсная траектории соответственно.

- Relative Pose Error (относительная ошибка положения) — метрика, показывающая разницу между референсным и оцененным положениями устройства на фиксированном временном интервале [46]:

$$F_i^\Delta := (P_{ref,i}^{-1} P_{ref,i+\Delta})^{-1} (P_{est,i}^{-1} P_{est,i+\Delta})$$

$$RPE_{trans}^{i,\Delta} = \left(\frac{1}{m} \sum_{i=1}^m \|\text{trans}(F_i^\Delta)\|^2 \right)^{\frac{1}{2}}$$

$$RPE_{rot}^{i,\Delta} = \frac{1}{m} \sum_{i=1}^m \angle(\text{rot}(F_i^\Delta))$$

где Δ — фиксированный временной интервал.

4. Обзор визуальной SLAM-системы

Ключевыми компонентами SLAM-системы являются модуль отслеживания (tracking) и модуль локального картографирования (local mapping). Система работает покадрово: на каждом кадре оценивается положение камеры, осуществляется управление динамической 3D-картой, а также принимается решение о добавлении нового ключевого кадра.

SLAM-система реализована на основе графов. Поддержание структурной согласованности карты обеспечивается за счёт графа совместной видимости (covisibility graph), в котором вершины соответствуют ключевым кадрам, а рёбра соединяют пары кадров, наблюдающих одни и те же 3D-точки карты. Для оптимизации этой структуры используется метод Левенберга-Марквардта, реализованный в фреймворке g2o [73].

4.1. Детектирование и ассоциация ключевых точек

В разрабатываемой системе реализован комплекс современных алгоритмов детекции и ассоциации ключевых точек, основанных на глубоком обучении. Для обнаружения ключевых точек были интегрированы два нейросетевых подхода: самообучаемая свёрточная нейронная сеть SuperPoint [14] и свёрточная нейросеть XFeat [70].

Для решения задачи ассоциации ключевых точек между кадрами в системе реализованы три современных алгоритма. LightGlue [42], основанный на архитектуре трансформеров, позволяет учитывать глобальный контекст изображения при сопоставлении особенностей, что существенно повышает точность по сравнению с традиционными методами. Алгоритм XFeat [70] для работы с одноименным детектором. GlueStick [28] — графовая нейронная сеть для сопоставления ключевых точек, линий.

4.2. Инициализация карты

Процедура инициализации предназначена для построения геометрически согласованной 3D-реконструкции на основе двух изображений с использованием сопоставленных ключевых точек. Сначала точки текущего кадра сравниваются с точками из предыдущего опорного кадра. Если число корректных соответствий оказывается недостаточным, а временной интервал между кадрами превышает допустимый порог, система сдвигает опорный кадр, чтобы обеспечить необходимый параллакс и достаточное перекрытие особенностей.

При наличии достаточного числа соответствий essential matrix (матрица, связывающая координаты соответствующих точек на двух изображениях) оценивается с помощью алгоритма MAGSAC [5], обеспечивающего устойчивое оценивание модели в условиях наличия выбросов. Относительное положение камеры между двумя кадрами восстанавливается путём декомпозиции essential matrix. При необходимости оценка положения дополнительно уточняется методом, описанным в статье [55]. Затем выполняется триангуляция 3D-точек на основе восстановленного положения и найденных соответствий. Все перечисленные этапы (оценка essential matrix, восстановление положения и триангуляция) реализованы с использованием стандартных функций библиотеки OpenCV [7]. Для повышения точности начальной реконструкции выполняется метод bundle adjustment, совместно оптимизирующий положения камеры и координаты 3D-точек.

4.3. Модуль отслеживания

Данный модуль оценивает положение камеры для каждого нового кадра, уточняет его с учётом данных карты, управляет добавлением ключевых кадров и поддерживает качество карты через фильтрацию 3D-точек.

4.3.1. Первоначальная оценка положения

Для каждого нового кадра система определяет ключевые точки и вычисляет их дескрипторы. Затем эти особенности сопоставляются с опорным ключевым кадром. На основе полученных соответствий выполняется первоначальная оценка положения камеры с помощью *motion-only bundle adjustment* — оптимизации положения камеры при неизменной 3D-структуре карты.

4.3.2. Уточнение положения

После первоначальной оценки система выполняет шаг уточнения, используя 2D–3D соответствия, полученные путём проекции всех существующих точек карты в текущий кадр. Каждая проекция проверяется на видимость по следующим критериям:

- Проекция точки находится в пределах изображения.
- Угол между текущим направлением обзора и средним направлением наблюдения точки не превышает определенного порога.
- Глубина точки лежит в заданном масштабно-инвариантном диапазоне видимости.

Если точка проходит проверки, она сопоставляется с ближайшими неиспользованными ключевыми точками на основе сходства дескрипторов. Полученные 2D–3D соответствия применяются во втором шаге *bundle adjustment* для более точного уточнения положения.

4.3.3. Фильтрация точек карты

Чтобы избежать ухудшения качества карты со временем, для каждой 3D-точки постоянно оценивается частота её успешных обнаружений. Точка считается надёжной, только если она успешно наблюдается как минимум в 25% кадров из тех, где она теоретически должна быть видна. Точки, не соответствующие этому критерию, удаляются из карты. Это обеспечивает использование исключительно стабильных и

достоверных ориентиров для оценки положения камеры и расширения карты.

4.3.4. Стратегия вставки ключевых кадров

Система постоянно оценивает необходимость добавления текущего кадра в качестве нового ключевого. Используются два критерия:

- Накопленное изменение угла поворота (в углах Эйлера) с последнего ключевого кадра превышает определенный порог.
- Нормированное смещение (отношение перемещения камеры к медианной глубине сцены) превышает заданный порог.

При выполнении любого из условий создаётся новый ключевой кадр, что обеспечивает точность отслеживания и поддерживает достаточную плотность карты.

4.4. Локальное картографирование

Данный модуль отвечает за интеграцию новых ключевых кадров, расширение карты посредством триангуляции и уточнение локальной карты с помощью bundle adjustment для обеспечения геометрической согласованности и точности отслеживания.

При добавлении нового ключевого кадра в систему он немедленно включается в граф совместной видимости (covisibility graph) в качестве нового узла. Новый узел соединяется с предыдущими узлами, если количество общих наблюдаемых 3D-точек превышает определенный порог.

Для расширения карты устанавливаются соответствия особенностей между новым ключевым кадром и его соседями в графе совместной видимости. Перед триангуляцией отбрасываются соответствия, нарушающие эпиполярное ограничение, чтобы гарантировать геометрическую согласованность. Оставшиеся соответствия используются для триангуляции кандидатов в 3D-точки, которые затем проверяются по следующим критериям: положительная глубина в обоих кадрах, достаточный

параллакс между камерами наблюдения, низкая ошибка репроекции, соответствие ожидаемому масштабу сцены. Точки, удовлетворяющие всем условиям, сохраняются в карте.

Для уточнения структуры локальной карты и повышения точности положений выполняется локальный bundle adjustment после добавления каждого нового ключевого кадра. В процессе оптимизации совместно корректируются: положение текущего ключевого кадра, положения его ближайших соседей в графе совместной видимости, позиции всех общих 3D-точек.

Дополнительные ключевые кадры, наблюдающие эти точки, но не связанные напрямую с текущим кадром, фиксируются для сохранения глобальной согласованности. Выбросы идентифицируются и исключаются как на промежуточных этапах, так и по завершении оптимизации.

5. Реализация

В качестве языка программирования был использован Python, так как SLAM-система, разрабатываемая в лаборатории, реализуется на этом языке, и все нейросетевые алгоритмы, используемые в данной работе, построены на Python фреймворках.

Архитектура SLAM-системы представлена на рисунке 2. Система состоит из компонентов: **System**, **FeatureDetectors**, **FeatureMatchers**, **Filters**, **Dataset**, **Algorithm**. В данной работе были реализованы компоненты **Dataset**, **Algorithm**, **Filters** и интегрированы алгоритмы детекции и ассоциации ключевых точек. Далее будут подробно описаны реализованные компоненты в данной работе.

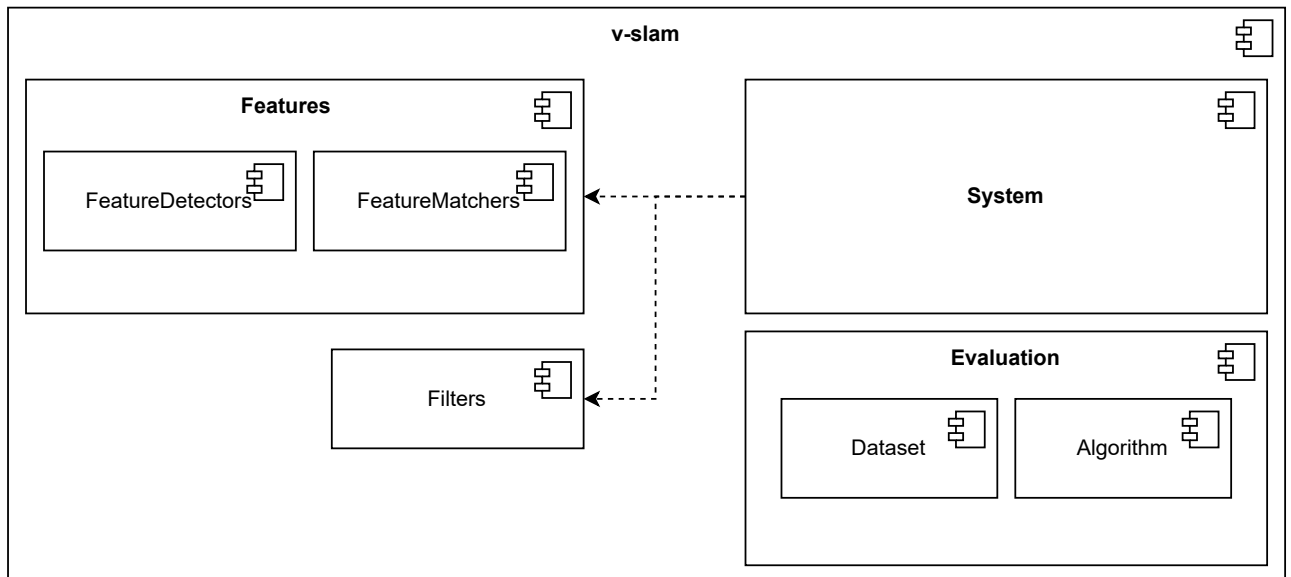


Рис. 2: Архитектура системы

Компонент **Filters** предоставляет алгоритмы для предобработки тепловизионных снимков. Были реализованы следующие алгоритмы: выравнивание гистограммы, CLАНЕ, двусторонний фильтр (bilateral filter), полосовой фильтр в частотной области (bandpass filter), медианный фильтр. Соответствующая диаграмма классов на рисунке 3 показывает зеленым цветом разработанные в данной работе классы.

Компонент **FeatureDetectors** включает алгоритмы детекции ключевых точек. В системе уже были реализованы такие алгоритмы,

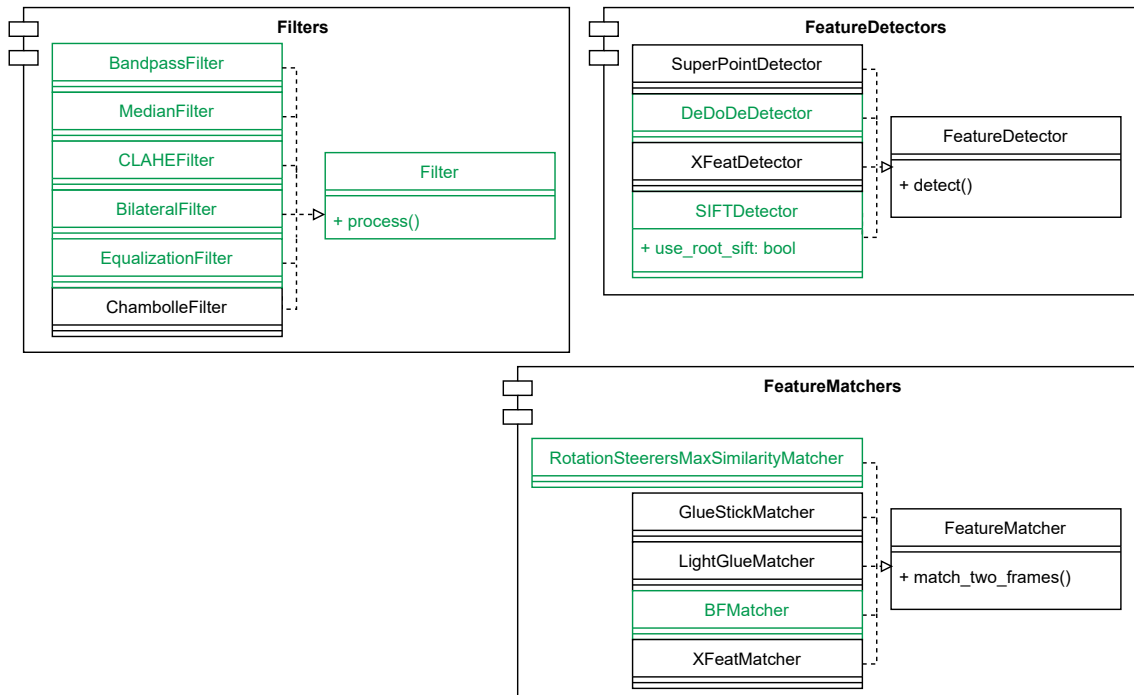


Рис. 3: UML-диаграмма добавленных классов

как SuperPoint [14] и XFeat [70]. В данной работе дополнительно интегрированы SIFT [44], RootSIFT [2] и DeDoDe [13], а также алгоритмы сопоставления: Brute-Force matcher из OpenCV [7] и RotationSteerersMaxSimilarity [60]. Ранее в системе уже были реализованы GlueStick [28], LightGlue [42] и XFeat [70]. На диаграмме классов 3 зеленым выделены добавленные в работе реализации. Служебные компоненты **Dataset** и **Algorithm** обеспечивают работу с входными данными. **Dataset** отвечает за чтение траекторий из различных наборов данных, в то время как компонент **Algorithm** обеспечивает их предобработку в зависимости от выбранного SLAM-алгоритма. Также данный компонент предоставляет вычисление метрик качества SLAM-алгоритмов с использованием библиотеки evo [31].

6. Экспериментальное сравнение

В данном разделе представлены экспериментальные сравнения алгоритмов сопоставления ключевых точек и методов предобработки, а также адаптированной для тепловизионных снимков SLAM-системы с другими алгоритмами. Поскольку для алгоритмов ассоциации ключевых точек отсутствуют эталонные данные, их оценка будет осуществляться как визуально, так и с использованием численных метрик, таких как количество найденных ассоциаций. Качество предобработки тепловизионных снимков будет определяться количеством сопоставлений ключевых точек.

6.1. Цель и вопросы эксперимента

Целью проводимого эксперимента является исследование эффективности методов предобработки тепловизионных снимков и сопоставления ключевых точек в рамках SLAM-алгоритмов. Для достижения данной цели были поставлены следующие исследовательские вопросы:

- RQ1: Какое сочетание алгоритмов детекции и сопоставления ключевых точек является наиболее эффективным для тепловизионных изображений?
- RQ2: Какой конфигурация алгоритмов предобработки тепловизионных изображений обеспечивает наибольшее количество сопоставлений ключевых точек?
- RQ3: Каково качество работы SLAM-системы после ее адаптации к тепловизионным снимкам по сравнению с другими алгоритмами?

6.2. Условия эксперимента

6.2.1. Наборы данных

В качестве тестовых наборов данных использовались датасет ViViD++ [69] и набор данных из [24]. Выбор таких тестовых данных обусловлен необходимостью всесторонней оценки работы алгоритмов в различных условиях, включая разные уровни освещенности, разнообразные траектории движения камеры, а также съемку как в помещениях, так и на улице. Тепловизионные снимки для датасета ViViD++ были получены с помощью камеры FLIR A65, в то время как изображения для второго датасета снимались на камеру FLIR Tau2 LWIR. Оба набора данных содержат снимки с разрешением 640x512 пикселей. Датасет из статьи [24] включает шесть последовательностей, каждая из которых разделена на несколько подпоследовательностей — в общей сложности шестьдесят пять. Экспериментальное сравнение алгоритмов предобработки тепловизионных снимков, детекции и ассоциации ключевых точек было проведено на одной последовательности из ViViD++ [69] и одной последовательности из датасета из статьи [24]. Экспериментальное сравнение SLAM-алгоритмов проводилось на шестидесяти пяти последовательностях из датасета [24] и на четырнадцати последовательностях (из тридцати шести имеющихся) из набора данных ViViD++ [69].

6.3. Результаты

6.3.1. RQ1

Была проведена оценка методов детекции и ассоциации ключевых точек для тепловизионных снимков. Для экспериментального сравнения алгоритмов были взяты следующие: SuperPoint [14] + LightGlue [42], SuperPoint [14] + Gluestick [28], RootSIFT [2] + Lightglue [42], SIFT [44] + LightGlue [42], DeDoDe [13] + MaxSimilarity [60], XFeat [70] + XFeatMatcher [70], SIFT [44] + BFMatcher, RootSIFT [2] + BFMatcher, ORB [49] + BFMatcher. Для сравнения были выбраны основные современные нейросетевые методы (SuperPoint [14], DeDoDe [13],

LightGlue [42], MaxSimilarity [60], GlueStick [28]) и классические алгоритмы (ORB [49], SIFT [44], RootSIFT [2], BFMatcher) детекции и ассоциации ключевых точек. Также были включены комбинированные подходы, сочетающие классические детекторы с нейросетевыми методами сопоставления (SIFT [44] + LightGlue [42], RootSIFT [2] + LightGlue [42]). Сопоставление ключевых точек происходило между парами кадров, разделенных временным интервалом (кадры i и $i + 10$). На рисунке 4 можно наблюдать, что нейросетевой алгоритм ассоциации XFeat [70] показывает большее количество сопоставлений. Следующие позиции по количеству сопоставлений занимают: DeDoDe [13] + MaxSimilarity [60], ORB [49] + BFMatcher, RootSIFT [2] + Lightglue [42] и SuperPoint [14] + LightGlue [42]. Хотя комбинация SuperPoint [14] + LightGlue [42] показывает меньшие результаты по количеству ассоциаций по сравнению с лидирующими методами, данный подход был выбран для дальнейшего рассмотрения, так как он является одним из популярных подходов к детекции и ассоциации.

На рисунках 5, 7, 6, 8, 9 можно наблюдать визуализацию ассоциаций. Можно заключить, что RootSIFT [2] + LightGlue [42], DeDoDe [13] + MaxSimilarity [60], ORB [49] + BFMatcher генерируют ложноположительные ассоциации даже на относительно простых тепловизионных сценах. В то же время, комбинации SuperPoint [14] + LightGlue [42] и XFeat [70] + XFeatMatcher демонстрируют большую устойчивость.

Далее было рассмотрено влияние XFeat [70] + XFeatMatcher и SuperPoint [14] + LightGlue [42] на качество адаптированной SLAM-системы. Для данного сравнения были использованы две последовательности: одна из датасета ViViD++ [69], а другая — из статьи [24]. Эти последовательности являются сложными: одна из них (lt3 5) имеет слабо выраженные контуры, а другая (outdoor_robust_night1) отличается большой длиной и замыкается. В качестве нормализации снимков использовалось выравнивание гистограммы. Результаты представлены в таблице 1. Можно сделать вывод, что SLAM-система, применяющая SuperPoint [14] + LightGlue [42], превосходит систему, использующую XFeat [70] + XFeatMatcher, по точности.

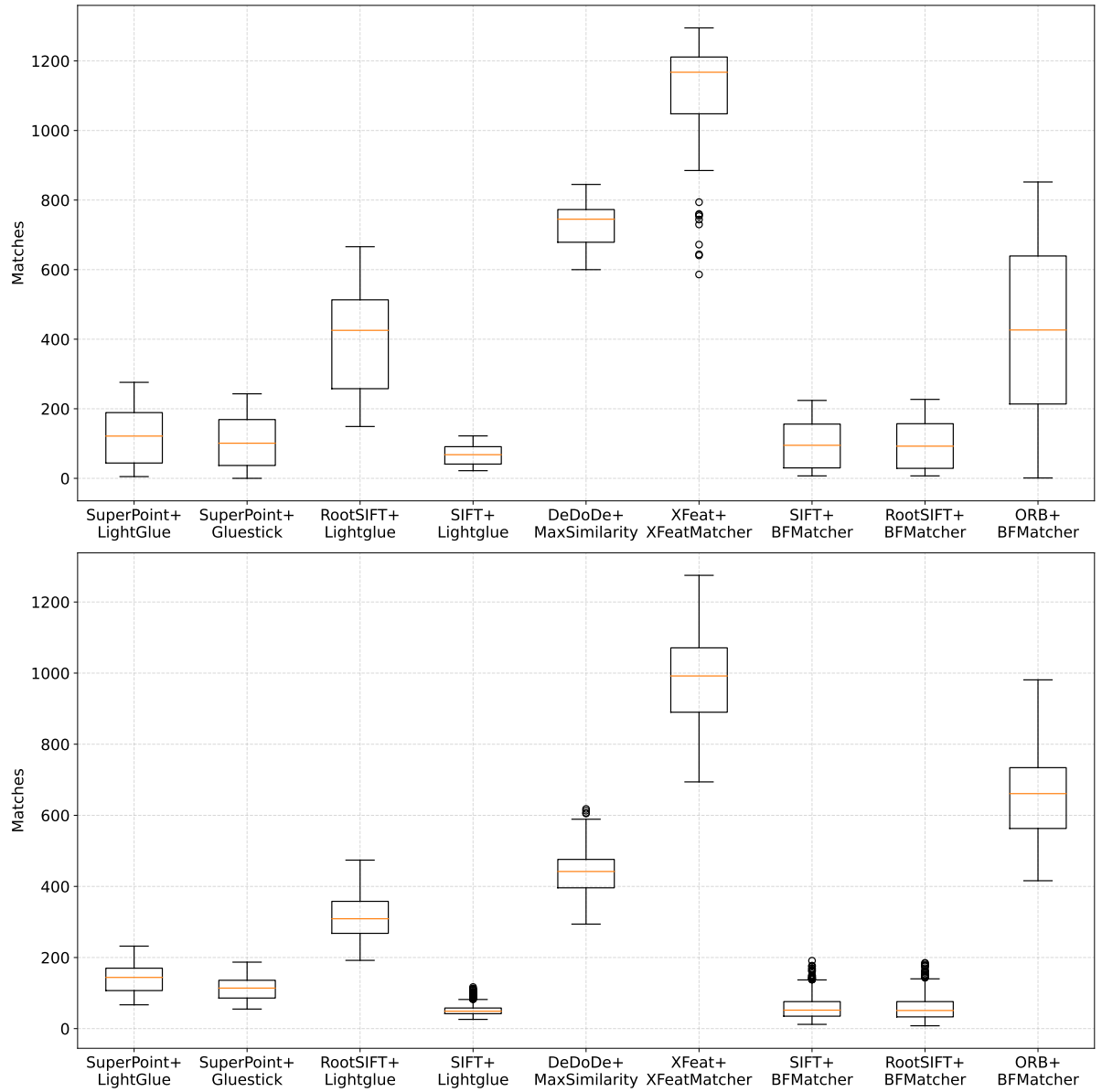


Рис. 4: Сравнение методов детекции и ассоциации. Вверху: результаты для набора данных из [24]. Внизу: результаты для ViViD++ [69]

Последовательность	XFeat [70] + XFeatMatcher				SuperPoint [14] + LightGlue [42]			
	ATE	RPE_{trans}	RPE_{rot}	%	ATE	RPE_{trans}	RPE_{rot}	%
lt3 5	1.04	0.07	2.40	100	0.01	0.06	0.78	100
outdoor_robust_night1	3.68	0.73	5.52	100	2.77	0.64	5.27	100

Таблица 1: Результаты метрик ATE , RPE_{trans} и RPE_{rot}

Таким образом, для дальнейшей работы были выбраны SuperPoint [14] и LightGlue [42].

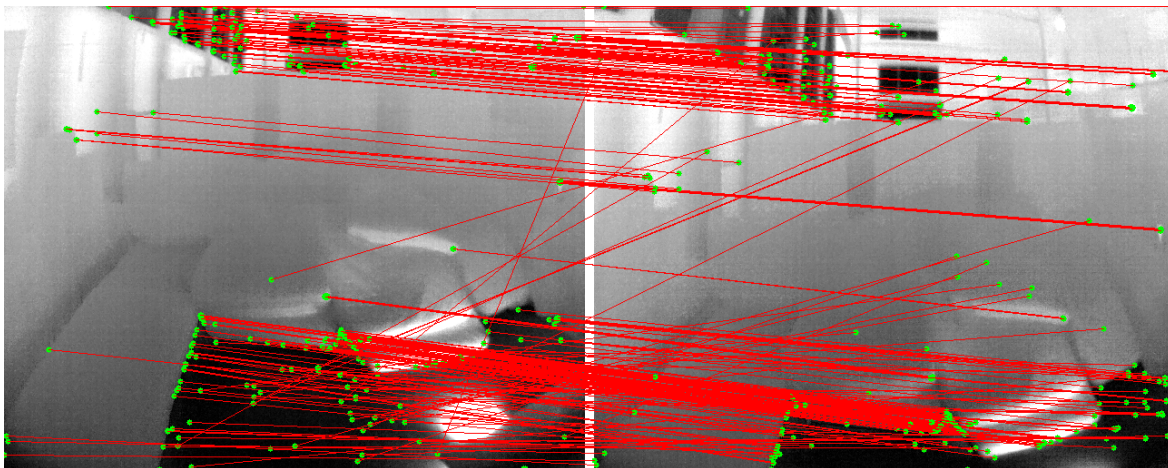


Рис. 5: Визуализация ассоциаций DeDoDe [13] + MaxSimilarity [60]

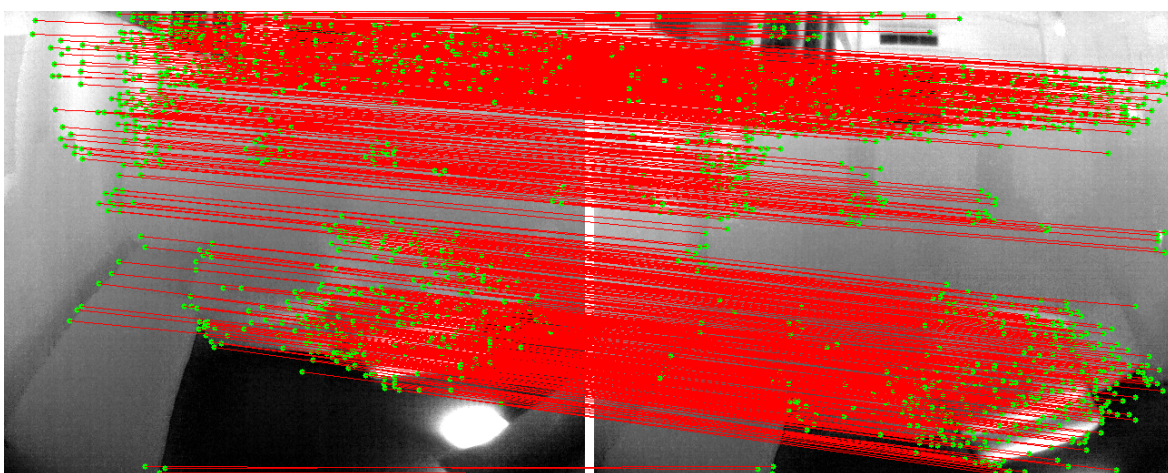


Рис. 6: Визуализация ассоциаций XFeat [70] + XFeatMatcher

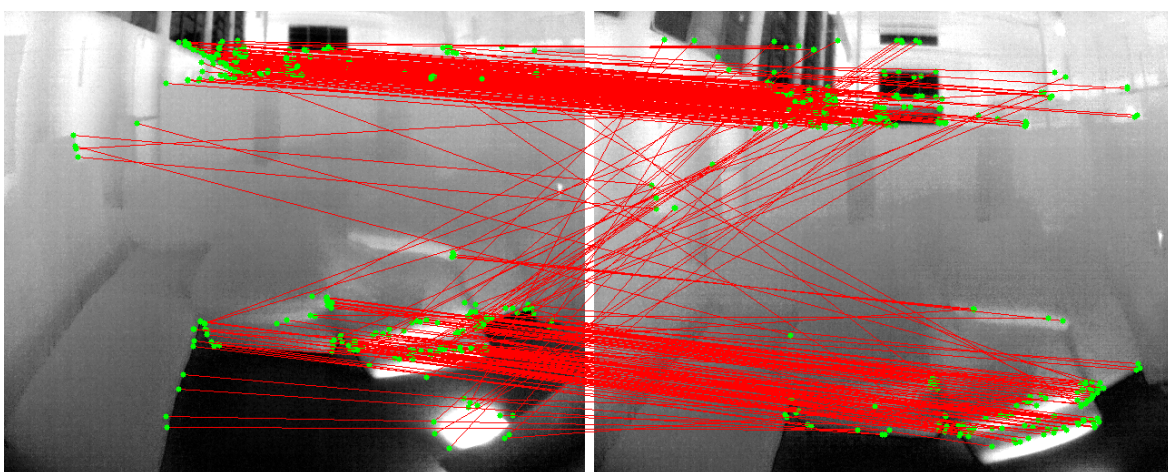


Рис. 7: Визуализация ассоциаций ORB [49] + BFMatcher

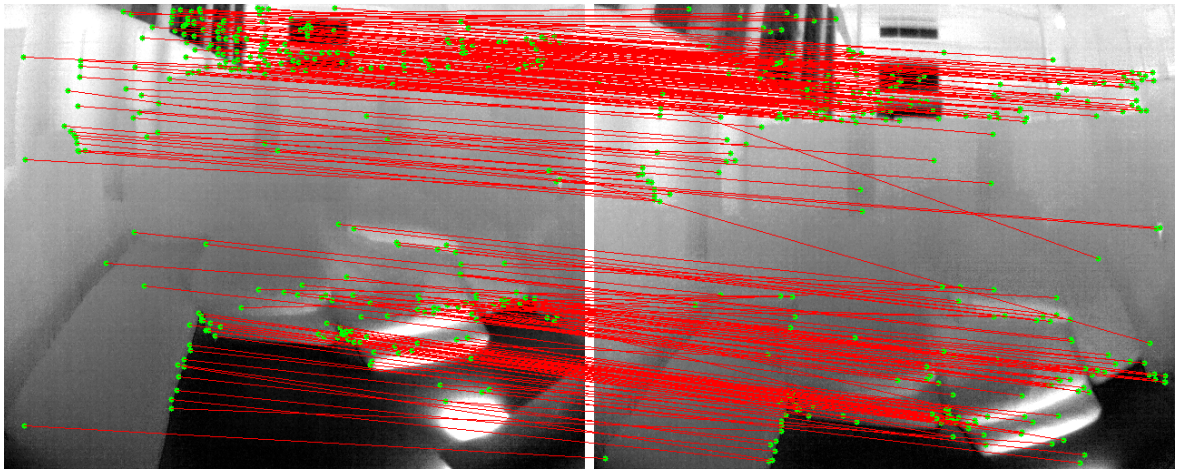


Рис. 8: Визуализация ассоциаций RootSIFT [2] + LightGlue [42]

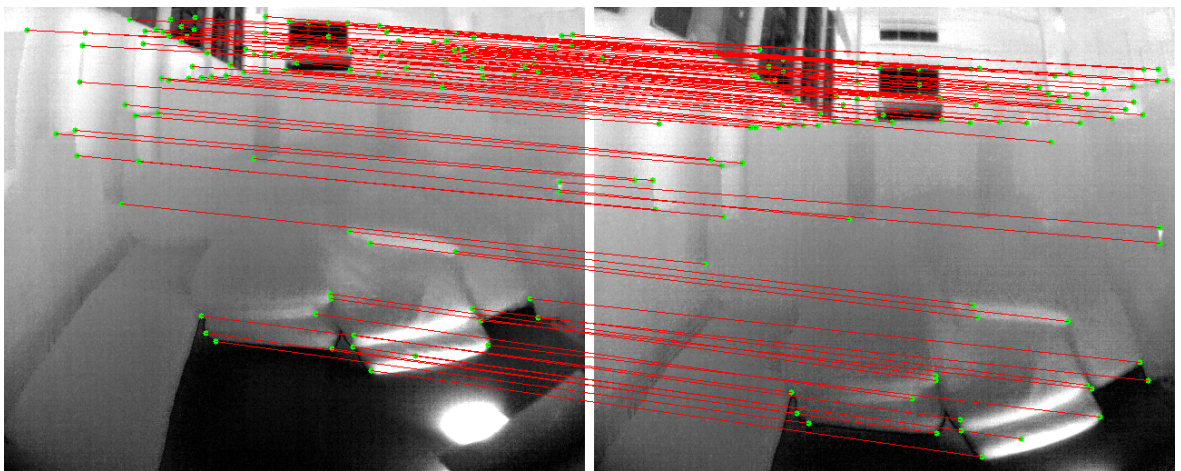


Рис. 9: Визуализация ассоциаций SuperPoint [14] + LightGlue [42]

6.3.2. RQ2

Была проведена оценка методов предварительной обработки тепловизионных изображений. Для проведения экспериментального сравнения в качестве алгоритмов детекции и ассоциации ключевых точек были использованы нейросети SuperPoint [14] и LightGlue [42] соответственно. Сопоставление ключевых точек происходило между парами кадров, разделенных временным интервалом (кадры i и $i + 10$). Этот процесс повторялся для различных конфигураций фильтров предварительной обработки. Следует отметить, что все методы удаления шума применялись к необработанным данным, за исключением медианного фильтра, который мог использоваться и после нормализации, так как его работа не зависит от конкретных значений пикселей. Настройка па-

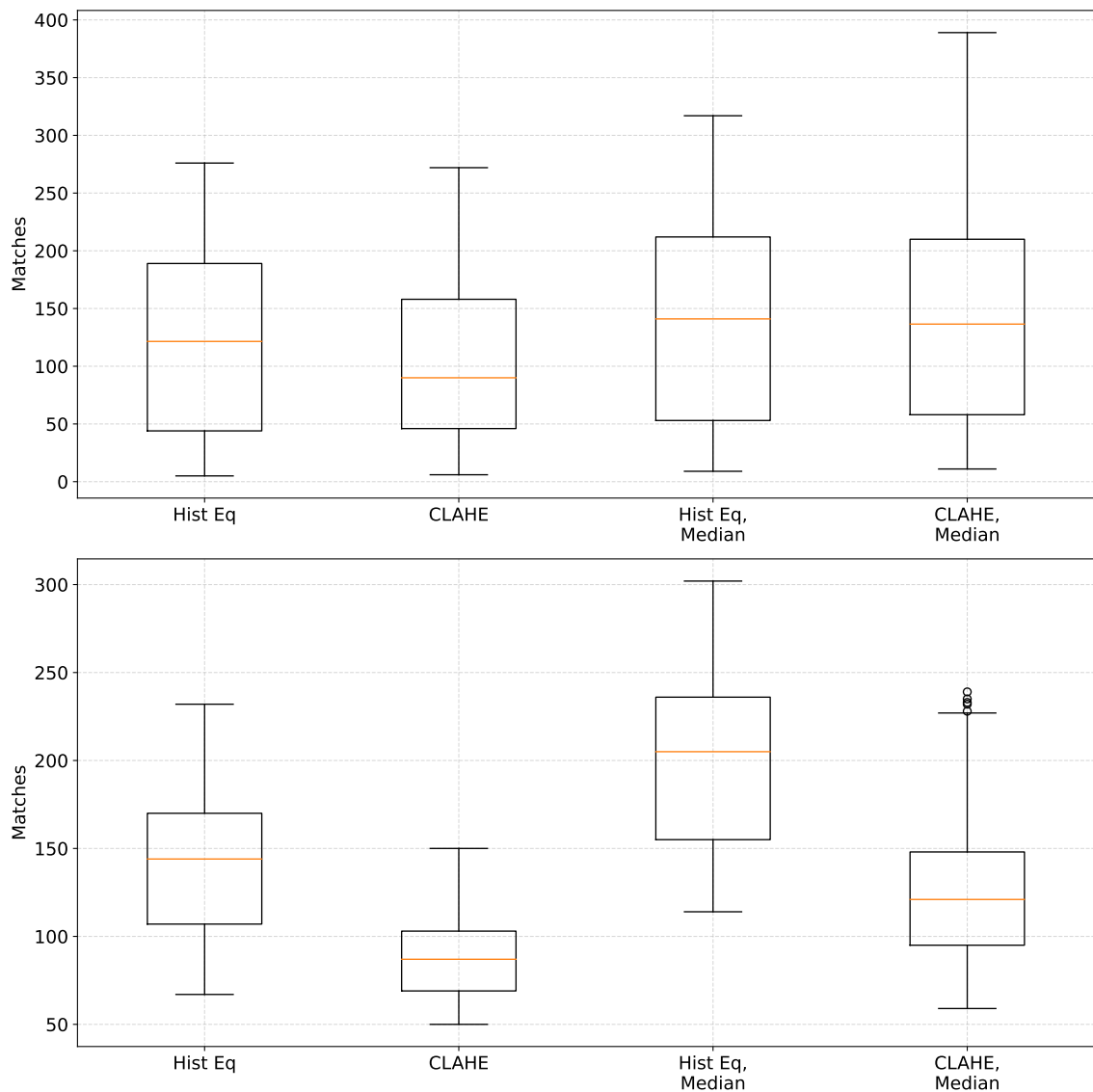


Рис. 10: Сравнение методов предобработки. Вверху: результаты для набора данных из [24]. Внизу: результаты для ViViD++ [69]

раметров фильтров проводилась итеративным методом. На первом этапе для десяти выбранных изображений из двух датасетов были определены диапазоны значений параметров, обеспечивающие визуально приемлемое качество обработки. На втором этапе выполнялась точная настройка параметров на основе количества ассоциаций ключевых точек.

В первую очередь были исследованы методы нормализации: выравнивание гистограммы (histogram equalization) и CLAHE. Результаты для датасета ViViD++ [69] и набора данных из [24] можно наблюдать на графиках 10. Можно сделать вывод, что классическое выравнивание гистограммы демонстрирует более высокую эффективность в сравне-

нии с CLANE по числу сопоставленных ключевых точек. Также стоит отметить, что после применения медианного фильтра число сопоставлений ключевых точек увеличивается.

Для дальнейших сравнений в качестве методов обработки были выбраны: стандартное выравнивание гистограммы для нормализации и медианный фильтр для подавления шума. На следующем этапе были протестированы следующие сочетания алгоритмов:

- Фильтр Chambolle [10] + выравнивание гистограммы + медианный фильтр (median).
- Полосовой фильтр (bandpass) + выравнивание гистограммы + медианный фильтр (median).
- Двусторонний фильтр (bilateral) + выравнивание гистограммы + медианный фильтр (median).
- Медианный фильтр + выравнивание гистограммы + медианный фильтр (median).

Результаты для датасета ViViD++ [69] и набора данных из [24] можно наблюдать на рисунке 11. Можно наблюдать, что конфигурация алгоритмов предобработки для тепловизионных снимков Chambolle фильтр + выравнивание гистограммы + медианный фильтр обеспечивает наибольшее количество сопоставлений ключевых точек. На рисунках 9, 12 можно наблюдать визуализацию ассоциаций только с выравниванием гистограммы и с вышеуказанной комбинацией соответственно. Можно сделать вывод, что конфигурация алгоритмов предобработки тепловизионных снимков Chambolle фильтр, выравнивание гистограммы, медианный фильтр обеспечивает как и наибольшее количество сопоставлений, также SuperPoint [14] + LightGlue [42] сохраняет свою точность.

6.3.3. RQ3

Было проведено экспериментальное сравнение SLAM-системы, адаптированной к тепловизионным снимкам путем подбора оптимальных

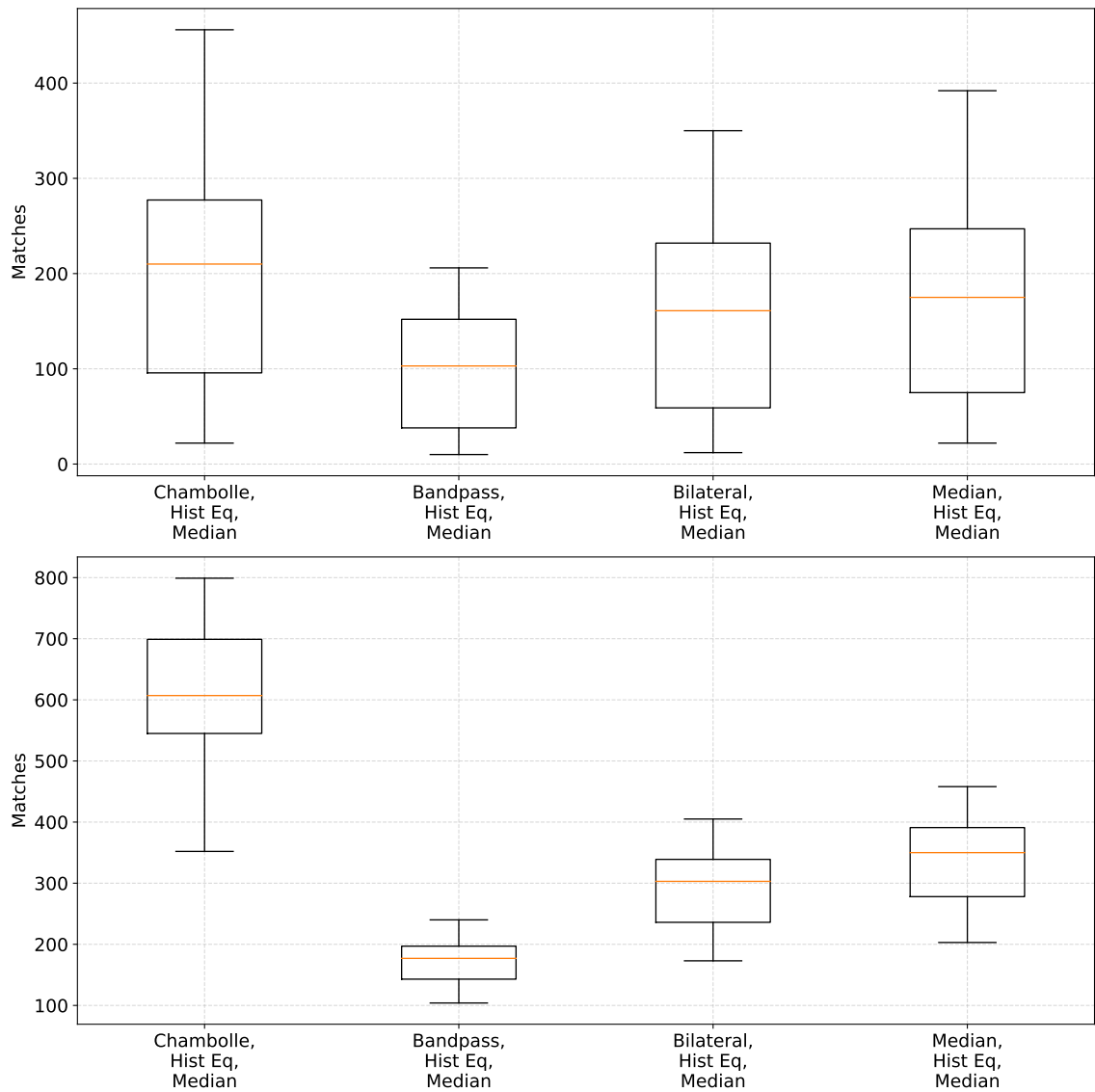


Рис. 11: Сравнение методов предобработки. Вверху: результаты для набора данных из [24]. Внизу: результаты для ViViD++ [69]

параметров, алгоритмов предобработки, детекции и ассоциации ключевых точек, с несколькими современными подходами, широко используемыми в области VIO и SLAM-систем, а именно: ORB-SLAM3 [51], ROTIO [52] и DSO [20]. Данные алгоритмы являются общепризнанными эталонами в современных исследованиях тепловизионных VIO/SLAM-систем. Далее разрабатываемая в лаборатории SLAM-система, адаптированная под тепловизионные снимки, будет обозначаться v-slam.

Все базовые методы оценивались с использованием их публично доступных реализаций и параметров по умолчанию. Для ROTIO был сохранен исходный метрический масштаб, полученный в результате обра-

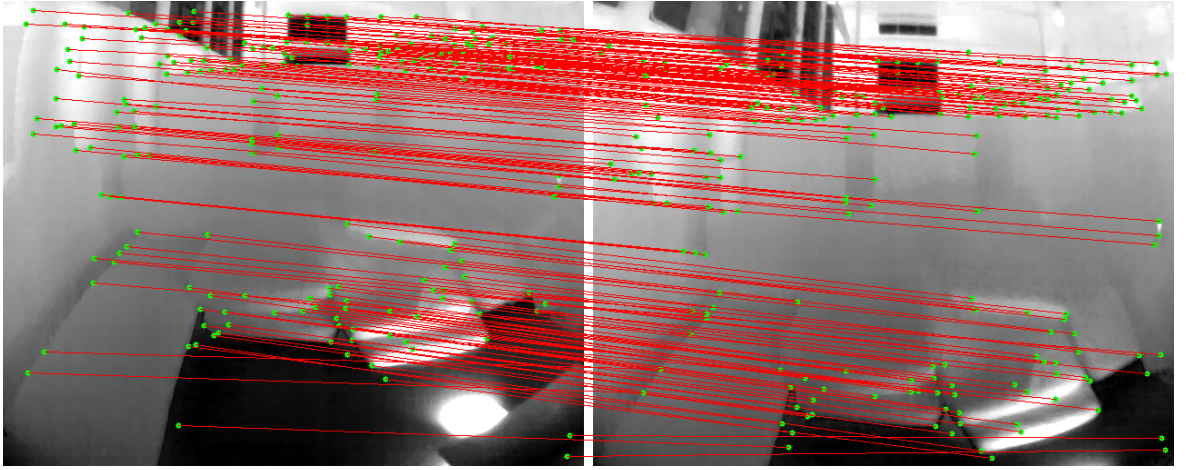


Рис. 12: Визуализация ассоциаций с предобработкой Chambolle фильтр + выравнивание гистограммы + медианный фильтр

Последовательность	Тип	ORB SLAM3				ROTIO				DSO				v-slam			
		<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%	<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%	<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%	<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%
alt1	Mean	0.76	0.16	5.39	93.06	22.33	1.13	0.84	100	0.29	0.15	5.31	92.38	0.14	0.16	5.16	100
	Median	0.66	0.14	4.74	100	0.26	0.05	0.86	100	0.24	0.16	4.80	100	0.12	0.15	4.59	100
alt2	Mean	0.85	0.17	5.88	100	0.31	0.05	0.91	100	0.84	0.17	5.79	100	0.30	0.17	5.27	100
	Median	0.71	0.17	5.99	100	0.25	0.05	1.00	100	0.59	0.15	5.48	100	0.13	0.17	5.07	100
sync	Mean	0.15	0.01	0.73	92.95	0.13	< 0.01	0.67	100	0.44	0.01	0.72	100	0.15	0.01	0.78	100
	Median	0.06	0.01	0.36	100	0.12	< 0.01	0.28	100	0.40	0.01	0.35	100	0.08	0.01	0.38	100
lt1	Mean	0.06	0.01	0.21	98.53	0.21	0.01	0.17	100	0.18	0.01	0.21	100	0.07	0.01	0.20	98.75
	Median	0.04	0.01	0.16	100	0.18	0.01	0.11	100	0.03	0.01	0.15	100	0.04	0.02	0.16	100
lt2	Mean	0.30	0.22	3.79	85.12	3.44	0.72	0.79	100	0.51	0.16	2.78	100	0.14	0.14	2.74	100
	Median	0.26	0.19	3.32	100	0.29	0.05	0.82	100	0.46	0.18	3.01	100	0.12	0.15	2.83	100
lt3	Mean	0.35	0.12	3.02	87.18	13.40	0.98	0.69	100	0.52	0.13	2.76	100	0.13	0.12	2.62	100
	Median	0.10	0.08	2.67	100	0.11	0.02	0.40	100	0.45	0.07	2.49	100	0.09	0.07	2.44	100

Таблица 2: Результаты метрик ATE , RPE_{trans} и RPE_{rot} на наборе данных из статьи [24]. Для каждой последовательности указаны среднее (Mean) и медианное (Median) значения по всем подпоследовательностям. % обозначает долю отслеживаемых кадров. Наилучшие результаты по каждой метрике выделены жирным шрифтом.

ботки данных его визуально-инерциальной системой. В случае монокулярных ORB-SLAM3 и DSO коррекция масштаба траектории выполнялась с использованием фреймворка evo [31], что также потребовалось для v-slam.

Для обеспечения объективного сравнения входные тепловизионные изображения для всех методов (за исключением ROTIO, который работает непосредственно с исходными тепловизионными данными) предварительно обрабатывались с использованием фильтра, предложенного Chambolle, выравнивания гистограммы и медианной фильтрации.

Качество работы систем оценивалось по трем стандартным метри-

Последовательность	ORB SLAM3				ROTIO				DSO				v-slam			
	<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%	<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%	<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%	<i>ATE</i>	<i>RPE_{trans}</i>	<i>RPE_{rot}</i>	%
dark_aggressive	0.05	0.06	5.47	26.70	72.43	0.92	4.74	100	0.65	0.04	5.01	100	< 0.01	< 0.01	0.21	100
dark_robust	0.14	0.01	0.90	28.5	0.14	0.01	0.64	100	0.35	0.01	0.82	100	0.48	0.01	0.93	100
dark_unstable	0.10	0.03	2.74	23.9	28.42	0.34	2.30	100	0.75	0.02	2.79	100	0.73	0.03	2.92	100
global_aggressive	—	—	—	—	7.15	0.13	3.16	100	0.68	0.04	3.85	100	0.01	< 0.01	0.24	100
global_robust	—	—	—	—	0.20	0.01	0.54	100	0.66	0.01	0.66	100	0.27	0.02	0.79	100
global_unstable	—	—	—	—	1.10	0.04	2.31	100	0.63	0.03	3.12	100	—	—	—	—
local_aggressive	0.11	0.05	6.02	17.5	64.45	1.01	4.03	100	0.26	0.17	19.74	27.8	0.01	0.01	0.47	100
local_robust	0.35	0.01	0.86	100	0.13	0.01	0.60	100	0.31	0.01	0.83	100	0.35	0.01	0.88	100
local_unstable	0.10	0.03	2.43	37.1	53.98	0.71	1.69	100	0.40	0.05	3.25	100	0.62	0.02	2.31	100
varying_robust	0.12	0.02	0.97	35.7	0.14	0.01	0.89	100	0.35	0.01	0.99	100	0.26	0.01	1.00	100
outdoor_robust_day1	1.43	2.17	12.47	100	2.92	1.11	12.28	100	2.55	1.16	11.62	100	2.15	1.76	15.38	100
outdoor_robust_day2	2.18	0.71	4.23	100	2664.00	79.64	5.02	100	5.48	0.61	4.28	100	8.54	0.53	4.23	100
outdoor_robust_night1	0.67	0.73	5.51	100	738.83	23.92	6.92	100	4.80	0.81	5.66	100	3.19	0.67	5.37	100
outdoor_robust_night2	0.50	0.66	4.73	100	605.62	15.58	6.96	100	3.17	0.80	5.38	100	6.53	0.81	5.27	100
Mean	0.52	0.41	4.21	60.85	302.82	8.82	3.72	100	1.50	0.27	4.86	94.84	1.78	0.30	3.08	100
Median	0.14	0.05	4.23	37.10	17.78	0.52	2.73	100	0.65	0.04	3.55	100	0.48	0.02	1.00	100

Таблица 3: Результаты метрик ATE , RPE_{trans} и RPE_{rot} на датасете ViViD++ [69]. Дополнительно вычислены среднее (Mean) и медианное (Median) значения по всем последовательностям. % обозначает долю отслеживаемых кадров. Наилучшие результаты по каждой метрике выделены жирным шрифтом.

кам: средняя абсолютная ошибка траектории (ATE), средняя трансляционная относительная ошибка положения (RPE) и средняя угловая RPE . Также фиксировалось количество успешно отслеженных кадров, поскольку некоторые SLAM-алгоритмы подвержены сбоям в трекинге.

Результаты метрик на датасете из [24] представлены в таблице 2. На наборе данных из [24] v-slam продемонстрировал наилучшие показатели по метрике ATE в большинстве случаев. По трансляционной составляющей RPE лучшие результаты, как правило, показывает метод ROTIO, хотя в целом все четыре метода демонстрируют сопоставимую точность. По угловой составляющей RPE снова лидирует ROTIO, тогда как остальные методы показывают меньшую точность. Кроме того, ORB-SLAM3 чаще других теряет трекинг. Следует отметить, что ROTIO часто не может точно восстановить масштаб траектории, что отражается в более высоких значениях ATE .

На наборе данных ViViD++ [69] ситуация иная. Результаты представлены в таблице 3. ORB-SLAM3 показывает наилучшие численные результаты по метрике ATE , но в ряде случаев не может инициализироваться или завершить трекинг последовательностей, что может приводить к искусственно заниженным ошибкам. v-slam и DSO демонстри-

руют схожее качество и успешно обрабатывают большинство последовательностей. По трансляционной составляющей RPE оба метода показывают сопоставимые и наилучшие результаты, тогда как по угловой составляющей RPE v-slam превосходит все остальные. Хотя ROTIO иногда показывает хорошие результаты по обоим метрикам RPE, он часто не может восстановить правильный масштаб траектории, что снижает его общую надежность.

Таким образом, v-slam демонстрирует стабильную работу и превосходит или соответствует по точности современным алгоритмам (ORB-SLAM3, ROTIO и DSO). В отличие от ORB-SLAM3, склонного к сбоям трекинга, v-slam обеспечивает надежную инициализацию и непрерывное отслеживание. При этом ORB-SLAM3 показывает преимущество лишь в сценариях с замыканием петель благодаря соответствующему модулю. ROTIO, несмотря на хорошие показатели относительной точности за счет использования инерциальных данных, часто ошибается в определении масштаба траектории. Алгоритм DSO, хотя и работает стабильно, уступает другим в точности другим методам. Таким образом, v-slam сочетает в себе устойчивость трекинга и точность, что делает его перспективным для практического применения.

6.4. Выводы

Таким образом, можно сделать вывод, что наиболее эффективными алгоритмами для детекции ключевых точек и их ассоциации на тепловизионных снимках являются нейросетевые методы, такие как SuperPoint [14], LightGlue [42]. Наиболее подходящая конфигурация алгоритмов предобработки из рассмотренных является фильтр Chambolle + выравнивание гистограммы + медианный фильтр. Также разрабатываемая визуальная SLAM-система в лаборатории «ИИ для автономных систем», адаптированная к тепловизионным снимкам, демонстрирует стабильную работу и превосходит или соответствует по точности современным алгоритмам (ORB-SLAM3, ROTIO и DSO).

7. Заключение

В результате работы были решены следующие задачи:

1. Выполнен обзор существующих решений SLAM-задачи с использованием тепловизионных снимков, включая наборы данных для экспериментов, методы предобработки, а также алгоритмы детекции и ассоциации ключевых точек.
2. Выполнен обзор визуальной SLAM-системы, разрабатываемой в лаборатории «ИИ для автономных систем».
3. Реализованы в рамках SLAM-системы различные алгоритмы предобработки тепловизионных снимков, а также устойчивые к ним детекторы и алгоритмы ассоциации ключевых точек.
4. Проведено экспериментальное сравнение алгоритмов предобработки, детекции и ассоциации ключевых точек.
5. Выбраны наиболее оптимальные алгоритмы предобработки снимков, детекции, ассоциации особенностей и проведено экспериментальное сравнение полученной системы с другими SLAM-алгоритмами.

Исходный код проекта закрыт, находится под соглашением о неразглашении.

Список литературы

- [1] 4DRT-SLAM: Robust SLAM in Smoke Environments using 4D Radar and Thermal Camera based on Dense Deep Learnt Features / Jun Zhang, Renxiang Xiao, Heshan Li et al. // 2023 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM) / IEEE. — 2023. — P. 19–24.
- [2] Arandjelović Relja, Zisserman Andrew. Three things everyone should know to improve object retrieval // 2012 IEEE conference on computer vision and pattern recognition / IEEE. — 2012. — P. 2911–2918.
- [3] BVT-SLAM: A Binocular Visible-Thermal Sensors SLAM System in Low-Light Environments / Liang Qin, Chang Wu, Xiaotong Kong et al. // IEEE Sensors Journal. — 2023.
- [4] Baker Simon, Matthews Iain. Lucas-kanade 20 years on: A unifying framework // International journal of computer vision. — 2004. — Vol. 56. — P. 221–255.
- [5] Barath Daniel, Matas Jiri, Noskova Jana. MAGSAC: marginalizing sample consensus // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. — 2019. — P. 10197–10205.
- [6] Bilateral filtering: Theory and applications / Sylvain Paris, Pierre Kornprobst, Jack Tumblin et al. // Foundations and Trends® in Computer Graphics and Vision. — 2009. — Vol. 4, no. 1. — P. 1–73.
- [7] Bradski G. The OpenCV library // Dr. Dobb's Journal of Software Tools. — 2000.
- [8] Brief: Binary robust independent elementary features / Michael Calonder, Vincent Lepetit, Christoph Strecha, Pascal Fua // Computer Vision—ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part IV 11 / Springer. — 2010. — P. 778–792.

- [9] Caltech Aerial RGB-Thermal Dataset in the Wild / Connor Lee, Matthew Anderson, Nikhil Ranganathan et al. // European Conference on Computer Vision / Springer. — 2024. — P. 236–256.
- [10] Chambolle Antonin. An algorithm for total variation minimization and applications // Journal of Mathematical imaging and vision. — 2004. — Vol. 20. — P. 89–97.
- [11] Cheeseman P, Smith R, Self M. A stochastic map for uncertain spatial relationships // 4th international symposium on robotic research / MIT Press Cambridge. — 1987. — P. 467–474.
- [12] De Pazzi Davide, Pertile Marco, Chiodini Sebastiano. 3D radiometric mapping by means of lidar SLAM and thermal camera data fusion // Sensors. — 2022. — Vol. 22, no. 21. — P. 8512.
- [13] DeDoDe: Detect, don't describe—Describe, don't detect for local feature matching / Johan Edstedt, Georg Bökman, Mårten Wadenbäck, Michael Felsberg // 2024 International Conference on 3D Vision (3DV) / IEEE. — 2024. — P. 148–157.
- [14] DeTone Daniel, Malisiewicz Tomasz, Rabinovich Andrew. Superpoint: Self-supervised interest point detection and description // Proceedings of the IEEE conference on computer vision and pattern recognition workshops. — 2018. — P. 224–236.
- [15] Design and evaluation of a generic visual slam framework for multi camera systems / Pushyami Kaveti, Shankara Narayanan Vaidyanathan, Arvind Tamil Chelvan, Hanumant Singh // IEEE Robotics and Automation Letters. — 2023.
- [16] Dvt-slam: Deep-learning based visible and thermal fusion slam / Ruochen Wang, Ling Pei, Lei Chu et al. // China Satellite Navigation Conference (CSNC 2021) Proceedings: Volume II / Springer. — 2021. — P. 394–403.

- [17] EIL-SLAM: Depth-enhanced edge-based infrared-LiDAR SLAM / Wenqiang Chen, Yu Wang, Haoyao Chen, Yunhui Liu // Journal of Field Robotics. — 2022. — Vol. 39, no. 2. — P. 117–130.
- [18] Edge-based monocular thermal-inertial odometry in visually degraded environments / Yu Wang, Haoyao Chen, Yufeng Liu, Shiwu Zhang // IEEE Robotics and Automation Letters. — 2023. — Vol. 8, no. 4. — P. 2078–2085.
- [19] Emilsson Erika, Rydell Joakim. Chameleon on fire—thermal infrared indoor positioning // 2014 IEEE/ION Position, Location and Navigation Symposium-PLANS 2014 / IEEE. — 2014. — P. 637–644.
- [20] Engel Jakob, Koltun Vladlen, Cremers Daniel. Direct sparse odometry // IEEE transactions on pattern analysis and machine intelligence. — 2017. — Vol. 40, no. 3. — P. 611–625.
- [21] Engel Jakob, Schöps Thomas, Cremers Daniel. LSD-SLAM: Large-scale direct monocular SLAM // European conference on computer vision / Springer. — 2014. — P. 834–849.
- [22] Fischler Martin A, Bolles Robert C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM. — 1981. — Vol. 24, no. 6. — P. 381–395.
- [23] Fix Evelyn. Discriminatory analysis: nonparametric discrimination, consistency properties. — USAF school of Aviation Medicine, 1985. — Vol. 1.
- [24] Flemmen Henrik Dobbe. Rovtio: Robust visual thermal inertial odometry. — 2021.
- [25] Forster Christian, Pizzoli Matia, Scaramuzza Davide. SVO: Fast semi-direct monocular visual odometry // 2014 IEEE international conference on robotics and automation (ICRA) / IEEE. — 2014. — P. 15–22.

- [26] Fusion of visible and thermal imagery improves situational awareness / Alexander Toet, Jan Kees IJspeert, Allen M Waxman, Mario Aguilar // Displays. — 1997. — Vol. 18, no. 2. — P. 85–95.
- [27] Geiger Andreas, Ziegler Julius, Stiller Christoph. Stereoscan: Dense 3d reconstruction in real-time // 2011 IEEE intelligent vehicles symposium (IV) / Ieee. — 2011. — P. 963–968.
- [28] Gluestick: Robust image matching by sticking points and lines together / Rémi Pautrat, Iago Suárez, Yifan Yu et al. // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2023. — P. 9706–9716.
- [29] Gonzales Rafael C, Wintz Paul. Digital image processing. — Addison-Wesley Longman Publishing Co., Inc., 1987.
- [30] Graph-based thermal-inertial SLAM with probabilistic neural networks / Muhamad Risqi U Saputra, Chris Xiaoxuan Lu, Pedro Porto B de Gusmao et al. // IEEE Transactions on Robotics. — 2021. — Vol. 38, no. 3. — P. 1875–1893.
- [31] Grupp Michael. evo: Python package for the evaluation of odometry and SLAM. — <https://github.com/MichaelGrupp/evo>. — 2017.
- [32] Huang Thomas, Yang GJTGY, Tang Greory. A fast two-dimensional median filtering algorithm // IEEE transactions on acoustics, speech, and signal processing. — 1979. — Vol. 27, no. 1. — P. 13–18.
- [33] Image denoising using a generative adversarial network / Abeer Al-saiari, Ridhi Rustagi, A’aeshah Alhakamy et al. // 2019 IEEE 2nd international conference on information and computer technologies (ICICT) / IEEE. — 2019. — P. 126–132.
- [34] Kennedy Howard V. Modeling noise in thermal imaging systems // Infrared Imaging Systems: Design, Analysis, Modeling, and Testing IV / SPIE. — Vol. 1969. — 1993. — P. 66–77.

- [35] Khattak Shehryar, Papachristos Christos, Alexis Kostas. Keyframe-based direct thermal-inertial odometry // 2019 International conference on robotics and automation (ICRA) / IEEE. — 2019. — P. 3563–3569.
- [36] Khattak Shehryar, Papachristos Christos, Alexis Kostas. Visual-thermal landmarks and inertial fusion for navigation in degraded visual environments // 2019 IEEE Aerospace Conference / IEEE. — 2019. — P. 1–9.
- [37] Khattak Shehryar, Papachristos Christos, Alexis Kostas. Keyframe-based thermal-inertial odometry // Journal of Field Robotics. — 2020. — Vol. 37, no. 4. — P. 552–579.
- [38] Krombach Nicola, Droschel David, Behnke Sven. Combining feature-based and direct methods for semi-dense real-time stereo visual odometry // Intelligent Autonomous Systems 14: Proceedings of the 14th International Conference IAS-14 14 / Springer. — 2017. — P. 855–868.
- [39] LSD: A fast line segment detector with a false detection control / Rafael Grompone Von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, Gregory Randall // IEEE transactions on pattern analysis and machine intelligence. — 2008. — Vol. 32, no. 4. — P. 722–732.
- [40] Lee Seong Hun, Civera Javier. Loosely-coupled semi-direct monocular slam // IEEE Robotics and Automation Letters. — 2018. — Vol. 4, no. 2. — P. 399–406.
- [41] Ligocki Adam, Jelinek Ales, Zalud Ludek. Brno Urban Dataset-The New Data for Self-Driving Agents and Mapping Tasks // 2020 IEEE International Conference on Robotics and Automation (ICRA) / IEEE. — 2020. — P. 3284–3290.
- [42] Lindenberger Philipp, Sarlin Paul-Edouard, Pollefeys Marc. Lightglue: Local feature matching at light speed // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2023. — P. 17627–17638.

- [43] LoFTR: Detector-free local feature matching with transformers / Jiaming Sun, Zehong Shen, Yuang Wang et al. // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. — 2021. — P. 8922–8931.
- [44] Lowe David G. Distinctive image features from scale-invariant keypoints // International journal of computer vision. — 2004. — Vol. 60. — P. 91–110.
- [45] Marr David, Hildreth Ellen. Theory of edge detection // Proceedings of the Royal Society of London. Series B. Biological Sciences. — 1980. — Vol. 207, no. 1167. — P. 187–217.
- [46] Measuring robustness of visual slam / David Prokhorov, Dmitry Zhukov, Olga Barinova et al. // 2019 16th International conference on machine vision applications (MVA) / IEEE. — 2019. — P. 1–6.
- [47] Mur-Artal Raul, Montiel Jose Maria Martinez, Tardos Juan D. ORB-SLAM: A versatile and accurate monocular SLAM system // IEEE transactions on robotics. — 2015. — Vol. 31, no. 5. — P. 1147–1163.
- [48] Mur-Artal Raul, Tardós Juan D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras // IEEE transactions on robotics. — 2017. — Vol. 33, no. 5. — P. 1255–1262.
- [49] ORB: An efficient alternative to SIFT or SURF / Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski // 2011 International conference on computer vision / Ieee. — 2011. — P. 2564–2571.
- [50] Odombeyondvision: An indoor multi-modal multi-platform odometry dataset beyond the visible spectrum / Peize Li, Kaiwen Cai, Muhamad Risqi U Saputra et al. // 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) / IEEE. — 2022. — P. 3845–3850.

- [51] Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam / Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez et al. // IEEE transactions on robotics. — 2021. — Vol. 37, no. 6. — P. 1874–1890.
- [52] Papachristos Christos, Mascarich Frank, Alexis Kostas. Thermal-inertial localization for autonomous navigation of aerial robots through obscurants // 2018 International Conference on Unmanned Aircraft Systems (ICUAS) / IEEE. — 2018. — P. 394–399.
- [53] Robust visual inertial odometry using a direct EKF-based approach / Michael Bloesch, Sammy Omari, Marco Hutter, Roland Siegwart // 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS) / IEEE. — 2015. — P. 298–304.
- [54] Rosten Edward, Drummond Tom. Machine learning for high-speed corner detection // Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9 / Springer. — 2006. — P. 430–443.
- [55] Rosten Edward, Reitmayr Gerhard, Drummond Tom. Improved RANSAC performance using simple, iterative minimal-set solvers // arXiv preprint arXiv:1007.1432. — 2010.
- [56] Shi Jianbo et al. Good features to track // 1994 Proceedings of IEEE conference on computer vision and pattern recognition / IEEE. — 1994. — P. 593–600.
- [57] Shin Young-Sik, Kim Ayoung. Sparse depth enhanced direct thermal-infrared SLAM beyond the visible spectrum // IEEE Robotics and Automation Letters. — 2019. — Vol. 4, no. 3. — P. 2918–2925.
- [58] Simple baselines for image restoration / Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, Jian Sun // European conference on computer vision / Springer. — 2022. — P. 17–33.

- [59] Speeded-up robust features (SURF) / Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool // Computer vision and image understanding. — 2008. — Vol. 110, no. 3. — P. 346–359.
- [60] Steerers: A framework for rotation equivariant keypoint descriptors / Georg Bökman, Johan Edstedt, Michael Felsberg, Fredrik Kahl // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2024. — P. 4885–4895.
- [61] Swinir: Image restoration using swin transformer / Jingyun Liang, Jiezhong Cao, Guolei Sun et al. // Proceedings of the IEEE/CVF international conference on computer vision. — 2021. — P. 1833–1844.
- [62] Thermal-depth odometry in challenging illumination conditions / Xingxin Chen, Weichen Dai, Jiajun Jiang et al. // IEEE Robotics and Automation Letters. — 2023. — Vol. 8, no. 7. — P. 3988–3995.
- [63] Thermal-inertial SLAM for the environments with challenging illumination / Jiajun Jiang, Xingxin Chen, Weichen Dai et al. // IEEE Robotics and Automation Letters. — 2022. — Vol. 7, no. 4. — P. 8767–8774.
- [64] Thermal-inertial odometry for autonomous flight throughout the night / Jeff Delaune, Robert Hewitt, Laura Lytle et al. // 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) / IEEE. — 2019. — P. 1122–1128.
- [65] Thermal stereo odometry for UAVs / Tarek Mouats, Nabil Aouf, Lounis Chermak, Mark A Richardson // IEEE Sensors Journal. — 2015. — Vol. 15, no. 11. — P. 6335–6347.
- [66] Towards Long Term SLAM on Thermal Imagery / Colin Keil, Aniket Gupta, Pushyami Kaveti, Hanumant Singh // arXiv preprint arXiv:2403.19885. — 2024.
- [67] Tp-tio: A robust thermal-inertial odometry with deep thermalpoint / Shibo Zhao, Peng Wang, Hengrui Zhang et al. // 2020 IEEE/RSJ

International Conference on Intelligent Robots and Systems (IROS) / IEEE. — 2020. — P. 4505–4512.

- [68] Visual-Inertial SLAM Technology Based on Monocular Infrared Camera / Chunming Lv, Leilei Li, Ranfeng Wei et al. // 2024 36th Chinese Control and Decision Conference (CCDC) / IEEE. — 2024. — P. 2009–2014.
- [69] Vivid++: Vision for visibility dataset / Alex Junho Lee, Younggun Cho, Young-sik Shin et al. // IEEE Robotics and Automation Letters. — 2022. — Vol. 7, no. 3. — P. 6282–6289.
- [70] XFeat: Accelerated Features for Lightweight Image Matching / Guilherme Potje, Felipe Cadar, André Araujo et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2024. — P. 2682–2691.
- [71] Zhang Lilian, Koch Reinhard. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency // Journal of visual communication and image representation. — 2013. — Vol. 24, no. 7. — P. 794–805.
- [72] Zuiderveld Karel J et al. Contrast limited adaptive histogram equalization. // Graphics gems. — 1994. — Vol. 4, no. 1. — P. 474–485.
- [73] g 2 o: A general framework for graph optimization / Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat et al. // 2011 IEEE international conference on robotics and automation / IEEE. — 2011. — P. 3607–3613.
- [74] A multi-spectral dataset for evaluating motion estimation systems / Weichen Dai, Yu Zhang, Shenzhou Chen et al. // 2021 IEEE International Conference on Robotics and Automation (ICRA) / IEEE. — 2021. — P. 5560–5566.
- [75] van Manen Benjamin Ronald, Sluiter Victor, Mersha Abeje Yenehun. FirebotSLAM: thermal SLAM to increase situational awareness

in smoke-filled environments // Sensors. — 2023. — Vol. 23, no. 17. — P. 7611.

- [76] A visible-thermal fusion based monocular visual odometry / Julien Pouljol, Cristhian A Aguilera, Etienne Danos et al. // Robot 2015: Second Iberian Robotics Conference: Advances in Robotics, Volume 1 / Springer. — 2016. — P. 517–528.