

УДК 004.89

ОТКРЫТЫЙ ПРОГРАММНЫЙ МОДУЛЬ ИЕРАРХИЧЕСКОЙ ЛОКАЛИЗАЦИИ РОБОТОВ НА РАЗРЕЖЕННОЙ 3D-КАРТЕ С ПРИМЕНЕНИЕМ МУЛЬТИМОДАЛЬНОГО РАСПОЗНАВАНИЯ МЕСТА

Юдин Д.А., к.т.н., зав. лабораторией, МФТИ, AIRI, Москва, Россия

E-mail: yudin.da@mipt.ru

Мелехин А.А., инженер, МФТИ, Москва, Россия

E-mail: melekhin.aa@mipt.ru

Линок С.А., мл.научный сотрудник, МФТИ, Москва, Россия

E-mail: linok.sa@phystech.edu

Аннотация

В интеллектуальном транспорте при построении 3D-карт часто возникает необходимость повышения качества глобальной локализации на построенной заранее карте, например, для первоначальной локализации при включении робота без дополнительной помощи в виде ручной установки его положения. Для решения этой проблемы в настоящей статье предлагается оригинальная архитектура подхода иерархической локализации, отличающаяся мультимодальной моделью распознавания места по данным видеокамер и лидара и быстродействующим нейросетевым алгоритмом регистрации облаков точек. Разработана и размещена в открытый доступ программная реализация предложенной архитектуры, которая включена в открытую библиотеку OpenPlaceRecognition. Проведены эксперименты с открытыми наборами данных реальных мобильных роботов NCLT и ITLP-Campus, которые продемонстрировали время работы менее 100 мс и достаточный уровень качества на встраиваемом устройстве Nvidia Jetson AGX Xavier. Показана перспективность подхода для практического применения.

Ключевые слова: интеллектуальное транспортное средство, иерархическая локализация, 3D-карта, распознавание места, нейронная сеть, мультимодальность

OPEN SOFTWARE FOR HIERARCHICAL LOCALIZATION OF ROBOTS ON A SPARSE 3D MAP USING MULTIMODAL PLACE RECOGNITION

Yudin D.A., Ph.D., Head of the laboratory, MIPT, AIRI, Moscow, Russia

E-mail: yudin.da@mipt.ru

Melekhin A.A., engineer, MIPT, Moscow, Russia

E-mail: melekhin.aa@mipt.ru

Linok S.A., Junior Researcher, MIPT, Moscow, Russia

E-mail: linok.sa@phystech.edu

Abstract

In intelligent transport, when constructing 3D maps, there is often a need to improve the quality of global localization on a pre-built map, for example, for initial localization when turning on a robot without additional assistance in the form of manual setting of its position. To solve this problem, this paper proposes an original architecture of the hierarchical localization approach, characterized by a multimodal model of place recognition based on video camera and LiDAR data and a high-speed neural network algorithm for point cloud registration. A software implementation of the proposed architecture has been developed and open sources as a part of OpenPlaceRecognition library. Experiments have been conducted with open datasets of real mobile robots NCLT and ITLP-Campus, which demonstrated an operating time of less than 100 ms and a sufficient level of quality on an embedded Nvidia Jetson AGX Xavier device. The prospects of the approach for practical application are shown.

Keywords: intelligent vehicle, hierarchical localization, 3D map, place recognition, neural network, multimodality

Введение

В мобильной робототехнике и интеллектуальном транспорте при построении 3D-карт часто возникает необходимость глобальной коррекции с помощью поиска петель (loop detection) - глобальной локализации на построенной заранее карте. Кроме того, без глобальной корректировки возможно дублирование объектов на построенных роботами картах, что может стать проблемой как для навигации, так и решения задач манипулирования. Еще одной практической проблемой часто является первоначальная локализация при включении робота без дополнительной помощи в виде ручной установки его положения.

Все эти проблемы может помочь преодолеть иерархическая локализация [1], которая на первом этапе использует нейросетевой подход распознавания места, позволяющий выучивать уникальные векторные представления (эмбеддинги), отличающие одно место на карте от другого, а на втором этапе позволяет уточнить найденную позу робота с помощью алгоритма регистрации облаков точек. Такой подход также позволяет экономить вычислительные ресурсы за счет построения разреженных 3D-карт, в которые сохраняется информация только от ключевых кадров, при этом сохраняя высокое качество локализации. Это дает им преимущество перед алгоритмами создания плотных карт местности [4], которые сложно масштабировать на большие пространства. В частности, этот факт подтверждают работы PRISM-ToroMap [11] и HPointLoc [5], демонстрирующие возможности локализации интеллектуального агента в помещениях.

В то же время, остаются ограничения, связанные с быстродействием методов иерархической локализации при использовании нескольких источников данных разной модальности. Они, как правило, имеют существенные сложности с запуском на встраиваемых бортовых устройствах роботов. В настоящей статье уделяется особое внимание преодолению этого ограничения и созданию программной реализации подхода нейросетевой иерархической локализации на основе мультимодальной модели распознавания места (см. Рис. 1).



Рисунок 1. Задача мультимодального распознавания места по разреженной 3D-карте

Основным вкладом настоящей статьи является:

- Предлагается оригинальная архитектура подхода иерархической локализации, отличающаяся мультимодальной моделью распознавания места по данным видеокamer и лидара и быстродействующим нейросетевым алгоритмом регистрации облаков точек.

- Разработана и размещена в открытый доступ программная реализация предложенной архитектуры, которая включена в открытую библиотеку OpenPlaceRecognition¹¹.
- Проведены эксперименты с открытыми наборами данных реальных мобильных роботов NCLT [21] и ITLP-Campus [22], которые продемонстрировали время работы менее 100 мс и достаточный уровень качества на встраиваемом устройстве Nvidia Jetson AGX Xavier.

1. Анализ предметной области

1.1. Анализ методов мультимодального распознавания места

Одновременное использование данных разных модальностей является отдельным направлением в задаче распознавания места. Метод MinkLoc++ [7] является развитием более раннего MinkLoc3D, в котором авторы добавили нейросеть на базе ResNet18 с GeM-пулингом для извлечения дескрипторов изображений с камеры. Дескриптор для облаков точек лидара извлекается с помощью уже описанной ранее сети MinkLoc3D, после чего два дескриптора объединяются в один с помощью конкатенации.

Отличительной чертой метода AdaFusion [8] является использование блоков с механизмом внимания в промежуточных слоях сверточных сетей, по-отдельности извлекающих признаки для изображений и облаков точек. Полученные карты внимания затем с помощью Global Average пулинга превращаются в вектора и с помощью полносвязных слоев кодируются в два числовых значения, имеющих смысл весов для дескрипторов изображения и лидара, соответственно.

Авторы метода MMDF [9] осуществляют более глубокое объединение признаков путем проецирования облаков точек на изображение и слияние промежуточных признаков внутри двух веток нейросети.

Существует также мультисенсорный мультимодальный подход MSSPlace [10], который осуществляет позднее комплексирование признаков, формируемых энкодеров различных модальностей, в том числе явно содержащих семантическую информацию. Его модульная архитектура позволяет расширять количество используемых сенсоров без дополнительного дообучения нейросетевой модели. Его вариант реализации на основе данных лидара и камеры хорошо себя зарекомендовал для задачи топологической одновременной локализации и построения карты, о чем свидетельствуют результаты в работе PRISM-ToroMap [11].

1.2. Анализ методов регистрации облаков точек

Классическим подходом к решению задачи регистрации облаков точек (оптимизации положения и ориентации) является нахождение соответствий между ключевыми точками и решение PnP-задачи оптимизации с помощью таких методов как RANSAC [12], ICP [13], графовый G2O [14], а также современный и устойчивый к выбросам TEASER [15]. Однако в случае, когда облака точек, между которыми нужно найти преобразование систем координат, слабо пересекаются, имеют шумы (в условиях присутствия динамических объектов на сцене), аналитические оптимизационные методы могут иметь существенные ошибки. Разработка и исследование обучаемых методов было призвано преодолеть такие ограничения.

¹¹ <https://github.com/OPR-Project/OpenPlaceRecognition/blob/main/src/opr/pipelines/localization/base.py>

Среди современных нейросетевых методов регистрации облаков точек можно выделить HRegNet [17], GeoTransformer [16] и DMS [18]. В Таб. 1 приведены метрики качества методов на датасете KITTI Odometry [19], указанные в оригинальных статьях. Данный датасет содержит последовательности лидарных данных, записанные с помощью автомобиля на улицах города. Датасет содержит несколько траекторий, часть из которых повторяется, благодаря чему его часто используют для обучения и валидации методов SLAM, визуальной одометрии, оптимизации позы и распознавания места. Модели GeoTransformer [16] и DMS [18] показывают более высокую точность, однако работают значительно медленнее и более требовательны к вычислительным ресурсам, что затрудняет их использование в режиме реального времени, когда скорость обработки одной пары облаков точек должна быть не более 100 мс.

Таблица 1. Сравнение современных нейросетевых методов регистрации облаков точек на датасете KITTI Odometry. Метрики из оригинальных статей.

Метод	Registration Recall ↑, %	RRE ↓, °	RTE ↓, м	Время итерации ↓, мс
HRegNet [17]	99,7	0,29	0,0120	~89
GeoTransformer [16]	99,8	0,24	0,0068	~234
DMS [18]	99,8	0,27	0,0067	~396

2. Предлагаемый подход иерархической локализации интеллектуального транспортного средства на разреженной 3D-карте

2.1. Общая постановка задачи и подход к ее решению

На Рис. 2 показана предлагаемая оригинальная архитектура подхода иерархической локализации, отличающаяся мультимодальной моделью распознавания места по данным видеокамер и лидара и быстродействующим нейросетевым алгоритмом регистрации облаков точек. Она отличается оригинальным нейросетевым алгоритмом мультимодального места и доработанной быстрой нейросетевой моделью вычисления трансформации для сопоставленных ключевых точек

В двух последующих подразделах приведены детали реализации архитектуры. Важными требованиями, которые предъявлялись к архитектуре являлись:

- использование на обоих этапах иерархической локализации обучаемых алгоритмов на основе нейросетевых моделей, что является в настоящее время одной из характерных тенденций;
- обеспечение обработки входных фреймов данных в реальном времени, причем рассматривается один из наиболее распространенных случаев в мобильной робототехнике, когда данные поступают с периодичностью в 100 мс. Частоты обработки входных данных в 10Гц обычно хватает, чтобы решать задачи навигации и реагирования на возникающие препятствия для медленно-движущихся наземных роботов (до 10 км/ч).

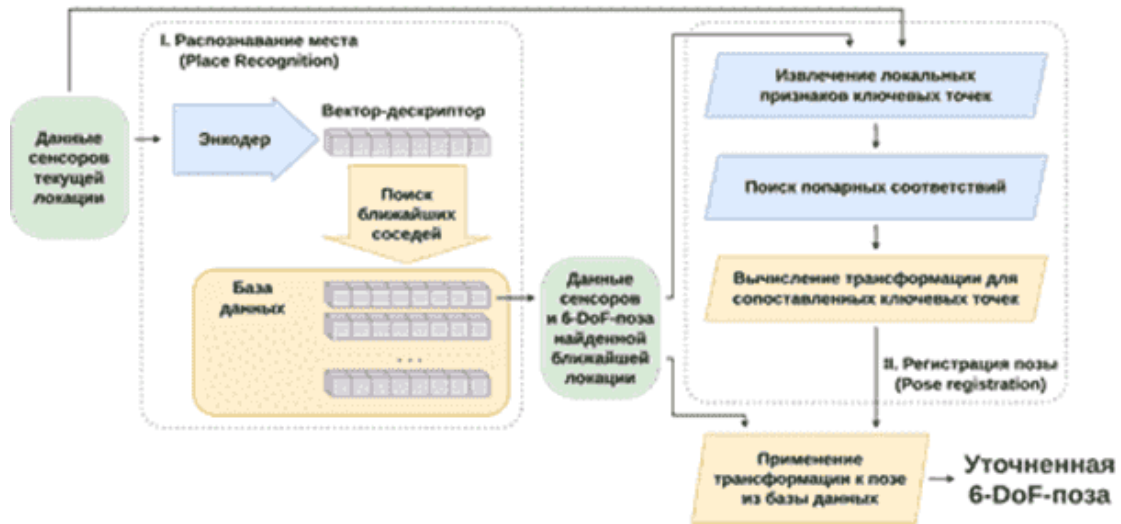


Рисунок 2. Укрупненная схема предлагаемого подхода иерархической локализации интеллектуального транспортного средства на разреженной 3D-карте

2.2. Модуль мультимодального распознавания места

На Рис. 3 изображен один из вариантов разработанной нейросетевой архитектуры метода формирования векторного представления текущего места с использованием изображений с нескольких камер, результатов их семантической сегментации, текстовых описаний и облака точек лидара. В настоящей статье использовалась модификация архитектуры модели MSSPlace [10], которая использует изображения двух камер и облака точек лидара.

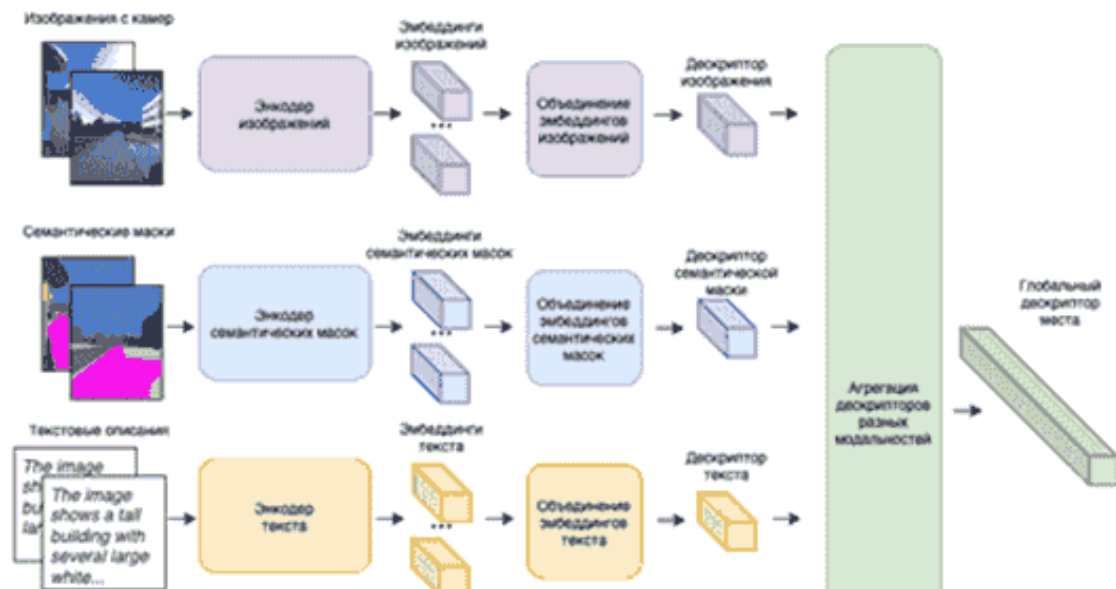


Рисунок 3. Архитектура применяемой модели мультимодального распознавания места MSSPlace, осуществляющего поиск в базе данных наиболее похожих записей по изображениям с нескольких камер, результатам их семантической сегментации, текстовым описаниям и облаку точек лидара

2.3 Нейросетевой модуль регистрации облаков точек в реальном времени

Для решения поставленной задачи модель HRegNet [17] была модифицирована путем замены модели оптимизации позы на более быструю архитектуру HRegNet Light. Данный метод использует иерархический подход локализации, при котором признаковые описания для облаков точек извлекаются на нескольких уровнях, и выполняется три шага: грубая (coarse) оптимизация с использованием высокоуровневых признаков и два шага уточненной (fine) оптимизации с использованием низкоуровневых признаков.

2.4. Программная реализация подхода в библиотеке OpenPlaceRecognition

В составе библиотеки OpenPlaceRecognition [20] разработан класс Localization-Pipeline, реализующий алгоритмы иерархической локализации в удобном для использования монолитном (end-to-end) интерфейсе. Он объединяет программную реализацию алгоритма мультимодального распознавания места (PlaceRecognitionPipeline) и алгоритм регистрации облаков точек (RegistrationPipeline) (см. Рис. 4).

2.5. Количественная оценка качества предлагаемого подхода

В настоящей статье оценка качества разработанного подхода осуществлялась на основе количественных метрик, которые вычислялись по двум комбинациям треков (где один из треков - query, а второй - database) и усреднялись. Метрика R@1 (25м) показывает процент правильно определенных мест с порогом попадания 25 метров, если в качестве места в базе брать top1 из найденных по запросу (чем больше метрика, тем лучше). Аббревиатуры RRE и RTE расшифровываются как Relative Rotation Error и Relative Translation Error (чем меньше метрики, тем лучше). Важно отметить, что медианные значения метрик RRE и RTE в данной задаче являются более репрезентативными, так как более устойчивы к выбросам.

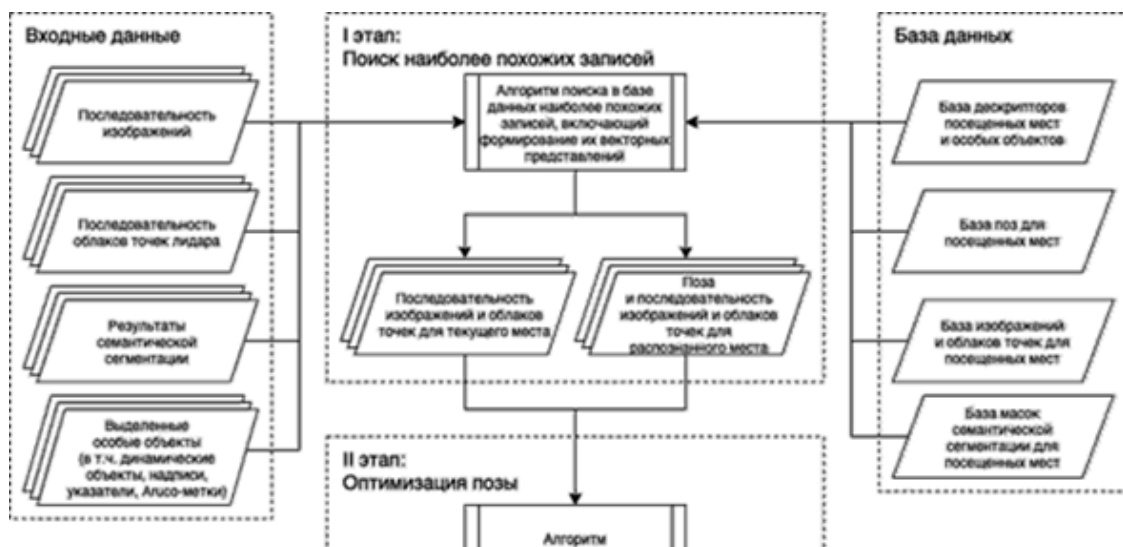


Рисунок 4. Схема алгоритма иерархической локализации, реализованная в библиотеке OpenPlaceRecognition

3. Результаты экспериментов

Эксперименты с предложенным подходом иерархической локализации транспортного средства осуществлялись на публичном наборе данных NCLT [21] (см. Таб. 2). Для тестирования были отобраны два трека – '2012-01-08' и '2012-01-22', использовался участок траектории, который отсутствовал в обучающей выборке, в качестве данных рассматривались изображения 2 камер (передней и задней), и облака точек лидара.

Таблица 2. Результаты иерархической локализации предлагаемым подходом на наборе данных NCLT

Метод	AR@1 ↑, %	Avg Median RRE ↓, °	Avg Median RTE ↓, m	Mean Time ↓, ms
Multimodal Place Recognition (Baseline)	92,80	6,65	4,47	25,85
Baseline + GeoTrans- former	92,51	2,36	0,23	234,73
Baseline + HRegNet- Light Coarse + outlier filtering	<u>92,78</u>	5,13	2,91	77,26

В составе подхода предложенная модель HRegNet-Light Coarse показывает менее высокие метрики в сравнении с ранее использованной GeoTransformer, однако качество остается в пределах допустимого (Recall@1 не менее 90%, ошибка позиционирования RTE не более 5 м, ошибка ориентации RRE не более 20°), при этом достигнута требуемая скорость работы (не более 100 мс на итерацию). Быстродействие указано для вычислителя с CPU AMD Ryzen 5 3500 6-Core, GPU NVIDIA RTX 3060 12 GB, а также для встраиваемой платформы Jetson AGX Xavier.

На Рис. 5 – 6 показаны характерные примеры успешной работы предложенного подхода иерархической локализации как в помещениях, так и на улице. Также на

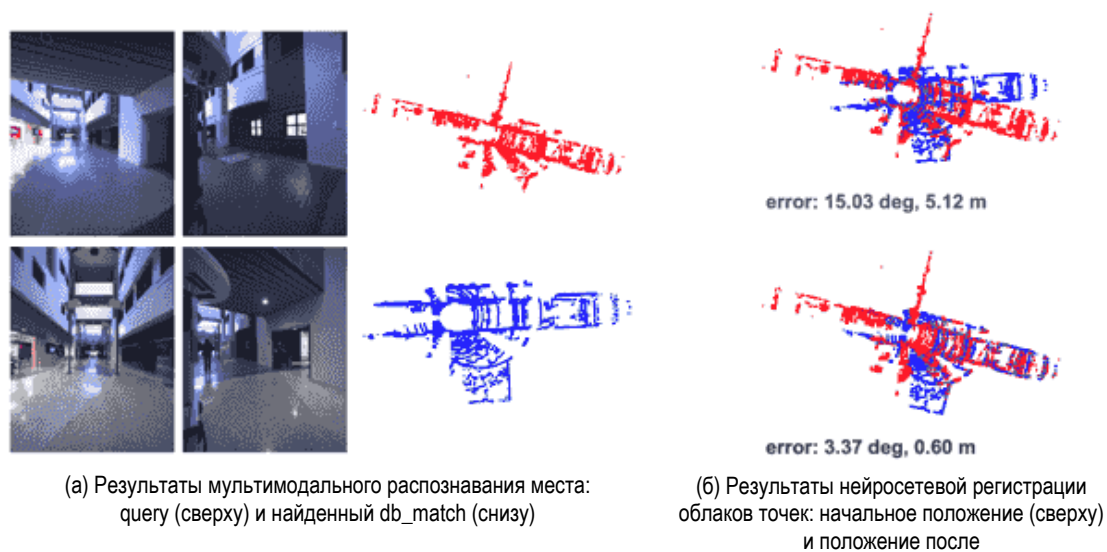


Рисунок 5. Результаты визуализации предлагаемого подхода иерархической локализации в помещениях

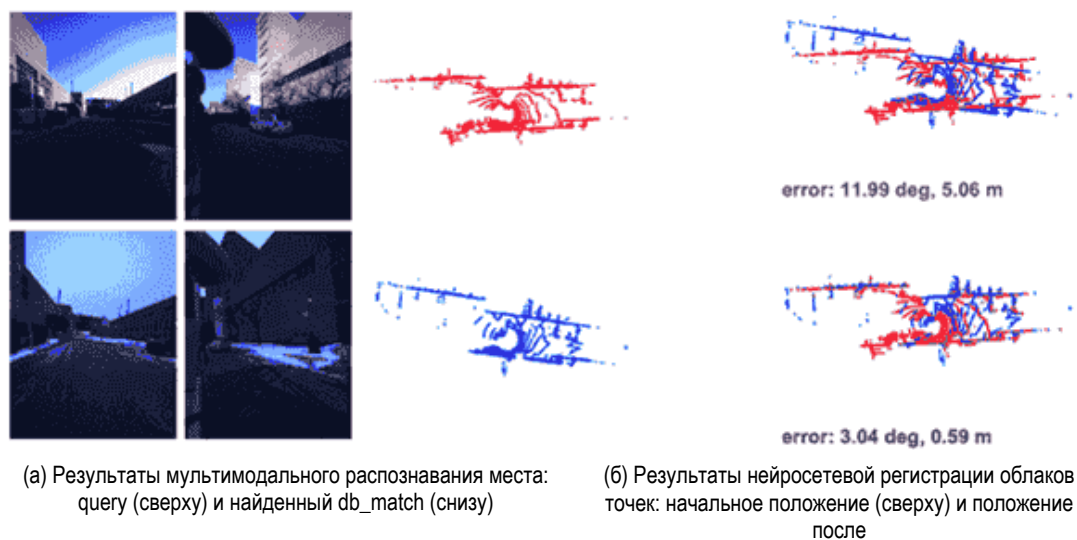


Рисунок 6. Результаты визуализации предлагаемого подхода иерархической локализации на уличной сцене

Рис. 7 показан один из примеров неудачной локализации в помещениях в длинных коридорах, который является ограничением предложенного подхода и может быть преодолен на основе использования последовательностей данных, текстовых надписей, графов сцены и иных нейро-символьных представлений сцен [6]. Это является объектом будущих исследований как продолжение настоящей работы.

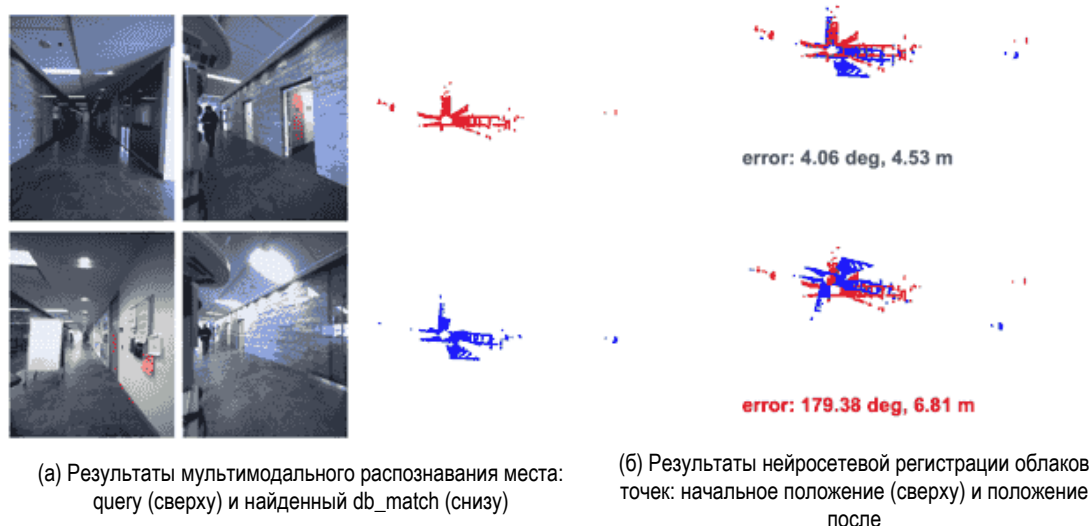


Рисунок 7. Пример ошибочной иерархической локализации в помещениях

Проведены эксперименты с четырьмя разработанными конфигурациями нейросетевой модели HRegNet, код реализации которых находится в репозитории OpenPlaceRecognition: Original¹², Light Coarse¹³, NoSim¹⁴, Light Fine¹⁵. Дополнительно реализован алгоритм фильтрации выбросов. Результаты этих экспериментов показаны в Таб.3. Каждая следующая конфигурация улучшает быстродействие метода ценой падения качества. Благодаря этому имеется возможность выбирать наиболее подходящую конфигурацию в зависимости от технических требований.

Таблица 3. Результаты экспериментов по повышению быстродействия нейросетевой модели регистрации облаков точек на основе архитектуры HRegNet с использованием базового метода мультимодального распознавания места на наборе данных NCLT

Нейросетевая модель регистрации облаков точек	Алгоритм фильтрации выбросов	AR@1↑, %	Avg Median RRE↓, °	Avg Median RTE↓, m	Mean Time ↓, ms (Jetson AGX Xavier)	Mean Time ↓, ms (RTX 3060)
HRegNet-Original	×	88,77	3,83	1,87	114,61	89,0
HRegNet-NoSim	×	90,74	5,02	3,44	92,02	83,0
HRegNet-NoSim	✓	<u>92,63</u>	<u>4,50</u>	<u>2,59</u>	92,02	83,0
HRegNet-Light Fine	×	90,99	4,86	3,56	<u>82,67</u>	<u>59,0</u>
HRegNet-Light Fine	✓	92,78	5,30	3,26	<u>82,67</u>	<u>59,0</u>
HRegNet-Light Coarse	×	91,63	7,44	4,20	77,26	50,0
HRegNet-Light Coarse	✓	92,78	5,13	2,91	77,26	50,0

Также было произведено тестирование разработанного подхода в условиях натуральных экспериментов, которое осуществлялось с помощью робота Clearpath Husky и узла LocalizationNode в составе ROS-2 реализации разработанной библиотеки OpenPlaceRecognition-ROS2 [23]. Для тестирования модуля был проведен эксперимент с проездом по территории кампуса МФТИ, который в том числе представлен в датасете ITLP-Campus [22]. В качестве базы данных во время эксперимента использовались данные из трека “05_2023-08-15” датасета ITLP-Campus-Outdoor.

Для визуализации работы узла использовался узел VisualizerNode. ROS2-сообщения с визуализациями просматривались с помощью утилиты RViz2. На представленных визуализациях изображения слева – данные эксперимента, изображения справа – найденные совпадения в базе данных. Примеры визуализаций приведены на Рис. 7.

¹² configs/model/registration/hregnet.yaml

¹³ configs/model/registration/hregnet_coarse.yaml

¹⁴ configs/model/registration/hregnet_nosim.yaml

¹⁵ configs/model/registration/hregnet_light_feats.yaml

Быстродействие нейросетевого модуля иерархической локализации, входящего в состав узла LocalizationNode, составило 91.16 ± 4.87 мс, что говорит о перспективах его практического применения.

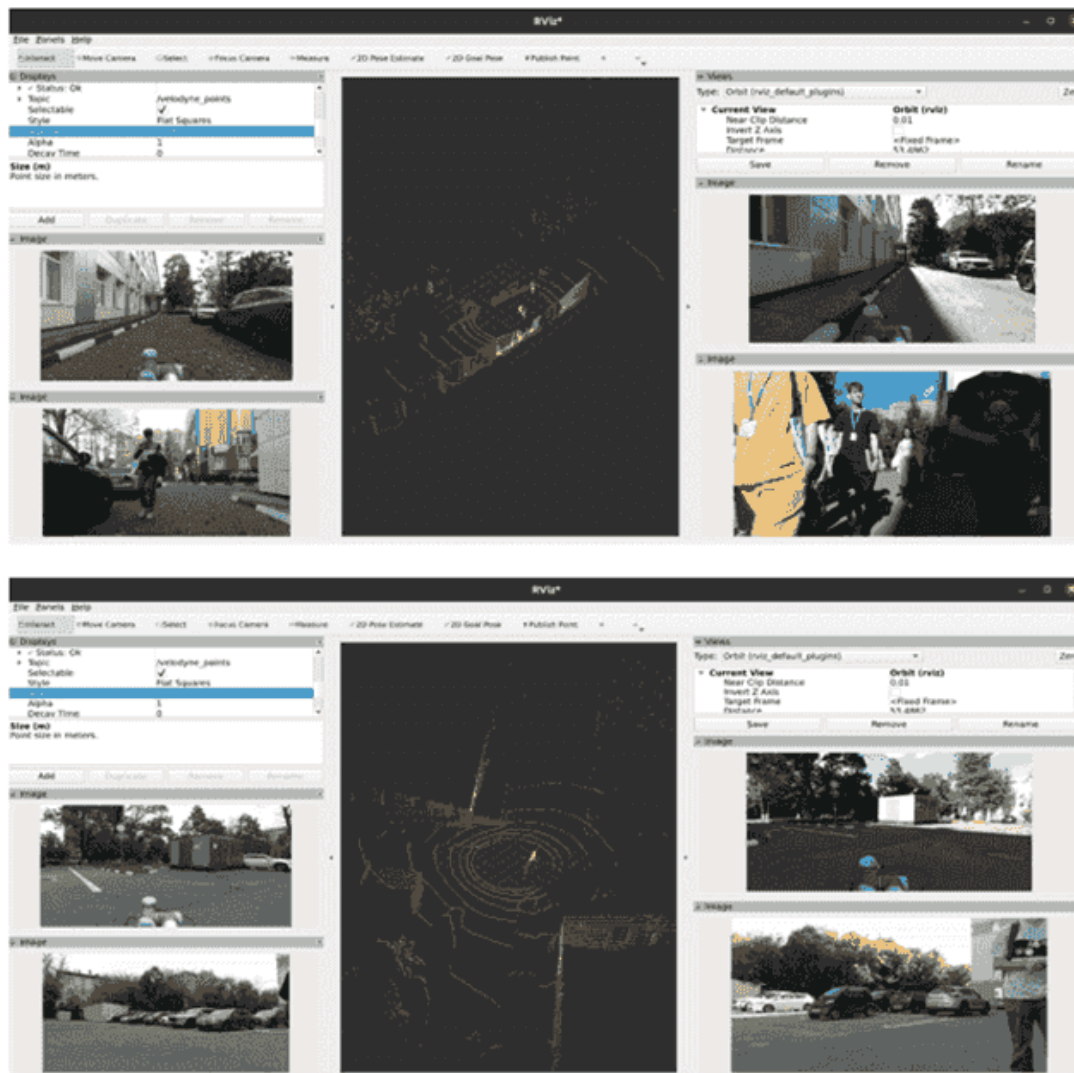


Рисунок 8. – Примеры корректной работы модуля иерархической локализации в составе узла LocalizationNode

Заключение

В настоящей статье рассмотрен нейросетевой вариант решения задачи иерархической локализации в реальном времени, в котором начальная оценка положения находится с помощью мультимодального нейросетевого метода распознавания места, а потом уточняется с помощью обучаемого алгоритма регистрации облаков точек. Разработанный подход демонстрирует время выполнения менее 100 мс на встраиваемых устройствах, что говорит о высоком потенциале применения подхода.

Разработана и успешно протестирована программная реализация иерархической локализации в составе библиотеки OpenPlaceRecognition, которая обеспечила удобный унифицированный интерфейс для использования различных открытых и оригинальных методов распознавания места и регистрации облаков точек, в том числе с использованием операционной системой роботов ROS2.

Благодарности

Работа выполнена при поддержке Фонда содействия инновациям для реализации проекта “Открытая программная библиотека мультимодальной нейросетевой глобальной локализации транспортного средства” Договор 34ГУКодИИС12-D7/81485 от 07.12.2022.

Список использованной литературы

1. Sarlin P. E. et al. From coarse to fine: Robust hierarchical localization at large scale //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – С. 12716-12725.
2. Garg S., Fischer T., Milford M. Where Is Your Place, Visual Place Recognition? //Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21). – International Joint Conferences on Artificial Intelligence, 2021. – С. 4416-4425.
3. Zhang Y. X. et al. A comprehensive survey and taxonomy on point cloud registration based on deep learning //Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. – 2024. – С. 8344-8353.
4. Belkin, I. V., Abramenko, A. A., Bezuglyi, V. D., Yudin, D. A. Localization of mobile robot in prior 3D LiDAR maps using stereo image sequence. Компьютерная оптика. – 2024. – № 48(3) – 406-417.
5. Yudin D. et al. Hpointloc: Point-based indoor place recognition using synthetic rgb-d images //International Conference on Neural Information Processing. – Cham : Springer International Publishing, 2022. – С. 471-484.
6. Kirilenko D. et al. Vector symbolic scene representation for semantic place recognition //2022 International Joint Conference on Neural Networks (IJCNN). – IEEE, 2022. – С. 1-8.
7. J. Komorowski, M. Wysoczańska, and T. Trzcinski, MinkLoc++: Lidar and Monocular Image Fusion for Place Recognition, in 2021 International Joint Conference on Neural Networks (IJCNN), Jul. 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9533373.
8. H. Lai, P. Yin, and S. Scherer, AdaFusion: Visual-LiDAR Fusion With Adaptive Weights for Place Recognition, IEEE Robotics and Automation Letters, vol. 7, no. 4, pp. 12038–12045, Oct. 2022, doi: 10.1109/LRA.2022.3210880.
9. X. Yu, B. Zhou, Z. Chang, K. Qian, and F. Fang, MMDF: Multi-Modal Deep Feature Based Place Recognition of Mobile Robots With Applications on Cross-Scene Navigation, IEEE Robotics and Automation Letters, vol. 7, no. 3, pp. 6742–6749, Jul. 2022, doi: 10.1109/LRA.2022.3176731.
10. Melekhin, A., Yudin, D., Petryashin, I., Bezuglyj, V. Mssplace: multi-sensor place recognition with visual and text semantics // arXiv preprint arXiv:2407.15663, 2024
11. Muravyev, K., Melekhin, A., Yudin, D., Yakovlev, K. PRISM-TopoMap: online topological mapping with place recognition and scan matching // IEEE Robotics and Automation Letters. 2025.
12. M. A. Fischler and R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM, vol. 24, no. 6, pp. 381–395, Jun. 1981, doi: 10.1145/358669.358692.
13. Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, Int J Comput Vision, vol. 13, no. 2, pp. 119–152, Oct. 1994, doi: 10.1007/BF01427149.
14. R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, G2o: A general framework for graph optimization, in 2011 IEEE International Conference on Robotics and Automation, May 2011, pp. 3607–3613. doi: 10.1109/ICRA.2011.5979949.
15. H. Yang, J. Shi, and L. Carlone, TEASER: Fast and Certifiable Point Cloud Registration, IEEE Transactions on Robotics, vol. 37, no. 2, pp. 314–333, Apr. 2021, doi: 10.1109/TRO.2020.3033695.

16. Qin Z. et al. Geotransformer: Fast and robust point cloud registration with geometric transformer //IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2023. – Т. 45. – №. 8. – С. 9806-9821.
 17. Lu F. et al. Hregnet: A hierarchical network for large-scale outdoor lidar point cloud registration //Proceedings of the IEEE/CVF International Conference on Computer Vision. – 2021. – С. 16014-16023.
 18. Cao H., Wang Y., Li D. Dms: Low-overlap registration of 3d point clouds with double-layer multi-scale star-graph //IEEE Transactions on Visualization and Computer Graphics. – 2024.
 19. Geiger A., Lenz P., Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite //2012 IEEE conference on computer vision and pattern recognition. – IEEE, 2012. – С. 3354-3361.
 20. OpenPlaceRecognition. URL: <https://github.com/OPR-Project/OpenPlaceRecognition>
 21. Carlevaris-Bianco N., Ushani A. K., Eustice R. M. University of Michigan North Campus long-term vision and lidar dataset //The International Journal of Robotics Research. – 2016. – Т. 35. – №. 9. – С. 1023-1035.
 22. Melekhin, A., Bezuglyj, V., Petryashin, I., Muravyev, K., Linok, S., Yudin, D., Panov, A. ITLP-Campus: A Dataset for Multimodal Semantic Place Recognition. In International Conference on Intelligent Information Technologies for Industry. – 2024. – pp. 185-195.
 23. ROS-2 реализация разработанной библиотеки OpenPlaceRecognition. <https://github.com/OPR-Project/OpenPlaceRecognition-ROS2>.
-