

Хомутов Денис Александрович, бакалавр
Уфимский университет науки и технологий
Khomutov Denis Aleksandrovich
Ufa University of Science and Technology

МЕТОДЫ ВИЗУАЛЬНОЙ ЛОКАЛИЗАЦИИ В СИСТЕМАХ АВТОНОМНОЙ МОБИЛЬНОЙ РОБОТОТЕХНИКИ VISUAL LOCALIZATION METHODS IN AUTONOMOUS MOBILE ROBOTIC SYSTEMS

Аннотация. Рассмотрены основные методы визуальной локализации мобильных робототехнических систем по данным камер. Описаны подходы на основе распознавания мест, локальных признаков, геометрической верификации, трёхмерных карт и визуального SLAM. Показаны их преимущества и ограничения с точки зрения автономной навигации.

Abstract. Major visual localization methods for mobile robotic systems using camera data are reviewed. Approaches based on place recognition, local features, geometric verification, three-dimensional maps, and visual SLAM are described. Their advantages and limitations from the standpoint of autonomous navigation are presented.

Ключевые слова: Визуальная локализация, мобильная робототехника, компьютерное зрение, локальные признаки, трёхмерная карта, автономная навигация.

Keywords: Visual localization, mobile robotics, computer vision, local features, three-dimensional map, autonomous navigation.

Введение

Автономная мобильная робототехника требует устойчивого определения положения робота относительно маршрута и окружающей среды. Одним из доступных источников такой информации является видеопоток с бортовой камеры. Визуальная локализация решает задачу оценки положения и ориентации камеры в известной сцене по текущему изображению и тем самым обеспечивает пространственную привязку, необходимую для навигации [1, 2].

Визуальные методы особенно актуальны внутри зданий и в других средах, где спутниковое позиционирование недоступно или нестабильно. Камера при этом может одновременно использоваться для локализации, обнаружения объектов и анализа свободного пространства. Основная сложность связана с изменением освещения, ракурса и масштаба, динамическими объектами, повторяющимися элементами интерьера и частичными изменениями сцены [2, 4]. Цель статьи – систематизировать методы визуальной локализации, применимые в автономных мобильных робототехнических системах, и определить особенности их практического использования.

Постановка задачи и основные подходы

Требуется оценить шестимерную позу камеры: положение и ориентацию. При наличии трёхмерной карты поза может вычисляться по соответствиям между двумерными точками кадра и пространственными ориентирами. В однородных координатах проекция имеет вид $s \cdot p = K(RP + t)$, где s – масштаб, K – матрица внутренних параметров камеры, R – матрица вращения, t – вектор переноса. По корректным соответствиям поза оценивается методом Perspective-n-Point с робастным исключением выбросов.

Наиболее простой класс методов основан на распознавании мест. Для кадра формируется глобальный дескриптор, который сравнивается с базой эталонных изображений. Такой подход быстро определяет вероятный участок маршрута, но сам по себе обычно не обеспечивает точную метрическую позу. В однотипных коридорах и помещениях визуально похожие изображения могут соответствовать разным местам. Поэтому распознавание мест целесообразно использовать как предварительный этап, после которого выполняются локальное сопоставление и геометрическая проверка [3].



Более точные методы используют локальные признаки. На изображении выделяются характерные точки и вычисляются дескрипторы их окрестностей. Затем признаки текущего и эталонного кадров сопоставляются между собой. Современные обучаемые детекторы и дескрипторы повышают устойчивость к изменению точки обзора и освещения; примером является SuperPoint [5]. Дополнительный прирост качества дают обучаемые сопоставители, учитывающие взаимный контекст локальных признаков. LightGlue реализует такой принцип с адаптивной вычислительной схемой [6].

Структурная локализация использует заранее построенную трёхмерную карту. По набору изображений методом Structure-from-Motion восстанавливаются положения камер и разреженные пространственные точки [1]. На рабочем кадре локальные признаки сопоставляются с ориентирами карты, после чего из двумерно-трёхмерных соответствий вычисляется поза. В отличие от этого, визуальный SLAM одновременно оценивает движение и поддерживает карту среды. Такое решение удобно при исследовании неизвестного пространства, однако для многократного движения по заранее известному маршруту офлайн-картографирование и последующая локализация могут быть проще и предсказуемее по вычислительным затратам [7].

Построение карты и геометрическая верификация

Качество локализации определяется качеством карты. Изображения маршрута должны иметь достаточное перекрытие, а ключевые кадры – представлять повороты, перекрёстки и участки с существенным изменением ракурса. При Structure-from-Motion выполняются сопоставление изображений, триангуляция пространственных точек и совместная оптимизация параметров камер [1]. Разреженная карта компактна по сравнению с плотной моделью поверхности и содержит именно те геометрические ориентиры, которые необходимы для оценки позы.

После сопоставления признаков текущего кадра с картой формируется набор пар «двумерная точка – трёхмерная точка». Часть совпадений неизбежно ошибочна, поэтому позу следует оценивать в робастной схеме, например с использованием RANSAC. Гипотеза считается надёжной, если с ней согласуется достаточное число независимых наблюдений. Количество геометрически подтверждённых соответствий может использоваться и как показатель доверия к результату.

Геометрическая проверка особенно важна в помещениях, где встречаются повторяющиеся двери, стены и элементы мебели. Ошибочные соответствия могут быть похожи по дескрипторам, но обычно не согласуются с одной пространственной позой. Для сложных внутренних сцен эффективность сочетания визуального сопоставления и геометрии показана в подходе InLoc [4]. Иерархическая схема «грубый поиск места – локальное сопоставление – оценка позы» позволяет дополнительно ограничить область поиска и масштабировать систему на большие карты [3].

Устойчивость и применение в навигации

Основными причинами деградации являются слабая текстура, размытие при движении, сильное изменение освещения, динамические объекты и большое отличие текущего ракурса от эталонных кадров. Устойчивость повышается при использовании нескольких наблюдений одного пространственного ориентира и достаточного разнообразия ключевых кадров. Полезен также контроль временной согласованности: положение между соседними кадрами не должно скачкообразно переходить на удалённый участок карты без физически возможного перемещения.

Для долговременной эксплуатации карта должна быть устойчива к частичным изменениям среды. Временные объекты не следует считать основными ориентирами, а решение желательно принимать по множеству независимых соответствий. При значительном изменении помещения карту можно обновлять, сохраняя стабильные геометрические элементы. Такой подход уменьшает зависимость от одного эталонного кадра и повышает вероятность повторной локализации после временной потери позиции.



В навигационном контуре оценённая поза преобразуется в положение относительно маршрута, направление движения и выбор следующей целевой точки. При этом визуальная локализация отвечает за глобальную привязку и не заменяет локальное восприятие. Новое препятствие может отсутствовать на карте, поэтому для безопасного движения необходим отдельный анализ свободного пространства и объектов перед роботом. Локализация и локальный анализ сцены решают взаимодополняющие задачи: первая определяет, где находится робот и куда проходит маршрут, второй – куда в данный момент физически допустимо двигаться.

Для бортового применения существенны задержка, объём памяти и стабильность времени обработки. Вычислительно тяжёлое построение и оптимизацию карты рационально выполнять заранее на производительном компьютере. На мобильной платформе остаются извлечение признаков текущего кадра, поиск соответствий, геометрическая оценка позы и фильтрация состояния. Такое разделение картографирования и локализации снижает требования к вычислителю робота и упрощает работу в реальном времени.

Заключение

Визуальная локализация обеспечивает пространственную привязку автономного мобильного робота по данным обычной камеры и может применяться в средах без спутниковой навигации. Распознавание мест эффективно для быстрого поиска области карты; локальные признаки обеспечивают точечные соответствия; структурные методы позволяют получать метрическую позу относительно трёхмерных ориентиров; визуальный SLAM используется, когда карту необходимо строить одновременно с движением.

Для повторяемой навигации по известным маршрутам рациональна структурная или иерархическая схема с предварительным построением карты и геометрической проверкой позы во время движения. Надёжность повышается за счёт множества независимых ориентиров, робастной фильтрации ошибочных соответствий и временной согласованности оценок. В полноценной системе результат локализации целесообразно объединять с анализом свободного пространства и препятствий

Список литературы:

1. Schönberger J. L., Frahm J.-M. Structure-from-Motion Revisited // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 4104–4113.
2. Sattler T., Maddern W., Toft C., Torii A., Hammarstrand L., Stenborg E., Safari D., Okutomi M., Pollefeys M., Sivic J., Kahl F., Pajdla T. Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 8601–8610.
3. Sarlin P.-E., Cadena C., Siegwart R., Dymczyk M. From Coarse to Fine: Robust Hierarchical Localization at Large Scale // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 12716–12725.
4. Taira H., Okutomi M., Sattler T., Cimpoi M., Pollefeys M., Sivic J., Pajdla T., Torii A. InLoc: Indoor Visual Localization with Dense Matching and View Synthesis // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 7199–7209.
5. DeTone D., Malisiewicz T., Rabinovich A. SuperPoint: Self-Supervised Interest Point Detection and Description // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2018. P. 224–236.
6. Lindenberger P., Sarlin P.-E., Pollefeys M. LightGlue: Local Feature Matching at Light Speed // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023. P. 17627–17638.
7. Campos C., Elvira R., Gómez Rodríguez J. J., Montiel J. M. M., Tardós J. D. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multi-Map SLAM // IEEE Transactions on Robotics. 2021. Vol. 37, No. 6. P. 1874–1890

