OpenPlaceRecognition: Распознавание места для локализации роботов



Александр Мелехин

Аспирант Лаборатории интеллектуального транспорта Центра когнитивного моделирования МФТИ

Дмитрий Юдин

Зав. Лабораторией интеллектуального транспорта Центра когнитивного моделирования МФТИ







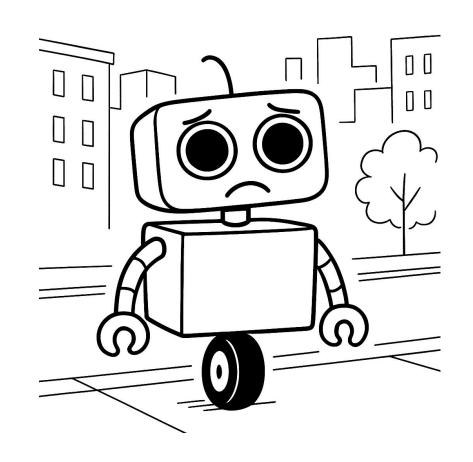




https://pikabu.ru/story/kot_popavshiy_na_google_maps_6034629?u=https%3A%2F%2Fwww.google.com%2Fmaps%2F%4041.895495%2C12.4773%2C3a%2C90y%2C341.76h%2C86.71t%2 Fdata%3D%213m7%211e1%213m5%211sAF1QipNOPqg6n5ql-PBdGoTcyKgCx0tiwJY4HN27VwjM%212e10%213e12%217i5376%218i2688&t=%D0%BA%D0%B0%D1%80%D1%82%D0%B 5&h=fb23d53040cf64c9f2fdd2ccff1e7d7dcbd89381

Когда робот теряется...

Как вы думаете, какие проблемы возникают, когда робот пытается узнать уже знакомое место в большом городе?



Цели

Чему мы научимся сегодня? **Поймём**, что такое Place Recognition и чем она отличается от Loop Closure Detection, Visual Localization и других задач

Изучим ключевые подходы и алгоритмы, применяемые для распознавания мест

© Обсудим роль этой задачи в робототехнике и компьютерном зрении

Узнаем, где и как применяется Place Recognition в реальных системах

Познакомимся с библиотекой OpenPlaceRecognition

Как всё это связано?

SLAM

topological localization

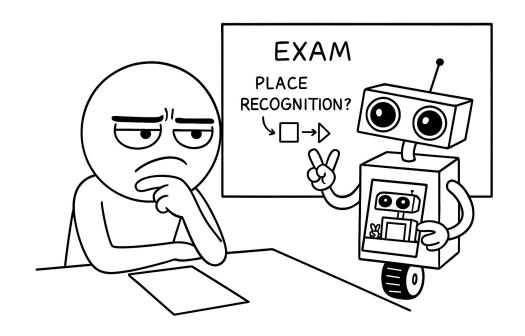
place recognition

metric localization

loop closure detection

Прежде чем углубляться... Давайте проверим, что мы уже знаем!

Небольшая разминка 🧠



1: Что такое SLAM?

А: Только построение карты

В: Построение карты и локализация

С: Только локализация

1: Что такое SLAM?

А: Только построение карты

В: Построение карты и локализация 🔽

С: Только локализация

2: Как вы понимаете термин «Place Recognition»?

А: Процесс определения точной позиции робота

В: Распознавание, что данное место уже было посещено

С: Метод построения 3D-модели окружения

2: Как вы понимаете термин «Place Recognition»?

А: Процесс определения точной позиции робота

В: Распознавание, что данное место уже было посещено 🔽



С: Метод построения 3D-модели окружения

3: Чем отличается Loop Closure Detection or Visual Localization?

(Открытый вопрос)

Кратко сформулируйте, в чем основная разница 🤔

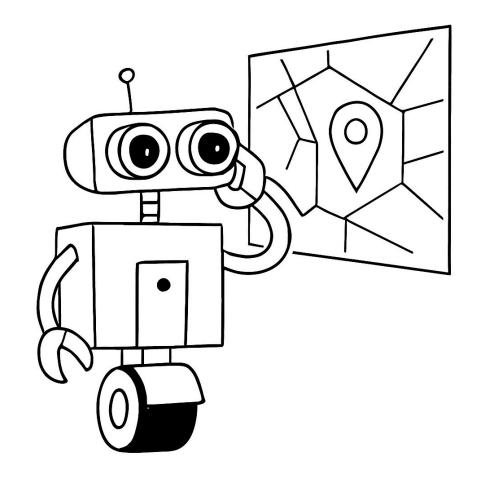
4: В каких приложениях вы встречали задачу распознавания мест?

(Открытый вопрос)

Назовите примеры 🤔

Что такое Place Recognition и зачем оно нужно?

↑ Прежде чем распознавать место, надо понять, что такое "место"



Что такое «место» для человека?

В 1971 году Джон О'Киф и Достровски открыли place cells — нейроны, которые активируются, когда животное находится в определённом месте

Так мозг формирует когнитивную карту пространства

В 2014 году Джон О'Киф был удостоен Нобелевской премии по физиологии или медицине за открытие клеток места, совместно с Мэй-Бритт и Эдвардом Мозерами, которые обнаружили «grid cells» или «клетки решетки», дополняющие систему позиционирования мозга

Posterior Hippocampus is LARGER in taxicab drivers and increases with experience

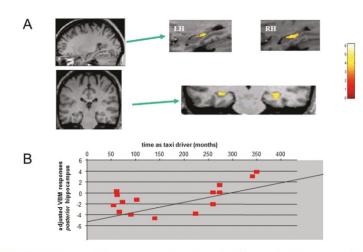


FIGURE 30. The posterior hippocampus is larger in taxicab drivers (A) than in controls and this effect increases with the length of time as a cab driver (B). After Maguire et al., 2000).

Что такое «место» для робота?

Робот также строит «внутреннюю карту» окружающей среды, сравнивая текущие наблюдения с тем, что он видел раньше

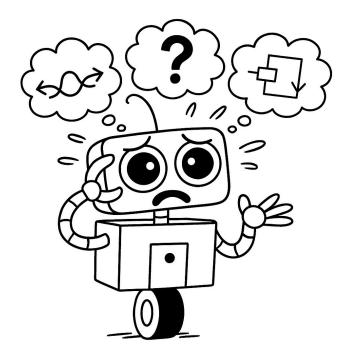
Но в отличие от мозга, у него нет "интуиции" — ему нужно определить, совпадает ли это место с каким-то из сохранённых



И человек, и робот узнают места. Но если мозг делает это "по ощущениям", то роботу приходится полагаться на алгоритмы и дескрипторы

«Место» в литературе

- Место это не фиксированная точка или область, а наблюдение, воспринимаемое через пересекающиеся поля зрения сенсоров [1]
- Определение места зависит от контекста навигации и может рассматриваться либо как точное положение, либо как более крупная область [2]
- Место это определённое положение в окружающей среде, которое может быть распознано при повторном посещении, независимо от положения или точки обзора агента [3]



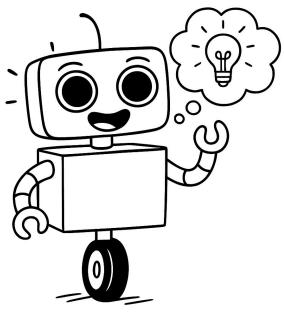
[1] S. Garg, T. Fischer, and M. Milford, 'Where is your place, Visual Place Recognition?', Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, pp. 4416–4425, Aug. 2021, doi: 10.24963/ijcai.2021/603.

[2] S. Lowry et al., 'Visual Place Recognition: A Survey', IEEE Transactions on Robotics, vol. 32, no. 1, pp. 1–19, Feb. 2016, doi: 10.1109/TRO.2015.2496823.

Что робот считает «местом»?

- Место это не точка, а результат пересечения полей зрения сенсоров [1]
- Определение места зависит от задачи: это может быть точка или область [2]
- Место это то, что может быть узнано независимо от ракурса [3]

Место — это фрагмент окружающей среды, который может быть распознан агентом по сенсорным данным при повторном посещении, независимо от ракурса или изменений



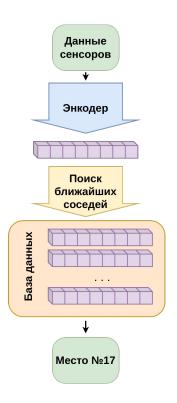
[1] S. Garg, T. Fischer, and M. Milford, 'Where is your place, Visual Place Recognition?', Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, pp. 4416–4425, Aug. 2021, doi: 10.24963/ijcai.2021/603.

[2] S. Lowry et al., 'Visual Place Recognition: A Survey', IEEE Transactions on Robotics, vol. 32, no. 1, pp. 1–19, Feb. 2016, doi: 10.1109/TRO.2015.2496823.

Как научная литература определяет Place Recognition?

- Мы определяем визуальное распознавание места (VPR) как способность определить собственное местоположение на основе двух наблюдений, воспринимаемых с пересекающимися полями зрения [1]
- Visual Place Recognition это вопрос о том, может ли человек, животное или робот, получив некое изображение места, определить, встречалось ли это место в его опыте ранее? [2]
- PR это способность распознавать посещённые области при различных условиях внешней среды и при различных ракурсах (точках обзора) [3]

Как работает система Place Recognition?



у Place Recognition — это задача определения, был ли
агент уже в текущей локации, путём сопоставления текущих
сенсорных данных с базой ранее сохранённых наблюдений

Важно: цель — распознать место, а не определить точную позу в координатах

Place Recognition ≠ Localization

задача узнавания, а не точного позиционирования

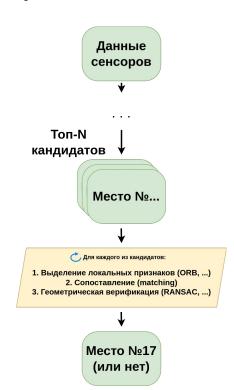
Place Recognition через сопоставление признаков

Не все системы распознают места по «целостному» признаку

В **геометрическом подходе** мы ищем локальные совпадения между кадрами: сопоставляем ключевые точки, убеждаемся в геометрической согласованности и только потом делаем вывод: «это то же место»

у Это ближе к методам локализации, но может использоваться и как Place Recognition — если цель не в позе, а в установлении факта повторного визита

Такой подход требует больше вычислений, но даёт более строгий критерий совпадения



※ Как мы определили "место"?

※ Как мы определили "место"?

Фрагмент среды, распознаваемый агентом при повторном посещении, независимо от ракурса и условий



Что делает система Place Recognition в топологическом подходе?

Фрагмент среды, распознаваемый агентом при повторном посещении, независимо от ракурса и условий



Что делает система Place Recognition в топологическом подходе?

Фрагмент среды, распознаваемый агентом при повторном посещении, независимо от ракурса и условий

Извлекает дескриптор текущего наблюдения, ищет наиболее похожее в базе и возвращает ID ранее посещённого места

А как говорят на практике?

Употребление терминов в статьях не всегда однозначно:

- Place Recognition ?
- Loop Closure Detection ?!
- Localization ?!?



Place Recognition? Loop Closure? Coarse Localization?

...VIN can be considered a reduced SLAM system, in which **the loop closure (or place recognition)** module is disabled [1]

Coarse visual localization, or place recognition, is a fundamental component in computer vision and robotics applications... [2]

Термины взаимозаменяются — но задачи разные

Чтобы сравнивать методы или системы, **нужно ясно** формулировать, что именно решается

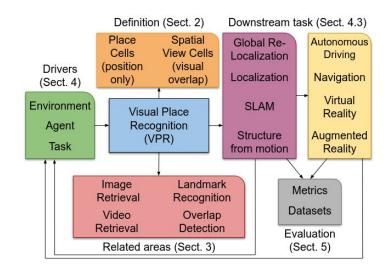


Figure 1: Visual Place Recognition (VPR) is the ability to recognize one's location based on two observations perceived from overlapping field-of-views. This figure illustrates the main sections of this paper and how they interrelate.

Pucyhok us [3]

[1] C. Cadena et al., 'Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age', IEEE Transactions on Robotics, vol. 32, no. 6, pp. 1309–1332, Dec. 2016, doi: 10.1109/TRO.2016.2624754.

[2] Y. Miao, F. Engelmann, O. Vysotska, F. Tombari, M. Pollefeys, and D. B. Baráth, 'SceneGraphLoc: Cross-Modal Coarse Visual Localization on 3D Scene Graphs', in Computer Vision – ECCV 2024, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, Eds., Cham: Springer Nature Switzerland, 2025, pp. 127–150. doi: 10.1007/978-3-031-73242-3 8.

Сравнение

Loop Closure Detection (LCD)

Суть: специальный случай применения Place Recognition в SLAM-задаче; указывает системе, что текущая позиция совпадает с какой-то ранее посещённой в рамках построенной карты

Отношение к PR: LCD напрямую опирается на механизм распознавания места (PR), но сфокусирован именно на «замыкании петли», то есть на поиске совпадения с конкретной точкой на карте

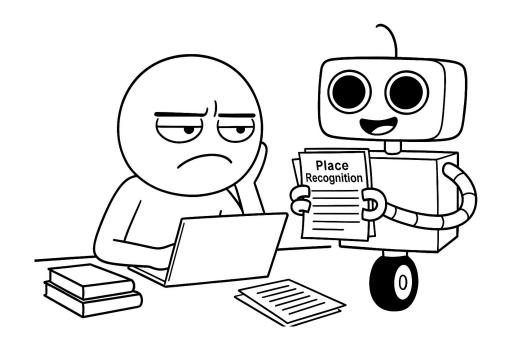
Coarse / Global / Visual / Hierarchichal Localization

Суть: определение координат или положения робота относительно некоторой карты (будь то заранее известная карта или строящаяся на лету)

Отношение к PR: распознавание места (PR) даёт возможность глобальной локализации — «где мы в известной карте?». Без знания карты (или хотя бы сохранённых описаний мест) локализация невозможна

А что внутри?

Как система Place Recognition находит знакомое место?



С чего всё начиналось: Bag-of-Words

Один из первых подходов к Place Recognition — адаптация текстовой модели Bag-of-Words:

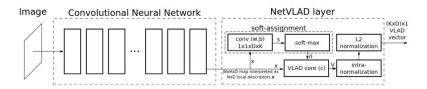
- Кадр представляется как набор визуальных "слов" (кластеров ключевых точек, например SIFT/ORB)
- Каждое изображение кодируется как вектор частот слов (histogram)
- Сравнение происходит через поиск ближайших векторов в базе

- ✓ Работает быстро, подходит для ограниченных условий
- Х Чувствителен к ракурсу, освещению, шуму
- _ Зависит от типа признаков (например, SIFT более устойчив, ORB быстрее)
- √ Использовался в системах типа FAB-MAP, DBoW2 (в ORB-SLAM)

NetVLAD: шаг от "мешка слов" к нейросетям

В 2016 году был предложен NetVLAD — модель, объединившая идеи VLAD и сверточных сетей

- Вместо набора визуальных слов → обучаемая агрегация признаков
- Вместо фиксированной схемы → end-to-end обучение
- На выходе глобальный дескриптор изображения
- Дескриптор можно сравнивать с другими — как в retrieval-задаче



- Устойчив к ракурсу и освещению
- 🔽 Можно дообучать под конкретные условия
- ✓ Работает как модуль можно заменить ВоW в SLAM-системах
- ↑ Требует GPU и предобучения, не всегда real-time

Современные методы

- AnyLoc (2024): используют предобученный DINOv2 для извлечения признаков, и агрегируют их с помощью VLAD [1]
- SALAD (2024): аналогично, но агрегируют признаки с помощью оптимального транспорта [2]
- SelaVPR++ (2025): обучают
 эффективные адаптеры для DINOv2 [3]

№ Общая тенденция: современные методы используют Foundation-модели (например, DINOv2) как универсальные экстракторы признаков

[1] N. Keetha et al., 'AnyLoc: Towards Universal Visual Place Recognition', IEEE Robotics and Automation Letters, vol. 9, no. 2, pp. 1286–1293, Feb. 2024, doi: 10.1109/LRA.2023.3343602.

[2] S. Izquierdo and J. Civera, 'Optimal Transport Aggregation for Visual Place Recognition', presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 17658–17668.

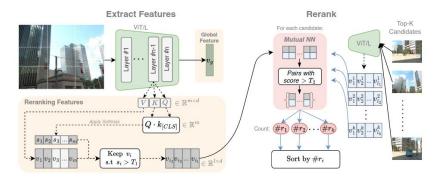
Глобальные + локальные признаки

✓ Идея: сначала — быстрый глобальный поиск, затем — проверка локальных совпадений (patch-level или keypoint-level)

Это снижает количество ложных совпадений и повышает надёжность

Методы:

- Patch-NetVLAD (2021) [1]
- TransVPR (2022) [2]
- EffoVPR (2025) [3] схема справа



- Используется токен [CLS] как глобальный дескриптор
- Извлекаются промежуточные признаки (через attention-карты)
- Выбираются ключевые точки на основе attention-карты S (по порогу Т1)
- Их дескрипторы берутся из компоненты V (Value)
- Кандидаты из top-К переранжируются по числу взаимных ближайших соседей (MNN), превышающих второй порог Т2

[1] S. Hausler, S. Garg, M. Xu, M. Milford, and T. Fischer, 'Patch-NetVLAD: Multi-Scale Fusion of Locally-Global Descriptors for Place Recognition', presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14141–14152.

[2] R. Wang, Y. Shen, W. Zuo, S. Zhou, and N. Zheng, 'TransVPR: Transformer-Based Place Recognition With Multi-Level Attention Aggregation', presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 13648–13657.

[3] I. Tzachor et al., 'EffoVPR: Effective Foundation Model Utilization for Visual Place Recognition', Feb. 02, 2025, arXiv: arXiv:2405.18065. doi: 10.48550/arXiv.2405.18065.

Place Recognition выходит за пределы RGB-изображений

В условиях плохой освещённости, повторяющихся структур или динамичных сцен одной камеры может быть недостаточно





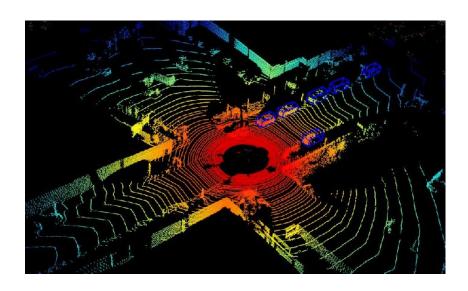
- Много визуальной информации
- Чувствительность к освещению и погоде

Использование облаков точек лидара

- Много пространственной информации
- Устойчивость к освещению
- Х Чувствительность к погоде (дождь, снег)
- **X** Нет информации о текстуре

Методы:

- TransLoc3D (2021)
- MinkLoc3Dv2 (2022)
- BEVPlace (2023)
- PTC-Net (2023)



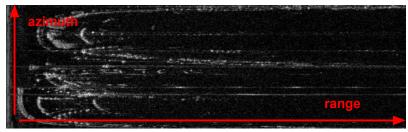
Пример облака точек лидара

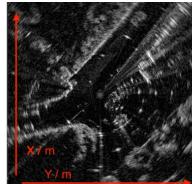
Использование данных радара

- Устойчивость к любой погоде (включая туман, дождь, снег)
- 🗙 Высокая разреженность данных
- Нет информации о текстуре

Методы:

- RadarLoc (2021)
- AutoPlace (2022)
- Off the Radar (2023)





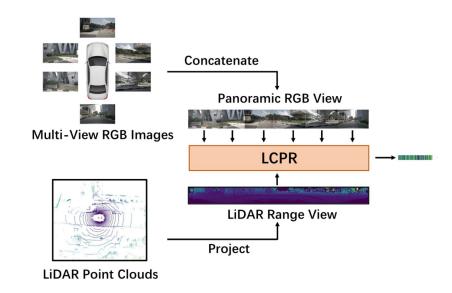
Форматы данных

Komorowski, J., Wysoczanska, M., & Trzcinski, T. (2021). Large-Scale Topological Radar Localization Using Learned Descriptors. In T. Mantoro, M. Lee, M. A. Ayu, K. W. Wong, & A. N. Hidayanto (Eds.), Neural Information Processing (pp. 451–462). Springer International Publishing.

Мультимодальные методы

Методы:

- MinkLoc++ (2021)
- AdaFusion (2022)
- LCPR (2023)



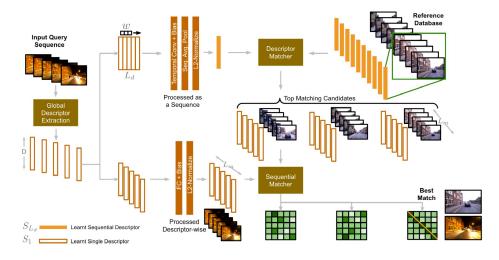
Метод LCPR

Использование последовательностей данных

- 🔽 Позволяют учитывать больше данных
- **Х** Сохраняют недостатки используемой модальности

Методы:

- SeqNet (2021)
- SeqMatchNet (2022)
- SeqOT (2022)



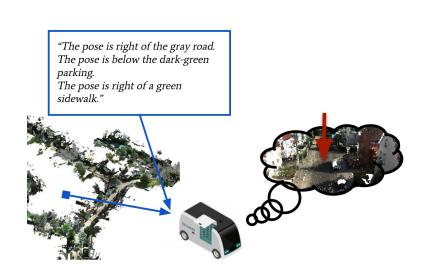
Метод SeqNet

Другие подходы

Text2Pos: поиск лидарных облаков точек по текстовым описаниям

Техt Spotting: выделение текста на изображениях с камеры

ТSVLoc: учёт масок семантической сегментации для ре-ранжирования результатов распознавания места



Метод Text2Pos

Какой подход вы бы выбрали для своей системы?

1. Мобильный робот в здании без GPS

условия: плохое освещение, повторяющиеся коридоры, ограниченные вычисления

2. Автономное авто в городе с переменными погодными условиями

условия: дождь, ночь, много динамики

3. AR-приложение для туристов в историческом центре

условия: наличие пользователей, важна объяснимость и взаимодействие

Вам нужно:

- Выбрать набор сенсоров
- Выбрать метод

Что вы выберете?

Выбор сенсоров и метода — всегда компромисс между ценой, точностью, устойчивостью и скоростью

Где Place Recognition работает в реальных системах?

Переходим от "как работает" к "зачем это нужно"

🧠 Мы узнали:

- как устроены методы распознавания мест
- какие данные и подходы можно использовать
- как выбирать метод под задачу

- в автономной навигации
- в SLAM и локализации
- в AR/VR-приложениях и даже взаимодействии с человеком

Loop Closure: когда PR помогает корректировать карту

В **SLAM**-системах **распознавание места** играет роль триггера: система замечает, что кадр похож на уже встречавшийся — и помечает его как потенциальное совпадение

у Это и есть этап Place Recognition — поиск кандидатов на повторное посещение

• Но чтобы замкнуть петлю, система должна убедиться: это действительно то же самое место, а не просто визуально похожее

Place Recognition даёт кандидатов, а Loop Closure Detection — включает в себя и PR, и последующую проверку и принятие решения

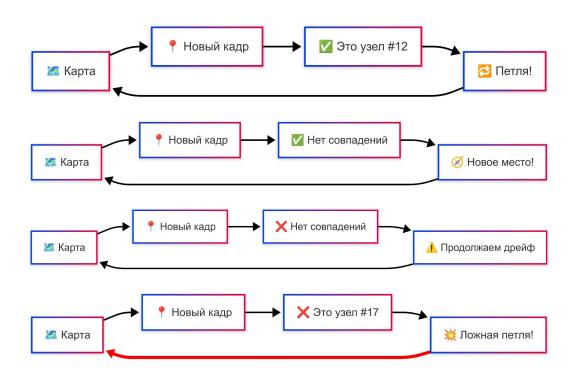
✓ Эти задачи часто путают, но они не равны

Почему важно проверять кандидатов от PR?

Разная цена ошибки:

♠ False negative – мы продолжим дрейф, накопление ошибки – но ничего не сломаем

False positive – мы можем "сломать" всю карту, если попытаемся скорректировать ее с учетом ложного цикла



А где я вообще? Глобальная локализация без GPS

В задачах **глобальной локализации** робот оказывается в незнакомом месте и должен определить: **«Где я нахожусь на карте, которую я видел раньше?»**

Для этого система **сравнивает текущее изображение с базой эталонных кадров** → и находит наиболее вероятное совпадение

Place Recognition здесь используется как:

- Этап поиска кандидатов
- Иногда финальное решение (если задача только топологическая)
- Иногда основа для запуска PnP / точной локализации

№ В отличие от Loop Closure Detection в SLAM, здесь нет карты, которую надо поправить

Задача — найти своё положение, опираясь на известные места

Иерархическая локализация с LiDAR

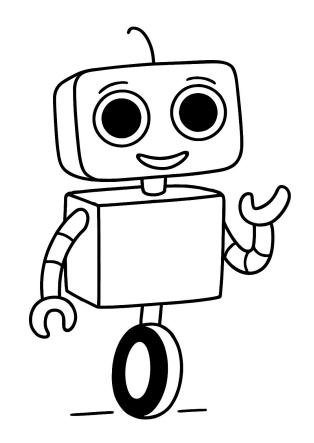


- PR даёт "грубое положение" похожее место в базе
- Сопоставление облаков (например, ICP) даёт точную трансформацию
- Результат локализация в глобальных координатах

А где бы Place Recognition пригодилась вам?

Подумайте:

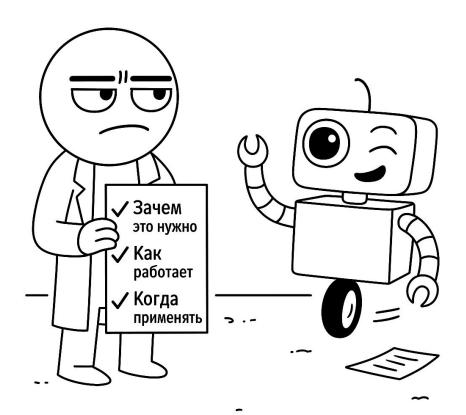
- Где нужно "узнавать" места без GPS?
- Где важно быстро понять, что ты "здесь уже был"?
- Где важно понимать, куда смотришь а не только где стоишь?



Отлично справились!

Теперь мы точно знаем, что разобрались в ключевых вопросах!

Давайте быстро подведём итог, чтобы закрепить, что сегодня было важного, и перейдём к самому интересному — практике!



Что мы теперь знаем о Place Recognition?

« Что такое «место» и как его определяют в робототехнике

Узнаваемый фрагмент среды, определяемый по сенсорным данным, даже при смене ракурса или условий

Что делает система Place Recognition (и что не делает)

Находит наиболее похожее место в базе — по текущему сенсорному наблюдению

- **!** Не гарантирует, что это то же самое место, и не вычисляет точную позу
- Какие подходы бывают: от BoW до мультимодальности

Om BoW и NetVLAD до ViT, LiDAR, последовательностей и мультимодальных моделей

Что мы теперь знаем о Place Recognition?

Чем PR отличается от Loop Closure и Localization

PR — поиск похожих мест

Loop Closure — верификация совпадения и коррекция карты

Localization — вычисление точной позы (XYZ + ориентация)

Ж Где применяется PR в реальных системах

В SLAM, глобальной локализации без GPS, навигации, AR/VR

Как это поможет вам на практике?

Ж Глубже понимать навигационные системы

→ Вы будете видеть, как PR встроен в SLAM и локализацию

X Осознанно выбирать алгоритмы под задачу

→ Понимание различий и компромиссов между подходами поможет принимать инженерные решения

🚀 Увереннее работать с библиотеками и своими проектами

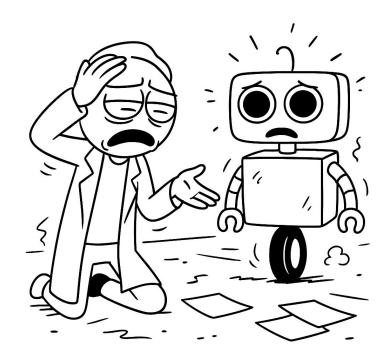
ightarrow Вы сможете применять PR в своих роботах, приложениях или исследованиях — не вслепую

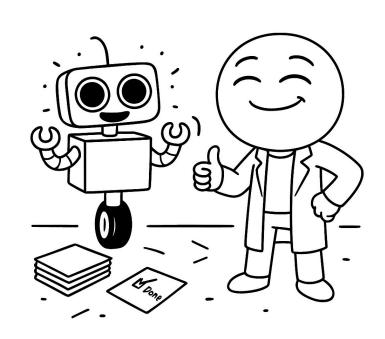
Теперь — попробуем это вживую!

Ответи практической части

- Познакомиться с библиотекойOpenPlaceRecognition
- Извлечь дескрипторы изображений
- Провести поиск мест по реальным данным
- **Ж** Посмотреть, как всё это работает "под капотом"

Два варианта финала





Давайте делать Open-Source!

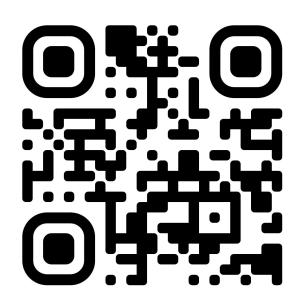
- Библиотека, без сомнений, нуждается в развитии и доработке
- Есть много идей и планов, что нужно сделать
- Давайте делать вместе!



github.com/OPR-Project/ OpenPlaceRecognition

Центр когнитивного моделирования МФТИ

- Стажировки
- Магистратура
- Аспирантура



cogmodel.mipt.ru

Контакты

Александр Мелехин

Tg: @alexmelekhin

Github: alexmelekhin

LinkedIn: a-melekhin

amelekhin96@gmail.com

Дмитрий Юдин

Tg: @yuddim

Github: yuddim

yuddim@yandex.ru