

ИНКРЕМЕНТАЛЬНЫЕ ПРЕДСТАВЛЕНИЯ СЦЕНЫ: ПРИНЦИПЫ ПОСТРОЕНИЯ И КРИТЕРИИ ПРИГОДНОСТИ ДЛЯ МОБИЛЬНОЙ МАНИПУЛЯЦИИ

Куркова Р. Е.¹

Научный руководитель – доктор технических наук, профессор Колюбин С. А.¹

¹Университет ИТМО

rekurkova@itmo.ru

Введение

В большинстве робототехнических сценариев список объектов и формулировок запросов не является закрытым, поэтому семантическое картографирование должно поддерживать работу с неограниченным набором классов и произвольными текстовыми описаниями. Типовой семантический слой строится на связке универсальных предобученных модулей: сегментация экземпляров (напр., SAM [1]) и сопоставление «изображение–текст» (напр., CLIP [2]). Однако эти модули не задают, как хранить и инкрементально обновлять пространственно согласованное представление сцены, сохранять экземплядность объектов и учитывать функциональные свойства, критичные для манипуляции.

Основная часть

Решение формулируется как построение инкрементального представления сцены, в котором геометрия, семантические признаки и экземплядность объектов объединяются в единую карту и обновляются по мере поступления наблюдений [3]. Ключевая операция состоит в переносе выделенных объектных признаков в трёхмерную систему координат, их ассоциации между кадрами и накоплении с контролем согласованности, чтобы сохранять устойчивые идентификаторы объектов и предотвращать ошибочные слияния и дублирование. Критерии пригодности такого представления для мобильной манипуляции задаются не только оценкой перцептивных метрик, но и способностью поддерживать локализацию функционально значимых участков, а также пространственные и функциональные отношения, влияющие на исполнимость действий.

Выводы

На практике такое представление может служить интерфейсом между восприятием и планированием: поддерживать локализацию объектов и функционально значимых участков, а также проверку достижимости целевого объекта и оценку возможных коллизий. Для верификации предлагается воспроизводимая методика испытаний в симуляционной среде с подготовкой тестовых эпизодов и контролируемыми вариациями наполнения сцены. План испытаний включает построение карты на полученных последовательностях и последующую оценку по критериям согласованности экземплядности и пригодности для манипуляционных запросов и сценариев [4].

Литература

1. Learning transferable visual models from natural language supervision / Alec Radford, Jong Wook Kim, Chris Hallacy et al. // International conference on machine learning / PmLR. — 2021. — Pp. 8748–8763.
2. Segment anything / Alexander Kirillov, Eric Mintun, Nikhila Ravi et al. // Proceedings of the IEEE/CVF international conference on computer vision. — 2023. — Pp. 4015–4026.

3. O2V-Mapping: Online Open-Vocabulary Mapping with Neural Implicit Representation / Muer Tie, Julong Wei, Ke Wu et al. // European Conference on Computer Vision / Springer. — 2024. — Pp. 318–333.
4. OSMa-Bench: Evaluating Open Semantic Mapping Under Varying Lighting Conditions / Maxim Popov, Regina Kurkova, Mikhail Iumanov et al. // 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). — 2025. — Oct. — Pp. 10786–10791.