УДК 004.896

ИЗВЛЕЧЕНИЕ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ ДЛЯ ПОСТРОЕНИЯ МУЛЬТИМОДАЛЬНЫХ КАРТ

Попов М.Ф. (ИТМО)**, Куркова Р.Е.** (ИТМО)

Научный руководитель - доктор технических наук, профессор Колюбин С.А. (ИТМО)

Введение. Построение мультимодальных карт предполагает не только сбор и хранение геометрической информации об окружении, но и представление семантической информации об объектах в картируемом пространстве и связей между ними в виде текстовых данных. Существуют разные подходы для извлечения семантической информации, однако не многие из них подходят для робототехники, так как решения должны быть эффективны по быстродействию и используемым памяти и вычислительным ресурсам.
Таким образом, основной целью данного исследования является сравнительный анализ современных методов извлечения семантической информации для последующего обоснованного выбора наиболее перспективных подходов для интеграции в систему построения мультимодальных карт.

Основная часть. При исследовании различных методов извлечения семантический информации для последующего построения мультимодальных карт оценивались несколько ключевых аспектов, которые включали точность выявления описаний и/или эмбеддингов, скорость работы алгоритма, эффективность хранения информации (в том числе использование памяти) и возможность применения на бортовых вычислителях мобильных роботов. Данные критерии помогают оценить качество будущих карт и их пригодность для реальных задач, где важны точность и оптимальное использование ресурсов.
Анализ преимуществ и недостатков изученных подходов проводился на датасетах ScanNet[1], Replica[2], а также с использованием симулятора AI2-THOR[3], и позволил выделить группу перспективных методов, которые могут быть использованы для улучшения качества и функциональности карт: Open-Fusion[4], ConceptGraphs[5] и другие.

Выводы. В ходе работы проведен обзор существующих методов построения мультимодальных карт, проанализированы подходы извлечения эмбеддингов и описаний. Выявлены преимущества и недостатки различных подходов к извлечению семантической информации, что позволяет предложить новый подход или улучшения на основе полученных результатов. Определены потенциальные области применения мультимодальных карт, а также указано ключевое направление для дальнейшего исследования и развития в этой области: использование связки из сегментационной и визуально-лингвистической моделей.

Список используемых источников:

1. Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.
2. Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The Replica dataset: A digital replica of indoor spaces. arXiv preprint arXiv:1906.05797, 2019.

3. Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli Vander-Bilt, Luca Weihs, Alvaro Herrasti, Daniel Gordon, Yuke Zhu, Abhinav Gupta, and Ali Farhadi. AI2-THOR: An Interactive 3D Environment for Visual AI. arXiv, 2017.
4. Yamazaki Kashu, Hanyu Taisei, Vo Khoa, Pham Thang, Tran Minh, Doretto Gianfranco, Nguyen Anh, and Le Ngan. Open-Fusion: Real-time Open-Vocabulary 3D Mapping and Queryable Scene Representation. arXiv. — 2023. — 2310.03923.
5. Gu Qiao, Kuwajerwala Alihusein, Morin Sacha, Jatavallabhula Krishna Murthy, Sen Bipasha, Agarwal Aditya, Rivera Corban, Paul William, Ellis Kirsty, Chellappa Rama, Gan Chuang, de Melo Celso Miguel, Tenenbaum Joshua B., Torralba Antonio, Shkurti Florian, and Paull Liam. ConceptGraphs: Open-Vocabulary 3D Scene Graphs for Perception and Planning. arXiv. — 2023