Toward Efficient Language-Grounded Semantic Exploration

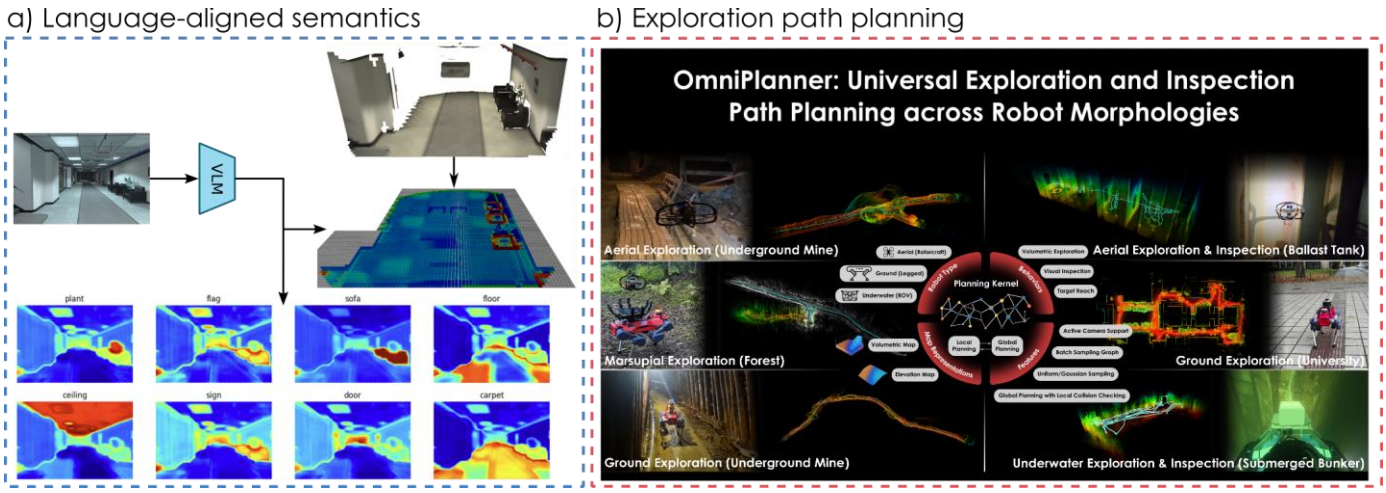

Figure 1: (a) Example of language-aligned feature extraction using a Vision Language Model. (b) Omiplanner[1][2][3], a framework for autonomous volumetric exploration and inspection.

Abstract:

Autonomous exploration in unknown environments is a key capability for autonomous robots. Existing exploration planners typically rely on geometric information, prioritizing areas that maximize the exploration of unseen space. In our current system, the **OmniPlanner**[1][2][3] exploration framework guides aerial and ground robots by generating collision-free trajectories that maximize a geometric exploration gain derived from the environment geometry. While effective for exploration coverage, this purely geometry-based approach does not account for semantic relevance, which is important for tasks such as object search.

This project aims to extend the current geometry-based exploration strategy by integrating **language-grounded semantic information** into the planning process. Specifically, we propose introducing a semantic gain component using a **Vision Language Model (VLM)** such as CLIP[4]. Image observations collected by the robot will be encoded by the VLM into language-aligned features and projected into the robot's map representation. These semantic features will then be transformed into a semantic gain, enabling the robot to prioritize the exploration or regions that are not only geometrically informative but also semantically meaningful with respect to a given task.

Tasks:

- Evaluate several VLMs for extracting language-aligned visual features.
- Extend the local environment representation to incorporate language-aligned semantic information.
- Design a semantic exploration gain based on the language-aligned feature map and integrate it with the existing volumetric exploration gain.
- Evaluate the proposed approach on embedded hardware using aerial and ground robots across multiple semantic exploration tasks.

Literature (indicative):

[1] M. Kulkarni *et al.*, "Autonomous Teamed Exploration of Subterranean Environments using Legged and Aerial Robots," *International Conference on Robotics and Automation (ICRA)*, 2022.

[2] Dang T, Tranzatto M, Khattak S, Mascarich F, Alexis K, Hutter M. "Graph-based subterranean exploration path planning using aerial and legged robots". *J Field Robotics*. 2020; 37: 1363–1388.

[3] A. Zacharia, M. Dharmadhikari, M. Singh, K. Alexis, "OmniPlanner: Universal Exploration and Inspection Path Planning across Robot Morphologies". *ArXiv*, 2026.

[4] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, "Learning Transferable Visual Models From Natural Language Supervision". *International Conference on Machine Learning*, 2021.

Relevant Projects: Norwegian Centre for Embodied AI, SYNERGISE

Main supervisor: Kostas Alexis, Professor, NTNU

Co-supervisors: Albert Gassol PhD Candidate, NTNU, **Angelos Zacharia** PhD Candidate, NTNU

Contact: konstantinos.alexis@ntnu.no, albert.g.puigjaner@ntnu.no, angelos.zacharia@ntnu.no