Lab Project Proposal: Open-Vocabulary Mobile Manipulation

Supervisor: Rohit Menon

Figure 1: Overview of an Open Vocabulary Mobile Manipulation Process

Goal

Enable a mobile manipulator (Neobotix MPO-700 with UR5e and Robotiq 2F gripper) to execute **open-vocabulary pick-and-place tasks** in real-world environments. The system should understand high-level, **robot-agnostic natural-language instructions** (e.g., "Bring me the red cup") and ground them in robot-specific functions for perception, planning, and manipulation. Students will first validate their approach in simulation (Isaac Sim) and then deploy on the real robot.

Motivation and Background

Robots in labs and industry often rely on rigid pipelines: they only recognize a fixed set of objects and follow pre-programmed tasks. This limits flexibility and generalization. In contrast, **open-vocabulary perception and robot-agnostic commands** allow robots to act on arbitrary user instructions without retraining for each new object or task.

Recent frameworks such as *Stretch-Compose* show how open-vocabulary reasoning, perception, and manipulation can be combined using vision-language models. For this 9 ECTS lab project, the focus is on the **open-space pick-and-place case**: mapping human instructions to robot actions and demonstrating execution with onboard sensors.

This project offers students:

- Hands-on experience with cutting-edge **robotics and AI integration**.
- Exposure to real robot platforms and large-scale simulation.
- Transferable skills in **ROS2, MoveIt2, deep learning models, and perception pipelines**.

Approach and Work Packages

The project is divided into work packages (WPs). Each student (2–3 students total, max. 4) will take responsibility for one WP, with integration and testing as a team effort.

WP1 – Perception Integration: Integrate open-vocabulary detectors/segmenters (e.g., OWL-ViT, GroundingDINO, SAM2) into the ROS2 pipeline. Convert detections into 3D using RGB-D sensors (L515, D405). Build point clouds via multi-view fusion in Isaac Sim and on the real robot.

WP2 – Command Grounding: Implement a mapping layer that converts robot-agnostic natural-language commands (e.g., "Bring me the red mug") into robot-specific action plans. Ensure modularity, so the same command layer could be reused on different robots.

WP3 – Manipulation and Control: Connect grounded detections to MoveIt2-based motion planning. Plan and execute grasps using GPD or analytical heuristics. Extend to **place** functionality (e.g., place object on table or bin).

WP4 – Deployment and Evaluation: Validate the pipeline in Isaac Sim first. Transfer to the Neobotix MPO-700 with UR5e and Robotiq 2F gripper. Conduct experiments with multiple open-vocabulary commands and report metrics (success rate, execution time, robustness).

Available Resources

- **Software:** Stretch-Compose framework (open-source), ROS2 (Humble), MoveIt2, Isaac Sim for simulation, Python and C++ integration.
- **Hardware:** Neobotix MPO-700 mobile base with UR5e arm and Robotiq 2F gripper, onboard RGB-D cameras (Intel L515, D405), LiDAR sensors.

Requirements

- Programming: Python (mandatory), Linux, shell scripting, and basic C++.
- Packages: Open3D, OpenCV, MoveIt2, ROS2, PyTorch-based perception models.
- Interest in robotics, machine learning, and working with real robot systems.

Expected Outcome

- A modular, ROS2-based pipeline that grounds robot-agnostic natural-language commands into robot-specific functions.
- Demonstration of open-vocabulary pick-and-place tasks for previously unseen objects in both Isaac Sim and on the real robot.
- Hands-on experience with real robot hardware and cutting-edge AI perception models.
- Well-documented results and reproducible code for future lab projects.