Enhancing Human-Robot Interaction using Large Language Models in Multi-Dimensional Image Capture and Robotic Manipulation

Background

Large Language Models (LLMs), which have traditionally been instrumental in text-based tasks, are now demonstrating capability in a wider spectrum of applications including dynamic interactions and complex task performances like voice control, image recognition, and object manipulation. Robots, serving as an intermediary between virtual agents and the real world, convert coded instructions into tangible actions. The synergy between LLMs and robotic systems can be harnessed to create more intuitive, accessible, and precise control systems, particularly in multi-dimensional image capture and object manipulation. The research group Realistic 3D has among its lab equipment an industrial robot that may act as a tool to theoretically and practically evaluate this exciting emerging field of research and engineering.

Problem Specification

Recent advancements in LLMs have showcased their potential in enhancing human-robot interaction, particularly through multimodal input handling, high-level reasoning, and plan generation.[1]

LLMs such as GPT-4 have been highlighted for its ability to guide different robotic based applications and thereby revolutionizing how robots interact with humans and their environment.[2] Moreover, the integration of Multimodal Large Language Models (MLLMs) in robotic vision applications presents a unified vision pipeline addressing object detection, segmentation, and identification.[3]

The advancements expressed above underline the potential use for an intelligent robotic arm capable of accurate positional capture, enhanced path control, and intuitive human-tech interaction. Both in everyday applications as well as a collaboration function when conducting scientific experiments.

This leads to the following research questions:

1. How can LLMs be seamlessly integrated to understand and execute voice commands, ensuring the robotic arm responds accurately and promptly?
2. How can an intuitive control interface be designed for the robotic arm, drawing inspiration from tools like Blender and RTtoolbox, to improve human-robot interactions?
3. How can a visual system using OpenCV be incorporated to enable more complex operations, as well as improved safety protocols that make the robot arm “aware” of its surroundings, particularly in multi-dimensional image capture and object manipulation?

Suggested Method

Conduct a thorough literature review to understand the state-of-the-art in LLMs, robotics, multi-dimensional image capture, and human-robot interaction, to ensure the project is grounded in current research and has the potential for meaningful contributions.

Explore and evaluate existing frameworks and tools such as Blender, RT Toolbox, and OpenCV, for designing a control interface and visual system of the robotic arm.

Design, implement, and evaluate a prototype robotic arm integrated with an LLM, focusing on real-world applicability, ease of interaction, and the precision of control systems.

Relevant Articles

[1] S. Vemprala*, R. Bonatti, A. Bucker and A. Kapoor, "ChatGPT for Robotics: Design Principles and Model Abilities," Microsoft, 2023.

[2] C. Zhang et al., “Large language models for human-robot interaction: A review,” Biomimetic Intelligence and Robotics, 2023

[3] “RoboLLM: Robotic Vision Tasks Grounded on Multimodal Large Language,” arXiv 2310.10221, 2023