Karlsruhe Institute of Technology

Intuitive Robots Lab (IRL)
Prof. Dr.-Ing. Rudolf Lioutikov

Project Type

- ☑ Master Thesis
- ☑ Bachelor Thesis
- ☑ Research Project

Supervisors

☞ Pankhuri Vanjani
@ pankhuri.vanjani@kit.edu
☞ Xinkai Jiang
@ xinkai.jiang@kit.edu

Difficulty

Algorithmic

Math

Application

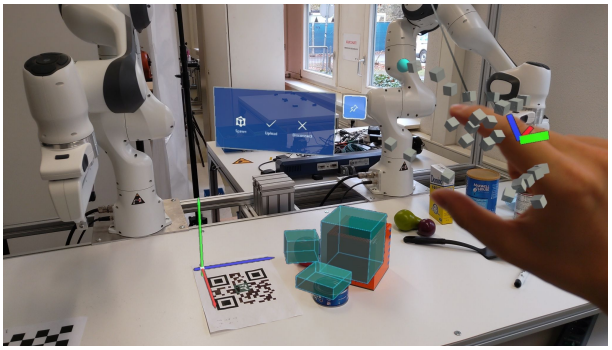
Requirements

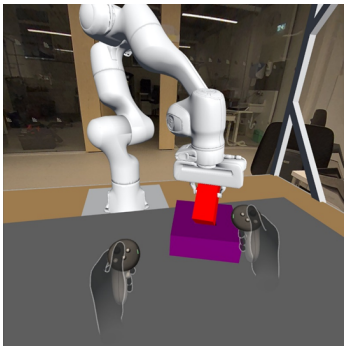
- ☑ Python
- ☑ C# & Unity
- ☑ Knowledge about pytorch and ML

Enhancing Imitation Learning with XR-Based Gaze and Gesture Tracking

Description

Gaze and gesture tracking play key roles in robot and imitation learning by revealing human intentions, attention, and decision-making cues [3]. These signals often highlight task-relevant objects and spatial relationships, helping robots better interpret context and prioritize actions. Incorporating them enables more adaptive learning models that mimic human strategies with improved accuracy and effectiveness. XR headsets support real-time gaze and gesture tracking in immersive environments, enabling naturalistic data collection in dynamic settings without restricting movement. In this project, we aim to advance the use of gaze and gesture as key modalities for improving robot manipulation. Rather than treating them solely as data sources, we focus on leveraging them for real-time intent recognition and feedback-driven correction during long-horizon tasks. By integrating these signals into the robot's perceptual and decision-making pipeline, we aim to enhance the interpretability, responsiveness, and adaptability of imitation learning models in complex, dynamic environments.





Tasks

- Hardware & Software Framework: Using gaze tracking setups (XR and vision based), develop a pipeline to generate a task-relevant efficient dataset.
- Gaze and gesture -Informed Demonstrations: Collect manipulation demonstrations using the IRIS framework [2], with integrated gaze and hand-tracking signals to inform task intent and attention dynamics [2].
- Policy Training and Evaluation: Design the architecture to integrate intent detection and feedback-based correction for SOTA methods [4], [1] on this dataset and Evaluate results in real world setting.

References

[1] Xiaogang Jia, Atalay Donat, Xi Huang, Xuan Zhao, Denis Blessing, Hongyi Zhou, Han A Wang, Hanyi Zhang, Qian Wang, Rudolf Lioutikov, et al. X-il: Exploring the design space of imitation learning policies. *arXiv preprint arXiv:2502.12330*, 2025.

[2] Xinkai Jiang, Paul Mattes, Xiaogang Jia, Nicolas Schreiber, Gerhard Neumann, and Rudolf Lioutikov. A comprehensive user study on augmented reality-based data collection interfaces for robot learning. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 333–342, 2024.

[3] Heecheol Kim, Yoshiyuki Ohmura, and Yasuo Kuniyoshi. Multi-task robot data for dual-arm fine manipulation. *arXiv preprint arXiv:2401.07603*, 2024.

[4] Moritz Reuss, Ömer Erdinç Yağmurlu, Fabian Wenzel, and Rudolf Lioutikov. Multimodal diffusion transformer: Learning versatile behavior from multimodal goals. In *First Workshop on Vision-Language Models for Navigation and Manipulation at ICRA 2024*, 2024.