

Research Statement

Yuming Chen

1 Introduction

My research interests lie in human-AI interactions in the real world. One challenge is that humans have different preferences and sometimes vague targets, which requires AI agents to reason about human strategic actions to provide better assistance. Moreover, real-world embodied interaction involves camera and sensor movement. It causes low-quality data such as blurry images. The AI agents need to correctly extract semantic information from low-quality data.

To address these challenges, my research focuses on *interaction modelling from partial observations* and *3D reconstruction from low-quality data*. Partial observation here does not only mean that the agent is not able to observe the whole environment. It also includes scenarios that the human target is not clear. For example, a person extends their hand to the robotic agent to ask for something. However, the agent does not know what is needed, causing ambiguity in target. To eliminate ambiguity, it is necessary to reason about personal preferences from the interaction history. Besides, vision perception is fundamental for real-world interactions. I will introduce my experience in these directions as follows.

2 Related Experience

2.1 Opponent Modelling

I have explored the problem of cooperating with an unseen partner in the setting of mixed cooperative-competitive games[2], from the perspective of multi-agent reinforcement learning (MARL) and game theory. Technically, we leverage contrastive learning to learn a consistent representation of the opponent's policy.

2.2 3D Hand Reconstruction from Blurry Monocular Images

Most current approaches focus on the blurriness problem in video, where the temporal information helps to reconstruct human motion. We recover the hand motion from a single blurry image to further utilize the temporal information in the image[1]. To overcome the ambiguity, we make multiple estimations for one image in a generative manner, and select the plausible ones with a learned selection module.

3 Future Plan

I currently plan to study the following problems:

- Making human motion estimation (HMP) with latent generative model. The generative model itself can be a basic model to predict future action in human-AI interaction.
- Learning to represent different preferences for people in the same interaction task. The representation can be used to better coordinate with humans.

The above are general and even rough ideas. I am willing to revise them to better fit the target of the team. I am also open to any other related topics.

References

- [1] Yuming Chen, Zhongqun Zhang, Yihua Cheng, Rongyu Chen, and Hyung Jin Chang. Multi-hypothesis 3d hand mesh recovering from a single blurry image. In *submission to International Conference on Multimedia and Expo (ICME)*, 2025.
- [2] Yuming Chen and Yuanheng Zhu. Policy representation opponent shaping via contrastive learning. In *International Conference on Neural Information Processing (ICONIP)*, 2023.