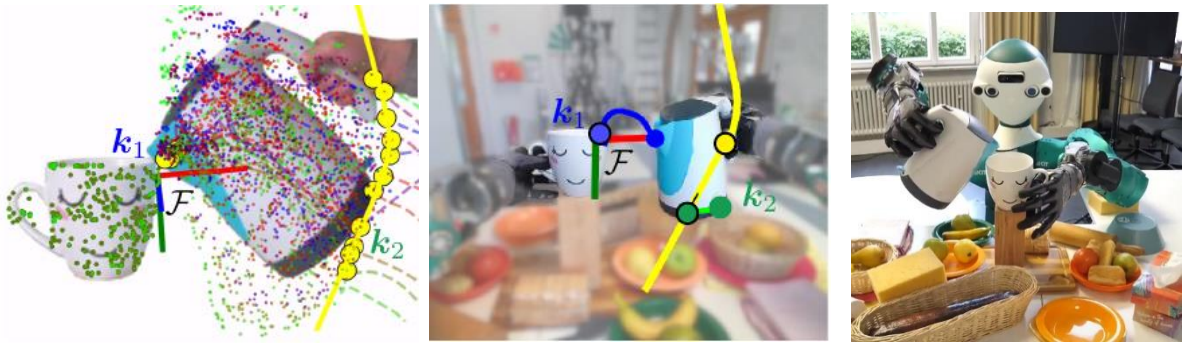
Keypoint Based Visual Imitation Learning Using Neural Radiance Fields



K-VIL of the pouring water task [1]

Visual imitation learning provides efficient and intuitive solutions for robotic systems to acquire novel manipulation skills. In our previous work [1], we proposed an approach for keypoint-based visual imitation (K-VIL) that jointly extracts explicit, sparse keypoints and endow them with geo-metric constraints of various types from a small set of demonstration videos. However, limited by the traditional 3D reconstruction methods and the visual perception models used in this work, K-VIL cannot be applied to transparent, reflective or think objects, which hinders K-VIL from being used in real-world applications. This challenge can be addressed by state-of-the-art scene representation models, e.g. the Neural Radiance Field (NeRF), which is able to reconstruct a scene in fine granular within a few seconds (see [2]) and to represent the aforementioned objects for ro-botic manipulation tasks (see [3]).

In this project, your will build the NeRF model for scene reconstruction and object representation based on [2], [3]. Then, human demonstration videos of manipulation tasks will be collected using RGB-D cameras to train the NeRF model, which will then be used to generate data to train the dense correspondence detection models used by K-VIL. Finally, to evaluate the visual perception models, you will apply K-VIL to learn the demonstrated cooking task using the trained models, and execute the learned task with ARMAR-6 humanoid robot to measure its generalization capability in novel situations.

Basic knowledge of computer vision, deep learning, Python and PyTorch is required. We provide the support and the opportunity to work in the research area of visual imitation learning.

Relevant research questions include:

- How to improve the dense correspondence model with the help of NeRF models?
- How to tackle occlusion and self-occlusion, especially in the case of bimanual manipulation?
- How to achieve extra-category generalization (transfer the learned skill to similar objects)?
- How to extract keypoints with uncertainties and relate them with affordances?

[1] Jianfeng Gao, et al. “K-VIL: Keypoints-based Visual Imitation Learning.” Preprint on arXiv, 2022
[2] Müller, Thomas, et al. “Instant neural graphics primitives with a multiresolution hash encoding.” ACM Trans. Graph. 2022.
[3] L. Yen-Chen, et al, “NeRF-Supervision: Learning dense object descriptors from neural radiance fields,” in ICRA, 2022.

Contact: Jianfeng Gao (jianfeng.gao@kit.edu)

Institut für Anthropomatik und Robotik | Lehrstuhl Prof. Asfour (H²T) | www.humanoids.kit.edu