Learning Spatio-Temporal Task Models from Human Demonstrations

In the future, humanoid robots are expected to operate autonomously in their environment in order to assist humans in their daily life. Therefore, robots must be able to learn and solve new and unknown tasks, such as setting the table and preparing a meal. Humans can transfer knowledge about a task simply by showing it to others, and correcting the other during imitation if needed. Robot assistants should have such capabilities when interacting with a human as well, especially since (re-)programming the robot by non-experts is generally not feasible.

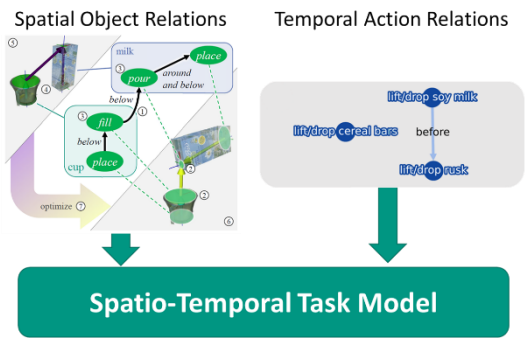

Figure 1: Combining spatial object relations with temporal action relations to obtain a spatio-temporal task model.

One integral component in the Programming by Demonstration cycle is concerned with the question of how to model what the robot perceives in such a way that it can reproduce the task at hand. These task models are representations of tasks, encoding all essential information needed to successfully execute the task in novel situations and contexts [1]. Especially symbolic or geometric constraints to be derived from demonstration are important since they must be obeyed to ensure a successful task execution. Apart from that, these constraints are essential for generalization to novel situations.

Two essential types of constraints that can be encoded in a task model are spatial relations between objects and temporal relations between actions manipulating these objects. Spatial relations symbolically encode the relative locations of objects, e.g. *a fork is on the right side of a plate*, or *the noodles are inside of the pot*. Temporal relations describe whether two actions must be executed sequentially or in parallel, e.g. *set the table before serving the meal* or *mix while adding water*. Such symbolic representations are typically understandable and transferable, but a robot must be able to extract them from subsymbolic human demonstrations, and generate subsymbolic action parameters for execution.

Recently, we have introduced task models based on spatial object relations [2] and temporal action relations [3] as well as methods for learning and execution. However, these task models are currently independent of each other. The goal of this work is to develop a **unifying task-model for representing and learning spatio-temporal relations** in manipulation tasks as well as methods to execute them by combining our previous works on spatial and temporal relations.

We encourage the student(s) to publish the results of this work on a highly prestigious international conference for robotics. Knowledge in C++ and Python is required for this work.

Contact: Christian Dreher (c.dreher@kit.edu), Rainer Kartmann (rainer.kartmann@kit.edu)

[1] Krüger, N., et al., *Object-Action Complexes: Grounded Abstractions of Sensorimotor Processes,* Robotics and Autonomous Systems, 2011
[2] Kartmann, R., Liu, D. and Asfour, T., *Semantic Scene Manipulation Based on 3D Spatial Object Relations and Language Instructions*, IEEE/RAS International Conference on Humanoid Robots (Humanoids), 2021
[3] Dreher, C. R. G. and Asfour, T., *Learning Temporal Task Models from Human Bimanual Demonstrations*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022