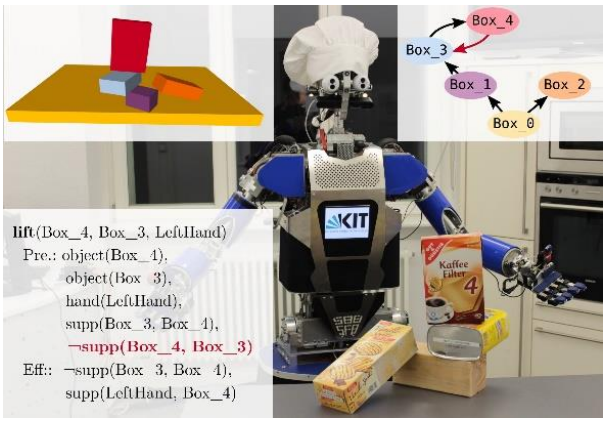
Learning Relational Scene Manipulation from Human Demonstration

Humans describe and manipulate scenes based on their relational understanding of the world (e.g. “The coffee filters are on the table”, “Put the cookies on top of the wooden block”). This relational reasoning is a key component of human intelligence. Robots perceive the world through a continuous stream of sensor data. In order to reason on this subsymbolic data, it is necessary to extract symbolic information about the objects and their relations.



The humanoid robot ARMAR-III extracts objects and physically plausible support relations between them from its depth camera. These relations are needed for planning and prediction.

To this end, the goal of this project is the design and collection of a new dataset about semantic scene manipulation. The dataset should contain videos of the manipulation actions executed by a human demonstrator, the relational changes in the scene and annotations about the task. A website will present the dataset and annotation tools which can be used to extend it in a crowd-sourced way. The dataset will be the basis for learning relational scene manipulation from human demonstration.

This work includes the design of the dataset, the data collection as well as the presentation in form of a website. For the data collection, the robotics framework ArmarX will be used (requires C++ or Python). Web development skills are required for the website design.

References

[1] Kartmann, Rainer, Paus, Fabian, Grotz, Markus and Asfour, Tamim, "Extraction of physically plausible support relations to predict and validate manipulation action effects." IEEE Robotics and Automation Letters 3.4 (2018): 3991-3998.

Kontakt: Fabian Paus (fabian.paus@kit.edu)

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