Recognizing Intention and Providing Assistance Based on Semantic Task Models

In order to provide effective and desirable help, assistive robots in household or workshop environments must understand what humans are doing – not only which actions they are performing at the moment but also what higher-level goal or task they are going to achieve, i.e. what the intention behind their actions is. Reasoning from perceived actions to intentions requires knowledge about high-level tasks: Which actions and what kind of objects are involved? In which order are actions performed and how is the scene changing as a consequence? Such knowledge can be encoded in Semantic Task Models which can be hand-crafted or learned from demonstration. After understanding the high-level task, the robot can reason about which actions may come next and which objects are required.
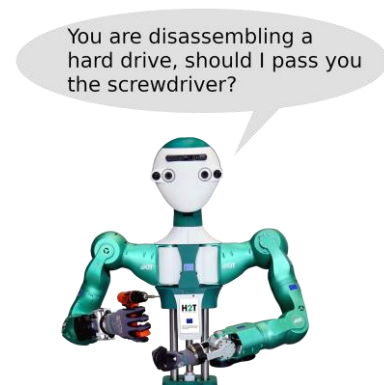

Fig. 1: The humanoid robot ARMAR-6 offering help.

Based on this understanding, the robot can offer assistance by asking whether it should perform an action or bring an object to the human to help. If the human confirms that they would appreciate help, the robot can finally execute the desired actions based on the task and action models.

In this work, you will investigate how action recognition, scene perception and semantic task models encoding temporal relations between actions (among other information) can be used to recognize the intention of a human performing an action, reason about which actions can be executed next, and offer assistance via a language dialog. To this end, recognized actions and objects in the scene can be used to identify the task a human is intending to accomplish. Together with the corresponding semantic task model, consecutive actions can be inferred or missing objects identified. These serve as candidates for the robot to offer help by suggesting to perform certain actions or fetch needed objects via a speech interface.

Relevant research questions include:

- What kind of knowledge is required in Semantic Task Models to allow recognition of higher-level tasks and intentions based on recognition of single actions?
- How can the robot decide which actions it can perform to assist?
- How can the robot verbalize its understanding in order to provide assistance?

This work will use the humanoid robot ARMAR-6 and/or the KIT multi-functional Gripper, as well as several robotics and machine learning tools:

- ArmarX (C++, Python): armarx.humanoids.kit.edu
- Tensorflow or PyTorch (Python): tensorflow.org, pytorch.org

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

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