Multi-modal Reasoning for Action Understanding in Household Environments

Filippos Gouidis[1,2#*], **Theodoros Patkos**[1], **Dimitirs Plexousakis**[1,2] and **Antonis Argyros**[1,2]

[1] Institute of Computer Science – FORTH Hellas

[2] Computer Science Department, University of Crete, Greece

Presenting author: Gouidis Filippos, email: gouidis@ics.forth.gr

* Corresponding author: Gouidis Filippos, email: gouidis@ics.forth.gr

ABSTRACT

Social Robotics (SR) is concerned with the development of entities that improve the daily life of humans, by exhibiting commonsense behavior and by offering intuitive means of interaction. Apart from its important theoretical aspects, SR holds also important practical interest, since due to the increase of like expectancy and to other socioeconomic factors, it is expected that the role of robots engaged in settings involving humans will become more and more important in the future.

A competent robotic household helper should be endowed with a host of capacities, such as locomoting smoothly, finding objects and manipulating them accordingly, understanding orders and instructions from humans and cooperating with them in the execution of tasks, when necessary. Moreover, the robot should be able to learn to perform new tasks and to improve, if possible, upon the execution of the skills that it already acquires.

Each of the previous requirements introduces a number of challenges involving, among others: reasoning and acting in the presence of incomplete or partially erroneous information; integrating information that stems from different sources of sensory input; utilizing different type of knowledge bases and resolving the possible knowledge inconsistencies and conflicts that occur.

In this work, we are interested in developing a reasoning framework suitable for robotic agents engaged in household tasks. The purpose of this mechanism is the inference of the constituents of actions (preconditions, partial ordering and effects). We aspire to implement this framework in a multi-encompassing manner that combines approaches stemming from different domains. Namely we would like to use prior knowledge in the form of KBs and rules (Symbolic AI), low-level features selection (Computer Vision) and web-mining enabling semantic extraction (Semantic Data Mining).