RoboCup@Home – Benchmarking Domestic Service Robots

Sven Wachsmuth
Bielefeld University
Germany
swachsmu@techfak.uni-bielefeld.de

Dirk Holz
University of Bonn
Germany
dirk.holz@ieee.org

Maja Rudinac
Delft University of Technology
The Netherlands
M.Rudinac@tudelft.nl

Javier Ruiz-del-Solar
Universidad de Chile
Chile
jruizd@ing.uchile.cl

Abstract

The RoboCup@Home league has been founded in 2006 with the idea to drive research in AI and related fields towards autonomous and interactive robots that cope with real life tasks in supporting humans in everday life. The yearly competition format establishes benchmarking as a continuous process with yearly changes instead of a single challenge. We discuss the current state and future perspectives of this endevour.

Research on autonomous robots and human-robot interaction have been core-fields of artificial intelligence since the very beginning. Domestic service robots combine both aspects in an application-oriented scenario. In order to solve typical tasks, like taking orders and serving drinks, welcoming and guiding guests, or just cleaning up, they require to integrate a large number of skills in a smooth and coordinated behavior. Measuring and comparing the performance of such robotic systems is notorously difficult. Reported research results are typically validated through experimental evaluation. In the last years, there have been several approaches that either focus on the re-producibility of such tests (Bonarini et al. 2006; Lier et al. 2014) or focus on live competitions and challenges (e.g. AAAI Robot Competition, ICRA Robot Challenges, DARPA Grand Challenges, ELROB, or RoCKIn@Home). The balanced definition of such tests remains a challenging task. RoboCup@Home is part of the RoboCup initiative (www.robocup.org) and defines a live competition of service robots that need to fulfill a series of tests in a domestic environment. Since 2006, the rulebook of the competition is changed on a yearly basis. Here, we discuss the main trends and how we use statistics to drive the development of the league.

RoboCup@Home

The RoboCup@Home league has been initiated by Tijn van der Zant and Thomas Wisspeinter (van der Zant and Wisspeintner 2005). The first competition was held at RoboCup 2006 in Bremen, Germany, as a demonstration league, before becoming an official league in 2007. The core idea was to go beyond the closed artificial environments of the soccer and rescue arenas, to allow human interaction and human

intervention, and to advance technologies by coping with real-life tasks in supporting humans in everyday life. Instead of defining a final challenge, the RoboCup@Home competition is defined as a benchmarking process that guides the system development from simple tasks towards more and more complex tasks in realistic environments. In the scoring scheme each task is broken down into different sub-tasks that focus on different skills and are scored in a binary manner. This scheme drives the league towards more general skill implementations across tasks and provides an instrument to analyze and control the development in the league.

The competition is organized in two stages plus a final of the top five teams. There are several tests in each stage that are defined by the technical committee (TC) covering a wide scope of skills (navigation, mapping, person recognition, person tracking, object recognition, object manipulation, speech and gesture recognition, as well as higher cognitive functions such as planning and decision making). These tests are incrementally adapted over the years setting challenges within reach of current state-of-the-art in artificial intelligence and related fields. The focus is on the flexible use, robustness, and coordination of skills, the robot autonomy, and the smoothness of human-robot interaction. In 2014, the stage-1 consisted of the following tests defined by TC:

- Basic Functionalities: This is a pipelined test focussing on a sequence of skills (object manipulation, navigation and obstacle avoidance, person detection and speech understanding).

- Follow Me: The robot needs to follow an unknown person through a public space dealing with narrow spaces and several interventions by other people blocking the way.

- Emergency situation: The robot needs to detect the event of a falling person and must flexibly react on this.

In stage-2 robots need to cope with more complex tasks:

- Cocktail Party: The robot needs to efficiently find persons, get their orders, fetch the orders from the kitchen, and deliver the orders to the correct persons.

- Enduring General Purpose Service Robot: An operator verbally specifies a complex, only partially defined, or inconsistent task to the robot. Any skill may be requested. The robot needs to perform it, report any problems and find alternative solutions.
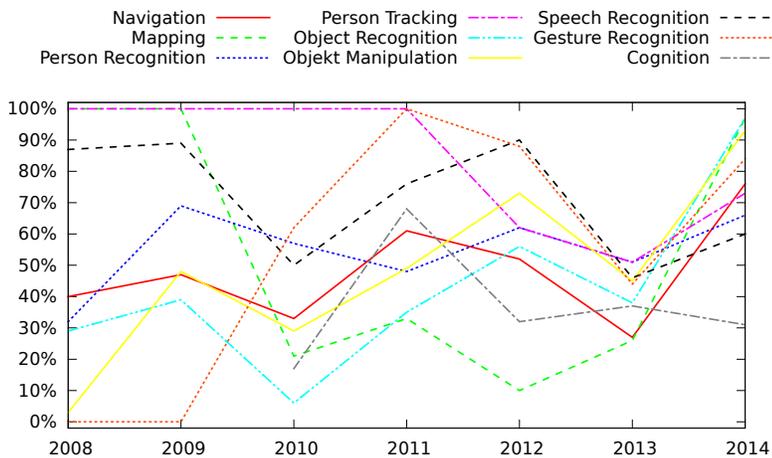
Figure 1: Statistics on the performance of the best team in each capability. 100% means that a team gained all points for those sub-tasks where this skill is relevant as a point of failure.

- Restaurant: An operator guides the robot though a previously unkown restaurant showing the important locations. After that the robot needs to fulfill verbal orders and deliver requested items to the specified tables.

Additionally, there are open challenges that are defined by the teams. Here, the teams are able to show specific performances beyond the state-of-the-art. Examples of the last years include robotic tool use (NimBro), 3D semantic mapping and scene understanding (TU/e and ToBI), speaker localisation in noisy environments (Golem), or multi-robot object manipulation (WrightEagle).

Statistics and league development

In Fig. 1 the performance of the best team in each of the tests defined by TC is analyzed over the last years with regard to the different skills (for more details see (Holz et al. 2014)). In order to illustrate how these statistics are used to drive certain rulebook changes, we will discuss two examples:

- In 2009, neither mapping nor speech recognition have been a typical point of failure. As a consequence, a new test (supermarket) was defined where robots need to deal with real unkown environments that require simultaneous localization and mapping approaches, and additional scores are introduced for on-board microphones instead of using head-phone devices.

- In 2012, the difficulty of person tracking in public spaces (Follow Me) was increased by introducing narrow spaces (elevators) and blocking person crowds.

Overall the league development is driven towards more realistic tasks by having less things decided by teams (set of objects, manipulation places, selection of operator, etc.), scaling up problems (larger object sets, more flexible spoken instructions, etc.), and less pre-knowledge (unknown objects, unkown environments, less pre-planned and more event-based acting).

In order to proceed on this track in the next years, we want to strengthen several aspects in the structure of the competition. In order to improve the benchmarking character of the competition, the number of tests per skill need to be increased. At the same time, setups need to be simplified in order to approach the application requirement of having robots running 'out of the box'. Getting closer to real world appications further requires longer operation times and more sophisticated ways of dealing with failure situations that currently are critical 'show stoppers'. Many of these goals (no setup, management of on-board resources, behavior monitoring) require more sophisticated methods from artificial intelligence. Thus, we hope that these aspects together with the improved benchmarking aspect will make the RoboCup@Home competition even more attrative for this research community.

References

Bonarini et al., A. 2006. Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets. In *IROS'06 Workshop on Benchmarks in Robotics Research*, volume 6.

Holz, D.; Ruiz del Solar, J.; Sugiura, K.; and Wachsmuth, S. 2014. On robocup@home – past, present and future of a scientific competition for service robots. In *Proceedings of the RoboCup International Symposium*.

Lier, F.; Wienke, J.; Nordmann, A.; Wachsmuth, S.; and Wrede, S. 2014. The cognitive interaction toolkit – improving reproducibility of robotic systems experiments. In *Simulation, Modeling, and Programming for Autonomous Robots*, 400–411. Springer International Publishing.

van der Zant, T., and Wisspeintner, T. 2005. Robocup x: A proposal for a new league where robocup goes real world. In *Proceedings of the RoboCup International Symposium*, 166–172.