RoboCup@Home: Results in Benchmarking Domestic Service Robots

Thomas Wisspeintner[1], Tijn van der Zan[2], Luca Iocchi[3], and Stefan Schiffer[4]

[1] Department of Mathematics and Computer Science
Freie Universität Berlin, Berlin, Germany
[2] Department of Artificial Intelligence
University of Groningen, Groningen, The Netherlands
[3] Dipartimento di Informatica e Sistemistica
University "La Sapienza", Rome, Italy
[4] Knowledge-Based Systems Group
RWTH Aachen University, Aachen, Germany

Abstract. Benchmarking robotic technologies is of utmost importance for actual deployment of robotic applications in industrial and every-day environments, therefore many efforts have recently focused on this problem. Among the many different ways of benchmarking robotic systems, scientific competitions are recognized as one of the most effective ways of rapid development of scientific progress in a field. The ROBOCUP@HOME league targets the development and deployment of autonomous service and assistive robot technology, being essential for future personal domestic applications, and offers an important approach to benchmarking domestic and service robots.

In this paper we present the new methodology for benchmarking DSR adopted in RoboCup@Home, that includes the definition of multiple benchmarks (tests) and of performance metrics based on the relationships between key abilities required to the robots and the tests. We also discuss the results of our benchmarking approach over the past years and provide an outlook on short- and mid-term goals of @HOME and of DSR in general.

1 Introduction

Creating a personal Domestic Service Robot (DSR) is a very complex task that requires cooperation between many scientific disciplines. DSRs have to operate in realistic and unconstrained environments which include humans. They must acquire on-line knowledge about both the animate and inanimate world and need to be very robust to unpredictable and changing environments and safe in the interactions with humans and the environment. This requires the integration of many abilities and technologies including: HRI, reasoning, planning, behavior control, object recognition, object manipulation or tracking of objects. Regarding artificial intelligence, the systems should contain adaptive but robust behaviors and planning methods, social intelligence, and on-line learning capabilities.

The recent increase in availability, accessibility, and compatibility of essential robot hardware and software components allows research groups not only to address a small

J. Baltes et al. (Eds.): RoboCup 2009, LNAI 5949, pp. 390–401, 2010.

subset of the mentioned above challenges in DSR, but also to address the problem as a whole, without having to sacrifice a focus in a specific research topic.

This progress is also confirmed by the presence of some rather specialized service robotic applications on the market. Such applications include floor cleaning (e.g. Roomba and Scooba), lawn mowing (e.g. Robomow), and surveillance (e.g. Robowatch). Still, these applications are missing some essential properties of a multi-purpose autonomous and intelligent domestic service robot. Prominent examples of domestic and personal assistant robot research projects are ReadyBot[1] and PR2[2], while Wakamaru[3] and PaPeRo[4] focus more on social interaction studies. Many of these projects address relevant aspects of DSR. Still, what appears to be missing is a joint continuous international and multi-disciplinary research and development effort which also includes the aspect of application-oriented benchmarking of systems in DSR.

The ROBOCUP@HOME league [1] targets development and deployment of autonomous service and assistive robot technology being essential for future personal domestic applications. It is part of the international RoboCup initiative and it is the largest annual service and home robotic competition world-wide. ROBOCUP@HOME aims to be a combination of inter-disciplinary community building, scientific exchange, and competition, that iteratively defines benchmarks and performance metrics on which service robots can be evaluated and compared in a realistic domestic environment.

Since the real world is not standardized, measuring the performance of non standardized robots acting in it is a difficult task. The experimental paradigm to evaluate complex robotic systems has to use consequent scientific analysis to improve on itself. Measuring the performance of the robots requires continuous reconsideration of the methodologies used, since both the robots (their capabilities) and their operation environment (and the robot's tasks) will definitively change over time. In our case, the tools are specific benchmarks testing certain robot abilities and the measurement of the robots' performance in the tests. We firmly believe that creating and applying this experimental paradigm can greatly improve and accelerate the development in DSR as it already is the case in other robotic areas like Rescue Robotics or the Robot Soccer leagues.

This paper presents a new methodology for benchmarking DSR. The proposed approach defines multiple benchmarks (tests) related to DSR and performance metrics based on the relationships between key abilities required to the robots and the tests. We also discuss the results of the ROBOCUP@HOME benchmarking in the past years and provide an outlook on short- and mid-term goals of @HOME and of DSR in general.

2 Benchmarking Domestic Service Robotics

Benchmarking has been recognized as a fundamental activity to advance robotic technology [2,3] and many activities are in progress, such as the EURON Benchmarking

[1] ReadyBot (http://www.readybot.com/)

[2] PR2 (http://www.willowgarage.com/)

[3] Wakamaru (http://www.mhi.co.jp/kobe/wakamaru/english/)

[4] PaPeRo (http://www.nec.co.jp/robot/english/robotcenter_e.html)

Initiative[5], the international workshops on Benchmarks in Robotics Research and on Performance Evaluation and Benchmarking for Intelligent Robots and Systems, held since 2006[6], the Rawseeds project[7], which aims at creating standard benchmarks specially for localization and mapping, the RoSta project[8], which focuses on standardization and reference architectures, etc.

Benchmarking can be distinguished in two classes: *system benchmarking*, where the robotic system is evaluated as a whole, and *component benchmarking*, where single functionality is evaluated. Component benchmarking is very important to compare different solutions to a specific problem and to identify the best algorithms and approaches. However, it is not sufficient to assess the general performance of a robot with respect to a class of applications. Indeed, the best solution for a specific problem may be unfeasible or inconvenient when integrated with other components that compose a robotic application.

On the contrary, system benchmarking offers an effective way to measure the performance of an entire robotic system in the accomplishment of complex tasks, as such tasks require the interplay of various sub-systems or approaches. Thus, standard reference environment, reference tasks and related performance metrics are to be defined. However, when defining standard benchmarks two common problems arise: Firstly, the difficulty of defining a benchmark that is commonly accepted by the community (this is due to different view points of different research groups about a problem) and secondly, the risk of causing specialized solutions for a certain benchmark or problem that can not be applied in real world applications. To avoid these problems, scientific competitions have proven to be a very adequate method, because benchmarks are usually discussed and then accepted by all the participants. Especially annual competitions provide a constant feedback on a yearly basis about the increase in performance and allow for setting up medium-term projects.

Among the many robotic competitions, AAAI Mobile Robot Competitions was one of the first, being established in 1992 [4]. It has provided significant scientific and technological contributions but its focus and benchmarks change domain every year.

RoboCup (founded in 1997) [5] has currently the largest number of participants (e.g., 440 teams with over 2600 participants from 35 countries in 2006). The RoboCup soccer competitions offer evaluation through competition in the robotic soccer domain and have contributed to significant scientific achievements in the last ten years. However, the special focus on soccer tends to steer the solutions towards specialized robotic architectures.

Robot rescue games started in 2000 within the AAAI Mobile Robot Competition [7] and then since 2001 within the RoboCup Rescue initiative [8]. There is a large focus on benchmarking robots in an abstract and standardized environment. Common metrics for HRI have been defined [9] and effective evaluation of HRI techniques have been carried out [10,11], but the type of HRI via an operator station is different to the more direct kind of interaction desired in DSR.

[5] http://www.euron.org/activities/benchmarks/index
[6] All these workshops are summarized in http://www.robot.uji.es/EURON/en/index.htm.
[7] http://www.rawseeds.org/
[8] http://www.robot-standards.eu/

The DARPA Grand Challenge[9] is probably the most recognized competition in terms of public and media attention, and the one that is most directly application oriented. However, there is little relation to DSR, the benchmarking setting is very difficult to reproduce (participation and organization were very costly), and the continuation of this competition is uncertain.

Finally, educational contests, such as EUROBOT[10] or RoboCup Junior[11], are organized with the main goal of presenting robotics to young students and thus they deal with simpler tasks and robotic platforms.

Competitions that have a more direct relation to DSR mainly focus on a single task. The AHRC Vacuum Contest[12] and the 2002 IROS Cleaning Contest[13] [12] are focused only on floor cleaning, while ROBOEXOTICA[14] focusses on robots preparing and serving cocktails. The ICRA HRI Challenge[15] has a broader scope, namely the effectiveness of HRI, but lacks evaluation criteria for the performance.

Following the successful lines of RoboCup competitions and the experiences offered by other competitions related to DSR, the ROBOCUP@HOME annual competition has been set up as a *system benchmarking* activity for domestic service robotics.

3 The @HOME Initiative

ROBOCUP@HOME is a combination of scientific exchange and competition that provides standard benchmarks (called ”Tests”) and performance metrics on which personal domestic service robots can be evaluated and compared in a realistic domestic environment. This section briefly summarize the main concepts behind the @HOME competitions, that are useful for the following analysis. Details on the rules and on the organization can be found in the ROBOCUP@HOME web site[16]. In particular, in this section we will discuss the key features that we identified to be relevant for DSR and the @HOME competitions, and the score system of @HOME.

3.1 Key Features

An initial set of robot key features (abilities and properties) was derived from an analysis of the state of the art in DSR and from experiences and observations of other robotic competitions. These features help to design the benchmarks and the score system for the competition. Furthermore, these features allow for a later analysis of the teams' performances and help to develop and later enhance the competition in a structured way. As the competition with its benchmarks is expected to evolve over time, also the key features and their weights in the competition are expected to be adapted. The key features are divided in two groups: *functional abilities* and *system properties*.

9 http://www.darpa.mil/grandchallenge/index.asp
10 http://www.eurobot.org/
11 http://rcj.sci.brooklyn.cuny.edu/
12 http://www.botlanta.org/
13 http://robotika.cz/competitions/cleaning2002/en
14 http://www.roboexotica.org/en/mainentry.htm
15 http://lasa.epfl.ch/icra08/hric.php
16 http://www.robocupathome.org/

Functional abilities. *Functional abilities* include specific functionalities that must be implemented on the robot in order to perform decently in the tests. Each test requires a certain subset of these abilities as they are also directly represented in the score system. Teams must thus decide which of these abilities to implement and up to which degree of performance, depending on their background and the kind of tests they intend to participate in. *Functional abilities* currently are:

– *Navigation*, the task of path-planning and safely navigating to a specific target position in the environment, avoiding (dynamic) obstacles
– *Mapping*, the task of autonomously building a representation of a partially known or unknown environment on-line
– *Person Recognition*, the task of detecting and recognizing a person
– *Person Tracking*, the task of tracking the position of a person over time
– *Object Recognition*, the task detecting and recognizing (known or unknown) objects in the environment
– *Object Manipulation*, the task of grasping or moving an object
– *Speech Recognition*, the task of recognizing and interpreting spoken user commands (speaker dependent and speaker independent)
– *Gesture Recognition*, the task of recognizing and interpreting human gestures

System properties. *System properties* include demands on the entire robotic system that we consider of general importance for any domestic service robot. They can be described as "Soft Skills" which need to be implemented for an effective system integration and a successful participation in the @HOME competition. *System abilities* currently are:

– *Ease of Use* - Laymen should be able to operate the system intuitively and within little amount of time
– *Fast Calibration and Setup* - Simple and efficient setup and calibration procedures for the system
– *Natural and multi-modal interaction* - Using natural modes of communication and interaction like, e.g. using natural language, gestures or intuitive input devices like touch screens.
– *Appeal and Ergonomics* - General appearance, quality of movement, speech, articulation or HRI
– *Adaptivity / General Intelligence* - Dealing with uncertainty, problem solving, online learning, planning, reasoning
– *Robustness* - System stability and fault tolerance
– *General Applicability* - Solving a multitude of different realistic tasks

The *system properties* can not be benchmarked as directly as the *functional abilities*, because it is not possible to relate actual portions of the score to them. However, they are considered as integral and implicit part of the competition and the tests, because teams are required to provide their robot with these properties. We thus believe that ROBOCUP@HOME tests allow to measure improvements in the *system properties* through improvements in the *functional abilities*.

3.2 Implementation

The competition is organized in a multi-stage system where teams perform from 5 to 10 tests. The tests are oriented along realistic and useful tasks for a domestic service robot. Each test evaluates certain key features, as shown in the next section.

Two types of tests exist: pre-defined tests, which are specified in terms of the task to solve and the scoring; open tests, in which teams can either freely choose what to show (the Open Challenge and the Finals) or a topic is given according to which teams can do a demonstration (Demo Challenge). The following pre-defined tests were implemented in the 2008 competitions:

- Fast Follow: A person guides the robot through the scenario using voice and gesture commands.
- Lost&Found: Find certain objects in the scenario.
- Fetch&Carry: Find and bring and object to the user.
- Who's Who: Find,remember and distinguish unknown persons.
- Partybot: Find persons, receive orders and serve a drink.
- Supermarket: An unknown user has make the robot to retrieve certain objects from a shelf.
- Walk&Talk: Teach in locations in an unknown environment by showing the robot around.
- Cleaning up: Detect and arrange unknown objects on the floor.

The scoring in the pre-defined tests is oriented along the key features mentioned earlier, while the scoring in the open tests is based on an evaluation by a jury and on a list of criteria along which the jury should evaluate.

Scoring for the pre-defined tests uses a *partial score system*, in which a team receives a part of the total score for showing a part of the task's specification. Each of the partial scores is connected to one or more of the functional abilities and/or system properties. This does not only allow for assessing the fulfillment of these features individually, but it is also an incentive for teams to participate in a test even if they know that they cannot solve it completely.

4 Evaluation of Results and Discussion

One important objective for an annual scientific competition is to provide a common benchmark to many teams that allows for measuring the advances of performance over time and to develop relevant scientific solutions and results. In this section we describe and discuss the results obtained by the ROBOCUP@HOME teams both in terms of performance in the tests and in terms of scientific achievements.

4.1 Representation of Key Features in the Benchmarks

The score system of ROBOCUP@HOME allows for directly relating the desired abilities of the robots with scores that are gained during the competition and adapting future benchmarks accordingly.

Table 1 relates the *functional abilities* defined in Section 3.1 with the pre-defined tests described above. It quantifies the maximum score distribution per test with respect to the contained functional abilities. For ease of notation, we use abbreviations

Table 1. Distribution of test scores related to functional abilities

Test	Nav	Map	PRec	PTrk	ORec	OMan	SRec	GRec	Total
FF	550	0	0	450	0	0	0	0	1000
FC	375	0	0	0	150	400	75	0	1000
WW	350	0	550	0	0	0	100	0	1000
LF	550	0	0	0	450	0	0	0	1000
PB	1000	0	700	0	0	300	0	0	2000
SM	0	0	0	0	400	1000	200	400	2000
WT	918	416	0	250	0	0	416	0	2000
CL	1000	0	0	0	550	450	0	0	2000
Tot	**4743**	**416**	**1250**	**700**	**1550**	**2150**	**791**	**400**	**16000**

as follows. For the tests we have Fast Follow (FF), Fetch & Carry (FC), Who is Who (WW), Lost & Found (LF), PartyBot (PB), Supermarket (SM), Walk & Talk (WT), and Cleaning Up (CL). The abilities are Navigation (Nav), Mapping (Map), Person Recognition (PRec), Person Tracking (PTrk), Object Recognition (ORec), Object Manipulation (OMan), Speech Recognition (SRec), and Gesture Recognition (GRec).

Since the competition involves mobile robots, navigation is currently the most dominant ability represented in the score. Object manipulation and recognition also play an important role since service robots are useful if they can effectively manipulate objects in the environment. Person recognition, tracking, and speech/gesture recognition are needed to implement effective human-robot interaction behaviors. As gesture recognition was introduced as a new (and optional) ability in 2008, its weight in the total score still is comparably low. Finally, mapping plays a more limited role: such an ability is used in the Walk & Talk test, where the environment is completely remodeled during the test, so the robot enters in an unknown environment, while for other tests only minor modifications of the environment are done right before the tests and thus pre-computed maps (either built off-line by the robot or manually drawn) can be used.

It is important to observe that these values have been chosen after discussion among the members of the Technical Committee, taking into account the feedback from the teams. Consequently, the values implicitly contains a compromise between pushing towards new capabilities and rewarding more difficult functionalities (Technical Committee) and measuring current performance of the robots (feedback from teams). It is even more important to notice that our benchmarking approach is not to look for an optimal set of values, but to make them evolve over time in order to gradually improve the performance of DSR.

This table is important in order to define the weight of each ability in a test and in order to distribute the abilities among the tests. Furthermore, one can actually measure and analyze the performance of the teams and the difficulty of the tests after a competition, allowing for an iterative and constant development of the benchmarks.

Similar relations between system properties and the tests exist. However, this relation can not be quantified in scores as easily, as the system properties are of more implicit meaning for the tests.

System properties are instead represented in the general rules, in overall requirements, and special properties in certain tests. By using laymen to operate the robots in the Supermarket test, the Who is Who test, and the PartyBot test, *Ease of Use* (EUse) is fostered. The restrictions on setup time and procedures demands for *Fast Calibration and Setup* (FCal). *Natural Interaction* (NInt) and *Multi-modal input* is currently rewarded in the Supermarket test and by the common use of speech and gestures. *Appeal and Ergonomics* (App) are part of the evaluation criteria in the Introduce test, the Open Challenge, and the Finals. *Adaptivity* (Adap) is especially requested in the Cleaning Up test. The limited amount of specifications in the tests and the environment and the fact that persons who interact with the robot are chosen randomly in many tests demands for *Robustness* (Rob). Finally, a team can only reach the *Finals* if their robot performs well in many tests with different tasks to solve. This stimulates the claim for *General Applicability* (GAppl).

4.2 Analysis of Team Performance

In the following, we analyze the performance of the teams in these abilities during ROBOCUP@HOME 2008 competition.

Table 2 presents the scores actually gained by the teams during the competition and the percentage with respect to the total score available, related to each of the desired abilities. The third column shows the best result obtained by some team, while the fourth one is the average of the results of the five finalist teams. This table allows for many considerations, such as: 1) which abilities have been mostly successfully implemented by the teams; 2) how difficult are the tests with respect to such abilities; 3) which tests and abilities need to be changed in order to guide future development into desired directions.

From the table it is evident that teams obtained good results in navigation, speech recognition, mapping, and person tracking. Notice that the reason for a low percentage score in navigation is not related to inabilities of the teams, but to the fact that part of the navigation score was available only after some other task was achieved. Speech recognition worked quite well, especially considering that the competition environment is much more challenging than a typical service or domestic application due to a large

Table 2. Available points for the desired abilities

Ability	Available points	Achieved score [max]	Achieved score [avg]
Navigation	4743 (40%)	1892 (40 %)	1178 (25%)
Object Manipulation	2150 (18%)	75 (3%)	15 (1%)
Object Recognition	1550 (13%)	450 (29%)	125 (8%)
Person Recognition	1250 (10%)	400 (32%)	190 (15%)
Speech recognition	791 (7%)	692 (87%)	293 (37%)
Person Tracking	700 (6%)	700 (100%)	570 (81%)
Mapping	416 (3%)	416 (100%)	183 (44%)
Gesture recognition	400 (3%)	0 (0%)	0 (0%)
Total	**12000 (100%)**	**4909 (41%)**	**2554 (21%)**

amount of people and a lot of background noise. The good achievements in mapping and person tracking may instead be explained by a limited difficulty of the corresponding tasks in the tests.

On the other hand, in some abilities, teams were not very successful. Object manipulation is a hard task, specially when an object is not known in advance and calibration time is limited or null. Because of the large proportion of score available, many teams have attempted manipulation but only a few were successful. A similar analysis holds for object and person recognition, that reported slightly better results with the same difficulties arising from operating under natural environment conditions (i.e., lighting) with small or null calibration time. Finally, gesture recognition has not been implemented by teams, probably for the small amount of points available.

An evaluation of system properties is more complicated since they are difficult to quantify precisely. Our current approach is to test for system properties through general requirements and by enforcing the combination of functional abilities.

An analysis of these results is very helpful for the future development of the @HOME competition. It gives direct, quantitative feedback on the performance of the teams with respect to the key abilities and tasks. This allows us to identify abilities and respective tests which need to be modified, to adjust the weights of certain abilities with respect to the total score. Possible modifications involve:

- Increasing the difficulty if the average performance is already very high
- Merging of abilities into high-level skills, more realistic tasks
- Keeping or even decreasing difficulty if the observed performance is not satisfying
- Introducing new abilities and tests

As the integration of abilities will play an increasingly important role for future general purpose home robots, this aspect should be especially considered in the future competition.

In addition to the analysis of the last competition we have conducted an analysis of presence and performance of teams over the years. Since 2006, a total of 25 teams (12 from Asia, 8 from Europe, 4 from America, 1 from Australia), have participated to the three editions of the annual ROBOCUP@HOME world championship. The percentage of @HOME teams in the RoboCup increased from 2.7% in 2006 to 3.7% in 2008. For RoboCup 2009 we expect 23 teams from 14 countries.

Moreover, it is interesting to notice that some teams adapted their robots designed and built for other RoboCup Leagues to compete in @HOME and that one team in 2006 and 2007 used the same robot in both the soccer Four-Legged and @HOME leagues and one team in 2008 used the same robot in both the Rescue and @HOME leagues.

Another important parameter to assess the results of an annual competition is the increase of performance of the teams over the years. Obviously, it is difficult to determine such measure in a quantitative way: the constant evolution of the competition with its iterative modification of the rules and of the partial scores do not allow a direct comparison.

However, it is possible to define some metrics of general increase of performance. In Table 3, the first row holds the percentage of unsuccessful tests, i.e., tests where no score was achieved at all, dropping from 83% in 2006 to 41% in 2008. The second row shows the increase in the total number of tests per competition. The third row holds the average number of tests that teams participated in successfully (i.e., with a

Table 3. Measures indicating general increase of performance

Measure	2006	2007	2008
Percentage of 0-score performance	83%	64%	41%
Total amount of tests	66	76	86
Avg. number of succ. tests p. team	1.0	2.5	4.9

non-zero score). The enormous increase from 1.0 tests in 2006 to 4.9 in 2008 is a strong indication for an average increase in robot abilities and in overall system integration.

4.3 Scientific Achievements

Besides numerical analysis about performance in the tests, relevant scientific achievements have been obtained by teams participating in the competition. Regarding the evolution of robot hardware and software architectures, we found special focus on Human-Robot-Interaction (e.g. [13]), on personal assistive robots (e.g. [14]) and on high level programming for domestic service robots (e.g. [15]). Regarding specific functionalities, speech recognition evolved from artificial and unnatural interaction with headset and portable laptop (2006-2007) to speaker independent speech recognition with effective noise cancellation using on-board microphones (2008) [16][17]. Face recognition has been made robust in presence of spectators standing at the border of the scenario [17,18] and tuned for real-time use [19]. Object recognition under natural and dynamic light condition has improved significantly: Techniques using different feature extractors and matching procedures have been tested (e.g. [20]), reaching a level in which the robot can reliably memorize an object shown by a user (by holding it in front of the robot) and then recognize it among several others (2008). Gesture detection and recognition has also been studied in order to communicate with the robot, using an effective approach based on active learning [21]. Finally, object manipulation has evolved from gathering a newspaper from the floor (2006), to grasping cups from a table (2007), and grasping different objects on various heights (2008).

5 Conclusion and Outlook

This paper presented the RoboCup@Home initiative as a community effort to iteratively develop and benchmark domestic service robots through scientific competitions, by evaluating robot performance in a realistic, complex and dynamic environment. Starting with the first competition in 2006, the overall development of the initiative with respect to the increase of performance, the growing community, knowledge exchange and public awareness is very promising. @Home has become the largest international competition for domestic service robots with currently 5 national competitions in China, Japan, Germany, Iran and Mexico besides the annual world championships.

The future development of the @Home competition is highly iterative, as it involves constant feedback from the community, adjustments on the focus of desired abilities

[17] Best Student Paper Award at RoboCup International Symposium 2008.

and changes of the rules. Tests, functional abilities and desired system properties will evolve over the years and will be combined to form more realistic high-level tasks. At the moment a focus lies on physical and sensory capabilities, such as manipulation, human recognition and navigation. In the future, more focus will be put on Artificial Intelligence, high-level autonomy and mental capabilities in the context of HRI. This includes situation awareness, online learning, understanding and modeling of the surrounding world, recognizing human emotions and having appropriate responses. Moreover, as one of the main issues for domestic robots is their safety, we will consider in the future also evaluation methods for robot safe operation in domestic environments.

Still, concrete goals are necessary as they help to identify and to approach specific problems in the large real-world problem space in a structured way.

Rule changes in 2009 will involve an increased focus on HRI, e.g combined use of speech and gestures, robot operation by laymen or following previously unknown persons. Application scenarios will become more realistic, e.g. the demo challenge will involve robots serving drinks and food at a real party setting involving many people unfamiliar with the robots.

Mid-term goals include the search, identification, design and use of a common robot software architecture or framework to better exchange and reuse of software components already developed in the community and beyond. The same holds true for hardware, where companies or groups with relevant hardware components like sensors, actuators, or even standard robot platforms will be identified and asked to join and to support the community. Gradually testing the robots in the real world like e.g. going shopping in a real supermarket or taking the public transport is another mid-term goal.

The future @HOME scenario will contain more ambient intelligence, which the robots can interact with. The use of the Internet as a general knowledge base, communication with household devices, TVs, or external video cameras are some examples. Moreover, usability, safety and appearance of the robots will be of higher importance if one wants to increase their public acceptance.

References

1. van der Zant, T., Wisspeintner, T.: RoboCup@Home: Creating and Benchmarking Tomorrows Service Robot Applications. In: Robotic Soccer, pp. 521–528. I-Tech Education and Publishing (2007)
2. del Pobil, A.: Why do We Need Benchmarks in Robotics Research? In: Proc. of the Workshop on Benchmarks in Robotics Research, IEEE/RSJ International Conference on Intelligent Robots and Systems (2006)
3. Sabanovic, S., Michalowski, M., Simmons, R.: Robots in the wild: observing human-robot social interaction outside the lab. In: 9th IEEE International Workshop on Advanced Motion Control, pp. 596–601 (2006)
4. Balch, T., Yanco, H.A.: Ten years of the aaai mobile robot competition and exhibition: looking back and to the future. AI Magazine 23(1), 13–22 (2002)
5. Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E., Matsubara, H.: RoboCup: A Challenge Problem for AI. AI Magazine 18(1), 73–85 (1997)
6. The RoboCup Federation: RoboCup (2008), `http://www.robocup.org/` (Retrieved 09/2008)

7. Meeden, L., Schultz, A.C., Balch, T.R., Bhargava, R., Haigh, K.Z., Bohlen, M., Stein, C., Miller, D.P.: The aaai 1999 mobile robot competitions and exhibitions. AI Magazine 21(3), 69–78 (2000)
8. Kitano, H., Tadokoro, S.: RoboCup Rescue: A Grand Challenge for Multiagent and Intelligent Systems. AI Magazine 22(1), 39–52 (2001)
9. Steinfeld, A., Fong, T., Kaber, D.B., Lewis, M., Scholtz, J., Schultz, A.C., Goodrich, M.A.: Common metrics for human-robot interaction. In: Proc. of HRI, pp. 33–40 (2006)
10. Yanco, H.A., Drury, J.L., Scholtz, J.: Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition. Human-Computer Interaction 19, 117–149 (2004)
11. Drury, J.L., Yanco, H.A., Scholtz, J.: Using competitions to study human-robot interaction in urban search and rescue. Interactions 12(2), 39–41 (2005)
12. Prassler, E., Hägele, M., Siegwart, R.: International Contest for Cleaning Robots: Fun Event or a First Step towards Benchmarking Service Robots (2006)
13. Savage, J., Ayala, F., Cuellar, S., Weitzenfeld, A.: The use of scripts based on conceptual dependency primitives for the operation of service mobile robots. In: Proceedings of the International RoboCup Symposium 2008 (CD-ROM Proceedings) (2008)
14. Ruiz-del-Solar, J.: Personal robots as ubiquitous-multimedial-mobile web interfaces. In: Proc. of 5th Latin American Web Congress (LA-WEB), pp. 120–127 (2007)
15. Schiffer, S., Ferrein, A., Lakemeyer, G.: Football is coming Home. In: Proc. of International Symposium on Practical Cognitive Agents and Robots (PCAR 2006). University of Western Australia Press (2006)
16. Doostdar, M., Schiffer, S., Lakemeyer, G.: Robust Speech Recognition for Service Robotics Applications. In: Proceedings of the International RoboCup Symposium 2008, July 14-18. LNCS. Springer, Heidelberg (2008)
17. Correa, M., Ruiz-del-Solar, J., Bernuy, F.: Face recognition for human-robot interaction applications: A comparative study. In: Proceedings of the International RoboCup Symposium 2008 (CD-ROM Proceedings) (2008)
18. Knox, W.B., Lee, J., Stone, P.: Domestic interaction on a segway base. In: Proceedings of the International RoboCup Symposium 2008 (CD-ROM Proceedings) (2008)
19. Belle, V., Deselaers, T., Schiffer, S.: Randomized trees for real-time one-step face detection and recognition. In: Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008), December 8-11. IEEE Computer Society, Los Alamitos (2008)
20. Loncomilla, P., Ruiz-del-Solar, J.: Robust object recognition using wide baseline matching for robocup applications. In: Visser, U., Ribeiro, F., Ohashi, T., Dellaert, F. (eds.) RoboCup 2007: Robot Soccer World Cup XI. LNCS (LNAI), vol. 5001, pp. 441–448. Springer, Heidelberg (2008)
21. Francke, H., Ruiz-del-Solar, J., Verschae, R.: Real-time hand gesture detection and recognition using boosted classifiers and active learning. In: Mery, D., Rueda, L. (eds.) PSIVT 2007. LNCS, vol. 4872, pp. 533–547. Springer, Heidelberg (2007)