RoboCup@Home 2008: Analysis of results

Thomas Wisspeintner
Department of Autonomous Robots
Fraunhofer IAIS
Schönstr. 8, 81543 München, Germany
thomas.wisspeintner@iais.fraunhofer.de

Tijn van der Zant
Department of Artificial Intelligence
University of Groningen
Nijenborgh 9, 9747AG
Groningen, The Netherlands
tijn@ai.rug.nl

Luca Iocchi
Dipartimento di Informatica e Sistemistica
Sapienza University of Roma
Via Ariosto 25, Roma 00185, Italy
iocchi@dis.uniroma1.it

Stefan Schiffer
Knowledge-Based Systems Group
RWTH Aachen University
Ahornstr. 55, 52056 Aachen, Germany
schiffer@cs.rwth-aachen.de

This report describes an analysis of performance of teams participating at the ROBOCUP@HOME 2008 competition, in Suzhou, China. The analysis has been performed by defining a set of key abilities that are required to RoboCup@Home teams, by relating them with sub-scores of the tests and by measuring team performance on such abilities.

Results are useful to evaluate performance of teams, difficulty of each ability in the tests and to plan changes in the tests.

1 Definition of key abilities

The key abilities are divided in two groups: *functional abilities* and *system properties*.

Functional abilities include specific functionalities that must be implemented on the robot in order to perform well in the tests. Each test requires a certain subset of these abilities as they are also directly represented in the score system. Teams 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 safely moving to a specific target position in the environment, avoiding possibly 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 possibly 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 and moving an object
- *Speech Recognition*, the task of recognizing and interpreting spoken user input
- *Gesture Recognition*, the task of recognizing and interpreting human gestures

System properties include demands on the entire robotic system, which 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 properties are difficult to quantify. The initiative should also contain research on how to measure them appropriately. Initial system abilities are:

- *Ease of Use* - Laymen should be able to operate the system within a little amount of time
- *Fast Calibration and Setup* - Simple and efficient setup and calibration procedures
- *Natural and multi-modal interaction* - Natural interaction, e.g., using natural language and gestures, no keyboard input
- *Attractiveness and Ergonomics* - Look and feel of the robot, e.g., general appearance, quality of movement, speech, articulation
- *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

Although some of these properties can not be benchmarked as directly as the *functional abilities*, they are meant as integral and implicit part of the competition.

2 Evaluation

The score system of ROBOCUP@HOME allows for relating the desired abilities of the robots with points that are gained during the competition. In contrast to other competitions (e.g., RoboCup soccer), where the score hides many factors, the @HOME score provides an actual way of measuring the performance of teams in terms of such abilities, and consequently to analyze performance over time and to update the rules in order to drive technological and scientific progress.

2.1 Representation of key features in the benchmarks

In the following we show where the key features, i.e., the functional abilities as well all system properties are being tested.

2.1.1 Functional abilities

Table 1 relates the *functional abilities* defined in Section 1 with the tests. It quantifies the maximum score distribution per test with respect to the contained functional ability. For ease of notation, we use abbreviations 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). Note that for the Introduce test, the Open Challenge, the Demo Challenge, and the Final values are not indicated because teams can freely choose what their robot does in these tests.

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 since the environment is quasi-static and the ability is only needed in a single test.

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 analyze the performance of the teams and the difficulty of the tests after a competition. This allows for an iterative and constant development of the tests.

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**

Table 1: Distribution of test scores related to functional abilities

2.1.2 System properties

Similar relations between system properties and the tests exist. As previously mentioned, this relation can not be quantified in scores as easily, as the system properties are of more implicit meaning for the tests. However, on the basis of the objective of the tests, we can estimate the importance of each of the system property. In Table 2 we relate tests with system properties by denoting a 'very important' relation with '++', an important relation with '+', and a minor relation with '-'.

System properties are further 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	EUse	FCal	NInt	Attr	Adap	Rob	GAppl
IN	-	+	-	++	-	-	-
FF	-	+	-	-	-	+	+
FC	+	+	+	-	+	+	+
WW	+	+	++	-	+	+	+
LF	-	+	+	-	+	+	+
OC	-	+	+	+	+	-	+
PB	+	+	++	-	+	+	++
SM	++	+	++	-	++	+	++
WT	+	+	++	-	+	+	++
CL	-	+	-	-	++	+	++
Dem	+	+	++	+	+	-	++
Fin	+	+	+	++	+	-	++

Table 2: Importance of system properties in each test

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%)**

Table 3: Available points for the desired abilities

test, the Who is Who test, and the PartyBot test, *Ease of Use* is enforced. The restrictions on setup time and procedures demands for *Fast Calibration and Setup*. We generally only allow for *Natural Interaction* and offer *Multi-modal input* in the Supermarket test. The *Attractiveness and Ergonomics* are part of the evaluation criteria in the Introduce test, the Open Challenge, and the Final. *Adaptivity* is especially requested in the Cleaning Up test. The fact that we do not have concrete specifications throughout the tests and that persons to interact with the robot are chosen randomly demands for *Robustness*. Finally, a team can only reach the final if their robot performs good in many tests with different tasks to solve. This stimulates the claim for *General Applicability*.

2.2 Analysis of team performance

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

Table 3 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 second column shows the best result obtained by some team, while the third 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 steer 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 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 difficult. System properties are difficult to describe and quantify precisely. RoboCup@Home is a research track by itself in how to quantify these properties. 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.

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.

2.3 League progress

The results obtained so far by the @HOME initiative can be measured over several directions: 1) increased number of participating teams, 2) increasing performance in the tests, 3) increasing number of scientific contributions, 4) increasing number of community members, 5) increase of public awareness (media, press, Internet). For some of these measures we will present a quantitative analysis over the years.

Table 4 describes the increased number of participating teams. The first column shows the number of teams that preregistered and delivered the necessary qualification material like videos and a team description paper, the second column shows the number of teams that qualified after a review from the Organizing Committee, the third column shows the number of teams that finally participated in the competitions.

Year	Preregistration	Qualification	Participation
2006	20	17	11
2007	16	13	11
2008	18	17	14

Table 4: Number of participating teams

Figure 1 shows the participating @Home teams at the RoboCup world championship 2008 in Suzhou, China.

Figure 1: Participants of RoboCup@Home 2008 in Suzhou, China

Another important parameter to assess the results of the competition is the increase of performance. Obviously, it is very difficulty to determine such measure in a quantitative way, mostly because of the change of the rules year by year and the introduction of the new partial score system in 2008. Indeed, comparing performance results (as those presented in the previous section) year by year is not very effective, because of the constant increase of difficulty of the tests.

However, it is possible to identify certain situations that indicate the success of the initiative in terms of general increase of performance. Table 5 gives some examples for this increase over the last three years. The first row holds the percentage of unsuccessful tests, i.e., test 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 non-zero score). The enormous increase from 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.

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

Table 5: Measures indicating general increase of performance

In addition, Table 6 summarizes the number of teams participating in each test and those which received a non-zero score. This table helps to evaluate team preferences and difficulty of the tests.

Test	Participating Teams	Teams with non-zero score
Introduce	12	12
Fast Follow	12	12
Fetch & Carry	9	5
Who's who	8	4
Comp. Lost & Found	8	2
Open Challenge	13	13
Party Bot	5	2
Supermarket	3	3
Walk & Talk	10	10
Robot Chef	4	4
Cleaning	3	1

Table 6: Number of teams participating and gaining score for each test.

Other advancements can be identified in team performance in a specific functionality. For example, 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). Speech recognition evolved from difficult interaction with headset and portable laptop (2006-2007) to speaker independent speech recognition with effective noise cancellation using on-board microphones (2008).

As for the scientific contributions, in 2008, five papers related to ROBOCUP@HOME have been presented to the International RoboCup Symposium, including one that received the best student paper award [Doostdar et al., 2008]. In comparison to all the other ten RoboCup leagues and sub-leagues, @HOME ranked third with respect to the number of papers presented at the RoboCup Symposium (together with Soccer Middle-Size and Soccer Simulation).

Also, the community is growing very fast, with the mailing list currently having 250 subscribers (September 2008). The number and kind of subscriptions indicate that the mailing list is not only used by the teams but by various people from research institutions, other communities, universities, media, and companies.

As of September 2008, the @HOME Wiki obtained about 15,000 page views and more than 250 page edits since it was setup end of 2007. The most popular pages are the software page (1,258 views) and the hardware page (1,132 views) which strongly indicates that knowledge is actually being exchanged in the community.

Finally, the presentation of the ROBOCUP@HOME activities to media and press is increased thanks to the many world-wide and regional events in which the competition has taken place.

References

[Doostdar et al., 2008] Doostdar, M., Schiffer, S., and Lakemeyer, G. (2008). Robust Speech Recognition for Service Robotics Applications. In *Proceedings of the International RoboCup Symposium 2008 (RoboCup 2008)*, LNCS. Springer. to appear.