

Trends, Challenges and Adopted Strategies in RoboCup@Home

— Extended version —

Mauricio MATAMOROS*, Viktor SEIB† and Dietrich PAULUS‡

Active Vision Group — University of Koblenz

May 30, 2018

Abstract — Scientific competitions are crucial in the field of service robotics. They foster knowledge exchange and allow teams to test their research in unstandardized scenarios and compare result. Such is the case of RoboCup@Home. However, keeping track of all the technologies and solution approaches used by teams to solve the tests can be a challenge in itself. Moreover, after eleven years of competitions, it’s easy to delve too much into the field, losing perspective and forgetting about the user’s needs and long term goals.

In this paper, we aim to tackle this problems by presenting a summary of the trending solutions and approaches used in RoboCup@Home, and discussing the attained achievements and challenges to overcome in relation with the progress required to fulfill the long-term goal of the league. Hence, considering the current capabilities of the robots and their limitations, we propose a set of milestones to address in upcoming competitions.

With this work we lay the foundations towards the creation of roadmaps that can help to direct efforts in testing and benchmarking in robotics competitions.

Remark: Information from 2018 participants couldn’t be considered since the list of attending teams to Montreal 2018 and their scoring in the RoboCup@Home league was not available at the writing time of this manuscript.

1 Introduction

In the eleven years since its foundation in 2006, the RoboCup@Home league has played an important role fostering knowledge exchange and research in service robotics. Moreover, nowadays the competition can influence –and sometimes direct– the course of research in the area of domestic service robotics.

Having such impact is not a minor thing. In consequence, the RoboCup@Home league has the responsibility of planning carefully what needs to be tested and when to introduce changes by establishing milestones for the competition. This, of course, can only be done after analyzing the grounds in which the league is standing.

In response, in this paper we present two main contributions. First, we present a survey of the approaches and

technical solutions reported by teams in each of the different basic functionalities or abilities to accomplish a task. The conducted overview serves as a basis for our second contribution since it reveals to us the capabilities of the competing robots. To achieve this, we tap on several information sources, including: a) claims made in the Team Description Papers (TDPs), b) relevant publications, c) rulebooks, d) multimedia material available on-line, and e) our cumulative experience as participants and referees in RoboCup@Home since 2009.

Second, we discuss the challenges yet to be overcome that we have identified throughout these first eleven years of RoboCup@Home. This discussion considers not only the robot’s current capabilities, but the feats attained since the foundation of the league. In addition, we also consider information from conducted polls targeting potential customers necessities all over the globe but specifically from people living in Germany, Japan, Mexico, and The Netherlands. Consequently, we propose several sets of features, tasks, or applications that have to be addressed to achieve the goal of RoboCup@Home, but making special emphasis in those relevant in the short term. Nonetheless, it is important to point out that the discussion focuses exclusively on service robots (referred hereinafter as robots for simplicity), with particular application for domestic environments. In the same sense, we are focusing only on RoboCup@Home without ignoring the existence of related robotics competitions like the European Robotics League (ERL) and World Robot Summit (WRS).

This manuscript is organized as follows: In Section 2, we present a brief introduction to RoboCup@Home and its history (experienced competitors might want to skip this section). In Section 3, we provide a brief summary of adopted hardware solutions. Once studied physical constrains, in Section 4, we address the strategies and software solutions used to cope with the trials of the competition. Later on, we discuss the extrinsic and intrinsic challenges to overcome which are presented in Section 5 and Section 6, respectively. Finally, in Section 7 we close by presenting our conclusions.

*mauricio[at]uni-koblenz.de

†vseib[at]uni-koblenz.de

‡paulus[at]uni-koblenz.de

2 RoboCup@Home

The RoboCup@Home league was created in 2006 with the goal of developing robots capable of realizing all domestic chores and bringing them from the labs into people's homes. As stated in its website¹, "the RoboCup@Home league aims to develop service and assistive robot technology with high relevance for future personal domestic applications. It is the largest international annual competition for autonomous service robots and is part of the RoboCup initiative".

2.1 Organization

The competition takes place in a test arena. This arena is configured to look like a typical apartment of the hosting country. However, the specifications and appearance of the arena are kept undisclosed to participants until they arrive.

The competition normally lasts 6 days. During the first day and a half, period known as setup, teams prepare their robots and familiarize them with the arena; while the last day is reserved for the final demonstration and the award ceremony.

In the competition, which is divided in two stages, each robot has to solve a set of tests. In most tests, the robots have to solve a household-related task while their abilities and performance are evaluated. Some tests may take place outside the competition arena. However, the competition scenarios are never specified beforehand. Finally, the best 50% of the participants in Stage I advance to the Stage II, from which only the very best would advance to the Finals.

2.2 Brief History of RoboCup@Home

During the first two years, the tests were scored with boolean criteria. A team would receive points only if the robot successfully accomplished the given task. This had a lot of setbacks, but the most important one was that it was barely impossible to analyze the robot's performance in each ability from the scores. Therefore, a new scoring system was introduced in 2008 that is formally introduced and analyzed in [42]. From this year on, tests were split in a sequential set of goals. In this schema, a robot can't advance to the next step unless (successfully) completing the current one. This allowed to delimit the degree of uncertainty, making possible to estimate performance by measuring the contribution of each ability to each reached goal; as performed by Iocchi et al. in their analysis in [15]. In addition, this schema considers increasing the difficulty every third year, fine-tuning only in between. This provides teams with enough time to tune their algorithms for the newly introduced challenges.

However, by 2013 the Technical Committee (TC) had noticed a generalized decrease in performance (see Figure 1).

Many robots failed in tasks that were considered more or less solved. At the same time, it wasn't rare that robots couldn't even try². For this reason, a test was specially designed to measure performance in basic skills for 2014. Unfortunately, most of the robots did not come off well despite the relative simplicity of the benchmarking tasks. Based on these results, the TC decided to modify the test scheme proposed by Wisspeintner et al. [42] in Stage I.

Therefore, by 2015 tests with a sequential execution scheme were replaced with tests focused on measuring performance. The former had the intention of attracting new participants by easing the Stage I while the difficulty level of the Stage II continued increasing. This new schema, in addition to include semi-isolated benchmarking of relevant abilities, also tried to tackle the luck factor by requiring each ability test to be run three times and considering only the average of the best two runs. The tested abilities were 1) navigation, localization, and mapping, 2) people recognition, 3) people tracking, 4) object recognition and manipulation, 5) speech recognition and sound-source localization, and 6) ability integration.

The introduced changes quickly paid off. Besides providing benchmarking data to the TC, the changes made evident to teams where their own weaknesses were. This led to an immediate increase in performance in 2016 as Figure 1 shows. Consequently, the TC decided to raise once more the difficulty for 2017. The gathered data was used to analyze strengths and weaknesses, and push in those areas that required more attention. Notwithstanding, this new schema came also with several setbacks. It widened the breach between stages, led to a performance decrease in tests relying on integration, and fostered score chasing.

Since 2015 a vast number of new challenges were introduced (many as bait) in the Stage II with little success. These new challenges include describing untrained objects, guiding people, avoiding tiny and light-reflecting objects, pouring, etc. However, the newly introduced challenges were barely addressed. Despite tests being sequential, the rules now allowed robots to receive help or skip a goal that wasn't a logic requirement to accomplish the next one. As a consequence, participants went after easy to solve goals, completely ignoring the newly introduced challenges. Increasing the reward didn't work either. Nonetheless, RoboCup@Home became more appealing to the audience since robots were allowed to show something, and team's frustration became almost inexistent.

²Since 2010, the standard procedure involved waiting for the door to be opened, enter the arena, and retrieve a command. Failing to detect when the door was opened, or misunderstanding the given command would make it impossible for the robot to advance any further. Loose USB connectors and ambient noise promoted these problems even more. This caused many teams with robust manipulation, people recognition, task planning, etc., to leave the arena without showing at what they excelled.

¹Source: <http://www.robocupathome.org/> Retrieved: Jan 1st, 2018.

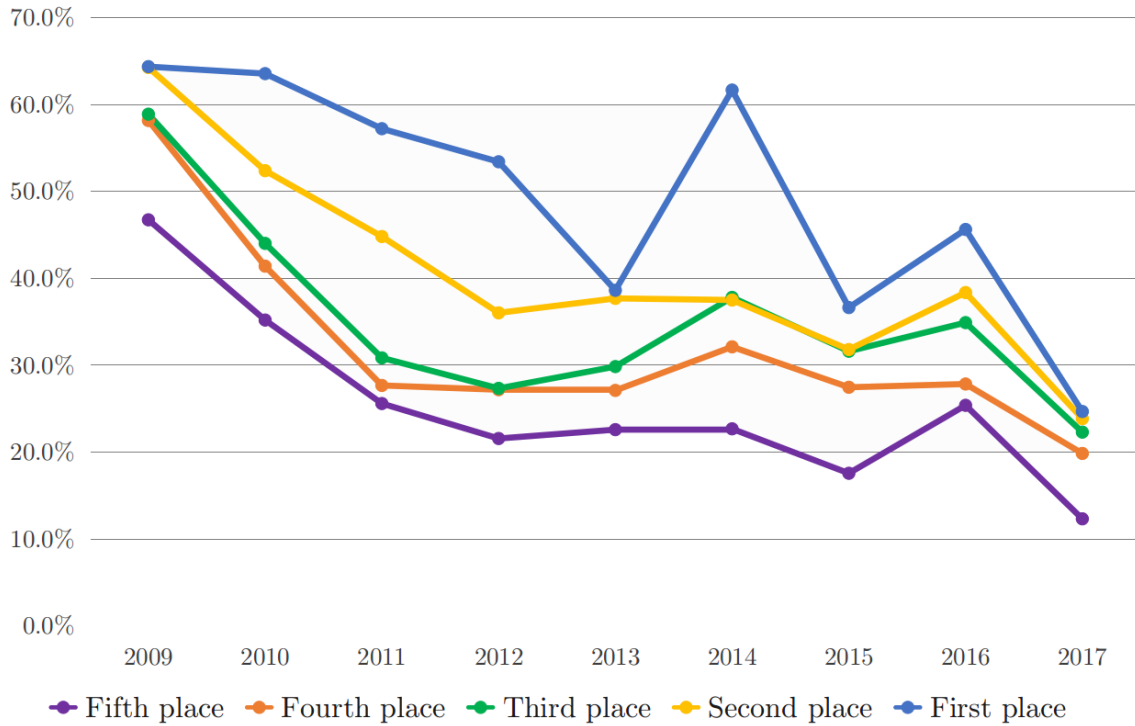


Figure 1: Performance of teams in the Top5 (2009 – 2017)
Obtained score with respect to the maximum attainable excluding special bonuses

Finally, 2017 introduced several important changes. RoboCup@Home was split in three leagues: 1) the legacy Open Platform League (OPL) with no hardware restriction, 2) the Domestic Standard Platform League (DSPL) akin to the OPL, but using standardized hardware, and 3) the Social Standard Platform League (SSPL) that focuses in high-level Human-Robot Interaction (HRI). This division caused several veteran teams to migrate to the Standard Platform Leagues (SPL), opening new spaces for new participants. The former, in addition with the increase in the difficulty of several tasks led to the lowest performances in the history of RoboCup@Home.

3 Summary of Hardware Solutions

In this section we present a summary of the hardware configurations most used in the OPL. We focus on five key aspects: 1) RGB-D sensor model, 2) drive mechanism of the base or locomotion type, 3) number of Degrees of Freedom (DoF) of the *head* 4) number of manipulators and their DoF, and 5) number of DoF of the *torso*. The aspects have been chosen for their potential influence in the robot's performance, either by expanding its perception or interaction with the environment.

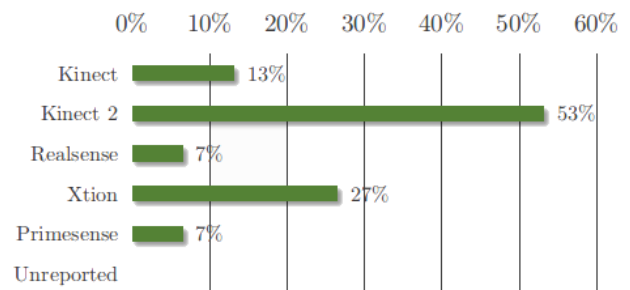


Figure 2: Most used RGB-D sensors (2017)

3.1 RGB-D Sensor

As of 2017, all teams report the use of at least one RGB-D sensor. From these, the preferred one seems to be the Microsoft Kinect 2 due to its incorporated Time of Flight sensor and better resolution, leaving the Asus Xtion in second place (see Figure 2).

3.2 Base

To the date, all robots in RoboCup@Home use wheels to move around. Therefore, it can be considered that all robots

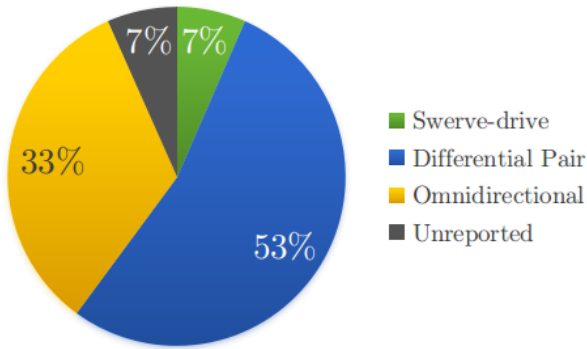


Figure 3: Reported locomotion types (2017)

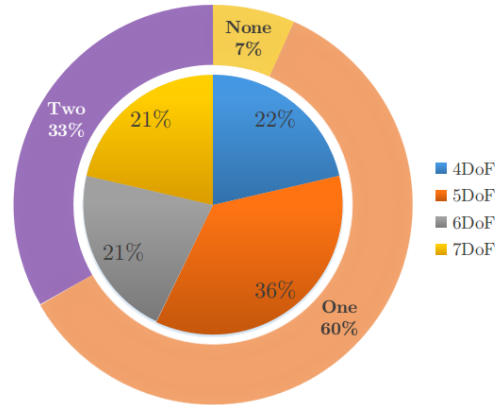


Figure 5: Adopted hardware solutions for manipulation

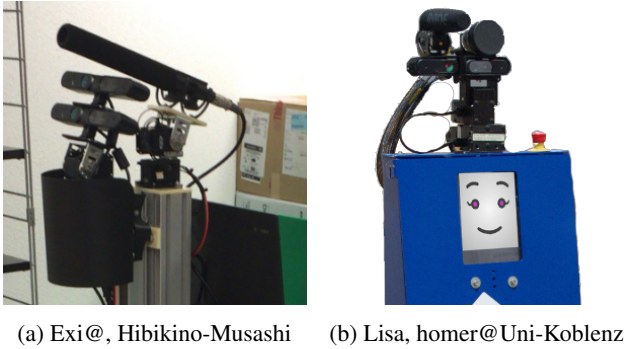


Figure 4: Addressing the right head may be deceiving.

have a base in which the driving mechanism lies. Furthermore, there is no evidence in the TDPs pointing at a possible change of paradigm in the preferred locomotion type.

From all possible configurations, the most used is the differential pair followed by omni-drive in either 3-wheeled or 4-wheeled configuration. There is only one reported use of *Swerve drive*, a special type of omni-directional configuration in which all four wheels rotate independently. This information is summarized in Figure 3.

3.3 Head

To humans, identify the head of something in nature is normally easy, intuitive, and straightforward. With robots, however, things change. An unfamiliarized operator can get easily confused when speaking to a bicephalous robot (Figure 4a), or a robot featuring a face located several centimeters away from its sensors (Figure 4b).

Hence, for the purpose of this summary we propose the following definition:

Robot head: Unit comprising at least a camera and a microphone mounted on a pan-tilt unit.

Based on that definition, we are confident to declare that 67% (10 out of 15) of the participant robots in Nagoya 2017 in the OPL featured a head. The established standard in the league are heads with 2 DoF.

3.4 Manipulator

Few people conceive a domestic service robot with no arms. In fact, most of the activities considered important by potential customers involve object handling. Moreover, most homes are designed to optimize spaces, while still being accessible to their inhabitants. Therefore, it would make sense that most robots had anthropomorphic configurations.

However, to handle objects, most team use either home-made or proprietary low-cost hardware [22, 41]. Although professional arms might be seen as the best option due to their strength and precision, their size makes them unfit for domestic narrow spaces. Therefore, they are rarely used in RoboCup@Home, with none of these present in Nagoya 2017. In contrast, home-made manipulators are usually anthropomorphic and much cheaper, although they lack the precision and strength of the former ones. In Nagoya 2017 the number of Degrees of Freedom for manipulators ranged from 4 to 7 with mode in 5 as Figure 5 depicts. Regarding the final effector strength, it is usually about 1.25kg, insufficient to lift a 1.5L bottle of soda or cutting food. Finally, since there is no mandatory task requiring two-handed manipulation, only few robots have more than one (see Figure 5).

3.5 Torso

Given an anthropomorphic configuration, a *torso* would provide a robot with a panning and variable elevation for its head and upper limbs. However, this is rarely the case. In Nagoya 2017, only 9 out of 15 teams (60%) reported a torso and, in all cases, it consisted of the elevator only. However,

in some cases, the elevator supported exclusively the manipulator. Therefore, it is necessary to clarify that, in strict terms, units providing an additional DoF to the manipulator shouldn't be considered torsos. Nonetheless, due to the lack of schematics, we decided to give our vote of confidence to teams, reporting the devices as alleged.

4 Adopted Strategies and Software Solutions

In this section, we summarize the solutions most commonly adopted by teams to address each of the basic functionalities involved in the tests.

It is important to remark that none of the extremes in the abstraction layer is addressed. On the high-level of abstraction side, all solutions differ, reflecting individual approaches of each team. On the other side of the spectrum, low-level solutions are in most cases either unreported or often vendor-specific, without the influence or means to alter them by the teams.

4.1 Frameworks and Middlewares

Frameworks and middlewares can operate at many levels of abstraction. This section refers exclusively to solutions for the intercommunication of the multiple modules that operate a service robot. In addition, there is no further reference to the platforms imposed by vendors such as NaoQui.

Nowadays, ROS has become a tacit standard in robotics with all teams declaring its use in their TDPs for Nagoya 2017 and Montreal 2018 [24]. Nonetheless, older frameworks like Orocos [40] are still in use. Moreover, some teams still make good use of their own solutions [35].

That being said, it is important to remark that, although it has been discussed in the past, the TC of the RoboCup@Home league has discarded the idea of making ROS a compulsory standard.

4.2 Audio, Speech, and Natural Language Processing

The most broadly adopted solution to deal with speech consists in a pipeline. In this pipeline, a filtered audio signal feeds an Automatic Speech Recognition (ASR) engine to get a text-transcript for further processing. Then, the transcript is sent to a natural language processor that extracts and conceals relevant information. Finally, the acquired information is consumed by a high-level task planner that triggers the pertinent behaviors. Although processing raw audio signals is technically possible [6, 33], this is still unexplored in RoboCup@Home.

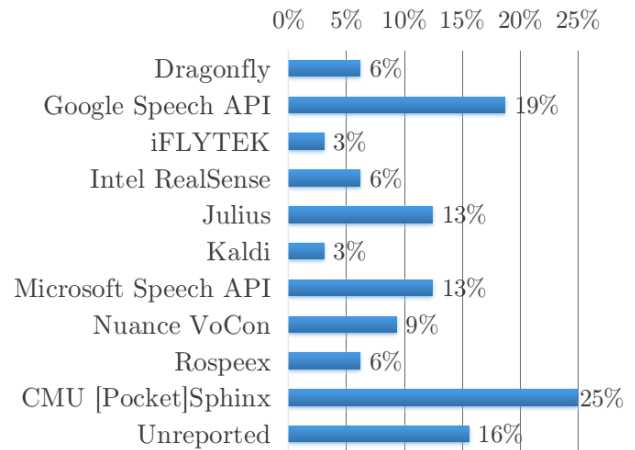


Figure 6: Trends in Automatic Speech Recognition (2017)

Typically, no external filters are used, leaving filtering to the microphone and the ASR engine [7]. When reported, the most recurrent solution is HARK³ [5, 13, 28], although it is intended for sound-source localization and separation.

Regarding ASR, the most commonly adopted off-line solutions include Julius [16], the Microsoft Speech API [35], and CMU Sphinx [39, 41] being the most popular solution as Figure 6 shows. Due to the limited computing power of their robots, most teams in the SPL used cloud services, where Google speech API is the most popular approach [28, 43]. However, network connectivity is often unreliable, a reason for which all teams using cloud services also have offline solutions as backup.

Moving forward to Natural Language Processing (NLP), despite the remarkable advances achieved in this area, little has been exploited in RoboCup@Home. To the date, we have found many teams still rely in keyword spotting and pattern matching to trigger the execution of a state machine [10, 36, 39], specially in simple tests. In fact, NLP and Natural Language Understanding (NLU) are mentioned in less than 50% of the TDPs. Nonetheless, robust A.I. solutions have always been in play and are now gaining strength. Among the approaches for processing language, we found probabilistic semantic parsers [12], Multimodal Residual Deep Neural Network (DNN) [20], ontology-based parsers over inference engines [35], and probabilistic parsers for syntax-tree extraction along with lambda calculus for semantic parsing [18]. Of special mention are the Stanford Parser [37], the most broadly adopted solution for POS-tagging and syntactic tree extraction, and LU4R [2], a Spoken Language Understanding Chain for HRI developed in

³HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) is an open-source robot audition software that includes modules for ASR and sound-source localization and sound separation. Source: <https://www.hark.jp/>

La Sapienza [27] by participants of RoboCup@Home which is also being used by several teams.

4.3 Manipulation

How teams address manipulation heavily depends on the platform. In the Social Standard Platform League, due to limitations of the robot, manipulation is barely addressed and often skipped. In direct opposition, the robot chosen for the Domestic Standard Platform League features a very precise and versatile manipulator that even incorporates a suction tip for lifting lightweight objects. Last, but not least, manipulators in the OPL range from home-made to professional ones.

Implementation solutions for manipulation are often based on direct-inverse kinematic models with a closed-loop control and camera feedback as an alternative to the ROS manipulation stack. However, nowadays many teams are migrating to *MoveIt!*

Out of 32 participant teams in Nagoya 2017, manipulation was reported in only 18 (56%) of the TDPs. From these, 61% are using *MoveIt!*, while the rest relies on custom solutions that include super-ellipsoid fitting [4], multiple deep auto-encoders fed with raw images and audio from the sensors [29], and the built-in software solution provided [27].

Other than picking and placing, robots are required to open doors, pour liquids, scrub spots from a table, handle a tray, and move tableware.

In the DSPL, handling a tray is particularly problematic since the robot has only one manipulator, while opening doors doesn't seem to be a problem for the robot, and the suction tip comes handy for moving tableware. In contrast, tray handling is bypassed in the OPL by mounting a custom tray in the robot, while the rest are usually skipped. Unfortunately, no Team Description Paper (TDP) addresses the solution of any of these particular challenges.

Finally, it seems to be a growing trend in the use of deep-learning-based methods for manipulation, especially in planning. In contrast with the traditional methods long-time used in industry, deep-learning-based approaches are computationally much more expensive, but also can be remarkably faster when a good-enough, non-optimal solution is acceptable. In addition, their supporters claim they require much less expertise and effort to code. Nonetheless, it is too early to know if these new methods will be a better solution for service robots.

4.4 Navigation

The basic functionality referred to as navigation in RoboCup@Home involves four research areas, namely 1) path planning, 2) obstacle avoidance, 3) localization, and 4) mapping. Navigation is fundamental in the competition. It is

Table 1: Trends in solutions for navigation. The first column, Report Ratio (R.R.) represents the percentage of TDP in which the solution used was specified. The second column shows the most adopted solution and its use percentage with respect to the number of reports.

	R.R.	Mode (cases)
Path planning	63%	A* (40%)
Obstacle avoidance	56%	ROS (40%) Occupancy grid (28%)
Localization	63%	AMCL (60%)
Mapping	78%	GMapping (72%)

assumed that all competing robots can safely navigate inside the arena. Therefore, path planning, and localization in known environments (i.e. inside the arena) are considered to be solved, while obstacle avoidance is not. Finally, on-line mapping becomes relevant only when robots are outside the arena, and nowadays is being extensively tested.

With a couple of exceptions, it can be said that all teams rely on the ROS navigation stack (see Table 1). However, it must be clarified that, in virtually all cases, this solution is adapted to the robot and the particular necessities of each team. In this regard, the most broadly adopted solution sums up OpenSlam's Gmapping and Adaptive Monte Carlo Localization (AMCL) with an A* path planner [22, 40, 41]. Also a recurrent solution involves incorporating a Kalman filter to Gmapping [8, 10]. Table 1 summarizes these trends showing the reported ratio, and the most adopted solution with the percentage of reported use cases.

Path Planning and Obstacle Avoidance

Other than A*, several teams just report the use of ROS. In these cases, as well as when nothing is said, we assume teams use ROS' default behavior. Other employed solutions include randomized path planners [28], and wave-propagation algorithms based on the Fast Marching Method [17]. Regarding obstacle avoidance, the most reported strategy is the use of occupancy grids. To build the occupancy map, several teams take information from both, the laser range finder and the RGB-D camera.

Localization and Mapping

These two abilities are often reported together. Regarding localization alone, there are two reported solution differing from the robot's built-in localization and AMCL. These solutions are addressed by different teams and implement Simultaneous Localization and Mapping (SLAM) using the Iterative Closest Point (ICP) algorithm [19, 34]. Both methods aim for accuracy and speed with limited resources. On

the mapping side, reported solutions other than GMapping include hector SLAM [16], Mobile Robot Programming ToolKit (MRPT) and ICP [26], and Omnimap[4].

4.5 Object Detection and Recognition

Object detection and object recognition are closely related, although they are often used separately. Object detection is much faster than object recognition, so a common strategy while looking for objects involves performing a continuous detection until a potential area of interest is found. A recurrent approach consists in taking the point-cloud of the RGB-D sensor and remove background, floor, and other surfaces (e.g. using Vectorial Quantization [35] or Random Sample Consensus (RANSAC) [17]). Later on, multiple color-depth images are extracted from the original RGB-D cloud using the detected clusters that can be further analyzed by the object recognizer. This approach is also a computationally inexpensive alternative to deep-learning-based approaches that consume raw data from the sensors.

Most teams rely in more than one software solution to recognize objects, either running them in parallel and using some consensus algorithm, or implementing a process pipeline. In addition, depth information is used for recognition based in contour and shape [4, 20, 35]. However, there is no clear tendency on how the related modules are coupled. For instance, some proposed solutions include You Only Look Once (YOLO) + Scale-Invariant Feature Transform (SIFT) + Binary Robust Independent Elementary Features (BRIF) [28], Speeded Up Robust Features (SURF) + Continuous Hough-space voting + Implicit Shape Models (ISM) [25], color, size and shape histograms + SIFT [35], Point Cloud Library (PCL) + RANSAC + YOLO [17], and contours using LINEMOD + HSV color histograms + SURF considering joint models for occlusion [20].

Detections and recognitions can take place either continuously or on demand. For simplicity, and considering the default implementations of most used packages, it is more likely that both processes are running all the time. However, unless the A.I. is designed to keep updating the world-model or knowledge-base of the robot with every new stimulus, acquired information will be simply discarded by the task planner. In this regard, very few teams have reported the mechanisms to take advantage of continuous detection and recognition.

It is important to point out, however, that object recognition in RoboCup@Home is not only about correctly labeling instances of previously trained objects. Recognizing an object shape and orientation is, in many cases, fundamental for grasping, let alone stacking and storing. Equally important are color, size, and relative position since one can always refer to an object by description. Therefore, robots must be able to identify features of untrained objects and

Table 2: Adopted software solutions for object recognition. Presented results are based on the 81% of reported strategies.

Solution	Reported uses
YOLO	23%
SIFT	19%
SURF	15%
OpenCV Caffe	12%
Tensorflow	12%

Table 3: Adopted software solutions for face detection and recognition. Presented results are based on the 67% of available TDPs.

Solution	Reported uses
Openface	23%
Haar-based algorithms	14%
Viola-Jones algorithm	14%
Caffe	9%
Microsoft Face	9%
OpenCV	9%

correctly categorize them by likeliness. Furthermore, since 2015 robots can be requested to describe objects they have never seen before. Unfortunately, the strategies used to address the aforementioned challenges are not documented in the TDPs.

4.6 People Detection & Recognition

People detection, recognition, and tracking are closely related but, at the same time, their approaches differ broadly. In RoboCup@Home, people detection and recognition means localizing a relatively static target, while tracking people involves a moving target (see Section 4.7). This section relates only to detection and recognition of people using exclusively visual information.

Regarding people detection, combining face and skeleton detection is the most popular approach since it reduces false positives and considers only people within range. In addition, hybrid techniques like combining 3D object recognition with face detection (e.g. OpenFace), or analysis of thermal images [15, 21, 40] are also being used. Focusing exclusively in face detection, all reported strategies are also recognition-capable, from which the more popular are Openface [40], Viola-Jones algorithm [32], and Haar-based algorithms [5].

On the other hand, all tests requiring to recognize a specific person either provide a description (see below), or provide the means to the robot to memorize the person's features and name.

Typically, getting to know a person for later recognition involves a facial recognition. The solutions used by teams to achieve this are as diverse as the teams themselves. Yet, some worth mentioning include Histogram of Oriented Gradients (HOG) descriptors with Support Vector Machine (SVM) classifiers [3], strands perception people [41], Viola-Jones and eigenfaces [32], Siamese Convolutional Neural Network (CNN)s [18], and Haar Cascades with either EigenFaces [35] or DNN [5].

Nonetheless, sometimes texture and color segmentation are also used as backup information. It is, nevertheless, important to mention that cloud services are gaining popularity because of their robustness and ease of implementation. However, they are also often unreliable due to connectivity problems so many teams prefer their own offline solutions.

When a description is provided, it may include height, gender⁴, age⁵, pose, relative position, and clothing. To overcome these challenges, most teams rely on DNN-based libraries and cloud services [14, 25]. Finally, it should be pointed out that some tasks may consist in requesting the robot to provide an accurate description of a person at the specified location.

4.7 People Tracking

People tracking is directly linked to two abilities: following and guiding people. This sets another fundamental difference with object recognition: the robot can't see the face of the people being tracked. Nonetheless, people tracking is always preceded by detection. The robot has to memorize the person to be followed or guided, and indicate when that person can start moving.

For Nagoya 2017, many teams relied on the same approaches that were implemented back in 2006, like leg detection using the robot's Laser Range Finder Scanner (LRFS), and color segmentation with a probabilistic tracker [11, 24]. Today, with better sensors and more powerful computers available, these methods have been improved and combined with other techniques. Examples of these approaches include leg-like clusterization with SVM and Unscented Kalman Filter (UKF) [28], 2D/3D leg detector fusing ellipse-fitting and 3D-windows obtained using a DNN [4], and LRFS-based leg-tracker with an upper-body detector [8].

Other approaches take advantage of the skeleton detection offered by RGB-D cameras [34]. However, this latter approach often produces unsatisfactory results due to vibrating sensors and people not fitting in the sensor's FoV. On the other hand, interesting approaches involve contour matching [40], head tracking using a monocular camera

and a Single Shot MultiBox Detector CNN [9], and the Track-Learning Detection (TLD) algorithm that learns from the difference between the estimated position of the target and the position detected in the analyzed frame [10, 19]. Unfortunately, people tracking is one of the least reported functionalities in TDPs, with a frequency of 34%.

Finally, at this point in time, there is no important difference between following and guiding. For guiding, teams resort on the same techniques used for following, but using information from sensors looking backwards.

5 Extrinsic Challenges

The challenges described in this section are considered to be extrinsic to the league. Although they operate from outside, the influence they exert is big enough to be taken into account. The discussed challenges are: 1) how to gain the general public interest and trust (Section 5.1), and 2) how to manage available resources and safety constraints (Section 5.2).

5.1 Gaining People's Interest and Trust

Acquiring and maintaining the sympathy and interest of the general public is critic for research in service robotics. Without people's trust, developing service robots would be pointless since there wouldn't be a market for them. Furthermore, the active participation of people is fundamental in several related areas.

Engaging People

It seems impossible to advance in natural HRI and NLU without eventually taking advantage of people's innate skills. For instance, latest advances in ASR and Machine Translation (MT) were possible thanks to the huge amount of data for training that became available with the extended use of smart-phones. However, the data acquired might be useless in HRI because it is very unlikely that people interact with robots in the same way they interact with their smart-phones. Although there may be similarities, the difference in interaction becomes critical when it comes to data sets for deep learning. In addition, while in a smart-phone a faulty transcription or an inaccurate translation are slight inconveniences, a robot that fails to correctly understand the given instructions becomes useless. Moreover, having a good understanding of the operator's intentions is crucial in tasks that heavily depend on HRI. Therefore, it is clear that instructing a robot must require significantly less time and effort than carry out the task.

The above suggests the need to build a corpus of interactions and, in particular, of spoken commands as a first step.

⁴To avoid gambling, incorrect labeling is penalized, for which acceptable solutions are male, female and unknown.

⁵For simplicity, age estimation in RoboCup@Home is fuzzy and maps to only three categories, child, adult, and elder.

Steps towards that direction have already been made [1] and the produced corpus is being used in the ERL, although not yet in RoboCup@Home⁶. The creation of corpora of that nature is not a trivial task. First, people with experience in robotics, ASR, or NLP cannot be involved in the process. We have found that the expertise of people with a strong background in any of these fields acts as a biasing factor during the interaction process. The expertise of experts makes them prone to alter pitch, speed, pauses, etc., when addressing robots. What is worse, is that the biasing occurs at a subconscious level, making it extremely difficult for a person to prevent such behavior. Therefore, non-experts shall be involved, and preferably volunteers with no experience in robotics and related fields. The only requirement, at least in the beginning, is that the operator should have a friendly attitude towards the robot to ease the interaction.

From these insights, the first logical step would be to attract the audience's attention. However, this is not as straightforward as it may seem. Unlike soccer, domestic chores are not exciting from most people's perspective. In addition, soccer playing robots benefit from the familiarity of the audience who knows what is happening. That's something we haven't achieved yet in RoboCup@Home. Most domestic chores are much more complex than any sport from a robotic's perspective due to the broadness of involved application domains of a domestic environment. In consequence, robots perform their tasks slowly and fail more often, making the league even less attractive. Therefore, new strategies are needed to attract audience's attention to RoboCup@Home, getting them actively involved to produce the aforementioned corpora.

Regaining Trust

RoboCup@Home is directly affected by the *Frankenstein Complex*. While soccer and rescue robots are designed to operate in specific environments and circumstances, it is expected that service robots blend with people and use the same appliances. This will be possible only if people consider them safe, even when using potentially harmful tools such as hammers, scissors, knives, and any other utensil of common use.

Media and entertainment industry are playing an important role in this matter. For instance, most successful blockbusters and series related to robotics feature anthropomorphic machines that are hostile to humans. Similarly, press and news constantly remind people that robots might not only turn evil, but also remark the high number of jobs that would be in jeopardy once robots are ready to be deployed. Nonetheless, quite often informative services also make good press to the robots. Moreover, sometimes their

optimism harms what they are trying to exalt, raising expectations beyond what robots can actually do today. We have witnessed on countless occasions the disenchantment of people who came to the competition to watch the prodigious robot of the news failing to grasp a simple object.

In both cases, competitions like RoboCup@Home set the perfect opportunity to vindicate robots as the useful and innocuous servants they are. It is during competitions, when people can effectively learn that robotics is a still-growing research area and why their participation is important to foster robot's development. Furthermore, is in this kind of events where transdisciplinary networks can be established. For instance, getting involved lawyers, sociologists, and philosophers can lead to the aperture of forums and research lines in topics of interest such as licensing and patents, liability in accidents caused by intelligent robots, role of robots as "affective artifacts", mid- and long-term impact of the integration of robots in human societies, when a machine can be considered conscious, and rights and freedom of conscious robots. Besides presenting the potential aperture of new research line in their fields, such forums can also contribute to engaging general public with the RoboCup@Home community, preparing the way for the future. We believe, in addition, that it's on our best interest to start discussing such matters sooner than later.

Conclusion

The League's organizers are aware of these challenges and are working to overcome them. One proposed solution is to have an expert moderating tests and explaining what is going on and why it is difficult. In our opinion, another possible solution could be to show to the audience the goal and challenges of the running test to make it more appealing and understandable.

To conclude, more interdisciplinary research is needed to find solutions to integrate robots into human society. It is also necessary to find ways to promote people's collaboration with the league to foster research in NLU and HRI. Often, hard questions come out, such as why should one acquire a robot while human assistants are much more efficient, cheaper, and hiring them helps creating jobs. Therefore, it is important to find ways to vindicate robots as the useful and innocuous servants they are, and not present them as a threat to humanity.

5.2 Available Resources and Safety Constraints

In robotics exists a close relationship between the available resources and a reliable safe execution. Today, with the internet becoming available everywhere, it becomes hard to resist to connect robots to the cloud. However, such

⁶The generation and use of a corpus for HRI in RoboCup@Home has been proposed and is discussed in [23].

decisions must be addressed with care since, unlike with other technologies, a faulty execution can lead to a human being harmed.

Available Resources

As the difficulty of the tests increases, it can be expected that the computational power required to solve them grew proportionally, especially when introduced changes are small. It is true that it is impossible to tell what *proportionally* means, especially in computer sciences and particularly in robotics. In this regard, even the slightest changes, neglectable from a human perspective, can increase the complexity of a task in several orders of magnitude for the robot. Nonetheless, in recent years we have detected a disproportional increase in the amount of computation needed to solve barely modified tasks. Moreover, we have found that solving similar tasks requires more resources today than it required half a decade ago.

We believe this disproportion is partially due to the popularization of Artificial Neural Network (ANN) and deep-learning-based solutions. Although it is true that these new approaches are often more robust and normally easier to deploy, it is also true that they require a huge amount of computational power. Besides, the enormous leap they caused in some disciplines have not fully reached the competition yet. Therefore, we are expecting to witness some remarkable advances in the open demonstrations⁷ in coming years.

Notwithstanding, experimenting with these technologies requires a computational power that laptops can't offer. In addition, robots must become what smartphones are today: affordable devices providing reliable long-term operation with very limited resources. This latter premise makes inadvisable to have robots performing heavy calculations all the time. In conclusion, robot's computers might be unfit for most DNN. In consequence, roboticists look at cloud services.

Safety Constraints

Cloud services offer several advantages besides reducing the amount of process a robot has to perform and increase the duration of its battery. They also process information remarkably faster while taking advantage of the data provided by their users to enhance.

This, however, comes with some risks. First, robots might experience a performance decrease due to lag and bandwidth problems during Internet rush-hours. This might make them slow and inefficient in the best case but, in worst-case scenarios, they might endanger people's lives while waiting

for the server's answer. This becomes critical when performing nursing or baby-sitting tasks, in which a delay of a couple of seconds can cause severe harm. Moreover, during emergencies the situation worsen. Having all robots becoming inoperative due to a disconnection caused by a power shortage could be just an annoyance. However, if the power shortage occurs during an earthquake or a fire, the mere presence of robots could hinder the evacuation instead of provide assistance.

Finally, and also a matter of security, but in a different sense, robots permanently connected to the internet have the potential risk of opening a window through which a malicious user could see and hear. Nonetheless, implications on this matter won't be analyzed here any further.

Conclusions

We believe that the robot's primary function should be helping people to have a better life. This necessarily implies that robots must never compromise people's security and integrity. Furthermore, the presence of a robot should help people to feel safe. In consequence, minimum performance standards must be established to ensure a reliable and safe operation. RoboCup@Home has already taken the first steps by allowing cloud services and external computers on the condition that robots must remain operational and have full obstacle avoidance.

6 Intrinsic Challenges

The challenges described in this section are considered to be intrinsic to the league. Most of them largely correspond to the role of the league in directing the research and development in robotics. First, we start discussing each of the most relevant abilities: navigation (Section 6.1), people recognition (Section 6.2), people tracking (Section 6.3), object recognition (Section 6.4), manipulation (Section 6.5), and NLU (Section 6.6). Finally, we close by addressing the necessity of planning tasks to be tested (Section 6.7).

6.1 Navigation

In general terms, navigation can be split into two categories: indoor and outdoor navigation. Although closely related, the categories need to be approached with disjoint ability sets. Therefore, they are discussed separately. Likewise and as in Section 4.4, navigation also involves obstacle avoidance, localization, and mapping.

Indoor-Navigation

Despite the continuous improvements, nowadays robots won't make it in most homes. Steps, wet floors, rough

⁷The open demonstrations differ from regular tests in that they do not follow a script nor is there a predefined task to complete. Instead, teams can show whatever they want.

carpets, and in general uneven surfaces are challenges to overcome that have not been addressed yet, let alone houses with staircases. Another minor detail important to consumers is quietness. We have found that people prefers robots that move silently and without damaging the floor, specially if they are meant to clean the house during the night.

Other than stairs, big houses and facilities present new challenges, specially when it comes to localization. The perception range of most robots is typically limited to about 4 meters, a distance short enough to impede localization in wide spaces and long corridors. Furthermore, the map of the environment is often a given condition (e.g. during startup) that uses a remarkable amount of geometrical data. Without semantic localization, robots might have it hard to localize themselves in buildings such as apartment towers and hospitals. Thus, without a pre-existing map, both, mapping and localization, can take advantage of robot's ability to read. This might prove to be specially useful in the long-term should a room or person be temporarily relocated.

Other important abilities that have been addressed in the past but haven't been solved⁸ include using an elevator, navigate in narrow corridors, and move furniture around. The latter is of special importance when cleaning and tidying up rooms. A robot should be able to move unattended objects that are blocking its path.

Last but not least, is functional touching. People often rely on their body to move or stop objects when moving around. For instance, it is common to see a person, carrying a lot of thing without a free hand, pushing a door with their hips, or holding a door with their foot. In RoboCup@Home, we call this kind of interaction *functional touching* to differentiate it from collisions and intended manipulations. Although allowed in the rulebook, functional touching hasn't been addressed yet by any team.

Outdoor-Navigation

Either to take out the garbage, or when going to buy groceries, eventually robots will be required to go outside. Urban environments come with a whole new set of challenges to overcome other than dealing with the elements. The first one affects directly the robot's ability to move. If in interiors the presence of unevenness in floors presents challenges, the terrain diversity outdoors seems more adverse. Conducted polls reveal that there is a potential market for robots in pet owners, specially when it comes to exercise the animals. Implementing such feature would require robots to be capable of moving in rough terrain like grass, sand, or gravel.

⁸Here it is important to remark that the aforementioned abilities haven't been solved in RoboCup@Home. That does not necessarily imply that they are not being extensively tested by research groups. For example, Hart et al. claim in [12] that their robot can localize itself after using an elevator.

However, not only the unevenness increases outside. Distances, and the number of stimuli to process, increases by a huge amount. To localize themselves in open spaces, robots would require not only the ability to see objects at 20 meters and beyond, but it would be also necessary to deal with occlusions, and correlate the perceived information in real time. As with autonomous cars, localization using point-clouds would be unreliable but, at the same time, the size difference and low movement speed would make it extremely hard to rely exclusively on GPS data. Therefore, we foresee many situations in which a robot will not only need to recognize signs, but also read street and shop names in order to reach its destination. In addition, in crowded streets a robot could encounter hundreds of pedestrians, including children, elders, dogs, and other robots, which it would have to evade.

Furthermore, speed is another important factor to take into account. While indoors, a robot can take its time to solve a task, outside, the world imposes its own schedule. Streetlights, public transport, and automatic doors are three examples of elements that require a quick response. Therefore, the reaction time of robots must be improved, even in conditions of insufficient information.

Conclusions

Although RoboCup@Home currently features wheeled-robots only, it is evident that this design is not suitable for most human environments. Therefore, we will have to choose between making our environments robot-friendly, or provide robots with better means of movement. Otherwise, robots will be restricted to those areas in which they can safely operate.

On the other hand, people can immediately recognize semantic information in their surroundings, having an innate ability to correlate the current environment with those previously known. Eventually, robots will need to integrate semantic mapping (or an equivalent ability) capable of mimicking such human abilities in order to efficiently integrate in human environments.

6.2 People Detection & Recognition

Detecting and recognizing people is fundamental for Human-Robot Interaction (HRI), therefore, it should be extensively tested in RoboCup@Home. Moreover, science-fiction has raised people's expectations, depicting robots as entities with an infallible, endless memory. In consequence, most users expect robots capable of remember or find anyone.

Advances and Current State

Although people detection is being extensively tested since it is involved in all types of HRI, people recognition is not.

Robots need to find a person to retrieve a command, deliver an object, answer a question, and detect a gesture, to name some examples. However, most of the time the detection occurs with the person standing. As of 2017, finding people lying or sitting was still problematic for most competitors.

In contrast, recognizing people hasn't been fully addressed. For our analysis, this feature is split in *people recalling*, when the robot needs to recall a person's name or order, and *people identification*, when the robot needs to identify a person's specific features.

Recalling, as stated in Section 4.6, usually consists in pairing the person's face with a name and, in some cases, some additional information like an order. This ability has been tested since the very first competitions, but always with a reduced number of people (less than five). In addition, although there shouldn't be a problem keeping the information longer, the face-name pairing lasts no longer than the test.

On the other hand, people identification is relatively new. Robots can be requested both, to describe a group of people, or find a person matching the given description (relevant features are described in Section 4.6). However, by 2017 the tests addressing this ability either randomized the feature selection, or left its selection to the robot. From our perspective, while most robots succeeded in finding their target, provided descriptions weren't very accurate, leading to the assumption that apart from relying in identification, heuristics were used to maximize the robot's chance of success. In consequence, it is impossible to know precise information regarding general performance.

Nonetheless, it must be acknowledged that most robots perform good at counting people grouped for a photo.

Next Steps

Considering the important number of features introduced in recent years, we believe the competition should focus in extending their usage. Moreover, features like estimated age, gender, and relative position can be used to test awareness, decision making and planning. Notwithstanding, focusing exclusively on detection and recognition, like with objects, the detection range needs to be increased, as well as the number of people in the memory of the robot.

On the features to be introduced in later stages and of importance in HRI we consider: emotions and moods, activities, health and vital signs (inebriation, fatigue, sickness, sleep, etc.), skin and hair color, clothing names and styles, and identification by voice. Also important, and shared with object recognition is addressing occlusions and translucent surfaces.

Conclusion

Much has been achieved in the past eleven years. However, HRI hasn't been tested as extensively as intended. Furthermore, several tests require to be redesigned to ensure benchmarking data is available for analysis.

Finally, we find that visual and auditory information are equally important when addressing people. Therefore, the integration of audio features in people detection and recognition has to be carefully planned.

6.3 People Tracking

People tracking is one of the first functionalities tested in RoboCup@Home. Although not necessarily useful when carrying out chores, it becomes relevant in daily life social events (e.g. parties) and during shopping. Its importance grows in non-domestic environments such as museums, hospitals, and restaurants.

Advances and Current State

People following has been greatly improved since 2006. Every year robots require less training time and are less prone to lose their target. In addition, they can robustly handle people crossing or standing in between. Moreover, back in 2014 robots were able to follow a person to an elevator, get inside, and continue following after leaving the elevator. And not only that, should the target go through a compact crowd blocking the sight, robots were able to go around it and meet their target. Whether robots are still capable of doing it is unclear, since elevators and crowds are not part of the tests since 2015.

Guiding, on the other hand, was first tested in 2006 with little success, leading to its removal by 2008. Since its reintroduction in 2015, the league has achieved good performance. Furthermore, this ability is tested along with navigation. After reaching a previously unknown location, the robot has to go back, find someone, and guide them to the location it just learned.

Next Steps

To humans, it is unnecessary to study someone in order to follow them. Moreover, it takes us only a quick glance to get all information required to track or even chase an object, let alone people. Therefore, one of the first improvements in this basic functionality would be removing training time, so it's important to start reducing it.

Another important aspect is speed, which must start to be gradually increased. Nowadays, robots follow and guide only *professional walkers*, who walk at constant pace to keep tests fair. In addition, professional walkers are robot-friendly, instructed to follow the instructions of the robot and even

slow down when it has lost track. However, people usually walk a lot faster than professional walkers, and are more likely to leave the robot behind if it can not follow their pace. Also, when following and guiding, people have no inconvenience in sliding through crowds or narrow spaces, temporary losing contact to each other. In consequence, it is important that robots can keep track of people walking fast in crowded environments and be able to predict or estimate local rendezvous points after optimizing trajectories.

Also of relevance in people tracking is side-by-side walking. When two people walk side by side, it is hard to tell who is guiding the march. It is one of the most common activities performed by humans. Thus, it is expected that robots learn how to walk with humans and not in front or behind them.

In direct relationship, walking holding hands is a natural way of interaction. Moreover, it is of special importance when interacting with children. Notwithstanding, physical contact is strictly forbidden in RoboCup@Home for security reasons. Therefore, this also needs to be addressed, but with caution.

Conclusion

Although unrelated to most domestic chores, people following and guiding are necessary abilities for many other applications. Yet, both abilities can be analyzed as a hybrid problem, namely people tracking and navigation. Moreover, tracking can be made extensive to animals and any kind of visual objects. Thus, tracking gets into the domain of object recognition, for which it becomes important to find the pertinent features that make it possible for robots to track something. Nonetheless, other problems like reducing training time and increasing speed need to be addressed with a carefully designed strategy.

6.4 Object Detection and Recognition

Mankind has transformed its environment giving priority to the sense of sight. Therefore, in order to effectively blend in human society, robots must be able to decode the visual stimuli of the world we have built around us. Furthermore, and from a totally utilitarian point of view, perceive objects is necessary for manipulation, a key feature to perform any chore.

Advances and Current State

Despite the remarkable advances in computer vision in the recent years, anyone who has witnessed recent competitions could safely argue that robots are shortsighted. On the one hand, is very unlikely that a robot can detect an object more than 4 meters away. On the other hand, robots continuously

fail to identify objects lying in direct line of sight. Both situations would deem a person as visually impaired and unable to perform in most environments without the proper aids. The former is mostly due to the short range and limited resolution of used RGB-D sensors. Nonetheless, this allows robots to do much more than years ago when sensors were limited to VGA cameras and Time-of-Flight sensors with a resolution of only a few thousands of pixels.

Limitations aside, successful detections are also used to update the grid of an occupancy map. This is an important step further from when visual information wasn't used for obstacle avoidance. In the case of recognition, features from several sensors are extracted and stored. Later on, the features of each detected candidate are analyzed (e.g. as in color and shape detection) or evaluated against the features of all trained objects, depending on the task requirements.

However, objects in the arena are always placed sparsely, and with partial occlusions at most. Also, the set of objects the robot should know is capped at 25, a very small quantity considering the number of objects people use to deal with at home. Furthermore, although robots have to identify objects *of a kind* (e.g. an apple, regardless the color, size and texture of the fruit), such recognition is still on very early stages.

Next Steps

Virtually all reported strategies process the input stream frame-by-frame, computing in a feed-forward manner. In other words, to track an object, a robot has to identify it and calculate its position in each frame with the possible assistance of a filter. Furthermore, the extracted visual information once used is discarded (i.e. not used for reinforcement). In contrast, advances in neurosciences suggest that the human brain uses contextual information to build scene representations, and it can choose to rely on memory search instead of visual search [30, 31]. Having robots using context to perform object detection and recognition, is an interesting area to explore. For instance, a robot could analyze a scene from a different perspective, calculating the positions of those objects that have been already identified and removing all their clusters from the frame, focusing only in those with a low recognition confidence.

Following the line of recognition as instance-class matching, there are more problems to address, like increasing the number of objects to hundreds and beyond, perform recognition of stacked objects, discriminate two identical objects based on its relative position, recognize translucent and transparent objects, and recognize objects in odd lighting conditions like direct sunlight and in the dark.

In a broader perspective, detection and recognition can be taken further on. Identifying the orientation of an object prior to grasping, as well as the best location for placement, or even infer the weight of an objects are requirements yet

to be addressed. Moreover, quite often robots will need to deal with occlusions and objects behind crystals, identify dirt, dust and spots on the floor, and objects that change over time (e.g. food).

Conclusions

Object recognition applied to robotics is in a very early stage. Similar to other abilities, a roadmap is necessary to direct advances on object recognition. However, recognizing things, more than characterizing features, is an A.I.-complete problem. The first logical steps would be those that help with manipulation, like identifying placing surfaces, grasping orientations, and containers. Other useful features that can be addressed first include occlusions and transparent objects. Finally, it is necessary to start integrating object recognition with high-level action planning and memory management so abstract constructs like spatial and temporal relationships can be also recognized.

6.5 Manipulation

Manipulation is perhaps the oldest and most mature research area in robotics, especially when it comes to industrial robots. With this background, it would be reasonable to expect a similar degree of precision and speed from service robots. However, mounting a manipulator in a mobile base and limiting available power and computational resources drastically increases the complexity of even the simplest tasks.

People's Expectations and Requirements

In our polls people are explicitly asking for robots able to clean the toilet, wipe windows, do the dishes (by hand), wash, iron, and fold clothes, open flasks and jars, brush and wash the dog, and take out the garbage to name some examples.

Also of importance, but not even considered by potential customers are abilities to open doors, move furniture, operate switches, and to operate the control panel of all electric and electronic appliances.

From these mandatory skills, opening doors has been addressed since 2006 as an optional challenge, but to the date has not been made compulsory in any test. Nonetheless, door opening was impressively solved by team eR@sers in 2016 with the proprietary robot now used in the Domestic Standard Platform League. Notwithstanding, it seems there is still a long way to go before this skill can be considered solved in OPL and SSPL.

Advances and Current State

Unfortunately, the advances in manipulation seem to have stalled [24]. Although challenges like pouring, stacking,

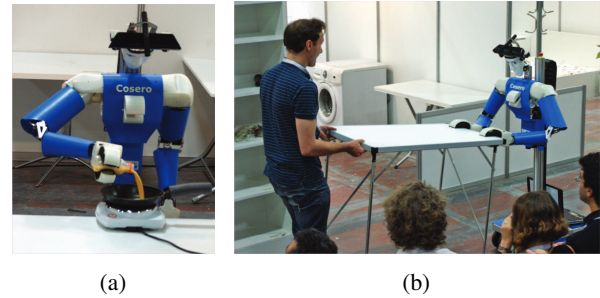


Figure 7: Team Nimbro's final demo in 2011 [38]. Robot Cosero (a) pouring pancake mixture into a pan, and (b) helping to move a table.

transporting a tray, and grasping small objects have been introduced, there have been no serious attempt to solve any of these during regular tests since 2014. What is worse, in 2017 during the storing groceries test, practically no team attempted to move any object. Furthermore, none of the aforementioned challenges are new in RoboCup@Home. In fact, most of them were already demonstrated several times by the three-times-champion team Nimbro between 2010 and 2014 (see Figure 7), and repeated later on by other teams during final demonstrations.

In addition and compliant with Section 6.4, the competition considers mostly moving regular-shaped, small, non-fragile lightweight objects. Apples, small cereal boxes, and water bottles are some examples of typical objects a robot can be requested to move. In numbers, most of the objects' weight ranges between 75 and 950 grams. Dimensions are between 5 and 25 centimeters, with a typical maximum volume of 2 liters.

Also in Stage I, all robots in the DSPL and the OPL are required to open the door of a cupboard. This optional task was skipped by all OPL participants in 2017. More complex tests in Stage II involve pouring, scrub spots from a table, handling a tray, and moving tableware (bowls, dishes, and cutlery). Once more, these tasks are often skipped or bypassed, as stated in Section 4.3.

Next Steps

The manipulation capabilities of the robots must be expanded in several directions, including reach, maneuverability, strength, and precision. Below we describe each of this directions in more detail.

Reach: As of today, robots are required to handle objects in heights ranging from 30cm to 1.8m, but eventually they should be able to reach the floor. More important is depth, since today robots are only required to grasp objects at most 5cm from the table border. Therefore, this distance should

be gradually increased until robots can reach objects within an average human grasping distance.

Maneuverability: Nowadays robot's movement is quite clumsy. Picking objects from a bag or box and stacking them all together in a narrow space is a trivial task for most people, but difficult even for industrial robots. At the same time, twisting, uncapping, shaking, folding, levering, and turning are features yet to be tested in RoboCup@Home.

Strength: In order to assist most people, a robot should be capable of carrying a 20 liters water bottle, hold a big dog by the leash, and uncap a flask of mayonnaise. Moreover, some applications require enough strength to gently lift a fallen elder. In consequence, weights and loads should start to gradually increase as soon as possible.

Precision: Today robots are very imprecise. To effectively operate in human environments, precision in movement and applied strength is required. Like a human, a robot should be able to store a needle and carefully take an egg without crushing it, but also being able to crack it in the right place. The same applies to acceleration and speed, either for applying an insulin injection or whipping cream.

Summarizing, some tasks that must be addressed soon within the reach of currently used manipulators include manipulating a switch, taking out the bag from the trash bin, using a ladle to serve soup, mopping the floor, dialing a phone number, picking groceries out of a box, grasping a towel from a hanger, and arranging cutlery into the drawer⁹.

Conclusion

To foster development in manipulation, several tests were modified for 2018. Nonetheless, there is still a long way to go before robots can effectively perform in domestic environments. Obviously, the RoboCup@Home league would enormously benefit from a manipulation road map with which teams could plan hardware improvements with sufficient time. For now, it seems sufficient to start by gradually incrementing the weight of most objects, locate them farther away from the corner, and request the robot to perform useful tasks using tools.

6.6 Automatic Speech Recognition and Natural Language Processing

Speech is a key aspect to people's communication, hence it is fundamental in HRI. Besides, it is widely acknowledged that language proficiency is related to intelligence. Therefore, it is expected that robots can understand people's orders expressed naturally, in contrast to the current state where

people learn how to talk to a robot. Regarding spoken interaction, we have identified three key aspects that require attention and are introduced below.

Noise

Noise is one of the most problematic aspects in spoken communication and, in fact, also in competitions like RoboCup@Home. The ambient noise produced by several hundred people greatly exceeds the noise levels of an average apartment. However, so far filtering is delegated to the ASR engine. Furthermore, eventually robots will be operating in airports, shopping malls, and other noisy environments. Thus, we think it is best to deal with this issue in an early stage.

Non-Native Biased Operators

Operators in RoboCup@Home have two fundamental characteristics: they are robot-friendly and they all have a strong background in robotics. This means that, unconsciously, operators in RoboCup@Home are trying to help the robot to succeed. In other words, they are biased. The use of biased operators helps robots to perform better and to give a better impression to the audience. However, it has the disadvantage of voiding the purpose of the competition of providing real-case scenarios for testing. Furthermore, few operators are native English speakers, which reduces the chances of a robot to deal with the richness of an unconstrained speech production.

Generators

Despite being a powerful tool during the competitions, the use of generators greatly limits the interactions a robot can experience. In drastical terms, reading verbatim sentences produced by command generators delves HRI into some sort of computer-computer interaction.

Next Steps

A solution to the aforementioned problems has already been proposed. The next steps consider the use of unbiased people to instruct the robot. Generators would still be used, but only to select the task, having the operator to command the robot using their own words. Besides, to address fairness, the interactions should be recorded beforehand, leaving robots to *listen* the recording, but always having the operator at hand for further interactions. This way, research would be pushed towards free-speech ASR engines, while at the same time, would enforce the integration of the latest techniques in NLP. But not only that, this process would lead to the creation of big annotated corpora of spoken HRI for training.

⁹Several of these tasks have already been proposed for TC discussion for the next rulebook.

Nonetheless, carefully planned restrictions shall be imposed to the interactions in the first year. For instance, while in the first years only imperative sentences are used, later on these can be replaced by declarative sentences that depict the desired goal. Further on, procedures are explained to the robot, reducing the amount of detail with every passing year until the interaction reaches the level of a fluent dialog.

Last, but not least, noise can be introduced gradually by superposing recordings of noisy environments to the ones containing the commands. This way, all robots would experiment the same initial conditions for the interaction, easing the benchmarking of components with higher levels of abstraction.

6.7 Roadmaps and milestones

We believe the RoboCup@Home league has the responsibility of carefully planning what needs to be tested and when to introduce changes by establishing milestones for the competition. Said in other works, the league should have a set of roadmaps and milestones to assess the TC when adapting and designing tests.

This is not an unfounded belief. In fact, this need arises from the teams and the members of the Technical Committee themselves. To the former, knowing in advance the challenges that will come in future competitions allows to prepare and direct their research. Moreover, some teams have even stated that they could have made a better use of their resources, should they have known in advance the direction of the league.

With regard to the latter, the TC renews every year with candidates elected by the community. In consequence, sometimes it happens that most people are replaced and the direction taken by former members is overridden. This situation worsens when most of the new members are relatively new in the league (e.g. less than three years), thus lacking of important undocumented empirical experience.

This lack of direction has led in several occasions to brainstorming in which valuable ideas appear. These ideas require to be analyzed, evaluated against the robot's capabilities, and condensed in milestones and test drafts that can be retaken in latter years. Unfortunately, this has not occurred until now.

In response, we have condensed not only ideas taken from the teams, but also the people's needs and, considering the current capabilities and limitations of the robots, present them as future steps or milestones in former sections. Nonetheless, this is just a first step, since roadmaps need to be still prepared and adopted by the league which, in the end, is peer-maintained.

7 Conclusions and Future Work

In this paper we conduct a thorough summary of the software solutions and strategies used by participating teams in RoboCup@Home to address the most important abilities required in the tasks of the competition's tests. Further on, we present an overview of the attained achievements since the league's foundation based on our experience as long-time participants, contributors, and referees in the league. Finally, also organized per ability and along with the overview, we discuss these achievements while addressing what is expected by potential consumers, what needs to be done, and would be the next logical steps based on the robot's current capabilities.

This study results in two contributions. First, we believe the presented summary can serve as quick reference guide for new competitors, or for experienced ones looking for alternatives to their current solutions. Second, our work sets the basis to build road maps that can help the RoboCup@Home league, as well as other competitions aiming at service robotics, to plan the direction of the competition towards its goal. Moreover, we believe that road maps are very important and that planning features to be tested can help teams to prepare in advance, set mid and long-term goals, and have a smarter resource management. In fact, based on this and previous work, we have designed a roadmap for spoken Human-Robot Interaction in RoboCup@Home which will be presented to the league in Montreal 2018 and in the RoboCup Symposium.

However, there is still work to be done. Not only the roadmaps have to be designed. This work has also allowed us to identify several important flaws that need to be addressed. For instance, the presence of certain rules might be undermining the development of certain features. At the same time, we have found that many successful approaches and strategies are never reported in the TDPs. This has two important setbacks. First, it makes it much harder for the scientific community to compare the performance of the different approaches when the best performers are missing. Second, it leads to an eventual loss of knowledge that worsens as the lifetime of a good team shortens.

These insights are left to the competition organizers to analyze as part of future work, for which we trust this manuscript can come handy.

References

- [1] E. Bastianelli, L. Iocchi, D. Nardi, G. Castellucci, D. Croce, and R. Basili. Robocup@home spoken corpus: using robotic competitions for gathering datasets. In *Robot Soccer World Cup*, pages 19–30. Springer, 2014.

- [2] E. Bastianelli, D. Croce, A. Vanzo, R. Basili, and D. Nardi. A discriminative approach to grounded spoken language understanding in interactive robotics. In *IJCAI*, pages 2747–2753, 2016.
- [3] G. Cheng, K. Ramirez-Amaro, I. Dianov, P. Lanillos, R. Guadarrama, E. Dean, E. Lozinska, G. Diez-Valencia, P. Grzywok, Q. Guo, G. Simonic, and X. Wang. Alle@home 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [4] H. I. Christensen, L. D. Riek, R. White, P. Parashar, S. Wang, T. Iqbal, A. Taylor, and D. Chan. Uc san diego 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [5] K. Demura, K. Demura, K. Nagashima, K. Enomoto, T. Yamakawa, R. Iwasaki, and S. Mashimo. Happy mini 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [6] P. F. Dominey. Learning grammatical constructions from narrated video events for human–robot interaction. In *Proceedings IEEE humanoid robotics conference, Karlsruhe, Germany*, 2003.
- [7] M. Doostdar, S. Schiffer, and G. Lakemeyer. A robust speech recognition system for service-robotics applications. In L. Iocchi, H. Matsubara, A. Weitzenfeld, and C. Zhou, editors, *RoboCup 2008: Robot Soccer World Cup XII*, pages 1–12, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-02921-9.
- [8] J. C. et al. Walking machine @home 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [9] T. Gerlach, T. Oschwald, M. Rättsch, F. Ostertag, T. Weber, P. Huber, J. Hoffarth, R. Hülsmann, M. Katzenmaier, P. Kopp, S. Braun, M. Foo, J. Engling, M. Litz, S. Roericht, T. Huschitt, C. Gedrat, R. Keinert, and A. Gopal. Rt lions team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [10] J. Guo, H. Yao, H. Ma, Y. Dong, Y. Zhu, J. Peng, X. Wang, and X. Ma. Tinker@home 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [11] W. Hans, S. Gräser, F. Jarmer, S. Schmitt, J. Bornemeier, P. Lambrecht, and D. Gossow. Robocup 2008 - homer@unikoblenz (germany), 2008.
- [12] J. W. Hart, P. Stone, A. Thomaz, and S. Niekum. Ut austin villa robocup@home domestic standard platform league team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [13] S. Hori, Y. Ishida, Y. Kiyama, Y. Tanaka, Y. Kuroda, M. Hisano, Y. Imamura, T. Himaki, Y. Yoshimoto, Y. Aratani, K. Hashimoto, G. Iwamoto, T. Morie, and H. Tamukoh. Hibikino-musashi@home 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [14] S. Hori, Y. Ishida, Y. Kiyama, Y. Tanaka, Y. Kuroda, M. Hisano, Y. Imamura, T. Himaki, Y. Yoshimoto, Y. Aratani, K. Hashimoto, G. Iwamoto, T. Morie, and H. Tamukoh. Hibikino-musashi@home spl 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [15] L. Iocchi, D. Holz, J. Ruiz-del Solar, K. Sugiura, and T. Van Der Zant. Robocup@home: Analysis and results of evolving competitions for domestic and service robots. *Artificial Intelligence*, 229:258–281, 2015.
- [16] S. Isobe, Y. Katsumata, Y. Tabuchi, A. Kinose, T. Ishimine, T. Fukui, H. Kobayashi, A. Taniguchi, A. Aly, Y. Hagiwara, and T. Taniguchi. Duckers 2017 @home dspl team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [17] T. Kelestemur, A. A. Allaban, and T. Padir. Frasier: Fostering resilient aging with self-efficacy and independence enabling robot team northeastern’s approach for robocup@home. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [18] B.-J. Lee, J.-Y. Choi, C.-Y. Lee, K.-W. Park, S. Choi, C. Baek, and B.-T. Zhang. 2017 aupair team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [19] T.-H. S. Li, C.-Y. Liu, Y.-T. Su, C.-H. Li, C.-W. Cheng, C.-Y. Hsieh, J.-J. Liang, C.-Y. Chen, H.-Y. Lin, and K.-C. Chang. airobots_ncku 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [20] J. Liu, Z. Zhang, B. Tang, and X. Chen. Wrighteagle@home 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [21] L. Martinez, M. Pavez, G. Olave, M. Correa, L. Sánchez, P. Loncomilla, and J. Ruiz-del Solar. Uchile homebreakers 2015 team description paper. 2015.
- [22] L. Martínez, R. Muñoz, G. Olave, G. Pais, G. Hernan, D. Gomez, L. Garrido, D. Campanini, P. Orellana, P. Loncomilla, and J. Ruiz-del Solar. Uchile homebreakers 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.

- [23] M. Matamoros, K. Harbusch, and D. Paulus. From commands to goal-based dialogs: A roadmap to achieve natural language interaction in robocup@home. 2018.
- [24] M. Matamoros, V. Seib, R. Memmesheimer, and D. Paulus. Robocup@home: Summarizing achievements in over eleven years of competition. In *2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 186–191, April 2018. doi: 10.1109/ICARSC.2018.8374181.
- [25] R. Memmesheimer, N. Y. Wettengel, F. Daniel Müller, Polster, M. Roosen, L. Buchhold, M. Moritz Löhne, Schnorr, I. Mykhalchyshyna, and D. Paulus. Robocup 2017 - homer@unikoblenz (germany). *RoboCup @Home 2017 Team Description Papers*, 2017.
- [26] K. Miyawaki, M. Sano, Y. Inoue, Y. Hiroi, S. Nishiguchi, and M. Suzuki. O.i.t-trial 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [27] L. M.T. Lázaro, Iocchi, D. Nardi, M. Hanheide, and J. P. Fentanes. Spqrel 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [28] S. Oishi, J. Miura, K. Koide, M. Demura, Y. Kohari, S. Une, L. Villamar-Gomez, T. Kato, M. Kojima, and K. Morohashi. Aisl-tut @home league 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [29] H. Okada, H. Yokoyama, T. Ogata, T. Inamura, N. Iwahashi, and K. Sugiura. Team er@sers[dspl] (toyota hsr) 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [30] A. Oliva and A. Torralba. The role of context in object recognition. *Trends in cognitive sciences*, 11(12):520–527, 2007.
- [31] A. Oliva, J. M. Wolfe, and H. C. Arsenio. Panoramic search: the interaction of memory and vision in search through a familiar scene. *Journal of Experimental Psychology: Human Perception and Performance*, 30(6):1132, 2004.
- [32] L. A. Pineda, C. Rascón, G. Fuentes, A. Rodríguez, H. Ortega, M. Reyes, N. Hernández, R. Cruz, I. Vélez, and M. Ramírez. The golem team, robocup@home 2017. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [33] D. K. Roy and A. P. Pentland. Learning words from sights and sounds: A computational model. *Cognitive science*, 26(1):113–146, 2002.
- [34] C. Sammut, M. Pagnucco, G. Castro, C. Flannagan, M. Gratton, B. Hengst, D. Rajaratnam, C. Schwering, M. Thielscher, M. Velonaki, and T. Wiley. Unsw robocup@home spl team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [35] J. Savage, M. Negrete, J. Cruz, J. Marquez, R. Martell, J. Cruz, E. Vazquez, M. Pano, J. Cruz, E. Silva, H. Estrada, H. Arce, M. Matamoros, A. Garzon, and O. Fuentes. Pumas@home 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [36] V. Seib, S. Manthe, R. Memmesheimer, F. Polster, and D. Paulus. Team homer@unikoblenz—approaches and contributions to the robocup@home competition. In *Robot Soccer World Cup*, pages 83–94. Springer, 2015.
- [37] Stanford. Corenlp, 2011. URL <http://nlp.stanford.edu:8080/corenlp/>.
- [38] J. Stuckler, D. Holz, and S. Behnke. Robocup@home: Demonstrating everyday manipulation skills in robocup @home. *IEEE Robotics & Automation Magazine*, 19(2):34–42, 2012.
- [39] J. T. C. Tan, F. Duan, Z. H. b. Ismail, and Y. Uchimura. Kamerider opl @home 2017 team description paper. *RoboCup @Home 2017 Team Description Papers*, 2017.
- [40] M. van der Burgh, J. Lunenburg, R. Appeldoorn, R. Wijndands, T. Clephas, M. Baeten, L. van Beek, R. Ottervanger, H. van Rooy, and M. van de Molengraft. Tech united eindhoven @home 2017 team description paper. *University of Technology Eindhoven*, 2017.
- [41] S. Wachsmuth, F. Lier, S. Meyer zu Borgsen, J. Kummert, L. Lach, and D. Sixt. Tobi-team of bielefeld: The human-robot interaction system for robocup @home 2017. 2017.
- [42] T. Wisspeintner, T. Van Der Zant, L. Iocchi, and S. Schiffer. Robocup@home: Scientific competition and benchmarking for domestic service robots. *Interaction Studies*, 10(3):392–426, 2009.
- [43] H. Yaguchi, B. Tran, M. Takeda, K. Kochigami, Z. Li, K. Sasabuchi, Y. Furuta, K. Nagahama, K. Okada, and M. Inaba. Jsk@home: Team description paper for robocup@home 2017. *RoboCup @Home 2017 Team Description Papers*, 2017.