RoboCup 2012, Robot Soccer World Cup XVI, Springer, LNCS

NimbRo@Home: Winning Team of the RoboCup@Home Competition 2012

Jörg Stückler, Ishrat Badami, David Droeschel, Kathrin Gräve,
Dirk Holz, Manus McElhone, Matthias Nieuwenhuisen, Michael Schreiber,
Max Schwarz, and Sven Behnke

Rheinische Friedrich-Wilhelms-Universität Bonn
Computer Science Institute VI: Autonomous Intelligent Systems
Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
{ stueckler | droeschel | graeve | holz | nieuwenhuisen | schreiber} @ ais.uni-bonn.de
{ badami | mcelhone | schwarz | behnke } @ cs.uni-bonn.de
http://www.NimbRo.net/@Home

Abstract. In this paper we describe details of our winning team NimbRo@Home at the RoboCup@Home competition 2012. This year we improved the gripper design of our robots and further advanced mobile manipulation capabilities such as object perception and manipulation planning. For human-robot interaction, we propose to complement face-to-face communication between user and robot with a remote user interface for handheld PCs. We report on the use of our approaches and the performance of our robots at RoboCup 2012.

1 Introduction

The RoboCup@Home league [16, 17] was established in 2006 to foster the development and benchmarking of dexterous and versatile service robots that can operate safely in everyday scenarios. The robots have to show a wide variety of skills including object recognition and grasping, safe indoor navigation, and human-robot interaction. At RoboCup 2012, which took place in Mexico City, 21 international teams competed in the @Home league.

With our team NimbRo@Home we compete in the RoboCup@Home league since 2009. We improved the performance of our robots in the competitions, from third place in 2009 to second place in 2010 to winning in 2011 and 2012.

So far, we focused on hardware design and a system that balances indoor navigation, mobile manipulation, and human-robot interaction. In this year, we further advanced object recognition, modelling, and pose tracking capabilities. We also integrated motion planning for manipulation in complex scenes into the system. Last but not least, we developed a novel remote user interface on handheld computers that allows the user to control the autonomous capabilities of the robots on three levels.

In the following, we will give a short overview on the ruleset of the RoboCup@Home competition 2012. We then detail our system with a focus on the novel components, compared to 2011. Finally, we will report on the performance of our robots at the 2012 competition.

2 Design of the RoboCup@Home Competition 2012

2.1 Overview

The competition consists of regular tests, i.e., tests with a predefined procedure, open demonstrations, and a technical challenge [5]. In two preliminary stages, the five best teams are selected for the final that is conducted as an open demonstration.

Regular tests cover basic mobile manipulation and human-robot interaction skills that all robots shall be able to demonstrate. The storylines of the regular tests are embedded in application scenarios. In these tests, the robots must act autonomously and fulfill the tasks within a limited amount of time. In the open demonstrations, the teams can choose their own task for the robot in order to demonstrate results of their own research. Finally, the technical challenge has been introduced to test a specific technical aspect in a benchmark. In this year, the robots had to demonstrate object recognition in cluttered scenes.

While the rules and the tests are announced several months prior to the competition, the details of the competition environment are not known to the participants in advance. During the first two days of the competition, the teams can map the competition arena, which resembles an apartment, and train object recognition on a set of 25 objects which are used as known objects with names throughout the recognition and manipulation tests. The arena is subject to minor and major changes during the competition and also contains previously unknown objects.

Performance is evaluated according to objective measures in the regular tests. Juries assess the quality of the open demonstrations based on score sheets. In the final, the jury consists of members of the league's executive committee and external jury members from science, industry, and media.

2.2 Tests and Skills

In Stage I, the teams compete in the tests *Robot Inspection and Poster Session*, *Follow Me*, *Clean Up*, *Who Is Who*, and the *Open Challenge*. During the *Robot Inspection and Poster Session*, the robots have to navigate to a registration desk, introduce themselves, and get inspected by the league's technical committee, while the team gives a poster presentation. In the *Follow Me* test, the robots must keep track of a previously unknown guide in an unknown (and crowded) environment. This year, the robots had to keep track of the guide despite a person blocking the line-of-sight. Then, they had to follow the guide into an elevator and demonstrate that they can find the guide after he/she went behind a crowd. *Clean Up* tests object recognition and grasping capabilities of the robots. They have to retrieve as many objects as possible within the time limit, recognize their identity, and bring them to their designated locations. The *Who Is Who* test is set in a butler scenario, where the robot first has to learn the identity of three persons. Then it has to take an order of drinks for each person, to grasp the correct drinks among others, and to deliver them to the correct person. The

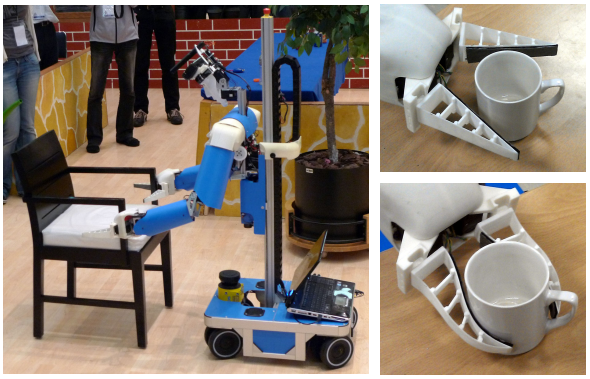

Fig. 1. The cognitive service robot *Cosero*. Left: Cosero moves a chair during the RoboCup@Home Final 2012 in Mexico City. Right: Cosero's grippers feature Festo FinRay fingers that adapt to the shape of objects.

Open Challenge is the open demonstration of Stage I. Teams can freely choose their demonstration in a 5 min slot.

Stage II consists of the *General Purpose Service Robot* test, the *Restaurant* test and the *Demo Challenge*. In the *General Purpose Service Robot* test, the robots must understand and act according to complex, incomplete or erroneous speech commands which are given by an unknown speaker. The commands can be composed from actions, objects, and locations of the regular Stage I tests. In the *Restaurant* test, the robots are deployed in a previously unknown real restaurant, where a guide makes them familiar with drink, food, and table locations. Afterwards, the guide gives an order to deliver three objects to specific locations. Finally, the *Demo Challenge* follows the theme "health care" and is the open demonstration of Stage II.

3 Hardware Design

We designed our service robots Cosero and Dynamaid [13] to cover a wide range of tasks in human indoor environments (see Fig. 1). They have been equipped with two anthropomorphic arms that provide human-like reach. Two torso joints extend the workspace of the arms: One joint turns the upper body around the vertical axis. A torso lift moves the whole upper body linearly up and down, allowing the robot to grasp objects from a wide range of heights—even from the floor. Its anthropomorphic upper body is mounted on a mobile base with narrow footprint and omnidirectional driving capabilities. By this, the robot can maneuver through narrow passages that are typically found in indoor environments, and it is not limited in its mobile manipulation capabilities by holonomic constraints.

In 2012, we improved Cosero's gripper design. We actuate two Festo FinGripper fingers using RX-64 Dynamixel actuators on two rotary joints (see Fig. 1).

When the gripper is closed on an object, the bionic fin ray structure of the fingers adapts its shape to the object surface. By this, the contact surface between fingers and object increases significantly, compared to a rigid mechanical structure. A thin layer of anti-skidding material on the fingers establishes a robust grip on objects.

For perceiving its environment, we equipped the robot with diverse sensors. Multiple 2D laser scanners on the ground, on top of the mobile base, and in the torso measure objects, persons, or obstacles for navigation purposes. The lasers in the torso can be rolled and pitched for 3D obstacle avoidance. We use a Microsoft Kinect RGB-D camera in the head to perceive tabletop objects and persons.

The human-like appearance of our robots also supports intuitive interaction of human users with the robot. For example, the robot appears to look at interaction partners while it tracks them with its head-mounted RGB-D camera. With its human-like upper body, it can perform a variety of gestures.

4 Mobile Manipulation

Some regular tests in the RoboCup competition involve object handling. Currently, objects are placed separated on horizontal surfaces such as tables and shelf layers. The robot needs to drive to object locations, to perceive the objects, and to grasp them.

We further advanced our mobile manipulation and perception pipelines. We developed means for object grasping in complex scenarios such as bin picking, and to track the pose of arbitrary objects in RGB-D images, for example, for moving chairs.

4.1 Motion Control

We implemented omnidirectional driving controllers for the mobile base of our robots [10]. The driving velocity can be set to arbitrary combinations of linear and rotational velocities. We control the 7-DoF arms using differential inverse kinematics with redundancy resolution. The arms also support compliant control in task-space [11].

4.2 Indoor Navigation

During the tests, the setup of the competition arena can be assumed static. We acquire 2D occupancy grid maps of unknown environments using GMapping [4]. We then employ state-of-the-art methods for localization and path planning in grid maps [10]. For obstacle-free driving along planned paths, we support the incorporation of all distance sensors of our robots. Point measurements are maintained in an ego-centric 3D map and projected into a 2D occupancy grid map for efficient local path planning.

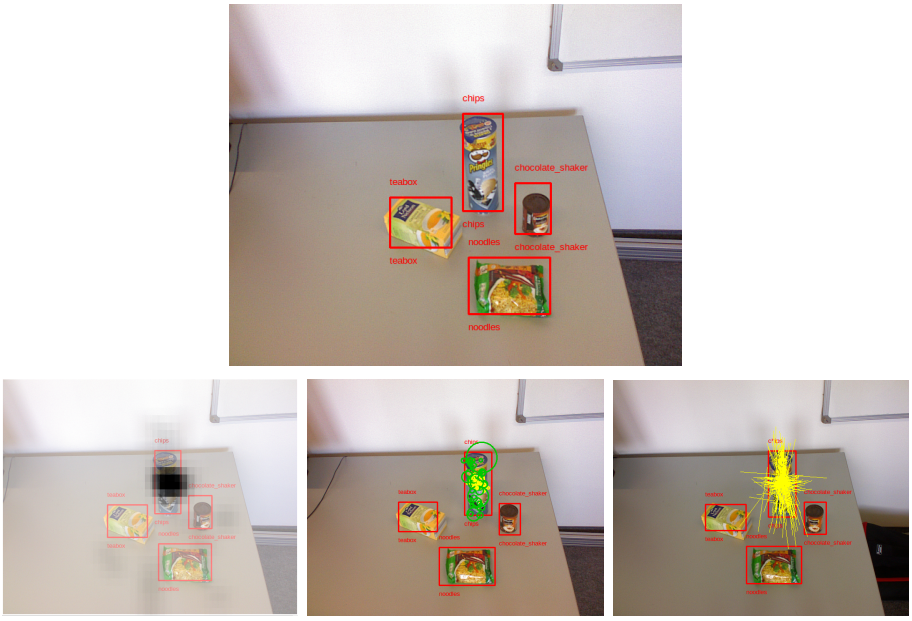

Fig. 2. Object recognition. Top: We recognize objects in RGB images and find location and size estimates. Bottom: Matched features vote for position in a 2D Hough space (left). From the features (middle, green dots) that consistently vote at a 2D location, we find a robust average of relative locations (middle, yellow dots) and principal directions (right, yellow lines).

4.3 Grasping Objects from Planar Surfaces

We developed efficient segmentation of RGB-D images to detect objects on planar surfaces [14]. On the raw measurements within the object segments, we plan top or side grasps on the objects. A collision-free grasp and reaching motion is then executed using parametrized motion primitives. Our method allows to grasp a large variety of typical household objects with cylindrical or box-like shapes. We implemented such highly efficient detection and motion planning to spend only little time for object manipulation during a test.

4.4 Object Recognition

Our robots recognize objects by matching SURF features [1] in RGB images to an object model database [10]. We improved our previous approach by enforcing consistency in the spatial relations between features (see Fig. 2).

In addition to the SURF feature descriptor, we store feature scale, feature orientation, relative location of the object center, and orientation and length of principal axes in the model. During recall, we efficiently match features between an image and the object database according to the descriptor using kd-trees.

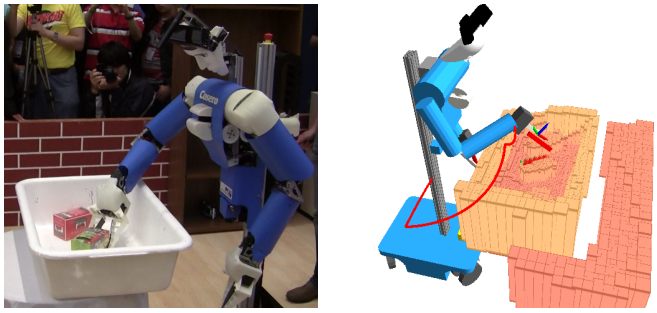

Fig. 3. Motion planning in a bin-picking scenario. We extend grasp planning on object segments with motion planning (reaching trajectory in red, pregrasp pose as larger coordinate frame) to grasp objects from a bin. For collision avoidance, we represent the scene in a multi-resolution height map. We decrease the resolution in the map with the distance to the object. This reduces planning time and models safety margins that increase with distance to the object.

Each matched feature then casts a vote to the relative location, orientation, and size of the object. We consider the relation between the feature scales and orientation of the features to achieve scale- and rotation-invariant voting.

With this object recognition method, our robots can recognize and localize objects in an RGB image as evaluated in this year's technical challenge. When unlabelled object detections are available through other modalities such as planar RGB-D segmentation (Sec. 4.3), we project the detections into the image and determine the identity of the object in these regions of interest.

4.5 Motion Planning in Complex Scenes

Our grasp planning module finds feasible, collision-free grasps at the object. The grasps are ranked according to a score which incorporates efficiency and stability criteria. The final step in our grasp and motion planning pipeline is now to identify the best-ranked grasp that is reachable from the current posture of the robot arm.

In complex scenes, we solve this by successively planning reaching motions for the found grasps ([9], see Fig. 3). We test the grasps in descending order of their score. For motion planning, we employ LBKPIECE [15].

To speed up the process of evaluating collision-free grasp postures and planning trajectories, we employ a multiresolution height map that extends our prior work on multiresolution path planning [2]. Our height map is represented by multiple grids that have different resolutions. Each grid has $M \times M$ cells containing the maximum height value observed in the covered area (Fig. 3). Recursively, grids with quarter the cell area of their parent are embedded into each other, until the minimal cell size is reached. With this approach, we can cover the same area as a uniform $N \times N$ grid of the minimal cell size with only $\log_2((N/M)+1)M^2$

Fig. 4. Object pose tracking. We train multi-view 3D models of objects using multi-resolution surfel maps. We estimate the pose of objects in RGB-D images through real-time registration towards the model. We apply object tracking, for instance, to track the model (upper right) of a watering can for approaching and grasping it.

cells. Planning in the vicinity of the object needs a more exact environment representation as planning farther away from it. This is accomplished by centering the collision map at the object. This approach also leads to implicitly larger safety margins with increasing distance to the object.

4.6 Object Modelling and Pose Tracking

Many object handling tasks assume object knowledge that cannot be deduced from a single view alone. If an object model is available, the robot can infer valid grasping points or use the model to detect objects and to keep track of them. For example, to implement the handling of a watering can or the moving of a chair with our robot, we teach-in grasping and motion strategies. These grasps and motions are specified in the local reference frame of an object model. To be able to reproduce the motions, the robot needs to perceive the pose of the object. While the robot moves, we register RGB-D images to the model at high frame rates to keep track of the object. This way, the robot does not require a precise motion model.

In our approach, we train a multi-resolution surfel map of the object ([12], see Fig. 4). The map is represented in an octree where each node stores a normal distribution of the volume it represents. In addition to shape information, we also model the color distribution in each node.

Our object modelling and tracking approach is based on an efficient registration method. We build maps from RGB-D images and register these representations with an efficient multi-resolution strategy. We associate each node in one map to its corresponding node in the other map using fast nearest-neighbor look-ups. We optimize the matching likelihood for the pose estimate iteratively to find the most likely pose.

We acquire object models from multiple views in a view-based SLAM approach. During SLAM, we generate a set of key frames that we register to each

other. We optimize pose estimates of the key frames to best fit the spatial relations that we obtain through registration. While the camera is moving, we register the current RGB-D image to the closest key frame. Each time the translational or angular distance is above a threshold, we include the current frame as a new key frame into the map. For SLAM graph optimization, we employ the g2o framework [6]. Finally, we merge all key frames based on their pose estimate in a multi-view map.

Once we have a model, we can register RGB-D camera images against it to retrieve the pose of the object. We initialize the pose of the tracker to a rough estimate using our planar segmentation approach.

5 Human-Robot Interaction

5.1 Intuitive Direct Human-Robot Interaction

Domestic service robots need intuitive user interfaces so that laymen can easily control the robots or understand their actions and intentions. Speech is the primary modality of humans for communicating complex statements in direct interaction. For speech synthesis and recognition, we use the commercial system from Loquendo [7]. Loquendo's text-to-speech system supports natural and colorful intonation, pitch and speed modulation, and special human sounds like laughing or coughing.

We also implemented pointing gesture synthesis as a non-verbal communication cue. Cosero performs gestures like pointing or waving. Pointing gestures are useful to direct a user's attention to locations and objects. The robots also interpret gestures such as waving or pointing [3].

5.2 Convenient Remote User Interfaces

We develop handheld user interfaces to complement natural face-to-face interaction modalities [8]. Since the handheld devices display the capabilities and perceptions of the robot, they improve common ground between the user and the robot (see Fig. 5). They also extend the usability of the robot, since users can take over direct control for skills or tasks that are not yet implemented with autonomous behavior. Finally, such a user interface enables remote interaction with the robot, which is especially useful for immobile persons.

The user interface supports remote control of the robot on three levels of autonomy. The user can directly control the drive and the gaze using joystick-like control UIs or touch gestures. The user interface also provides selection UIs for autonomous skills such as grasping objects or driving to locations. Finally, the user can configure high-level tasks such as fetch and delivery of specific objects.

The user interface is split into a main interactive view in its center and two configuration columns on the left and right side (see Fig. 5, top). In the left column, further scaled-down views are displayed that can be dragged into the main view. In this case, the dragged view switches positions with the current

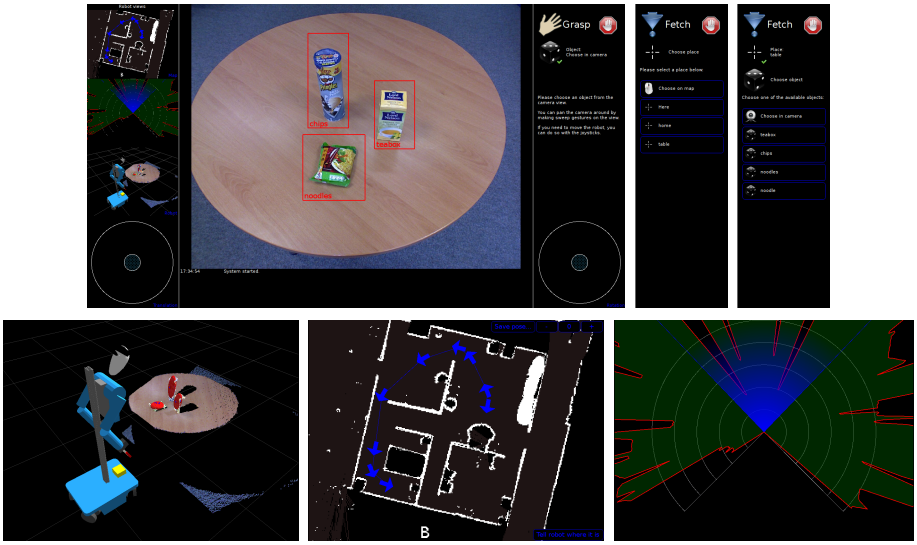

Fig. 5. Handheld User Interface. The user interface provides controls on three levels of autonomy. Top: Complete GUI with a view selection column on the left, a main view in the center, and a configuration column on the right. We placed two joystick control UIs on the lower and right corners for controlling motions of the robot with the thumbs. Lower right: 3D external view generated with Rviz. Lower middle: The navigation view displays the map, the estimated location, and the current path of the robot. Lower right: The sensor view displays laser scans and the field-of-view of the RGB-D camera in the robot's head.

main view. One view displays live RGB-D camera images with object perception overlays (Fig. 5, top). The user may change the gaze of the robot by sweep gestures, or select objects to grasp. A further view visualizes laser range scans and the field-of-view of the RGB-D camera (Fig. 5, bottom right). The navigation view shows the occupancy map of the environment and the pose of the robot (Fig. 5, bottom center). The user can set current pose and goal pose. While the robot navigates, the view shows the current path. Finally, we also render a 3D external view (Fig. 5, bottom left).

On the right (Fig. 5, top), high-level tasks such as fetch and delivery can be configured. For fetching an object, for instance, the user either selects a specific object from a list, or chooses a detected object in the current sensor view.

6 Competition Results at RoboCup 2012

With our robot system, we achieved scores among the top rankings in almost every test of the competition[1]. In Stage I, Cosero and Dynamaid registered for the competition in the *Robot Inspection and Poster Session*. In the new *Follow*

[1] A video can be found at http://www.NimbRo.net/@Home

Fig. 6. Left: Cosero follows a guide into an elevator during the *Follow Me* test. Middle: In the *Restaurant* test, a guide shows Cosero drink and food locations in a real and previously unknown restaurant. Right: Cosero waters a plant in the final.

Me test, Cosero learned the face of the guide and was not disturbed later by another person blocking the line-of-sight. It followed the guide into the elevator (see Fig. 6) and left it on another floor. Unfortunately, it falsely detected a crowd of people and could not finish the test. In *Who Is Who*, Cosero learned the faces of three persons, took an order, fetched three drinks in a tray and each of its arms, and successfully delivered two of them within the time limit. In the *Clean Up* test, our robot Cosero had to find objects that were distributed in the apartment, recognize them, and bring them to their place. Our robot detected three objects, from which two were correctly recognized as unknown objects. It grasped all three objects and deposited them in the trash bin. In the *Open Challenge*, we showed a "housekeeping" scenario. Cosero demonstrated that it could recognize a waving person. It took over an empty cup from this person and threw it into the trash bin. Afterwards, it approached a watering can and watered a plant. After finishing all tests of Stage I, our team lead the competition with 5,071 points, followed by WrightEagle (China) 3,398 points and ToBi (Germany) 2,627 points.

In the second stage, Cosero recognized speech commands from two out of three categories in the *General Purpose Service Robot* test. It recognized a complex speech command consisting of three actions. While it successfully performed the first part of the task, it failed to recognize the object in a shelf. It also understood a speech command with incomplete information and posed adequate questions to retrieve missing information. The third speech command was not covered by the grammar and, hence, could not be understood. Overall, Cosero achieved the most points in this test. In the *Demo Challenge* with the theme "health care", an immobile person used a handheld PC to teleoperate the robot. The person sent the robot to fetch a drink. The robot recognized that the requested drink was not available and the user selected another drink in the transmitted camera

image. After the robot delivered the drink, it recognized a pointing gesture and navigated to the referenced object in order to pick it up from the ground. In the *Restaurant* test, our robot Cosero was guided through a previously unknown bar (see Fig. 6). The guide showed the robots where the shelves with items and the individual tables were. Our robot built a map of this environment and took an order. Afterwards, it navigated to the food shelf to search for requested snacks. The dim lighting conditions in the restaurant, however, prevented Cosero from recognizing the objects. After both stages, we accumulated 6,938 points and entered the final with a clear advantage towards WrightEagle (China, 4,677 points) and eR@sers (Japan, 3,547 points).

In the final, our robot Cosero demonstrated the approaching, bi-manual grasping, and moving of a chair to a target pose. It also approached and grasped a watering can with both hands and watered a plant (see Fig. 6). After this demonstration, our robot Dynamaid fetched a drink and delivered it to the jury. In the meantime, Cosero approached a transport box, from which it grasped an object using grasp planning. This demonstration convinced the high-profile jury, which awarded the highest number of points in all categories (league-internal jury: scientific contribution, relevance, presentation and performance; external jury: originality, usability, difficulty and success). Together with the lead after Stage II, our team received 100 normalized points, followed by eR@sers (Japan, 74 points) and ToBi (Germany, 64 points).

7 Conclusion

In this paper, we presented the contributions of our winning team NimbRo to the RoboCup@Home competition 2012 in Mexico City. Since the 2011 competition, we improved object recognition, developed model learning and tracking, and implemented motion planning to further advance the mobile manipulation capabilities of our robots. We also developed a novel remote user interface on handhelds to complement natural face-to-face interaction through speech and gestures.

Our robots scored in all the tests of the competition and gained a clear advantage in the preliminary stages. In the final, our robots convinced the high profile jury and won the competition.

In future work, we will further develop robust object recognition in difficult lighting conditions. More fluent and flexible speech and non-verbal cues will improve the naturalness of human-robot interaction. Finally, we also plan to investigate tool-use and learning for object handling.

Acknowledgments

This project has been partially supported by the FP7 ICT-2007.2.2 project ECHORD (grant agreement 231143) experiment ActReMa.

References

1. H. Bay, T. Tuytelaars, and L. Van Gool. SURF: speeded up robust features. In *9th European Conference on Computer Vision*, 2006.
2. Sven Behnke. Local multiresolution path planning. *Robocup 2003: Robot Soccer World Cup VII, Springer LNCS*, pages 332–343, 2004.
3. D. Droeschel, J. Stückler, D. Holz, and S. Behnke. Towards joint attention for a domestic service robot – Person awareness and gesture recognition using time-of-flight cameras. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2011.
4. G. Grisetti, C. Stachniss, and W. Burgard. Improved techniques for grid mapping with Rao-Blackwellized particle filters. *IEEE Trans. on Robotics*, 23(1), 2007.
5. D. Holz, F. Mahmoudi, C. Rascon, S. Wachsmuth, K. Sugiura, L. Iocchi, J. R. del Solar, and T. van der Zant. RoboCup@Home: Rules & regulations. http://purl.org/holz/rulebook.pdf, 2012.
6. R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A general framework for graph optimization. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2011.
7. Loquendo S.p.A. Vocal technology and services. http://www.loquendo.com, 2007.
8. S. Muszynski, J. Stückler, and S. Behnke. Adjustable autonomy for mobile teleoperation of personal service robots. In *Proc. of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2012.
9. M. Nieuwenhuisen, J. Stückler, A. Berner, R. Klein, and S. Behnke. Shape-primitive based object recognition and grasping. In *Proc. of the 7th German Conference on Robotics (ROBOTIK)*, 2012.
10. J. Stückler and S. Behnke. Integrating indoor mobility, object manipulation, and intuitive interaction for domestic service tasks. In *Proc. of the IEEE Int. Conf. on Humanoid Robots (Humanoids)*, 2009.
11. J. Stückler and S. Behnke. Compliant task-space control with back-drivable servo actuators. *Robocup 2011: Robot Soccer World Cup XV, Springer LNCS*, pages 78–89, 2012.
12. J. Stückler and S. Behnke. Model learning and real-time tracking using multi-resolution surfel maps. In *Proc. of the AAAI Conference on Artificial Intelligence (AAAI-12)*, 2012.
13. J. Stückler, D. Droeschel, Kathrin Gräve, Dirk Holz, Jochen Kläß, M. Schreiber, R. Steffens, and S. Behnke. Towards robust mobility, flexible object manipulation, and intuitive multimodal interaction for domestic service robots. In *RoboCup 2011: Robot Soccer World Cup XV*, Lecture Notes in Computer Science. 2012.
14. J. Stückler, R. Steffens, D. Holz, and S. Behnke. Efficient 3D object perception and grasp planning for mobile manipulation in domestic environments. In *Robotics and Autonomous Systems*, 2012.
15. I. A. Sucan and L. E. Kavraki. Kinodynamic motion planning by interior-exterior cell exploration. In *Algorithmic Foundation of Robotics VIII (Workshop Proceedings)*, 2009.
16. Tijn van der Zant and Thomas Wisspeintner. RoboCup X: A proposal for a new league where RoboCup goes real world. In *RoboCup 2005: Robot Soccer World Cup IX*, LNCS 4020, pages 166–172. Springer, 2006.
17. Thomas Wisspeintner, Tijn van der Zant, Luca Iocchi, and Stefan Schiffer. RoboCup@Home: Scientific competition and benchmarking for domestic service robots. *Interaction Studies*, 10(3):393–428, 2009.