NimbRo@Home 2012 Team Description

Jörg Stückler, David Droeschel, Kathrin Gräve,
Dirk Holz, Michael Schreiber, 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 | schreiber | behnke} @ ais.uni-bonn.de
http://www.NimbRo.net/@Home

Abstract. This document describes the RoboCup@Home league team NimbRo of Rheinische Friedrich-Wilhelms-Universität Bonn, Germany, for the competition to be held in Mexico City in June 2012.
Our team uses self-constructed humanoid robots for mobile manipulation and intuitive multimodal communication with humans. The paper describes the mechanical and electrical design of our robots Cosero and Dynamaid. It also covers perception and behavior control.

1 Introduction

Since 2009, our team NimbRo competes with great success in the @Home league. In our first year at RoboCup 2009 in Graz we came in third and won the innovation award for "Innovative robot body design, empathic behaviors, and robot-robot cooperation". At RoboCup 2010 in Singapore, we could further advance to the second overall place. Last year at RoboCup 2011 in Istanbul, we demonstrated many novel approaches like fast and flexible grasp planning, real-time object tracking, and cooperative human-robot manipulation. We successfully participated in most of the tests in stages I and II, and reached the Finals with the highest score. Our final demonstration achieved the highest scoring by the jury, such that we won the RoboCup 2011 competition.

Our robots, Dynamaid and Cosero, have been designed to balance indoor navigation, mobile manipulation, and intuitive human-robot interaction. We equipped the robots with omnidirectional drives for robust navigation, two anthropomorphic arms for object manipulation, and with communication heads. In contrast to many other service robot systems, our robots are lightweight, inexpensive, and easy to interface.

In the next section, we detail the mechanical and electrical design of our domestic service robots. Sections 3 and 4 cover perception and behavior control, respectively.

2 Mechanical and Electrical Design

We equipped our robots Cosero and Dynamaid (see Fig. 1) with omnidirectional drives to maneuver in the narrow passages found in household environments.

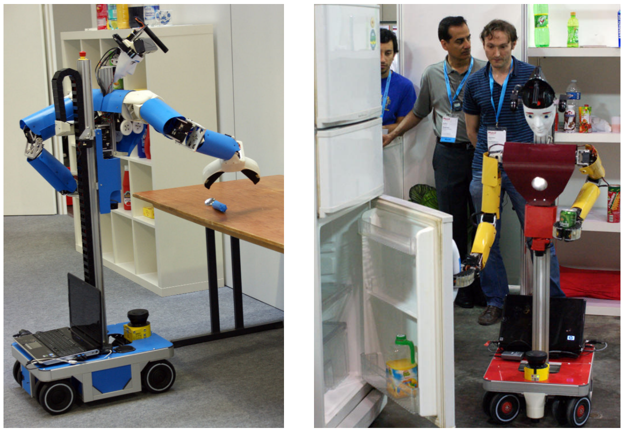

Fig. 1. Left: Cognitive service robot *Cosero* grasps a spoon. Right: *Dynamaid* manipulates the fridge.

Their two anthropomorphic arms resemble average human body proportions and reaching capabilities. A yaw joint in the torso enlarges the workspace of the arms. In order to compensate for the missing torso pitch joint and legs, a linear actuator in the trunk can move the upper body vertically by approx. 0.9 m. This allows the robots to manipulate on similar heights like humans.

The robots have been constructed from light-weight aluminum parts. All joints are driven by Robotis Dynamixel actuators. These design choices allow for a light-weight and inexpensive construction, compared to other domestic service robots. While each arm of Cosero has a maximum payload of 1.5 kg (Dynamaid: 1 kg) and Cosero's drive has a maximum speed of 0.6 m/sec (Dynamaid: 0.5 m/sec), Cosero's low weight of ca. 32 kg (Dynamaid: ca. 20 kg) requires only moderate actuator power. This makes the robots inherently safer than a heavy-weight industrial-grade robot.

Compared to its predecessor Dynamaid [10], we increased payload and precision of Cosero by stronger actuation. Cosero is mainly driven by Dynamixel EX-106+ (10.7 Nm holding torque, 154 g) and RX-64 (6.4 Nm holding torque, 116 g) actuators. The strongest joints in the robot are the shoulder pitch joints with a holding torque of 42.8 Nm. Each of these joints is actuated by two EX-106+ in parallel via a 2:1 transduction. We also improved safety and appearance of the robot with 3D-printed covering for joints and an energy chain in the torso.

The robots perceive their environment with a variety of complementary sensors. A SICK S300 laser scanner measures the distance to objects in a height of approx. 24 cm within 30 m maximum range and with a 270° field-of-view. It is primarily used for 2D mapping and localization. In order to detect small obstacles on the floor in front of the robots, a Hokuyo URG-04LX laser scanner is mounted between the front wheels. It scans in a height of 3 cm. The robots also sense the environment in 3D with a tilting Hokuyo UTM-30LX in their chest (max. range 30 m) and a Microsoft Kinect RGB-D camera in their head that is attached to the torso with a pan-tilt unit in the neck. A second URG-04LX laser scanner is attached through a roll joint to the torso. In horizontal alignment, its scan plane is adjusted to be 2 cm above the surface height when the robot manipulates on tables or in shelves. Its height above the ground can be adjusted from ca. 0.13 m to 1.03 m with the linear joint in the trunk.

We mounted the RGB-D camera on the head for several reasons: First, since the robots have a similar body height (1.6 m default height) like humans, faces can be viewed from the front. The fact, that we as humans design our environment to be easily perceivable with our own sensing capabilities, further supports to perceive the world from human eye height. The placement of the sensor on a pan-tilt neck enables the robot to point its sensors towards targets in a human-like way, i.e., humans can easily interpret the robot's gaze. We use all laser scanners and the depth camera for obstacle detection. For robust manipulation, the robots can measure the distance to obstacles directly from the grippers.

Finally, the sensor head also contains a shotgun microphone for speech recognition. By placing the microphone on the head, the robots point the microphone towards human users and at the same time direct their visual attention to her/him.

3 Perception

3.1 Continuous People Awareness

For human-robot interaction, a key prerequisite for a robot is awareness of the whereabouts of people in its surrounding. We combine complementary information from laser range finders (LRFs) and vision to continuously detect and keep track of people. In LRF scans, the measurable features of persons like the shape of legs are not very distinctive, such that parts of the environment may cause false detections. However, LRFs can be used to detect person candidates, to localize them, and to keep track of them at high rates. In camera images, we can verify that a track belongs to a person by detecting more distinctive human features like faces and upper bodies on the track.

Using the VeriLook SDK, we implemented a face enrollment and identification system. In the enrollment phase, our robots approach detected persons and ask them to look into the camera. The extracted face descriptors are stored in a repository. If the robot meets a person later, it compares the new descriptor to the stored ones, in order to determine the identity of the person.

3.2 Gesture Recognition

Gestures, like pointing or showing are a natural way of communication in human-robot interaction. A pointing gesture, for example, can be used to draw the robot's attention to a certain object in the environment. We implemented the recognition of pointing gestures, showing of objects, and stop gestures. The primary sensor in our system for perceiving a gesture is the RGB-D camera mounted on the robot's pan-tilt unit. We determine the position of the head, hand, shoulder, and elbow which allows us to interpret gestures. The perception is based on the detection of body parts in amplitude images as well as body segmentation in three-dimensional point clouds of the camera.

We interpret gestures for their parameters. For example, we seek to interpret the intended target of a pointing gesture. Especially for distant targets, the line through eyes and hand yields a good approximation to the line towards the target. We also applied Gaussian Process regression to learn a better interpretation of the pointing direction [2] using all body features. We train Hidden Markov Models to recognize and segment gestures into preparation, holding, and retraction phases.

3.3 Semantic Speech Understanding

We apply the commercial Loquendo [7] system for speech recognition and synthesis. Loquendo's speech recognition is grammar-based and speaker-independent. Its grammar definition allows rules to be tagged with semantic attributes. For instance, one can define keywords for actions or attributes like "unspecific" for location identifiers such as "room". When Loquendo recognizes a sentence that fits to the grammar, it provides the recognized set of rules together with a semantic parse tree. Our task execution module then interprets the resulting semantics and generates appropriate behavior.

3.4 Self-Localization and Mapping

For simple scenarios, we use state-of-the-art methods for localization and mapping in 2D representations of the environment. In order to acquire 2D maps of unknown environments, we apply GMapping [5], a FastSLAM2 approach to the Simultaneous Localization and Mapping (SLAM) problem. We use adaptive Monte Carlo Localization (MCL) to estimate the robot's pose in a given occupancy grid map.

Full 3D representations of the environment, however, provide several advantages over 2D maps. For instance, they allow for fully judging traversability for navigation. Since many changing elements of the environment like furniture and people do not reach fully to the ceiling, localization can be made more robust by focussing on the upper part of the static environment. We therefore investigated Monte Carlo localization in 3D surfel grid maps [6] which uses the tilting laser scanner in the robot's chest.

3.5 Perception of Objects

For object perception we develop approaches that combine depth sensing and vision (see Fig. 2). From Kinect depth images, we extract the surface on which the objects are located through efficient RANSAC methods [14]. We cluster the remaining measurements to obtain a segmentation into objects and keep track

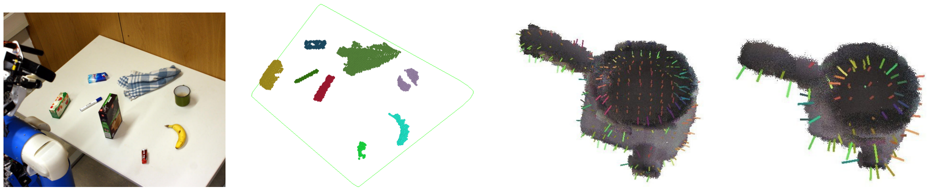

Fig. 2. From left to right: Table-top scene, segmentation into objects, object model at 5 cm resolution, and object model at 10 cm resolution.

of these detections. In order to identify the detected objects, we extract SURF features [1] and color histograms on the segments. For each object class, multiple descriptors are recorded from different view points during training, in order to achieve a view-independent object recognition.

During mobile manipulation, the pose of objects needs to be retrieved and tracked in real-time to robustly compensate for the motion of the robot. We thus developed model learning and real-time tracking of objects (see Fig. 2). We successfully applied our method for tracking a table during human-robot cooperative carrying of the table [12]. Core to our approach is the compact representation of RGB-D images in multi-resolution surfel maps [13]. We extract such maps from 640×480 VGA images in just a few milliseconds. We devised a robust registration method that allows for registering two VGA images in real-time at about 10 Hz [13].

4 Behavior Control

The autonomous behavior of our robots is generated in a modular control architecture. We employ the inter process communication infrastructure and tools of the Robot Operating System (ROS) [9].

We implement task execution, mobile manipulation, and motion control in hierarchical finite state machines. The task execution level is interweaved with human-robot interaction modalities. For example, we support the parsing of natural language to understand and execute complex commands.

Tasks that involve mobile manipulation trigger and parametrize sub-processes on a second layer of finite state machines. These processes configure the perception of objects and persons, and they execute motions of body parts of the robot. The motions themselves are controlled on the lowest layer of the hierarchy and can also adapt to sensory measurements.

4.1 Control of the Omnidirectional Drive

We developed a control algorithm for the mobile base that enables the robots to drive omnidirectionally. Their driving velocity can be set to arbitrary combinations of linear and rotational velocities.

4.2 Control of the Anthropomorphic Arms

The arms are controlled using differential inverse kinematics to follow trajectories of either the 6 DOF end-effector pose or the 3 DOF end-effector position. Redundancy is resolved using nullspace optimization of a cost function that favors convenient joint angles and penalizes angles close to the joint limits. We also developed compliant motion for the arm exploiting properties of the configurable position controllers in the Dynamixel actuators [11]. Compliance can be set for each direction in task or joint space separately. For example, the end-effector can be kept loose in both lateral directions while it keeps the other directions at their targets.

Cosero can perform a variety of parameterizable motions like grasping, placing objects, and pouring out containers. For example, the robot can perform pointing gestures towards a location given relative to the robot. We further investigate learning of motion primitives by imitation and reinforcement learning [4].

4.3 Robust Indoor Navigation

For navigation, we implemented path planning in occupancy grid maps and 3D obstacle avoidance using measurements from the LRFs in the robot and the depth camera [3]. To enlarge the narrow field-of-view of the depth camera, we implemented active gaze control strategies.

4.4 Mobile Manipulation

To robustly solve mobile manipulation tasks we integrate object detection, safe navigation, and motion primitives. Our robots can grasp objects on horizontal surfaces like tables and shelves. They can also carry the object, and hand it to human users. We also developed solutions to pour-out containers, to place objects on horizontal surfaces, to dispose objects in containers, to grasp objects from the floor, and to receive objects from users.

When handing an object over, the arms are compliant in upward direction so that the human can pull the object, the arm complies, and the object is released. Based on compliant control, we also developed mobile manipulation controllers to open and close doors, when the door leaf can be moved without the handling of an unlocking mechanism.

For mobile manipulation in complex settings, we also integrate grasp and motion planning with advanced object perception capabilities. In table-top settings, we derive grasps from the top and the side directly from the raw point clouds [14]. For grasping objects from piles or out of a box, we learn shape-primitive based object models and find collision free grasps and reaching-motions on the shape decomposition of the objects [8] (see Fig. 3).

4.5 Intuitive Human-Robot Interfaces

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, we use the commercial system from Loquendo.

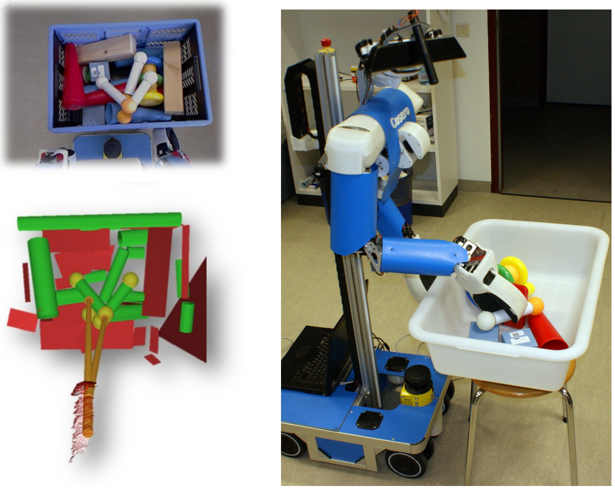

Fig. 3. Cosero grasps an object out of a box. We recognize objects based on shape primitives and derive suitable grasps on the shape decomposition.

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 for the robot. Cosero performs gestures like pointing or waving. Pointing gestures are useful to direct a user's attention to locations and objects.

5 Conclusion

The described system has been evaluated for three years now at RoboCup German Open and RoboCup competitions in 2009 to 2011. In all competitions, it performed very well. In 2009, we successfully participated in the tests *Introduce*, *Follow Me*, *Fetch&Carry*, *Who-Is-Who*, *Open Challenge*, *Walk&Talk*, *Supermarket*, *PartyBot*, and the *Demo Challenge*. With the new rules in 2010, we could participate with Dynamaid in all tests. She was the first robot to grasp an object from a shelf in a previously unknown shopping mall and to open and close the fridge at RoboCup. At RoboCup 2011, we could further improve the performance of Cosero and Dynamaid in the tests. We could show convincing open demonstrations and won first place in 2011.

We plan to equip Dynamaid and Cosero with an expressive communication head similar to Robotinho. We will continue to improve the system for RoboCup 2012 and to integrate new capabilities. The most recent information about our team (including videos) can be found on our web pages www.NimbRo.net/@Home.

Team Members

Currently, the NimbRo@Home team has the following members:[1]

- Team leader: Jörg Stückler, Prof. Sven Behnke
- Staff: David Droeschel, Kathrin Gräve, Dirk Holz, and Michael Schreiber
- Students: Ishrat Badami and Ricarda Steffens

References

1. H. Bay, T. Tuytelaars, and L. Van Gool. SURF: speeded up robust features. In *9th European Conference on Computer Vision*, 2006.
2. D. Droeschel, J. Stückler, and S. Behnke. Learning to interpret pointing gestures with a time-of-flight camera. In *In Proc. of the 6th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2011.
3. D. Droeschel, J. Stückler, D. Holz, and S. Behnke. Using time-of-flight cameras with active gaze control for 3D collision avoidance. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, 2010.
4. K. Gräve, J. Stückler, and S. Behnke. Improving imitated grasping motions through interactive expected deviation learning. In *Proc. of the International Conference on Humanoid Robots (Humanoids)*, 2010.
5. G. Grisetti, C. Stachniss, and W. Burgard. Improved techniques for grid mapping with Rao-Blackwellized particle filters. *IEEE Trans. on Robotics*, 23(1), 2007.
6. J. Kläß, J. Stückler, and S. Behnke. Efficient mobile robot navigation using 3d surfel grid maps. In *Proc. of the 7th German Conference on Robotics (ROBOTIK)*, 2012, to appear.
7. Loquendo S.p.A. Vocal technology and services. http://www.loquendo.com, 2007.
8. M. Nieuwenhuisen, A. Berner, J. Stückler, R. Klein, and S. Behnke. Shape-primitive based object recognition and grasping. In *Proc. of the 7th German Conference on Robotics (ROBOTIK)*, 2012, to appear.
9. M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng. ROS: an open-source Robot Operating System. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2009.
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. In *Proc. of the RoboCup International Symposium*, Istanbul, Turkey, 2011.
12. J. Stückler and S. Behnke. Following human guidance to cooperatively carry a large object. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Bled, Slovenia, 2011.
13. J. Stückler and S. Behnke. Robust real-time registration of RGB-D images using multi-resolution surfel maps. In *Proc. of the 7th German Conference on Robotics (ROBOTIK)*, 2012, to appear.
14. J. Stückler, R. Steffens, D. Holz, and S. Behnke. Real-time 3D perception and efficient grasp planning for everyday manipulation tasks. In *Proc. of the European Conf. on Mobile Robots (ECMR)*, Örebro, Sweden, 2011.

[1] Funding for the project is provided by Deutsche Forschungsgemeinschaft (DFG) under grant BE 2556/2 and Rheinische Friedrich-Wilhelms-Universität Bonn.