NimbRo@Home 2011 Team Description

Jörg Stückler, David Dröschel, Kathrin Gräve,
Dirk Holz, Michael Schreiber, and Sven Behnke

Rheinische Friedrich-Wilhelms-Universität Bonn
Computer Science Institute VI: Autonomous Intelligent Systems
Römerstr. 164, 53117 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 Istanbul in July 2011.
Our team uses self-constructed humanoid robots for object 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 the first year, we participated at RoboCup German Open 2009 and at RoboCup 2009 in Graz, where we came in second and third, respectively. We also received the innovation award for "Innovative robot body design, empathic behaviors, and robot-robot cooperation". In 2010, we continued our success and reached second places at RoboCup German Open and at RoboCup 2010 in Singapore.

In the project NimbRo – Learning Humanoid Robots – we investigate not only humanoid soccer, but also intuitive multimodal communication between humans and robots. Our test scenario for human-robot interaction is a museum tour guide. This application requires interacting with multiple unknown persons. In January 2010, our museum tour guide Robotinho has been successfully tested in the Deutsches Museum Bonn, Germany [8].

Since 2009, we develop domestic service robots. Our robots, Dynamaid [11] 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 a communication head. 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.

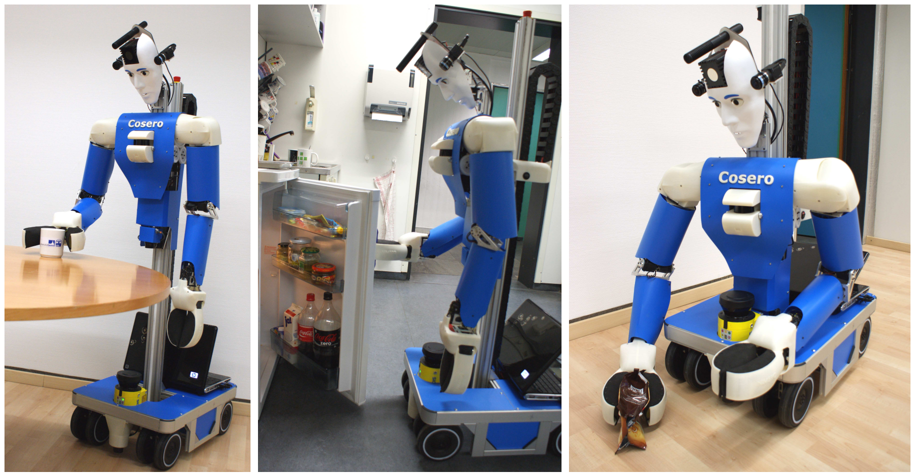

Fig. 1. The cognitive service robot Cosero with anthropomorphic arms and omnidirectional base.

2 Mechanical and Electrical Design

We equipped our robots Cosero and Dynamaid (s. Fig. 1 and Fig. 2) with omnidirectional drives to maneuver in the narrow passages found in household environments. 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

Fig. 2. Our domestic service robot Dynamaid with anthropomorphic arms and omnidirectional base.

(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 [11], 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 an RGB-D camera in their head (max. depth 10 m) 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.

The RGB-D camera is composed of a MESA SR-4000 time-of-flight camera and two PointGrey Flea2 color cameras with a resolution of 1280×960. We mounted the camera system on the head for several reasons: First, since the robots have a similar body height (1.6 m default height) than 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 time-of-flight sensor 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 informa-

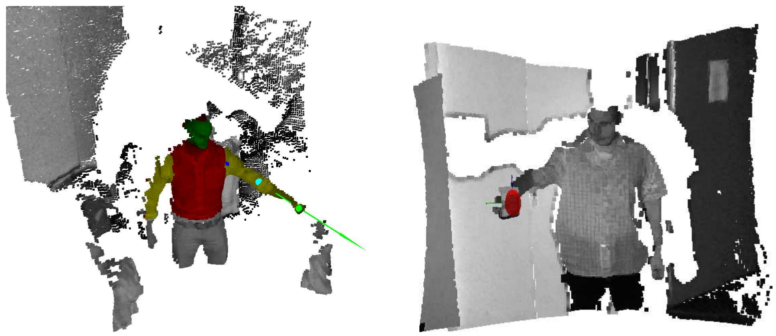

Fig. 3. Recognizing pointing and showing gestures. Left: the user points to an object in the scene. Right: the user shows an object to the robot.

tion 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 (s. Fig. 3). The primary sensor in our system for perceiving a gesture is the ToF 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.

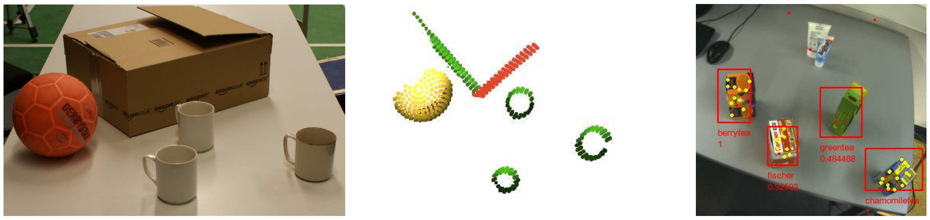

Fig. 4. Object Perception. Left: tabletop scene. Middle: points on extracted geometric shape primitives for the left scene (yellow sphere, green cylinder, red plane). Right: camera image (different scene) with rectangular object regions that are computed from object detections. The image also shows extracted SURF features (yellow dots) and recognized object class.

3.3 Auditory Perception

Speech is recognized using a commercial ASR system from Loquendo [7]. This system is speaker-independent and uses a small-vocabulary grammar which changes with the dialog state.

The grammar definition of the Loquendo speech recognition system allows to tag rules with semantic attributes. When speech is recognized, a semantic parse tree is provided that we process further. We use semantic parse trees to interpret sentences for complex commands and to generate appropriate behavior.

3.4 Perception of Objects

For object perception we develop approaches that combine ToF sensing and vision. From ToF depth measurements of manipulation scenes, we extract the surface on which the objects are located through efficient RANSAC methods [6]. We cluster the remaining measurements to obtain a segmentation into objects and extract geometric shape primitives.

The locations of the detected objects are mapped into the image plane (see Fig. 4a). In the rectangular regions of interest, both color histograms and SURF features [1] are extracted. For each object class, multiple descriptors are recorded from different view points during training, in order to achieve a view-independent object recognition.

3.5 Self-Localization and Mapping

To acquire 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. In the standard MCL approach, the map is assumed static. Movable objects like doors violate this assumption which may lead to poor localization performance. Also, the knowledge about doors and their state could be considered for navigational planning. For these reasons, we developed an extension to the MCL approach to simultaneously localize the robot and estimate the state of doors [9].

4 Behavior Control

The autonomous behavior of our robots is generated in a modular control architecture. We employ the inter process communication infrastructure of the Robot Operating System (ROS) [10]. The control modules are organized in four layers.

On the *sensorimotor layer*, data is acquired from the sensors and position targets are generated and sent to the actuating hardware components. The kinematic control module, for example, processes distance measurements of the IR sensors in the gripper and feeds back control commands for the omnidirectional drive and the actuators in torso and arm.

The *action-and-perception layer* contains modules for person and object perception, safe local navigation, localization, and mapping. These modules use sensorimotor skills to achieve reactive action and they process sensory information to perceive the state of the environment. E.g., the local navigation module perceives its close surrounding with the LRFs and the ToF camera to drive safely to target poses.

Modules on the *subtask layer* coordinate sensorimotor skills, reactive action, and environment perception to achieve higher-level actions like mobile manipulation, navigation, and human-robot-interaction. For example, the mobile manipulation module combines motion primitives for grasping and carrying of objects with safe omnidirectional driving and object detection.

Finally, at the *task layer* the subtasks are further combined to solve complex tasks that require navigation, mobile manipulation, and human-robot-interaction. One such task in the RoboCup@home competition is to fetch an object from a location in the environment after a human user gives a hint on the object location through spoken commands.

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. 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 firm 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 obstacle avoidance using measurements from LRFs and the ToF camera [3]. To enlarge the narrow field-of-view of the ToF 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. Fig. 2 (middle) shows how Dynamaid hands an object to a human user during RoboCup 2009. For receiving an object from a person, we localize the object that is extended towards the robot by the person with the depth camera and drive towards it. As soon as the object is reachable with the arms, the object is grasped. 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.

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. 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 two years now at RoboCup German Open and RoboCup competitions in 2009 and 2010. 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.

We plan to equip Dynamaid and Cosero with an expressive communication head similar to Robotinho. We will continue to improve the system for RoboCup 2011. 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: Prof. Sven Behnke, Jörg Stückler
- Staff: David Dröschel, Kathrin Gräve, Dirk Holz, and Michael Schreiber
- Students: Jochen Kläß, Ricarda Steffens, and Oliver Tischler

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, to appear.
3. David Droeschel, Jörg Stückler, Dirk Holz, and Sven 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. D. Holz, R. Schnabel, D. Droeschel, J. Stückler, and S. Behnke. Towards semantic scene analysis with time-of-flight cameras. In *Proc. of the 14th Int. RoboCup Symposium*, 2010.
7. Loquendo S.p.A. Vocal technology and services. http://www.loquendo.com, 2007.
8. M. Nieuwenhuisen, J. Gaspers, O. Tischler, and S. Behnke. Intuitive multimodal interaction and predictable behavior for the museum tour guide robot Robotinho. In *Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2010.
9. Matthias Nieuwenhuisen, Jörg Stückler, and Sven Behnke. Improving indoor navigation of autonomous robots by an explicit representation of doors. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, 2010.
10. 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.
11. 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.

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