NimbRo @Home 2010 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 Singapore in June 2010.
Our team uses self-constructed humanoid robots for object manipulaion and intuitive multimodal communication with humans. The paper describes the mechanical and electrical design of our robots Dynamaid and Robotinho. It also covers perception and behavior control.

1 Introduction

Our team NimbRo competes now for the second time 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 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.

Fig. 1. Our robot Dynamaid with anthropomorphic arms and omnidirectional base.

In January 2010, Robotinho has been successfully tested as a museum tour guide in the Deutsches Museum Bonn, Germany.

Since 2009, we develop Dynamaid [13], a domestic service robot that balances indoor navigation, mobile manipulation, and intuitive human-robot interaction. We equipped Dynamaid with an omnidirectional drive for robust navigation, two anthropomorphic arms for object manipulation, and with a communication head. In contrast to many other service robot systems, Dynamaid is lightweight, inexpensive, and easy to interface.

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

2 Mechanical and Electrical Design

2.1 Robotinho

Our humanoid robot Robotinho, shown in Fig. 2, has been originally designed for playing soccer in the RoboCup Humanoid League TeenSize class [3]. It is 110cm tall and has a weight of only 6kg. Its body has 25 degrees of freedom (DOF): six per leg, four per arm, three in the trunk, and two in the neck.

The mechanical design focused on simplicity, robustness, and weight reduction. The body is driven by 35 Dynamixel DX-117 actuators. All joints in the legs and the trunk, except for the yaw axes, are driven by two parallel actuators. The hip and trunk yaw and roll axes are reinforced by external spur gears.

For the use as communication robot, we equipped Robotinho with an expressive 15DOF head, visible in the right part of Fig. 2. The eyes are small (22×26 mm) high-resolution (765×582 pixels) color cameras with high-speed USB 2.0 interface (Videology 21K155). One camera has a narrow field-of-view of 21.4° while the other camera is equipped with a wide-angle lens, yielding a horizontal field-of-view of 64°. All head joints are driven by small digital servos. Four servos move the eyes in two axes. While the lower eye lid moves together with the eyeballs, the upper eye lid can be moved independently. Six servo motors animate jaw and mouth and four servos animate the eyebrows.

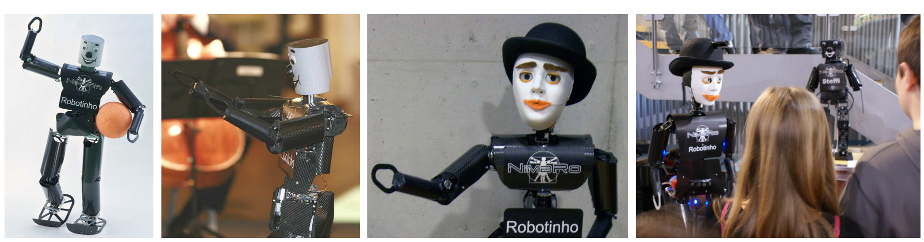

Fig. 2. Our robot Robotinho was used as soccer player in the Humanoid League. He conducted The 12 Cellists of the Berlin Philharmonic. The robot is now equipped with an expressive communication head, a wheeled base, and navigation sensors.

For navigation in indoor environments, Robotinho is equipped with a small laser-range finder (Hokuyo URG-04-LX), located in its neck. In addition, we added eight ultrasonic distance sensors (Devantech SRF02) around the hip.

For the @Home competitions, we placed Robotinho on a wheeled base with four individually steerable axes. This base uses four Roomba 530 differential drives, four Dynamixel joints, and a Sick S300 LRF.

2.2 Dynamaid

Dynamaid's mobile base (see Fig. 1) also consists of four individually steerable differential drives, which are attached to corners of an rectangular chassis. Each pair of wheels is connected to the chassis with a Robotis Dynamixel RX-64 actuator. Both wheels of a wheel pair are driven individually by Dynamixel EX-106 actuators.

Dynamaid's anthropomorphic arm has seven joints which are also driven by Dynamixel actuators. We designed its size, joint configuration, and range of motion to resemble human reaching capabilities. The arm is equipped with a 2 degree of freedom (DOF) gripper, which contains four infrared distance sensors. Its maximum payload is 1kg.

In the trunk, Dynamaid is equipped with a Hokuyo URG-04LX laser range finder, which is mounted on an actuated roll joint to measure in the horizontal and in the vertical plane. The trunk is equipped with two more joints. One trunk actuator can lift the entire upper body by 1m. This allows for object manipulation in different heights. In the lowest position, the trunk laser is only 4cm above the ground. Hence, Dynamaid can pick up objects from the floor. In the highest position, Dynamaid is about 180cm tall and can grab objects from 100cm high tables. The second actuator allows to twist the upper body by $\pm 90°$. This extends the working space of the arms and allows the robot to face persons with its upper body without moving the base.

The head of Dynamaid contains a directional microphone, a Swissranger SR-4000 time-of-flight camera, and a stereo camera (PointGrey Flea2-13S2C-C) on a pan-tilt neck built from 2 Dynamixel RX-64 actuators. Overall, Dynamaid currently has 35 joints which can be accessed from the main computer via USB. The robot is powered by Kokam 5Ah Lithium polymer cells.

3 Perception

3.1 Visual Perception of Communication Partners

To detect and track people in the environment, we combine laser range finders (LRFs) and vision. In order to keep track of persons even when they are temporarily outside the robot's field of view, the robots maintain a probabilistic belief about the people in their surroundings.

Our face detection system uses a boosted cascade of Haar-like features [7]. For Robotinho, we use a Kalman filter to track the position of a face. To account

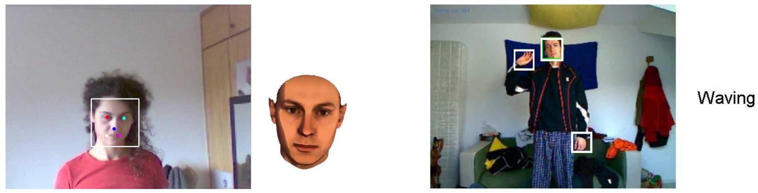

Fig. 3. Feature-based head pose estimation (left) and gesture recognition (right).

for false classifications of face/non-face regions and association failures, we use a recursive Bayesian update scheme [9] to estimate the existence probability of a face. Dynamaid tracks multiple persons in its vicinity with a multi-hypothesis tracker.

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

In order to judge the focus of attention of Robotinho's interaction partners, we developed a module for head-pose estimation from camera images [14]. Starting from the detected faces, we search for facial features, such as the nose tip, the mouth corners, the eyes, and the ears (see Fig. 3). We track these features and estimate the 3D head pose from the relative feature positions using a neural function approximator.

We also developed a module for the recognition of human gestures (also shown in Fig. 3) [1]. Starting from the detected faces, we localize the hands by using the facial skin color in combination with luminance-based hand detectors. We track the hands and model the trajectories of hands and head using Hidden Markov Models. These HMMs are trained to recognize gestures. For parametric gestures, such as pointing and size indicating gestures, we localize the hold phase in time using the inferred HMM states. We estimate the respective gesture parameter from the relative positions of head and hands during the hold phase.

3.2 Auditory Perception

Speech is recognized using a commercial ASR system from Loquendo [8]. This system is speaker-independent and uses a small-vocabulary grammar which changes with the dialog state.

3.3 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. We cluster the remaining measurements to obtain a segmentation into objects and extract geometric shape primitives.

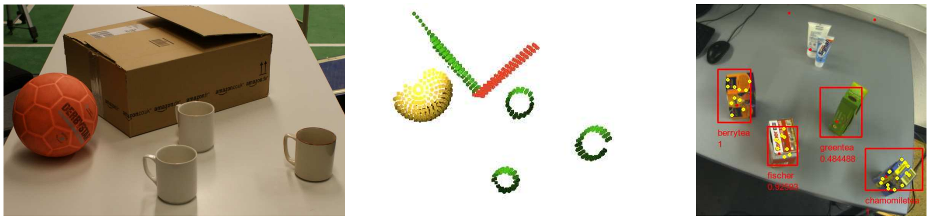

Fig. 4. Object Perception. Left: tabletop scene. Middle: points on extracted geometric shape primitives for the left scene (yellow sphere, green cylinder, red plane). Right: camera image (different scene) with rectangular object regions that are computed from object detections. The image also shows extracted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

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 [2] 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.

3.4 Self-Localization and Mapping

To acquire maps of unknown environments, we apply GMapping [6], 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. To overcome this problem, we developed an extension to the MCL approach to simultaneously localize the robot and estimate the state of doors [10].

4 Behavior Control

The autonomous behavior of our robots is generated in a modular multi-threaded control architecture. We employ the inter process communication infrastructure of the Robot Operating System (ROS) [11]. 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 LRF 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 implemented several motion primitives for grasping, carrying, and handing over of objects. We also investigate learning of motion primitives by imitation and reinforcement learning [5].

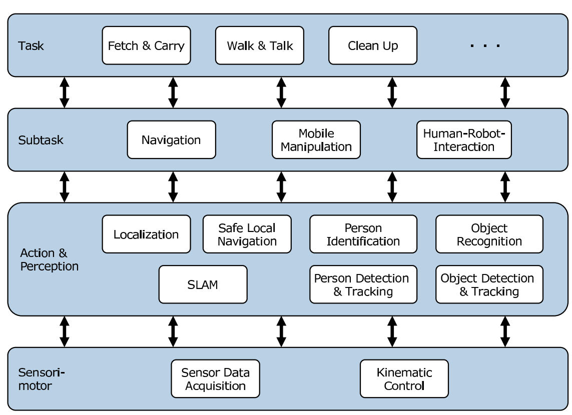

Fig. 5. Sketch of the control architecture.

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 [4]. 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. Dynamaid can grasp objects, carry them, and hand them to human users.

When Dynamaid hands-over an object to a human, the user can trigger the release of the object either by a speech command or by simply taking the object. We measure the pulling from the displacement of the actuators. When the pulling persists, Dynamaid opens her gripper and releases the object. Fig. 1 (middle) shows how Dynamaid hands an object to a human user during RoboCup 2009.

4.5 Attentional System

Robotinho shows interest in multiple persons in its vicinity and shifts its attention between them so that they feel involved into the conversation. To determine the robot's focus of attention, we compute an importance value for each person.

4.6 Multimodal Dialogue System

The dialogue system covers a restricted domain only. It is realized using a finite-state automaton. Some transitions are triggered by timeouts. Possible dialog actions include spoken utterances, changes in the emotional state, and gestures.

4.7 Arm and Head Gestures

Our robot uses arm and head movements to generate gestures and to appear livelier. It performs symbolic, batonic, and pointing gestures. To communicate the robot's mood, we use a face with animated mouth and eyebrows to display facial expressions (see Fig. 5). The robot's mood is computed in a two-dimensional space, using six basic emotional expressions (joy, surprise, fear, sadness, anger, and disgust) [12]. In combination with facial expressions, we use emotional speech to express the robot's mood.

5 Conclusion

The described system was evaluated for the first time at RoboCup German Open 2009 during Hannover Fair and at RoboCup 2009 in Graz. In both competitions, the robots performed very well. They successfully participated in the tests *Introduce*, *Follow Me*, *Fetch&Carry*, *Who-Is-Who*, *Open Challenge*, *Walk&Talk*, *Supermarket*, *PartyBot*, and the *Demo Challenge*.

At the time of writing, February 10th, 2010, in addition to the described robots, we made good progress in the development of a successor to Dynamaid with a more powerful upper body. We also plan to equip Dynamaid and its successor with an expressive communication head similar to Robotinho. We will continue to improve the system for RoboCup 2010. 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, Oliver Tischler, and Ralf Waldukat

References

1. T. Axenbeck, M. Bennewitz, S. Behnke, and W. Burgard. Recognizing complex, parameterized gestures from monocular image sequences. In *Proc. of the IEEE/RSJ Int. Conf. on Humanoid Robots (Humanoids)*, 2008.
2. H. Bay, T. Tuytelaars, and L. Van Gool. SURF: speeded up robust features. In *9th European Conference on Computer Vision*, 2006.
3. S. Behnke, M. Schreiber, J. Sückler, R. Renner, and H. Strasdat. See, walk, and kick: Humanoid robots start to play soccer. In *Proc. of the IEEE/RSJ International Conference on Humanoid Robots (Humanoids)*, page 497—503, 2006.
4. 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.
5. K. Gräve, J. Stückler, and S. Behnke. Learning motion skills from expert demonstrations and own experience using gaussian process regression. 2010. Accepted for the joint conference of the 41st Int. Symp. on Robotics (ISR 2010) and the 6th German Conf. on Robotics (ROBOTIK 2010).
6. G. Grisetti, C. Stachniss, and W. Burgard. Improved techniques for grid mapping with Rao-Blackwellized particle filters. *IEEE Trans. on Robotics*, 23(1), 2007.
7. R. Lienhard and J. Maydt. An extended set of haar-like features for rapid object detection. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2002.
8. Loquendo S.p.A. Vocal technology and services. http://www.loquendo.com, 2007.
9. H.P. Moravec and A.E. Elfes. High resolution maps from wide angle sonar. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1985.
10. 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.
11. 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.
12. Z. Ruttkay, H. Noot, and P. Hagen. Emotion Disc and Emotion Squares: Tools to explore the facial expression space. *Computer Graphics Forum*, 22(1):49–53, 2003.
13. 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.
14. T. Vatahska, M. Bennewitz, and S. Behnke. Feature-based head pose estimation from images. In *Proc. of Int. Conf. on Humanoid Robots (Humanoids)*, 2007.

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