NimbRo @Home 2009 Team Description

Sven Behnke, Jörg Stückler, and Michael Schreiber

Rheinische Friedrich-Wilhelms-Universität Bonn
Computer Science Institute VI: Autonomous Intelligent Systems
Römerstr. 164, 53117 Bonn, Germany
{ behnke | stueckler | schreiber} @ 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 Graz in July 2009.
Our team uses self-constructed humanoid robots for intuitive multimodal communication with humans and for object manipulation. The paper describes the mechanical and electrical design of our robots Robotinho and Dynamaid. It also covers perception and behavior control.

1 Introduction

Our team NimbRo is new to the @Home league. The only competition so far was RoboCup German Open 2009, where we came in second in the @Home league. However, since 2004, NimbRo competes with great success in the Humanoid league. Our KidSize robots [3] reached all humanoid soccer finals played so far and won the 2007 and 2008 soccer tournaments. Our TeenSize robots [4] also performed very well. They won the penalty kick competitions in 2005 and 2007.

In the project NimbRo – Learning Humanoid Robots – we investigate not only humanoid soccer, but also intuitive multimodal communication between humans and robots. The general idea is that by mimicking the way humans interact

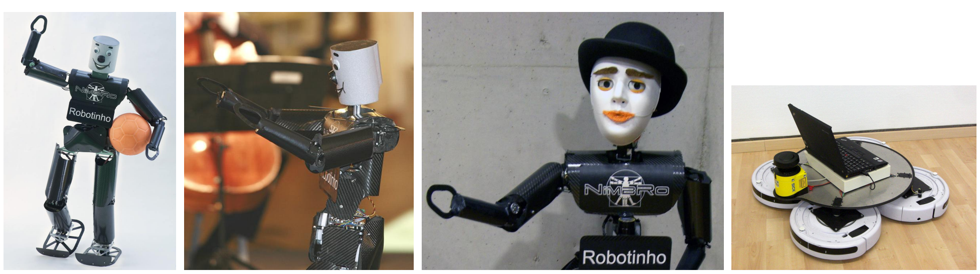

Fig. 1. Our robot Robotinho was used as soccer player in the Humanoid League. He conducted The 12 Cellists of the Berlin Philharmonic. The robot is now equipped with an expressive communication head, a wheeled base, and navigation sensors.

with each other, it will be possible to transfer the efficient and robust communication strategies that humans use in their face-to-face interactions to the man-machine interface. This includes the use of multiple modalities like speech, facial expressions, gestures, body language, etc. If successful, this approach yields a user interface that leverages the evolution of human communication and that is intuitive to naive users, as they have practiced it since early childhood.

Our test scenario for human-robot interaction is a museum tour guide. This application requires interacting with multiple unknown persons. The project moved in 2008 from Freiburg to Bonn. Here, we started a project group on mobile manipulation to extend the application domains of our robots.

Our goal is to design a balanced system for the @Home competitions, capable of robust navigation in a home environment, object manipulation using two anthropomorphic arms, and intuitive multimodal human-robot interaction.

Here, we present the humanoid robot Robotinho that we developed as successor to the communication robot Fritz [5] and our robot Dynamaid that we develop for RoboCup 2009. 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

Our humanoid robot Robotinho, shown in Fig. 1, has been originally designed for playing soccer in the RoboCup Humanoid League TeenSize class [2]. 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. 1. The eyes are small (22×26

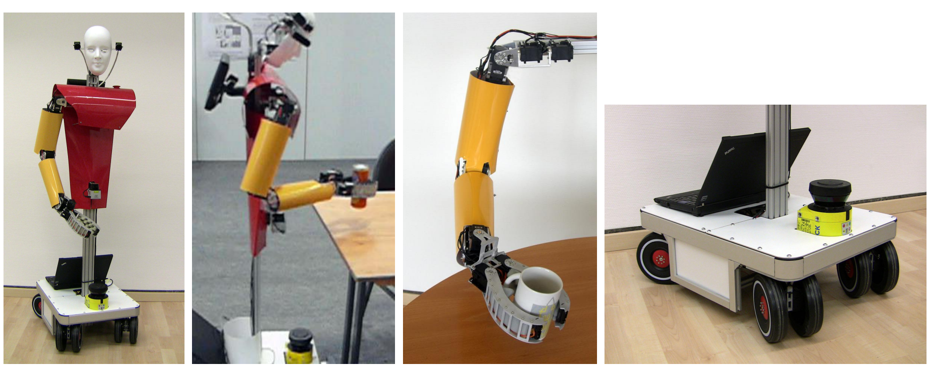

Fig. 2. Our robot Dynamaid with anthropomorphic arm and omnidirectional base.

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.

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.

Dynamaid's mobile base (see Fig. 2) 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 can measure in the horizontal and in the vertical plane.

The head of Dynamaid contains a directional microphone and a stereo camera (PointGrey Flea2-13S2-C) on a pan-tilt neck built from 2 Dynamixel RX-64 actuators. Overall, Dynamaid currently has 24 joints, which can be accessed from the main computer via USB. The robot is powered by Kokam 5Ah Lithium polymer cells.

The Dynamixel actuators used in the robots have a flexible interface. Not only target positions are sent to the actuators, but also parameters of the control loop, such as the compliance. In the opposite direction, the current positions, speeds, loads, temperatures, and voltages are read back.

3 Perception

• **Visual Perception of Communication Partners**

To detect and track people in the environment of Robotinho, we use the two cameras. In order to keep track of persons even when they are temporarily outside the robot's field of view, the robot maintains a probabilistic belief about the people in its surroundings.

Our face detection system uses a boosted cascade of Haar-like features [7]. We use a Kalman filter to track the position of a face. To account for false

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

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.

In order to judge the focus of attention of Robotinho's interaction partners, we developed a module for head-pose estimation from camera images [12]. 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.

• **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.

We also implemented a speaker localization system that uses a stereo microphone. We apply the cross-power spectrum phase analysis to calculate the time difference between left and the right channel (ITD). This yields an estimate of the horizontal angle between the speaker and the microphones.

• **Self-Localization and Mapping**

To acquire maps of unknown environments, we apply a FastSLAM2 approach to the Simultaneous Localization and Mapping (SLAM) problem. We use the GMapping implementation which is contained in the OpenSLAM open source repository. Fig. 4 shows a generated map of the RoboCup@home arena at GermanOpen.

We apply a variant of the adaptive Monte Carlo Localization to estimate the robot's pose in a given occupancy grid map. We use the end-point model for laser range scans, as it can be implemented efficiently through look-up tables and it is more robust than the ray-cast model to small changes in the environment.

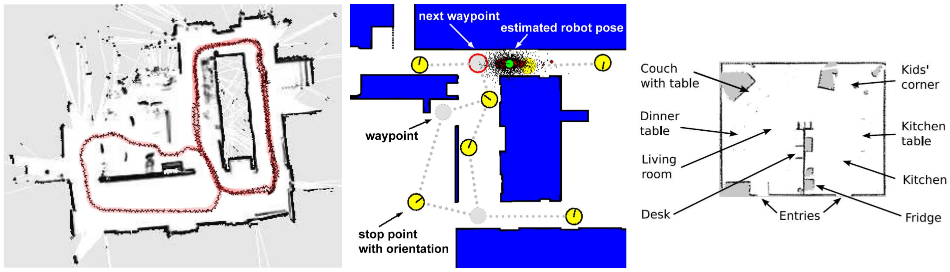

Fig. 4. Left: Map learned from walking Robotinho [11]. Middle: Topological map used for waypoint navigation. Right: Map of German Open 2009 arena.

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 Player/Stage. 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 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 manip-

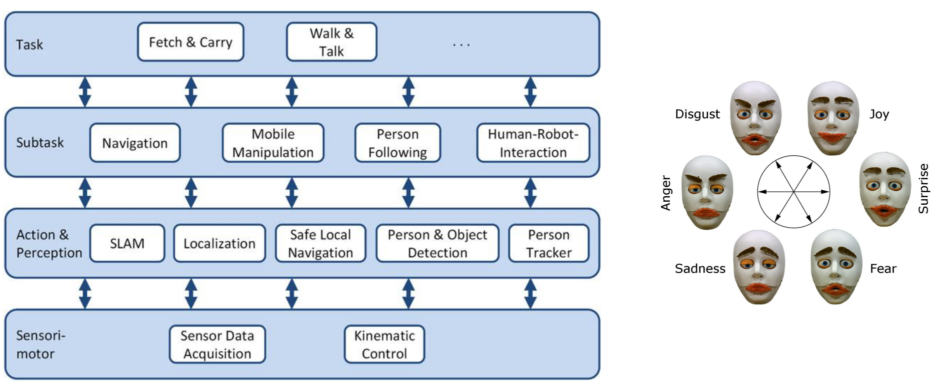

Fig. 5. Sketch control architecture and facial expressions generated.

ulation 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.

- **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. The orientation of the four drives and the linear velocities of the eight wheels are controlled kinematically such that their instantaneous centers of rotation (ICRs) coincide with the ICR that results from the commands for the center of the base.

- **Control of the anthropomorphic arm**

The arm is controlled using differential inverse kinematics to follow trajectories of either the 6 DOF end-effector pose or the 3 DOF end-effector position $\dot{\theta} = J^{\#}(\theta)\ \dot{x} + \alpha\ (I - J^{\#}(\theta)J(\theta))\ \frac{dg(\theta)}{d\theta}$, where θ are the joint angles, x is the end-effector state variable, J and $J^{\#}$ are the Jacobian of the arm's forward kinematics and its pseudoinverse, respectively, and α is a step size parameter. Redundancy is resolved using nullspace optimization of the cost function $g(\theta)$ 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. These motion primitives are either open-loop motion sequences or use feedback like the distance to objects as measured by the IR sensors, e.g. to adjust the height of the gripper over surfaces or to close the gripper when an object is detected between the fingers.

- **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 focus of attention of the robot, we compute an importance value for each person. Turning towards communication partners is distributed over three levels: the eyes, the neck, and the trunk with different time constants [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.

- **Arm and Head Gestures**

Our robot uses arm and head movements to generate gestures and to appear livelier. Arm gestures consist of a preparation phase, where the arm moves slowly to a starting position, the stroke phase that carries the linguistic meaning, and a retraction phase, where the hand moves back to a resting position. The stroke is synchronized to the speech synthesis.

◦ *Symbolic Gestures:* The symbolic gestures in our dialog system include a single-handed greeting gesture that is used while saying hello to newly detected people. The robot performs a come-closer gesture with both arms when detected persons are farther away than the normal conversation distance. There robot accompanies certain questions with an inquiring gesture, where it moves both elbows outwards to the back. In appropriate situations, the robot performs a disappointment gesture by moving, during the stroke, both hands quickly down. To confirm or to disagree, the robot also nods or shakes its head, respectively.

◦ *Batonic Gestures:* Humans continuously gesticulate to emphasize their utterances while talking to each other. Robotinho also makes small emphasizing gestures with both arms when he is generating longer sentences.

◦ *Pointing Gestures:* To draw the attention of communication partners towards objects of interest, our robot performs pointing gestures, as shown in Fig. 5. When the robot wants to draw the attention to an object, it simultaneously moves the head and the eyes in the corresponding direction and points in the direction with the respective arm while uttering the object name.

◦ *Non-Gestural Movements:* Our robot performs small movements with its arms to appear livelier. We also implemented a regular breathing motion and pseudo-random eye blinks.

• **Emotional Expression**

Showing emotions plays an important role in inter-human communication. During an interaction, the perception of the mood of the conversational partner helps to interpret his/her behavior and to infer intention.

◦ *Facial Expressions:* 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) [10].

◦ *Emotional Speech Synthesis:* In combination with facial expressions, we use emotional speech to express the robot's mood. Most speech synthesis systems do not support emotional speech directly; neither does the system we use (Loquendo TTS [8]). However, in this system, we can influence the parameters average pitch, speed, and volume and thereby communicate the robot's emotional state.

5 Conclusion

The described system was evaluated for the first time RoboCup German Open 2009 during Hannover Fair. In Stage I, we used Robotinho for the introduce task. It explained itself and Dynamaid and interacted with a human in a natural way. For the follow me task, we used Dynamaid. She was able to quickly follow an unknown human though the arena, outside, and back into the arena. She also could be controlled by voice commands. Dynamaid also did the fetch and carry task very well. She took spoken orders from the human user and delivered reliably the requested object. In Stage II, Dynamaid did the walk and talk task perfectly. A human showed her places in the apartment that she could visit afterwards as requested by spoken commands. In the demo challenge, the users could order different drinks that she fetched quickly and reliably from different places in the apartment. In the final, Robotinho gave a tour though the apartment while Dynamaid fetched a drink for a guest.

At the time of writing, May 15th, 2009, we made good progress in preparation for the @Home competition in Graz. We will continue to improve the system for RoboCup 2009. 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: Matthias Nieuwenhuisen, Michael Schreiber, and Gideon Wenniger
– Students: David Dröschel, Kathrin Gräve, Dirk Holz, Jochen Kläß, Sebastian Muszynski, Ralf Waldukat, and Oliver Tischler

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. 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.
3. S. Behnke, J. Stückler, and M. Schreiber. NimbRo KidSize 2009 team description. In *Humanoid League Team Descriptions, Graz, Austria*, to appear 03/2009.
4. S. Behnke, J. Stückler, and M. Schreiber. NimbRo TeenSize 2009 team description. In *Humanoid League Team Descriptions, Graz, Austria*, to appear 03/2009.
5. M. Bennewitz, F. Faber, D. Joho, and S. Behnke. Fritz - A humanoid communication robot. In *Proceeding of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, page 1072—1077, 2007.
6. F. Faber, M. Bennewitz, and S. Behnke. Controlling the gaze direction of a humanoid robot with redundant joints. In *Proc. of 17th IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2008.
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. 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.
11. C. Stachniss, M. Bennewitz, G. Grisetti, S. Behnke, and W. Burgard. How to learn accurate grid maps with a humanoid. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2008.
12. 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.