

# Delft Robotics RoboCup@Home 2013

## Team Description Paper

Floris Gaisser, Aswin Chandarr A.B., Maja Rudinac, Machiel Bruinink, Susana Pons Rueda, Mukunda Bharatheesha, Guus Liqui Lung, Martijn Wisse and Pieter Jonker

Delft University of Technology,  
Department of Bio-Mechanical Engineering,  
Mekelweg 2, 2628 CD, Delft, The Netherlands ,  
[info@delftrobotics.nl](mailto:info@delftrobotics.nl),  
Website: <http://www.delftrobotics.nl>

**Abstract.** This paper provides an introduction to the Delft Robotics RoboCup@Home team and the Delft Personal Robot Robby designed at the Delft Biorobotics Laboratory (DBL)<sup>1</sup>. Robby is an affordable service robot with simplistic design which keeps number of degrees of freedom and sensors to minimum while still allowing for a complex vision based control of robot behaviour in domestic settings. Besides novel design, main innovation lies in robot software architecture which allows learning of novel objects and users. We propose new methods for real time learning of users as well as the robust object description method using multimodal features. Finally, manipulation of unknown objects is achieved using novel vision based grasping method which utilizes an underactuated gripper developed in our lab.

**Keywords:** affordable robot, vision based control, real-time face learning, underactuated gripper

## 1 Introduction

The Delft Robotics RoboCup@Home team has been founded in September 2011 as part of the Delft Biorobotics Lab and Delft Robotics Institute. The lab has a long experience with the Robocup competition in the Clockwork Orange Middle size league team, Dutch AIBO team and Dutch Robotics Humanoid robots team. From 2012 we joined the Robocup@Home competition and participated in the Dutch Open 2012 in Eindhoven and RoboCup 2012 in Mexico City. The team consists of a group of Bachelor, Master and PhD students, advised by post doctoral researchers and professors from the Delft Robotics Institute at the Delft University of Technology.

The strong side of the Institute and the DBL lab is combining multidisciplinary research on smart mechanical design with advanced vision based robot control. This allowed us to design an affordable robot Robby which we describe in more detail in the following text. Our robot has achieved much recognition through its demonstrations in several popular shows at Dutch national TV, healthcare fairs and children manifestations.

At first the hardware of Robby will be introduced, followed by a detailed description of the software architecture. In the following text we will explain the details of the modules of speech, object and face recognition, object manipulation and navigation as well as how they interact with each other.

---

<sup>1</sup> <http://robotics.tudelft.nl>

## 2 Robot platform

In this section, we describe the mechanical design and sensor choices of the robot with a main focus on robot affordability.

### 2.1 Hardware

The mechanical design used for the mobile base results in a significant reduction in the overall robot cost and also provides us with the scientific challenge of dealing with non-holonomicity during navigation. The entire hardware of the robot is composed of the following components and is illustrated in the figure 1.

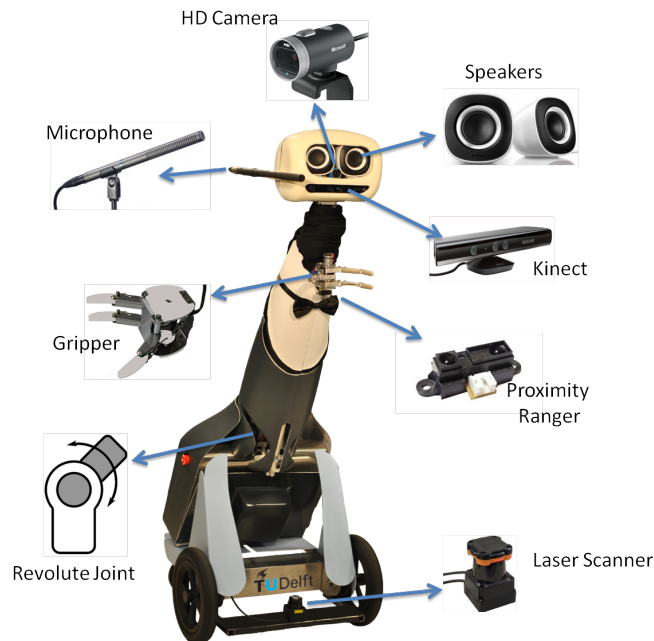


Fig. 1: Robot hardware design

**Base** A mobile base consisting of two wheels coupled to 12V/3.6A DC motors is the platform over which the entire system has been developed. The motors are controlled by 3Mxel, an advanced Motor Controller developed in the Laboratory.

**Arm** The arm of the robot consists of a four bar linkage mechanism. The head and the neck are mounted on top of the arm. This is coupled to the mobile base through a revolute joint in the hip. The mechanism ensures the head remains horizontal during the bending of the arm.

**Head** The perception and human interaction modules are housed in the head of the robot. This consists of the following:

**RGB-D Camera** The Kinect developed by *Microsoft* apart from having a standard RGB camera, uses interpretation of structured light to estimate the distance. The *Depth Map* is synchronized with color map and it provides a RGB-D

output in realtime speed of 30 Fps. This depth map is transformed into a Point-Cloud data using Projection matrices. By this, the environment is seen by the robot discrete points( $[x, y, z]$  in  $m$ ) containing RGB data.

**HD Camera** Though, the Kinect provides spatial and chromatic data of the environment, the resolution is not high for certain applications. Hence a *Microsoft LifeCam* has been used in conjunction to the kinect to get a High Resolution image for more reliable perception. The Kinect and the camera together are used sense the chromatic and the spatial aspects of the visual field of the environment.

**Neck Joint** The Camera system encapsulated within the robot head, has to be in different orientation for different circumstances like varying height of the person/table, etc. Hence a *Pan-Tilt* unit comprising of Dynamixel RX-64 is used.

**Microphone and Speakers** A highly directional (Shotgun Audiotechnica AT-8035) Microphone is used to robustly receive the speech in noisy real-time conditions and a Philips (SPA2201) speakers are used for the natural Human robot interaction.

**Gripper** The objects present in the everyday situations have varied properties like shape, texture, rigidity etc. It is quite complex if a standard fully actuated gripper is used as the finger configuration has to be computed for each and every circumstances. Hence an under-actuated gripper [6] developed in the Delft BioRobotics Lab which adapts itself to different object properties has been used to simplify computational complexity.

**Proximity Sensor** Since an adaptive gripper is used, the only feedback to control the gripper is the presence/absence of the object. Hence a SHARP(0A41SK F15) proximity sensor is used to detect the presence of object within the manipulable range of the gripper.

**Laser Scanner** A planar laser scanner (HOKUYO URG - 04LX) with a range of 4m with a precision of 1mm is used along with the mobile base to navigate in dynamic environments avoiding the obstacles.

### 3 Software Architecture

Robot Robby has been equipped with a basic client-server software architecture. The software has been divided into 4 layers, the core, subcores, individual modules and low-level control depicted in figure 2. Every layer in the system is designed to be modular and has its own specific task, allowing any part of the system to be changed or updated without affecting the other parts of the system unless the functionality of such a module has been changed. The software architecture has been designed to work with Robot Operating System (ROS). ROS provides many existing package and drivers built specifically for robots and it provides a client-server communication interface for passing messages between different modules [1].

#### 3.1 Core

The core acts as the brain of the robot, it is responsible for making the decisions and planning the requested task. The robot receives the task commands from the user through speech. When the robot receives a task from the user the robot will formulate a list of planned steps in a specific order to complete the required task. The robot must also know what its current state is in order to further formulate the steps. A

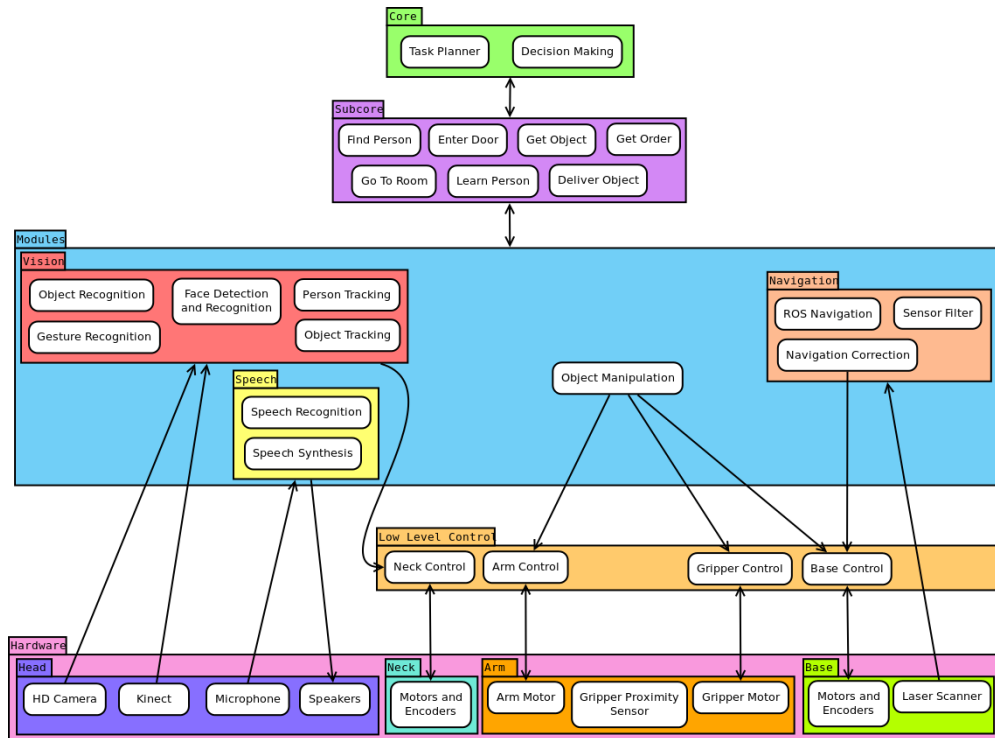


Fig. 2: Software architecture

state machine has been implemented to keep track of the robots current state and to initiate the modules to perform the planned steps. Once a step is completed it will inform the core so that the core can initiate the next step. Once all steps have been completed the core informs the user of the task being completed and waits for the next user command.

### 3.2 Subcore

A subcore is a sequence of steps that are reusable for different tasks. For example, finding a person subcore can be used in the tasks involving bringing a person a drink or also finding a person in emergency situations. Each subcore has been created as generic as possible so that it can be used as part of many different complex tasks. A subcore receives the inputs from the core and it initiates the individual modules and provides feedback to the core.

### 3.3 Individual Modules

Individual modules are responsible for main actions of the robot. These include navigation, manipulation, speech recognition, face recognition, object recognition and tracking as well as person tracking. Each individual module directly interacts with the low-level control layer.

### 3.4 Low-Level

Low-level control layer consists of drivers for the robot hardware such as wheels, gripper, arm, cameras and distance sensors. This layer facilitates the interaction of the individual modules with the hardware.

## 4 Speech recognition and synthesizing

Speech processing is a client software module that performs the tasks of speech recognition and synthesis to enable a human to interact with Robby using voice commands. One of the challenges in the Robocup@home competition is the robot's ability to communicate with an human and to understand his instructions, requests and questions. To achieve this goal, a dialogue system has been implemented [11] as is described in figure 3. The structure of the Dialogue System is following.

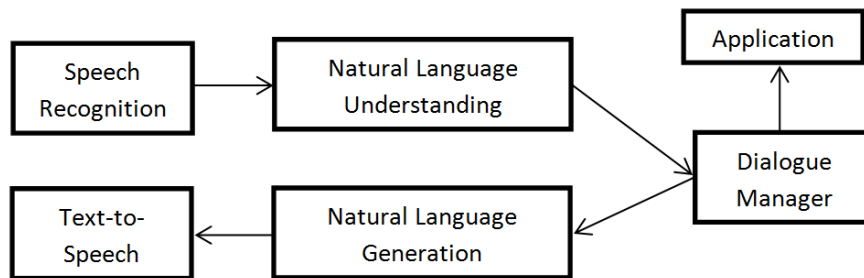


Fig. 3: Speech processing module

- **Speech recognition:** To convert voice to text, the system called SPHINX has been used. The speech recognizer selects the three most probable words and further passes them to the parser.
- **Natural Language Understanding:** A probabilistic parser is used to increase the performance of the Speech Recognizer based on the grammar. Once the result is parsed, all the words that do not have the potential to provide any added value to the recognition are removed.
- **Dialogue Manager:** A Frame-Based Dialogue system is implemented. The system first checks if it has all the information necessary to complete the tasks. The robot also confirms that what it has understood is correct. Finally, it asks for the missing information.
- **Text-to-Speech:** The interface FESTIVAL has been used to convert the desired feedback to speech.

## 5 Vision framework

The vision module is novel and specifically developed to allow Robby to localize, manipulate and learn unknown objects, persons and faces in unstructured environments.

### 5.1 Object recognition

In order to obtain robust object localization with a single Kinect sensor even in the case of large illumination change when Kinect data is not reliable, we have combined 3D plane segmentation from the Point Clouds Library (PCL) with 2D saliency based object localization [10].

To describe object appearances in dynamic scene environments, multimodal representations of properties are required. In order to represent object appearances, which can significantly vary in color, texture and 2D/3D shape, we extract all corresponding object feature vectors from multiple viewpoints and automatically calculate the most dominant features over all the robot's viewpoints [9]. This dominance weighting makes the object descriptors more robust to changes in viewpoint and illumination conditions and leads to better object recognition. It also automatically calculates the most efficient way to register the representative features of an unknown object [8]. To allow manipulation of unknown objects we are constantly tracking the object and updating the information on its shape model as proposed in [12]. Such a vision module allows us to localize, learn and further manipulate both known and unknown objects.

### 5.2 Face Recognition

Before service robots become part of every household, they need to be able to adapt to novel environments and new users. For good human-robot interaction facial appearance as well as the name of novel users has to be learned. One of the problems in the state of the art methods is that face learning was either too slow to be applied in a real-time application or had too low performance for efficient interaction. Therefore novel methods have been developed that will enable real-time face recognition and learning with a very high performance applicable on a service robot [4]. At first to improve the performance as well as the speed of the state of the art PCA method, we have introduced the Class Average PCA method. This method improves the performance by creating a description space, which is more easily separable, by increasing the inter-class distances while decreasing the intra-class distances. The description performance is improved by reducing the description variance within a class and obtaining a dimension for each class. To significantly reduce the speed, the projection matrix is reduced in size by not taking all training samples as in PCA, but by finding a single representation for each class.

Further to allow for classification of unknown faces, the novel CertKNN method has been introduced. The main benefit over the state of the art methods is finding the relation between the distance of classification and the certainty of that classification. This relation is automatically calculated from the data belonging to each class. In that way nearly optimal unknown classification can be done. Finally to further improve recognition performance a method has been developed that utilizes multiple frames in classification.

These novel methods combined provide with a face recognition module that can recognize faces robustly and at the same time classify unknown faces to learn these online in a real-time manner. Using this module Robby can adapt itself to new users in a robust and fast way as is required in real-world situations.

### 5.3 Person following

Following a person in a cluttered and dynamic environment is required for tasks like shopping and helping people in emergency situations in public places. For the robot

to follow a person, it first has to recognize and track a person over multiple frames as well as to regain tracking in situations of sudden occlusion and target loss. For person tracking a novel method has been developed which utilizes multi-model features of color, depth and motion for detection combined with a particle filter [5] and TLD [7] for tracking. The new location of the tracked person is used to update the local planner of the navigation module. To allow safe navigation through cluttered environment, this module will plan a path avoiding fixed objects as well as dynamic obstacles like other people.

## 6 Navigation

Autonomous navigation of the robot is performed by adaptation of the ROS navigation stack [2]. The navigation works based on a 2D projection of the environment. The global path planning is done by Dijkstra algorithm based on a occupancy grid map, the local obstacle avoidance and the localization of the robot is achieved by the combination of multimodal sensory information from the kinect, planar laser scanner and the robots inherent wheel odometry. This combined with the versatile motor controller (3Mxel) provides smooth navigation with the non-holonomic constraint of the mobile base.

## 7 Manipulation

Two novel methods have been used for the manipulation of the objects. An adaptive underactuated gripper is used [6]. This can grasp most of the common household objects with only a single actuation. Also a novel active grasping technique based on continuous object tracking has been implemented on the robot [3, 8]. It uses continuous object tracking based on TLD [7] as a visual feedback of the object's current position with respect to the robot. This method is named active as the robot is not passively executing a pre-generated motion, but actively moving towards the object. This method in the current state uses only a 1 DoF revolute joint in the hip augmented by a 2 DoF mobile base to grasp objects from heights varying from a height of 150 cm to the ground. Any position in 3 dimensions with a single (fixed) orientation can be reached by the end-effector with the available 3 degrees of freedom.

## 8 Conclusion and future work

In this paper we have introduced the Delft Personal Robot platform Robby designed at the Delft Biorobotics Laboratory. Main benefits of this robot are following:

1. simplistic and affordable design
2. novel algorithm for manipulation of unknown objects using an underactuated gripper
3. novel method for robust object recognition and learning using multimodal features
4. novel robust and fast methods for face recognition and online learning for human-robot interaction
5. novel method for safe person following in cluttered and dynamic environments
6. novel dialog based system with missing information retrieval for efficient human-robot interaction

These features will allow our robot to perform the required tasks at the RoboCup 2013. Since these modules are generic and platform independent, we intend to expand the existing platform with increased hardware sophistication.

## References

1. [Online]. Available: [www.ros.org](http://www.ros.org)
2. [Online]. Available: <http://www.ros.org/wiki/navigation>
3. A. A.B., "Towards bootstrapping robotic perception of indoor environments," Master's thesis, Delft University of Technology, 2012.
4. F. Gaisser, "Face recognition for cognitive robots," Master's thesis, Delft University of Technology, 2013.
5. D. A. Klein, D. Schulz, S. Frintrop, and A. B. Cremers, "Adaptive real-time video-tracking for arbitrary objects," in *IROIS*, 2010, pp. 772–777.
6. G. A. Kragten, "Underactuated hands: Fundamentals, performance analysis and design," Ph.D. dissertation, Delft University of Technology, 2011.
7. K. Mikolajczyk, Z. Kalal, and J. Matas, "Online learning of robust object detectors during unstable tracking," in *3rd Online Learning for Computer Vision Workshop, Kyoto, Japan, IEEE CS*, 2009.
8. M. Rudinac, "Exploration and learning for cognitive robots," Ph.D. dissertation, Delft University of Technology, 2013.
9. M. Rudinac and P. P. Jonker, "A fast and robust descriptor for multiple-view object recognition," in *Proceedings of the International Conference on Control Automation Robotics & Vision (ICARCV)*, 2010, pp. 2166–2171.
10. M. Rudinac and P. P. Jonker, "Saliency detection and object localization in indoor environments." in *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE, 2010, pp. 404–407.
11. S. P. Rueda, "A speech-based dialog system for household robot," Master's thesis, Delft University of Technology, 2012.
12. X. Wang, M. Rudinac, and P. P. Jonker, "Robust online segmentation of unknown objects for mobile robots," in *VISAPP*, Rome, 2012.