ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2013

Leon Ziegler, Jens Wittrowski,
Matthias Schöpfer, Frederic Siepmann, Sven Wachsmuth

Faculty of Technology, Bielefeld University,
Universitätstraße 25, 33615 Bielefeld, Germany

Abstract. The Team of Bielefeld (ToBI) has been founded in 2009. The RoboCup activities are embedded in a long-term research history towards human-robot interaction with laypersons in regular home environments. The RoboCup@Home competition is an important benchmark and milestone for the overall research goal. For RoboCup 2013, the team concentrates on an easy to use programming environment, semantically annotated maps and deeper scene analysis via Implicit Shape Models.

1 Introduction

The RoboCup@Home competition aims at bringing robotic platforms to use in domestic environments including human-robot interaction and service robotic tasks. Thus, the robot needs to deal with unprepared real environments, perform autonomously in them and interact with laypersons.

Todays robotic systems obtain a big part of their abilities through the combination of different software components from different research areas. To be able to communicate with humans and interact with the environment, robots do not only need to perceive their surrounding, they also have to interpret the current scene. This ability becomes even more important for more complex scenarios, such as domestic service environments.

Team of Bielefeld (ToBI) has been founded in 2009 and successfully participated in the RoboCup German Open from 2009-2012 as well as the RoboCup World Cup from 2009-2012. The robotic platform and software environment has been developed based on a long history of research in human-robot interaction [1, 2]. The overall research goal is to provide a robot with capabilities that enable interactive teaching of skills and tasks through natural communication in previously unknown environments. The challenge is two-fold. On the one hand, we need to understand the communicative cues of humans and how they interpret robotic behavior [3]. On the other hand, we need to provide technology that is able to perceive the environment, detect and recognize humans, navigate in changing environments, localize and manipulate objects, initiate and understand a spoken dialog and analyse the different scenes to gain a better understanding of the surrounding. Thus, it is important to go beyond simple feature detection in the environment and e.g. recognize different object instances to exploit the

fusion of different information cues to improve the robots behavior in complex environments.

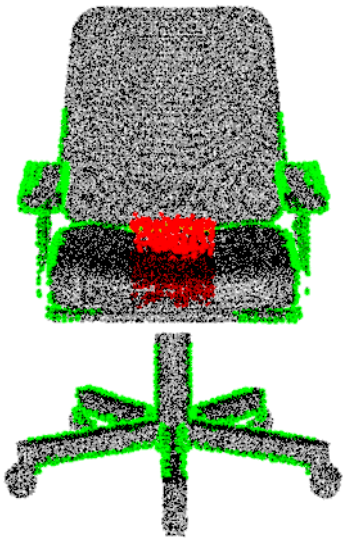

Fig. 1. ISM example of a chair.

In this year's competition, we extend the capabilities for scene-analysis of our robot. Additionally to our spatial attention system [4], we have added an Implicit Shape Model (ISM), which is able to learn spatial relationship of typical object regions from a set of artificial 3D models.

Another focus of the system is to provide an easy to use programming environment for experimentation in short development-evaluation cycles. We further observe a steep learning curve for new team members, which is especially important in the RoboCup@Home context. The developers of team ToBI change every year and are Bachelor or Master students, who are no experts in any specific detail of the robots software components. Therefore, specific tasks and behaviors of the robot need to be easily modeled and flexibly coordinated. In concordance with common robotic terminology we provide a simple interface that is used to model the overall system behavior. To achieve this we provide an abstract sensor- and actuator interface (BonSAI) [4] that encapsulates the sensors, skills and strategies of the system and provides a SCXML-based [5] coordination engine.

2 The ToBI Platform

The robot platform *ToBI* is based on the research platform *GuiaBot*^{TM} by adept/mobilerobots[1] customized and equipped with sensors that allow analysis of the current situation. ToBI is a consequent advancement of the *BIRON* (**BI**elefeld **R**obot compani**ON**) platform, which is continuously developed since 2001 until now. It comprises two piggyback laptops to provide the computational power and to achieve a system running autonomously and in real-time for HRI.

The robot base is a PatrolBot^{TM} which is 59cm in length, 48cm in width, weighs approx. 45 kilograms with batteries. It is maneuverable with 1.7 meters per second maximum translation and 300+ degrees rotation per second. The drive is a two-wheel differential drive with two passive rear casters for balance. Inside the base there is a 180 degree laser range finder with a scanning height of 30cm above the floor (SICK LMS, see Fig.2 bottom right).

[1] www.mobilerobots.com

In contrast to most other PatrolBot bases, ToBI does not use an additional internal computer. The piggyback laptops are Core i7 © (quadcore) processors with 8GB main memory and are running Ubuntu Linux. The cameras that are used for person and object detection/recognition are 2MP CCD firewire cameras (Point Grey Grashopper, see Fig.2).

For room classification, gesture recognition and 3D object recognition ToBI is equipped with an optical imaging system for real time 3D image data acquisition, one facing down (objects) and an additional one facing towards the user/environment.

Additionally the robot is equipped with the Katana IPR 5 degrees-of-freedom (DOF) arm (see Fig.2); a small and lightweight manipulator driven by 6 DC-Motors with integrated digital position encoders. The end-effector is a sensor-gripper with distance and touch sensors (6 inside, 4 outside) allowing to grasp and manipulate objects up to 400 grams throughout the arm's envelope of operation.

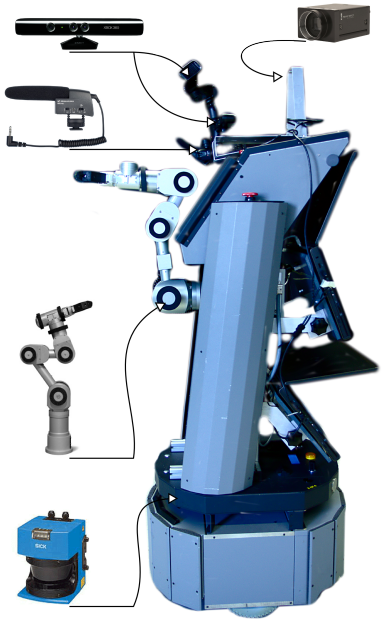

Fig. 2. ToBI with its components: camera, 3D sensors, microphone, KATANA arm and laser scanner.

To improve the control of the arm, the inverse kinematics of the Katana Native Interface (KNI) was reimplemented using the Orocos [6] Kinematics and Dynamics Library (KDL). This allowed further exploitation of the limited workspace compared to the original implementation given by the vendor. This new implementation also enables the user to use primitive simulation of possible trajectories to avoid obstacles or alternative gripper orientations at grasp postures, which is important due to the kinematic constraints of the 5 DoF arm.

The upper part of the robot houses a touch screen ($\approx 15in$) as well as the system speaker. The on board microphone has a hyper-cardioid polar pattern and is mounted on top of the upper part of the robot. The overall height is approximately 140cm.

3 Reusable Behavior Modeling

For modeling the robot behavior in a flexible manner ToBI uses the *BonSAI* framework. It is a domain-specific library that builds up on the concept of *sensors* and *actuators* that allow the linking of perception to action [7]. These are organized into robot *skills* that exploit certain *strategies* for an informed de-

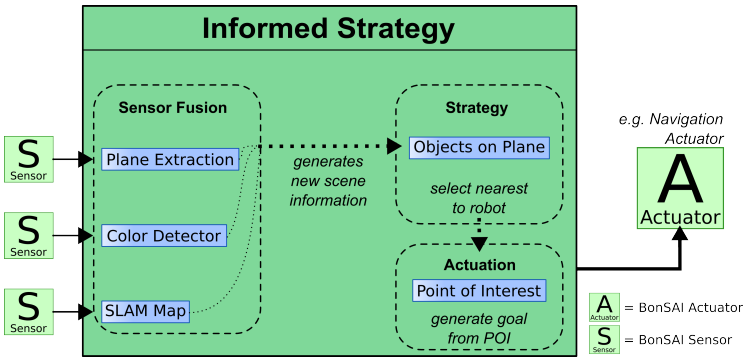

Fig. 3. Schema of an *Informed Strategy*: *Sensor Fusion* on the left generates information for the *Strategy*. The *Actuation* generates, e.g., a goal to which the robot can navigate.

cision making. In the following we will concentrate on two new aspects of the *BonSAI* modeling framework: *Informed strategies* for reusable robot behaviors and the SCXML-based coordination engine. We facilitate BonSAI in different scenarios: It is used for the robot BIRON which serves as a research platform for analyzing human-robot interaction [3] as well as for the RoboCup@Home team ToBI, where mostly unexperienced students need to be able to program complex system behavior of the robot in a short period of time. In both regards, the BonSAI framework has been improved such that system components are further decoupled from behavior programming and the degree of code re-use is increased. This has especially been achieved by the introduction of *strategies* and *State-Charts*.

3.1 Informed Strategies

The construct within the BonSAI framework that can be used to affect the robots behavior to on the one hand enhance the re-usability of the code and on the other hand accomplish an enriched interpretation of the scene, is depicted in Fig. 3: The *Informed Strategies*.

In BonSAI such a *strategy* only makes use of available *sensors* and produces an output that can be used by one specific *actuator* of the framework. This means that one certain way of processing information, possibly in software components from layers underneath the BonSAI layer, is modeled through a *strategy*. This allows to reuse this *strategy* at different points in one skill or in different skills and react to unexpected situations during the processing. Assuming one of the sensors does not provide correct data, the *strategy* may detect an error and the behavior can react to that, e.g. by trying another *strategy*. With behavior code enclosed in the software components, the processing would fail leaving no chance to react to it.

```
<?xml version="1.0" encoding="UTF-8"?>
<scxml xmlns="http://www.w3.org/2005/07/scxml" version="1.0" initial="ready">
  <state id="ready">
    <transition event="watch.start" target="running"/>
  </state>
  <state id="running">
    <transition event="watch.split" target="paused"/>
    <transition event="watch.stop" target="stopped"/>
  </state>
  <state id="paused">
    <transition event="watch.unsplit" target="running"/>
    <transition event="watch.stop" target="stopped"/>
  </state>
  <state id="stopped">
    <transition event="watch.reset" target="ready"/>
  </state>
</scxml>
```

Fig. 4. SCXML example of a stop watch.

3.2 SCXML-based Coordination Engine

To support the easy construction of more complex robot behavior we have improved the control level abstraction of the framework. BonSAI now supports modeling of the control-flow, as e.g. proposed by Boren [8], using State Chart XML (see Fig. 4, taken from [2]). The coordination engine serves as a sequencer for the overall system by executing *BonSAI skills* to construct the desired robot behavior. This allows to separate the execution of the skills from the data structures they facilitate thus increasing the re-usability of the skills. The BonSAI framework has been released under an Open Source License and is available under `http://opensource.cit-ec.de/projects/bonsai`.

4 Spatial Awareness

Our robot builds up different kinds of spatial representations of its environment using 2D and 3D sensors. This improves the robot's situation awareness and supports its searching abilities.

4.1 Semantic Map Annotation

In order to improve the effectiveness of search tasks, the robot performs a scene analysis of its environment and builds up a 2D representation of the possibly most interesting regions. The basis for the semantically annotated map is an occupancy grid representing the spatial structure of the environment generated by a SLAM implementation [9]. This map contains only physical obstacles that can be detected by the laser range finder, such as walls and furniture. Additional grid map layers on top of the SLAM obstacle map are introduced by our "Semantic Annotation Mapping" approach (SeAM) to encode the low-level visual

[2] `http://commons.apache.org/scxml`

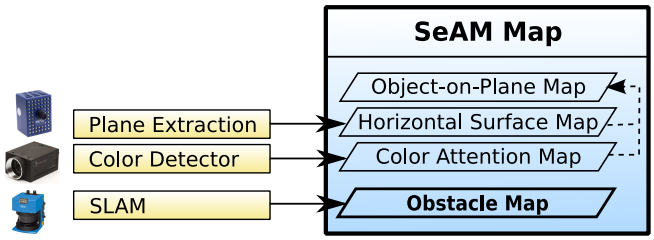

Fig. 5. Layout of the SeAM map.

cues calculated while the robot explores its environment (see Fig. 5). Hence, the combination of these information can be considered as a mechanism for mapping spatial attention that constantly runs as a subconscious background process.

In order to establish the attention map potential object positions are detected within the robots visual field by using simple and computationally efficient visual features. Additionally we detect horizontal surfaces in the perceived environment, because potential target objects of a search are most probably sitting on such a surface.

4.2 Spatial Mapping

In order to register information-rich regions into the grid maps, the visual information need to be spatially estimated relatively to the robot's current position. The actual mapping of the found regions is done by raising or lowering the cell values of the corresponding layer in the SeAM map. The encoding is similar to the representation of the SLAM results.

Because of the layer structure of the grid maps representing the same spatial area, information from multiple layers can be fused to generate more sophisticated data. We introduce an additional grid map layer that fuses information from the color detector and the horizontal surface detector. Semantically this map represents object hypotheses on horizontal surfaces above the floor (*object-on-plane* map). The probabilities are only raised if both detectors vote for the same cell. More details can be found in [4].

5 Implicit Shape Models

For robots acting in domestic environments especially the correct recognition of furniture objects is a very important task. The applied approach in [10] uses the idea of Implicit Shape Models to model furniture categories, together with a 3-dimensional Hough-voting mechanism to detect object instances in scenes captured with a 3D camera. An Implicit Shape Model (ISM) learns the spatial relationship of typical object regions from a set of artificial 3D models. It consists of a codebook and a list of vote vectors linked to each codebook entry. To generate the codebook, keypoints in the training models are detected and described using a SHOT descriptor [11]. The codewords are then extracted by running a k-means clustering algorithm over the whole set of point descriptions.

To learn the spatial relationship of these identified regions the ISM stores descriptors for the appearance of these object regions in relation to a unique object reference point. Therefor each detected keypoint in the training models is matched against the generated codebook. For each matching codeword a 3D vector is added describing the relationship between the detected point an the object centroid.

To detect object instances of the learned category in test scenes captured with a 3D camera a probabilistic Hough voting is performed. This enables the vision system to simultaneously recognize and localize object instances. The procedure is divided into a hypotheses generation and a hypotheses selection step. For hypotheses generation keypoints with corresponding descriptors are calculated. Then the codewords are matched using a codeword activation strategy (k-Nearest Neighbour or Euclidean distance threshold). For each activated codeword the list of vote vectors is received. Each received vector casts a vote for the position of the object centroid. The existence of local maxima in the vote space gives good hypotheses for object centers of the requested object category. To further verify the existence of an object instance found in the scene, the hypotheses selection step is performed. Therefor the spatial layout of the keypoints corresponding to the descriptors that voted for the chosen hypothesis is validated. Figure 6 shows the detection of a dining chair by the described Hough-space voting approach.

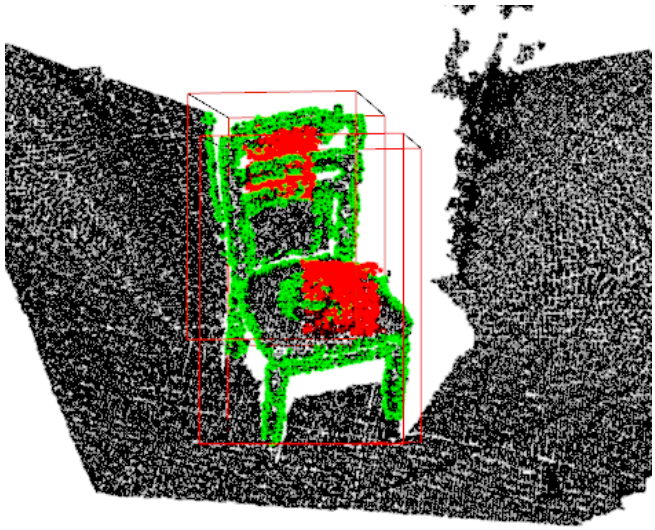

Fig. 6. Hough space voting: Red dots show a local maximum of votes, green dots show the points that casted their votes into that maximum

6 Conclusion

We have described the main features of the ToBI system for RoboCup 2013 including sophisticated approaches for semantic map annotation and 3D object recognition. BonSAI represents a flexible rapid prototyping environment, providing capabilities of robotic systems by defining a set of essential skills for such systems.

Especially BonSAI with its abstraction of the robot skills proved to be very effective for designing determined tasks, including more script-like tasks, e.g. *Follow Me* or *Who is Who*, as well as more flexible tasks including planning and dialog aspects, e.g. *General Purpose Service Robot*. By deploying the mapping and recognition capabilities, we hope to improve ToBI's performance in searching tasks like *Clean Up* and also to show that it improves the robot's orientation in unknown environments.

References

1. Wrede, B., Kleinehagenbrock, M., Fritsch, J.: Towards an integrated robotic system for interactive learning in a social context. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems - IROS 2006, Bejing (2006)
2. Hanheide, M., Sagerer, G.: Active memory-based interaction strategies for learning-enabling behaviors. In: International Symposium on Robot and Human Interactive Communication (RO-MAN), Munich (2008)
3. Lohse, M., Hanheide, M., Rohlfing, K., Sagerer, G.: Systemic Interaction Analysis (SInA) in HRI. In: Conference on Human-Robot Interaction (HRI), San Diego, CA, USA, IEEE (2009)
4. Siepmann, F., Ziegler, L., Kortkamp, M., Wachsmuth, S.: Deploying a modeling framework for reusable robot behavior to enable informed strategies for domestic service robots. Robotics and Autonomous Systems (2012)
5. Barnett, J., Akolkar, R., Auburn, R., Bodell, M., Burnett, D., Carter, J., McGlashan, S., Lager, T.: State chart xml (scxml): State machine notation for control abstraction. W3C Working Draft (2007)
6. Bruyninckx, H.: Open robot control software: the OROCOS project. In: Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation, IEEE (2001) 2523–2528
7. Siepmann, F., Wachsmuth, S.: A Modeling Framework for Reusable Social Behavior. In De Silva, R., Reidsma, D., eds.: Work in Progress Workshop Proceedings ICSR 2011, Amsterdam, Springer (2011) 93–96
8. Boren, J., Cousins, S.: The smach high-level executive. Robotics & Automation Magazine, IEEE **17**(4) (2010) 18–20
9. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B.: FastSLAM 2.0: an improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In: Proceedings of 18th IJCAI, Acapulco, Mexico (2003) 1151–1156
10. Wittrowski, J.: Furniture recognition using implicit shape models on 3d data. Masters Thesis, Faculty of Technology, Bielefeld University (2012)
11. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III. ECCV'10, Berlin, Heidelberg, Springer-Verlag (2010) 356–369