ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2012

Sven Wachsmuth, Frederic Siepmann, Leon Ziegler,
Florian Lier, Matthias Schöpfer

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 2012, the team concentrates on mixed-initiative scenarios, a generic interaction-pattern based dialog, an easy to use programming environment and semantically annotated maps.

1 Introduction

Todays robotic systems obtain a big part of their abilities through the combination of different software components from various 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.

The RoboCup@Home competition aims at bringing robotic platforms to use in these kinds of environments: Realistic home scenarios. Here the robot needs to deal with unprepared domestic environments, perform autonomously in them and interact with laypersons. Team of Bielefeld (ToBI) has been founded in 2009 and successfully participated in the RoboCup German Open from 2009-2011 as well as the RoboCup World Cup from 2009-2011. The robotic platform and software environment has been developed based on a long history of research in human-robot interaction [1–3]. 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 [4]. 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. Thus, it is important to go beyond typical command-style interaction and to support mixed-initiative learning tasks. In the ToBI system this is managed by a sophisticated dialog model that enables flexible dialog structures [5].

In this year's competition, we extend the dialog capabilities of our robot. While current techniques for human-robot interaction modeling are typically limited to restrictive command-control style, traditional dialog modeling approaches are not directly applicable to robotics due to the lack of real-world integration. Our approach combines insights from dialog modeling with software engineering demands that arise in robotics system research to provide a generalizable framework that can easily be applied to new scenarios. This goal is achieved by defining interaction patterns that combine abstract task states with robot dialog acts (e.g. *assertion* or *apology*) [6].

Furthermore we have extended the scene-analysis functionality of our robot. We apply a spatial attention system that uses different visual cues which are mapped in a SLAM-like manner in order to identify hypotheses for possible object locations [3]. In order to improve the construction of an accurate semantic 3D model of the indoor scene, we exploit human-produced verbal descriptions of the relative location of pairs of objects.

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) that encapsulates the sensors, skills and strategies of the system and provides a simple SCXML-based [7] coordination interface.

2 The ToBI Platform

The robot platform *ToBI* is based on the research platform *GuiaBot*$^{\mathrm{TM}}$ by 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$^{\mathrm{TM}}$ 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.1 bottom right). In contrast to most other PatrolBot bases, ToBI does not use an additional internal computer.

[1] www.mobilerobots.com

The piggyback laptops are Core2Duo © processors with 2GB 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.1).

One is facing down for object detection/recognition, the second camera is facing up for face detection/recognition. For room classification and 3D object positions ToBI is equipped with an optical imaging system for real time 3D image data acquisition (Kinect).

Additionally the robot is equipped with a Katana IPR 5 degrees-of-freedom (DOF) arm (see Fig.1 second from bottom on the right); 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. The upper part of the robot houses a touch screen (≈ $15in$) 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.

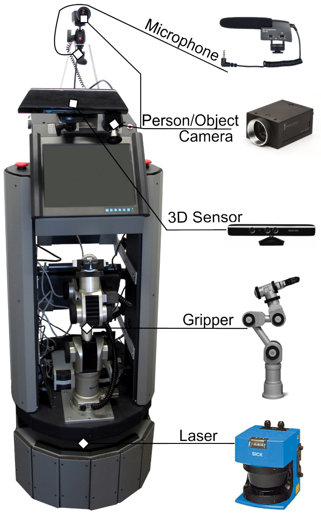


Fig. 1. ToBI with components on the right: microphone, cameras, Kinect$^{\mathrm{TM}}$, KATANA arm and laser scanner.

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 [8]. These are organized into robot *skills* that exploit certain *strategies* for an informed decision 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 [4] 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

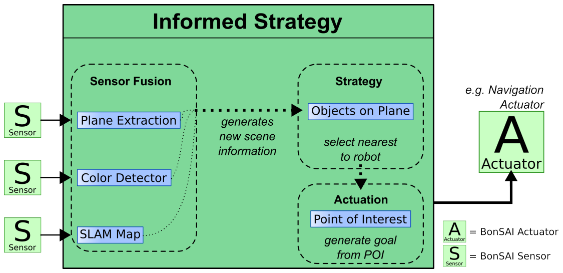


Fig. 2. Schema of an *Informed Strategy*: *Sensor Fusion* on the left generates information for the *Strategy*. The *Actuation* generates, e.g., a goal to which the robot can navigate.

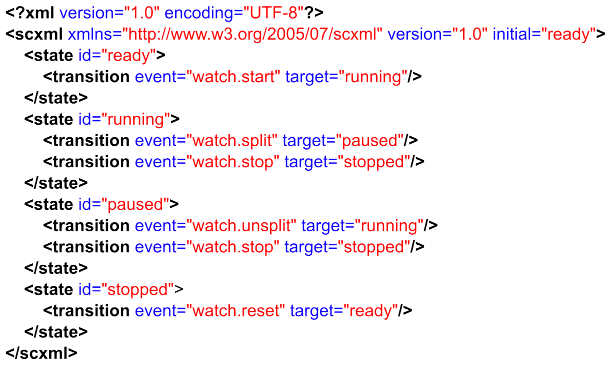
The construct within the BonSAI framework that can be used to affect the robots behavior to on the one hand enhance the re-usability of the code and on the other hand accomplish an enriched interpretation of the scene, is depicted in Fig. 2: The *Informed Strategies*.

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.

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 [9], using State Chart XML (see Fig. 3, 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`.

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



```
<?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. 3. SCXML example of a stop watch.

4 Dialog: The Pamini Framework

For modeling the dialog in more complex interaction scenarios, ToBI facilitates the Pamini framework [6], that especially accounts for maintainability and reusability of the dialog by using generic interaction patterns that support rapid prototyping of human-robot interactions. This enables us to flexibly combine different dialog acts to solve more complex dialog scenarios.

4.1 Dialog Manager and Interaction Patterns

The *dialog manager* offers dialog tasks for other components, e.g. greeting the human, informing the human about tasks of the robot or requesting (new) information from the human. The dialog manager also takes care of managing sub-dialogs to e.g. request missing information from the user. During an interaction dialog acts are not unrelated events, but form coherent sequences, e.g. a question normally is followed by an answer. Hence the *interaction patterns* describe recurring dialog structures on a high level. During the interaction, *interaction patterns* are employed in a flexible way by admitting patterns to be interrupted by other patterns and possibly be resumed later, enabling interleaving patterns. More simple patterns, e.g. greeting the user, are permitted to be nested within temporally extended patterns. The concept of interaction patterns constitutes configurable (and thus reusable) building blocks of interaction. With interleaving patterns, flexible dialog modeling is achieved on robotic platforms. Further, the dialog manager reduces the complexity of component integration and the efforts for task programmers dealing with the robots behavior.

5 Spatial Awareness

ToBI 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.

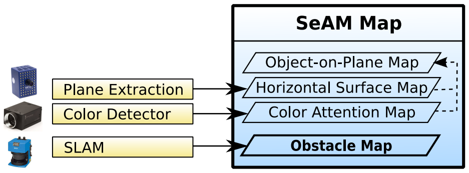


Fig. 4. Layout of the SeAM map.

5.1 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 [10]. 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 cues calculated while the robot explores its environment (see Fig. 4). These overlays are used for a more detailed analysis later on. 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 the case of *lost-and-found* tasks, the annotation component relies on two low-level visual cues to establish the attention map. At first, potential object positions are detected within the robots visual field by using simple and computationally efficient visual features. The learned color distribution of a target object is the *rg chromaticity space* [11] histogram of the colorful regions in the training set. For detection the current camera image is scanned with a sliding window approach on different scales. The calculated histograms are then compared with the codebook of one of the known objects using a histogram intersection algorithm [12].

Additionally we detect horizontal surfaces in the perceived environment, because potential target objects of a search are most probably sitting on such a surface. The detector uses implementations of the *Point Cloud Library* (PCL)[3] and works on the 3D point clouds from a 3D sensor.

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 3D plane description can be easily transformed into a 2D aerial view representation. In case of the color distribution cue, the direction of the detected location can be calculated using several facts about the camera’s properties like FoV and resolution, as well as how it is mounted on the robot.

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

[3] http://www.pointclouds.org/

representation of the SLAM results. While values near 0.5 mean unknown area, higher values represent free space and lower values stand for detected attention regions (corresponding to obstacles in SLAM).

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 [3].

5.2 Scene Segmentation And Grounding

For orientation in a complex indoor environment the robot must have an internal representation of rooms and prominent locations in them. This is crucial for successful communication and for fulfilling complex tasks that involve location changes. This kind of model can not only evolve from visual observation but also from spatial descriptions in human utterances during a conversation. We try to combine both sources in a system that generates spatial relations of furniture in the robot's environment.

In the implemented system, the visual scene is initially scanned by a 3D camera system. From the 3D point cloud, we extract grouped planar patches that suggest the presence of certain furniture objects. The correspondingly clustered sub-cloud is used to guess an initial probability distribution for the furniture category of the perceived object.

Further we developed a computational model that extracts partial orderings of spatial arrangements between furniture items from verbal descriptions. This model has to deal with ambiguities in the reference of objects in natural language and has to have knowledge about the distribution of different spatial "frames of reference" (FOR) that humans use interchangeably for furniture-predicate combinations.

We then integrate the partial orderings extracted from the verbal utterances incrementally and cumulatively with the estimated probabilities about the identity and location of objects in the scene, and also estimate the probable orientation of the objects. This allows to improve both the accuracy and richness of the visual scene representation significantly.

6 Conclusion

We have described the main features of the ToBI system for RoboCup 2012 including sophisticated approaches for dialog management and semantic map annotation. BonSAI represents a flexible rapid prototyping environment, providing capabilities of robotic systems by defining a set of essential skills for such systems. The RoboCup@HOME competitions in 2009 to 2011 served for as a continuous benchmark of the newly adapted platform and software framework (achieved

8th, 7th, and 5th place). 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' or 'Open-Challenge'. We are confident that the newly introduced features and capabilities will further improve the overall system performance. By deploying the new mapping capabilities, we hope to improve ToBI's performance in searching tasks like "Go Get It" and also to show that it improves the robot's orientation in unknown environments. The latter espacially applies for the combination of the dialog system and grounding of locations in the environment, too.

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 (01/08/2008 2008)
3. Ziegler, L., Siepmann, F., Kortkamp, M., Wachsmuth, S.: Towards an informed search behavior for domestic robots. In: Domestic Service Robots in the Real World. (2010)
4. 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 (11/03/2009 2009)
5. Peltason, J., Wrede, B.: Modeling human-robot interaction based on generic interaction patterns. In: AAAI Fall Symposium: Dialog with Robots, Arlington, VA, USA, AAAI Press (11/11/10 2010)
6. Peltason, J., Wrede, B.: Pamini: A framework for assembling mixed-initiative human-robot interaction from generic interaction patterns. In: SIGDIAL 2010 Conference, Tokyo, Japan, Association for Computational Linguistics (24/09/10 2010)
7. 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)
8. 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
9. Boren, J., Cousins, S.: The smach high-level executive. Robotics & Automation Magazine, IEEE **17**(4) (2010) 18–20
10. 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 the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI), Acapulco, Mexico, IJCAI (2003)
11. Caetano, T., Olabarriaga, S., Barone, D.: Do mixture models in chromaticity space improve skin detection? Pattern Recognition **36**(12) (12 2003) 3019–3021
12. Swain, M., Ballard, D.: Indexing via color histograms. In: Computer Vision, 1990. Proceedings, Third International Conference on. (dec 1990) 390–393