ToBI - Team of Bielefeld
A Human-Robot Interaction System for
RoboCup@Home 2024

Leroy Rügemer, David Leins, and Sven Wachsmuth

Research Center for Cognitive Interaction Technology (CITEC),
Bielefeld University, Inspiration 1, 33615 Bielefeld, Germany
`https://clf.pages.ub.uni-bielefeld.de/Robocup/website/`

Abstract. Team of Bielefeld (ToBI) was founded in 2009. The RoboCup team's activities are embedded in a long-term research agenda towards human-robot interaction with laypersons in regular and smart home environments. The RoboCup@Home competition is an important benchmark and milestone for this goal in terms of robot capabilities as well as the system engineering approach. In order to achieve a robust and stable system performance, we apply a methodical approach for reproducible robotic experimentation including automated tests. Another focus is reusability which means both, the reuse in different robot tasks and the reuse across different platforms. For RoboCup 2024, we introduce several changes in our speech, dialog and behavior systems, improvements for our world model, as well as a new simulation framework with realistic contact physics.

1 Introduction

The RoboCup@Home competition aims at bringing robotic platforms to use in realistic domestic environments. Today's robotic systems obtain a big part of their abilities through the combination of different software components from different research areas. Robots need to coordinate and dynamically configure their components to generate an appropriate overall robot behavior that fulfills parallel goals such as gathering scene information, achieving a task goal, communicating its internal status, and being always responsive to humans. This is especially relevant for complex scenarios in domestic settings which involve communication with humans and interaction with the environment.

ToBI was founded in 2009 and successfully participated in the RoboCup World Cup (2009-2018) and 2023 as well as the RoboCup German Open from 2009 to 2019 on different robot platforms within RoboCup@Home. Each year a team of students has been newly formed and supervised by senior researchers and student members of the previous years.

In 2016, the team won the global @Home-competition [2] and finished first in several of the individual tests (Navigation, Person Recognition, GPSR, EE-GPSR, Restaurant). At RoboCup 2018, the team achieved first place in the Social

Standard Platform League (SSPL) [8]. Although the team newly acquired the
TIAGo platform late in 2018, the Team placed 1st at the 2019 RoboCup@Home
German Open. After the COVID-induced break, ToBI rejoined the competition
for the 2023 RoboCup with a 5th place finish. Over the years, multiple team
members have been actively engaged in comittee work.

Bielefeld University has been involved in research on human-robot interaction
for more than 20 years, especially gaining experience in experimental studies with
integrated robotic systems [10,14]. Within this research, strategies are utilized
for guiding the focus of attention [12] as well as behaviors for object reference
and object handovers [14].

The introduction of a systematic approach towards reproducible robotic
experiments [7] has been turned out as a key factor to maximally stabilize
basic capabilities like, e.g., navigation or person following. Further aspects —
regarding the RoboCup@Home — refer to the usage of systematic world models
that introduce knowledge to the robotic system on how to interact with its
environment. Based on these models, capabilities are needed to re-configure the
navigation and grasping skills on the fly considering environmental factors. This
is implemented within our system as a navigation plug-in and through the usage
of the Task Constructor framework for `MoveIt!` [4].

2   Robot Platforms

In 2016, ToBI participated in RoboCup@Home with the two service robots Biron
and Floka, in 2017 with Biron and Pepper. Although focusing on the TIAGo
since 2019, we still aim at the development of platform-independent, as well as
multi-platform robot capabilities.

The TIAGo Platform (Fig. 4) from PAL Robotics is a service robot for indoor
environments. The technical specifications can be found online[1]. We equipped
the robot with a Lenovo P15 Gen 2 notebook running Ubuntu Linux 22.04 on
a 3D-printed mount. Additional computing power is provided by an NVIDIA
Jetson AGX ORIN. For improved recognition, we added a Sennheiser MKE 400
shotgun microphone along with a RealSense L515 LiDAR Camera to the head.
Our version of the robot is equipped with a Schunk WSG32 gripper with a
maximum opening width of 68mm. As this is not enough to grasp the objects
utilized in the competition, we augmented the robot with a custom expansion.
This non-invasive add-on turns the parallel into a scissoring motion to increase the
maximum opening width. The gripper is designed to mount up to four adaptive
fingers.

3   System Architecture

Our service robots employ distributed systems with multiple clients sharing
information over the network. On these clients, there are numerous software

---

[1] http://pal-robotics.com/wp-content/uploads/2019/07/Datasheet_TIAGo_
Complete.pdf

components written in different programming languages. Such heterogeneous systems require abstraction on several levels.

3.1   Reusable Behavior Modeling

For modeling the robot behavior flexibly, 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 [13]. These are organized into robot *skills* that exploit certain *strategies* for informed decision-making.

To support the easy construction of more complex robot behavior, BonSAI supports modeling of the control flow, as e.g. proposed by [1]. The BonSAI `behavior-scxml` extension implements this modeling using State Chart XML (SCXML). It serves as a sequencer for the overall system by executing *skills* to construct the desired robot behavior. This allows separating execution of the skills from the data structures they facilitate, thus increasing the re-usability of the skills. Through the usage of BonSAI, the same skill could also be reused on different systems that use XCF, ROS1, ROS2, RSB or NaoQi as middleware.

The concurrent support for ROS1 and ROS2 allows the transition of individual components in a hybrid configuration. The BonSAI framework has been released under an Open Source License and is available online[2].

3.2   Development and Deployment Tool-Chain

In order to foster reproducibility, traceability, and potential software (component) re-use of the ToBI system, we provide a complete specification of the 2024 system in our online catalog platform[3]. The catalog provides detailed information about the soft- and hardware system, including all utilized software components. [7]. Our development and deployment process is illustrated in Fig. 1 (red numbers) [9].

Since setting up a CI server and the required plugins takes time and requires expert knowledge, we provide prepackaged installations in the [Research & Robotics] Development Toolkit (RDTK)[4]. The RDTK further allows deploying RDTK-based systems using Linux containers, like Docker. System descriptions and their metadata, e.g. source code locations, wiki pages, issue tracker, current build status, experiment descriptions, and so forth are frequently synchronized to a web-based catalog Cognitive Interaction Toolkit Catalogue (CITKat) that also implements the RDTK data model providing a global, human-readable, and search-able platform which is a prerequisite for open research.
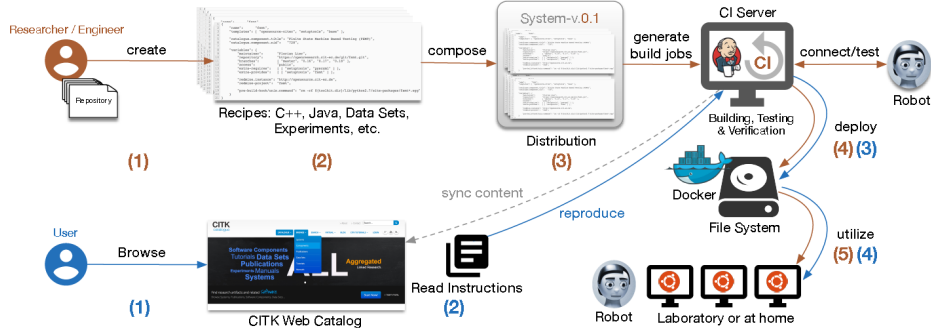
4   Entity Component World Model

To facilitate sophisticated reasoning about the environment, a representation of the world is needed. The Entity component World Model (ECWM) framework

---

[2] https://github.com/CentralLabFacilities/bonsai
[3] https://citkat-citec.bob.ci.cit-ec.net/distribution/tiago-jammy-one-nightly.xml
[4] https://rdtk.github.io/documentation/

Fig. 1. Cognitive Interaction Toolkit: tool chain and workflow. The red numbers show the workflow of the system developer, while the blue numbers represent the workflow of a researcher reproducing the system.

provides an Entity component system (ECS) based World model for quick prototyping. Known information is encoded as different `components` in an entity pool of `types`. Most of these are geometry-based to describe rigid objects, such as furniture, with basic primitives or 3D meshes and more advanced semantic information, such as descriptions for named regions, e.g. the different shelves of a bookcase. Some `components`, such as point clouds and egocentric maps, providing suitable positions for interactions, can be calculated offline or during startup by utilizing the `plugin` system. All `type components` are static during the runtime of the robot task.

The world state comprises the instances of the `models` in the current `entity` pool, where each `entity` contains `components`, e.g. position, observations, themselves. ECWM provides the world as a list of entities and the ability to add, remove and update as well as query the associated (`model`)`components` of entities via ROS services. Additionally, it also provides interfaces for runtime plugins to collect sensor information and update `entity` properties such as the current position. As an example, the `icp-matcher` plugin associates robot sensor data with the geometries enabling the position refinement with Iterative Closest Point (ICP).

Interaction Maps are model components that map robot positions to robot poses. Each interaction map represents an arbitrarily sized 2D map centered at the object with action-dependent size and resolution. Each cell of the map has a value for suitability and the desired robot joint configuration. The `geometries` can have multiple interaction maps for different actions, like *recognize objects on the lower shelf* or *grasp from upper shelf* as shown in Fig. 2. As the capabilities of robots can be different, the maps are specifically generated based on the description of the robot.

The *recognize objects* interaction, for a specific planar surface of the `geometry`, calculates suitability scores by computing how much of the surface is visible for

(a) View the bookcase   (b) Valid camera poses for interaction   (c) Grasp from the upper shelf
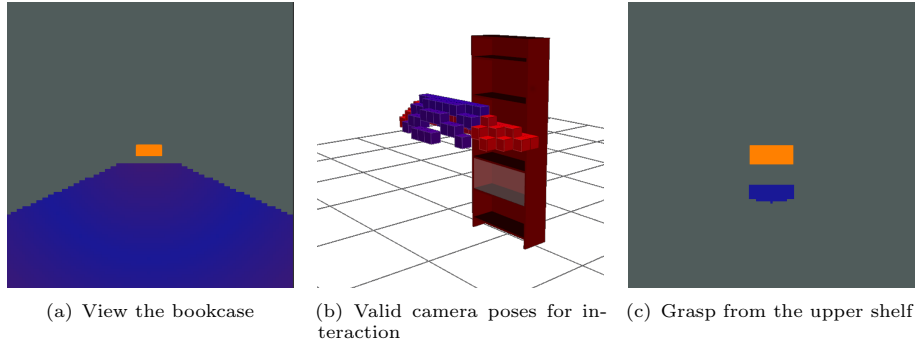
Fig. 2. Visualization of different interaction maps for the bookshelf (in orange). The blue to red shading encodes the suitability of the cell, other space is insufficient for the interaction

different camera sensor positions using ray casting. This reduces the occlusion of objects during robot operation. It also applies a factor calculated from measuring the deviation to an optimal view angle and distance to reproduce the settings applied during collection of the object recognition training data.

Even slight position changes may take a considerable amount of time, especially for robots with differential drives. Thus, different maps can be combined to find a valid position for multiple interactions, possibly minimizing the required robot movement.

5   Further Software Contributions

Speech and Dialog Our former dialogue system facilitated good performance across diverse competition tasks since 2009. However, it fails to match the performance of state-of-the-art research. Thus, we increased the compute-capabilities of our robot with an NVIDIA Jetson which enables us to utilize more complex algorithms for different robot capabilities. The increased computation capabilities also enable us to explore features such as speaker recognition and addressee detection. A new speech recognition system based on `whisper` [11] was developed as part of these advancements. We have made our implementation available on GitHub[5]. Looking ahead to the forthcoming RoboCup@Home 2024 competition, other parts of the dialogue management system will undergo redesign, wherein the integration of `flexdiam` [16] concepts will be implemented to enhance responsiveness throughout interactions.

Navigation To improve navigation capabilities for our different platforms, we maintained a fork of the Robot Operating System (ROS) `navigation`[6] packages.

---

[5] https://github.com/CentralLabFacilities/speech_recognition
[6] https://github.com/CentralLabFacilities/navigation

(a) Object Recognition visualization   (b) Grasp Task
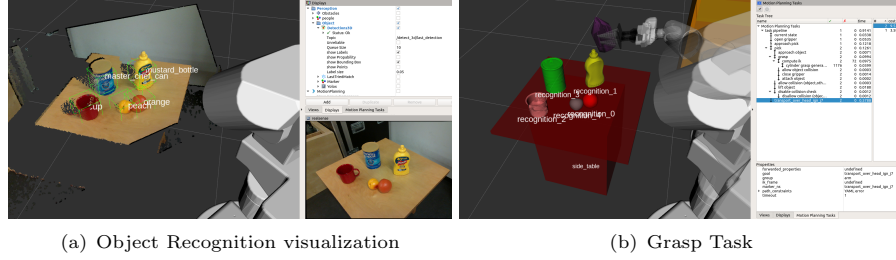
Fig. 3. Object Recognition visualization and resulting object poses for grasping

To facilitate the reuse of our changes, we decided to repackage the changes into different plugins in a separate package[7]. As navigation tasks differ in their requirements for path planning and obstacle avoidance, our navigation stack is able to select different configurations for each navigation goal. While we currently are in the process of porting all functionalities to ROS2 `navigation2` we utilize our BonSAI framework to choose the navigation framework depending on the task.

Object Recognition Our object recognition is based on `YoloX` [3] and utilizes the ROS `vision_msgs` package. Figure 3(a) shows our visualization plugin for `vision_msgs`.

We improve the 2D recognition results with 3D segmentation and fitting of known object geometries from ECWM. The `clf_object_recognition` packages implement the described functionality and are hosted on our GitHub[8].

Manipulation TIAGo utilizes the Task Constructor framework for `MoveIt!` which provides a way to solve manipulation tasks by defining multiple interdependent subtasks organized in hierarchical containers, allowing for sequential as well as parallel compositions. The framework enables us to work on the improvements of subtasks. During manipulation, grasps are selected by specific quality metrics depending on the information of the known model or the inferred primitive shape of the object obtained from the perception pipeline. The hierarchy of the `simple grasp` task can be seen in Fig. 3(b).

Person Perception Pose detection is done with `OpenPose`. We build a customized ROS wrapper[9] that can compute gestures (pointing, waving, raising arms) and posture (sitting, standing, lying) of detected people as well as their position using the RGBD camera of TIAGo.

---

[7] https://github.com/CentralLabFacilities/clf_navigation_plugins
[8] https://github.com/CentralLabFacilities/clf_object_recognition
[9] https://github.com/CentralLabFacilities/openpose_ros

(a) Tiago in MuJoCo   (b) Person tracking
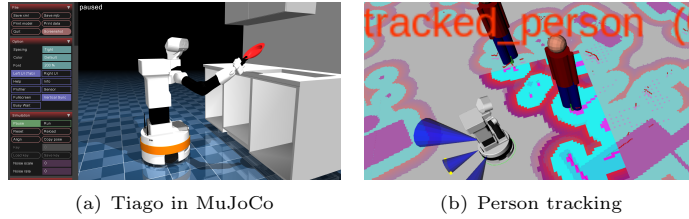
Simulation We found the current standard of simulating ROS-enabled robots in Gazebo to be too inaccurate and complex to extend. Thus, we developed `mujoco_ros`, an open source[10] ROS wrapped simulation environment based on the MuJoCo physics engine [15]. MuJoCo's constraint-based solver approach enables the use of soft body contacts, which yield a more realistic simulation of contact physics compared to rigid body contacts. For easy and flexible extension of the simulation with custom plugins `mujoco_ros` integrates ROS `pluginlib` and provides a set of standard plugins for robot navigation and perception. The framework has already been successfully used for realistic tactile sensor simulation [6] and reinforcement learning [5].

In the future, we plan to integrate HRI via the simulator's VR and motion capture integration capabilities and port the automated benchmarking from Gazebo.

6   Conclusion

We have described the main features of the architecture and technical solutions of the ToBI system for the RoboCup@Home Open Platform League (OPL) 2024. Based on the already achieved development state and an analysis of the robot's performance at previous competitions, we improved the software architecture and development cycle in several aspects. The architecture allows to program and use robot skills across multiple ecosystems on both, internal and external computing resources of the robot. The incremental system development stages are completely reproducible by using the RDTK environment. We are confident to further improve the capabilities of the TIAGo robot to a significant degree utilizing the experiences from very successful RoboCup@Home competitions from 2009 to 2023.

References

1. Boren, J., Cousins, S.: The smach high-level executive. Robotics & Automation Magazine, IEEE **17**(4), 18–20 (2010)
2. Meyer zu Borgsen, S., Korthals, T., Lier, F., Wachsmuth, S.: ToBI – Team of Bielefeld: Enhancing Robot Behaviors and the Role of Multi-Robotics in RoboCup@Home, vol. 9776. Springer (2016)

---

[10] https://github.com/ubi-agni/mujoco_ros_pkgs

3. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: exceeding YOLO series in 2021. CoRR **abs/2107.08430** (2021), https://arxiv.org/abs/2107.08430

4. Görner, M., Haschke, R., Ritter, H., Zhang, J.: Moveit! task constructor for task-level motion planning. 2019 International Conference on Robotics and Automation (ICRA) pp. 190–196 (2019)

5. Härtel, M., Leins, D., Lach, L.M., Haschke, R. (eds.): Learning safe placement of objects based on tactile feedback (2023), https://nbn-resolving.org/urn:nbn:de:0070-pub-29838801,https://pub.uni-bielefeld.de/record/2983880

6. Leins, D., Patzelt, F., Haschke, R.: More accurate tactile sensor simulation with hydroelastic contacts in mujoco. In: IROS 2023 Workshop on Leveraging Models for Contact-Rich Manipulation (2023), https://openreview.net/forum?id=gCdemyFIVe

7. Lier, F., Hanheide, M., Natale, L., Schulz, S., Weisz, J., Wachsmuth, S., Wrede, S.: Towards Automated System and Experiment Reproduction in Robotics. In: Burgard, W. (ed.) 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2016)

8. Lier, F., Kummert, J., Renner, P., Wachsmuth, S.: ToBI - Team of Bielefeld Enhancing the Robot Capabilities of the Social Standard Platform Pepper, vol. 11374, pp. 524–535. Springer (2019). https://doi.org/10.1007/978-3-030-27544-0_43, https://pub.uni-bielefeld.de/record/2940019

9. Lier, F., Wienke, J., Nordmann, A., Wachsmuth, S., Wrede, S.: The Cognitive Interaction Toolkit – Improving Reproducibility of Robotic Systems Experiments. In: Brugali, D. (ed.) SIMPAR 2014, LNAI. pp. 400–411. Springer International Publishing Switzerland (2014)

10. Lohse, M., Siepmann, F., Wachsmuth, S.: A Modeling Framework for User-Driven Iterative Design of Autonomous Systems. International Journal of Social Robotics **6**(1), 121–139 (2014). https://doi.org/10.1007/s12369-013-0209-8

11. Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I.: Robust speech recognition via large-scale weak supervision (2022)

12. Richter, V., Carlmeyer, B., Lier, F., Meyer zu Borgsen, S., Kummert, F., Wachsmuth, S., Wrede, B.: Are you talking to me? Improving the robustness of dialogue systems in a multi party HRI scenario by incorporating gaze direction and lip movement of attendees. In: Proceedings of the Fourth International Conference on Human-agent Interaction. ACM Digital Library (2016)

13. 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. pp. 93–96. Springer, Amsterdam (2011)

14. Simmering, J., Meyer zu Borgsen, S., Wachsmuth, S., Al-Hamadi, A.: Combining static and dynamic predictions of transfer points for human initiated handovers. In: ICSR-2019 (2019)

15. Todorov, E., Erez, T., Tassa, Y.: Mujoco: A physics engine for model-based control. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 5026–5033. IEEE (2012). https://doi.org/10.1109/IROS.2012.6386109

16. Yaghoubzadeh, R., Kopp, S.: Enabling robust and fluid spoken dialogue with cognitively impaired users. In: Jokinen, K., Stede, M., DeVault, D., Louis, A. (eds.) Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. pp. 273–283. Association for Computational Linguistics, Saarbrücken, Germany (Aug 2017). https://doi.org/10.18653/v1/W17-5533, https://aclanthology.org/W17-5533

Robot Description

Our TIAGo specifications are as follows:

- Base: Differential drive 1m/s max speed.
- Torso: 35cm lift.
- Arm: 7DOF, 87cm reach, Maximum load: 3kg.
- Head: Pan and tilt motion with an RGB-D camera.
- Dimensions: height: 110-145cm, footprint: ø54cm.
- Robot weight: 70kg.
- 25m Laser Range, 3×1m sonar.

  Additional Hardware:

- Schunk WSG32 gripper with a custom extension.
- Lenovo P15 Gen 2 Notebook.
- Sennheiser MKE 400.
- NVIDIA JETSON AGX ORIN 64GB
- Intel RealSense L515 LiDAR Camera.
- Intel RealSense D435i Depth Camera.

Fig. 4. TIAGo

Robot's Software Description

A exhaustive list of our software and all its dependencies can be found online[11].

- Operating Systems: Ubuntu 18.04 with ROS melodic and Ubuntu 22.04 with One and ROS Humble
- Navigation: karto, movebase, CLF Planner[12], navigation2
- Object Recognition: YOLOX [3][13]
- Manipulation: MoveIt, MoveIt Task Constructor [4]
- People Detection: Strands people[14], Openpose, CFTLD based Tracking, Bayes Tracking
- Speech Recognition: PocketSphinx with context-dependent ASR, whisper[15]
- Speech Synthesis: Acapela TTS
- Behavior Control: BonSAI with SCXML [16]

External Devices and Services

TIAGo uses neither.

---

[11] https://citkat-citec.bob.ci.cit-ec.net/distribution/tiago-jammy-one-nightly.xml

[12] https://github.com/CentralLabFacilities/clf_navigation_plugins

[13] https://github.com/CentralLabFacilities/clf_object_recognition

[14] https://github.com/strands-project/strands_perception_people

[15] https://github.com/CentralLabFacilities/speech_recognition

[16] https://github.com/CentralLabFacilities/bonsai

Robot software and hardware specification sheet

Team Information

Name of Team:

Team of Bielefeld (ToBI)

Contact information:

Leroy Rügemer
Research Center for Cognitive Interaction Technology (CITEC)
Bielefeld University
Inspiration 1, 33619 Bielefeld, Germany
{tobi,lruegeme}@techfak.uni-bielefeld.de

Website:

https://clf.pages.ub.uni-bielefeld.de/Robocup/website/

Team video:

https://clf.pages.ub.uni-bielefeld.de/Robocup/website/q2024

Team members:

Sven Wachsmuth, Leroy Rügemer, David Leins, Aiko Galetzka, Alexander Esau, Alexander Seltmann, Anargh Viswanath, Ben Julius Rogalski, Christoph Kowalski, Kristin Willms, Laurin Gräsner, Michal Ljubljanac, Michel Wagemann, Mohamed Amine Hammami, Niklas Schmitz, Robin von Collani