

RoboCup 2019 - homer@UniKoblenz (Germany)

Raphael Memmesheimer, Ivanna Mykhalchyshyna, Niklas Yann Wettengel,
Tobias Evers, Lukas Buchhold, Patrik Schmidt, Niko Schmidt, Ida Germann,
Mark Mints, Greta Rettler, Christian Korbach, Robin Bartsch, Isabelle
Kuhlmann, Thomas Weiland, Dietrich Paulus

Active Vision Group
University of Koblenz-Landau
Universitätsstr. 1
56070 Koblenz, Germany
raphael@uni-koblenz.de
<http://homer.uni-koblenz.de>

Abstract. This paper describes the robots *TIAGo* and *Lisa* used by team homer@UniKoblenz of the University of Koblenz-Landau, Germany, for the participation at the RoboCup@Home 2019 in Sydney, Australia. A special focus is put on novel system components and the open source contributions of our team. We have released packages for object recognition, a robot face including speech synthesis, mapping and navigation, speech recognition interface, gesture recognition and imitation learning. The packages are available (and new packages will be released) on <http://homer.uni-koblenz.de>.

1 Introduction

Our team won the RoboCup@Home Open Platform League two times in a row in Montreal (Canada) and Nagoya (Japan) [4]. Before we won the RoboCup@Home in Hefei, China in 2015. Further we won the RoboCup GermanOpen in the @Home track two years in a row. Furthermore we were awarded with four out of five prizes of the European Robotics League two years in a row. In the World Robot Summit this year we ended up 3rd in the Customer Interaction Task of the Future Convenience Store Challenge.

Beside these successes our team homer@UniKoblenz has already participated successfully as finalist in Suzhou, China (2008), Graz, Austria (2009) in Singapore (2010), where it was honored with the RoboCup@Home Innovation Award, in Mexico-City, Mexico (2012), where it was awarded the RoboCup@Home Technical Challenge Award and in Eindhoven, Netherlands (2013). Further, we participated in stage 2 at the RoboCup@Home World Championship in Istanbul, Turkey (2011). Our team achieved several times the 3rd place in the RoboCup GermanOpen (2008, 2009, 2010 and 2013) and participated in the GermanOpen finals (2011, 2012 and 2014).

Apart from RoboCup, team homer@UniKoblenz won the best demonstration award at RoCKIn Camp 2014 (Rome), 2015 (Peccioli), the 1st place in the overall

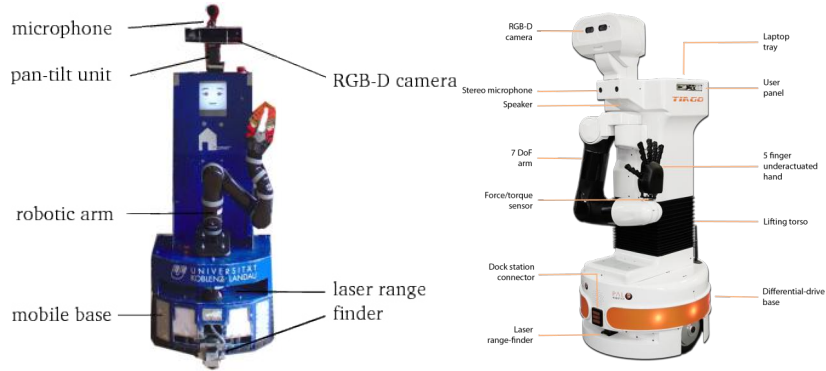


Fig. 1. Lisa (left) and TIAGo (right)

rating, as well as the 2nd place in the Object Perception Challenge in the RoCKIn Competition (Toulouse, 2014). In the RoCKIn 2015 competition (Lisbon) team homer@UniKoblenz won the 1st overall rating together with SocRob, the Best Team Award, 1st place in the Navigation Challenge, 1st place in the Getting to Know my home task benchmark.

The current research focus of the team is in Imitation Learning where we recently proposed to imitate behaviours of human demonstrations purely on visual input.

In 2019 we plan to attend the RoboCup@Home in Sydney (Australia), with two robots: a sponsored PAL Robotics TIAGo¹ robot and a custom built Lisa (Fig. 1). Our team will be presented in the next Section. Section 3 describes the hardware used for Lisa. In Section 4 we present the software components that we contribute to the community. The following Section 5 presents our recently developed and improved software components. Finally, Section 6 will conclude this paper.

2 Team homer@UniKoblenz

The Active Vision Group (AGAS) offers practical courses for students where the abilities of Lisa are extended. In the scope of these courses the students design, develop and test new software components and try out new hardware setups. The practical courses are supervised by a research associate, who integrates his PhD research into the project. The current team is lead and supervised by Raphael Memmesheimer.

Each year new students participate in the practical courses and are engaged in the development of Lisa. These students form the team *homer@UniKoblenz* to participate in the RoboCup@Home. *Homer* is short for “home robots” and is one of the participating teams that entirely consist of students.

¹ <http://tiago.pal-robotics.com/>

2.1 Focus of Research

The current focus of research is imitation learning by observation.

Additionally, with large member fluctuations in the team, as is natural for a student project, comes a necessity for an architecture that is easy to learn, teach and use.

3 Hardware

In this year's competition we will use two robots (Fig. 1). The blue Lisa is built upon a CU-2WD-Center robotics platform². Furthermore we will use a PAL Robotics TIAGo robot that is able to higher and lower its torso has an wider working range. Currently, we are using a Workstation Notebook equipped with an Intel Core i7-6700HQ CPU @ 2.60GHz $\times$ 8, 16GB RAM with Ubuntu Linux 16.04 and ROS Kinetic. Each robot is equipped with a laser range finder (LRF) for navigation and mapping. The most important sensors of the blue Lisa are set up on top of a pan-tilt unit. Thus, they can be rotated to search the environment or take a better view of a specific position of interest. Apart from a RGB-D camera (Microsoft Kinect2) a directional microphone (Rode VideoMic Pro) is mounted on the pan-tilt unit. A 6 DOF robotic arm (Kinova Mico) is used for mobile manipulation. The end effector is a custom setup and consists of 4 Festo Finray-fingers. Finally, an Odroid C2 inside the casing of the blue Lisa handles the robot face and speech synthesis.

4 Software Contribution

We want to share stable components of our software with the RoboCup and the ROS community to help advancing the research in robotics. All software components will be released on the homer project webpage: <http://homer.uni-koblenz.de>. The contributions are described in the following paragraphs.

Mapping and Navigation

Simultaneous Localization and Mapping To know its environment, the robot has to be able to create a map. For this purpose, our robot continuously generates and updates a 2D map of its environment based on odometry and laser scans. Figure 2 shows an example of such a map.

Navigation in Dynamic Environments An occupancy map that only changes slowly in time does not provide sufficient information for dynamic obstacles. Our navigation system, which is based on Zelinsky's path transform [8, 9], always merges the current laser range scans into the occupancy map. A calculated path is checked against obstacles in small intervals during navigation. If an object blocks the path for a given interval, the path is re-calculated.

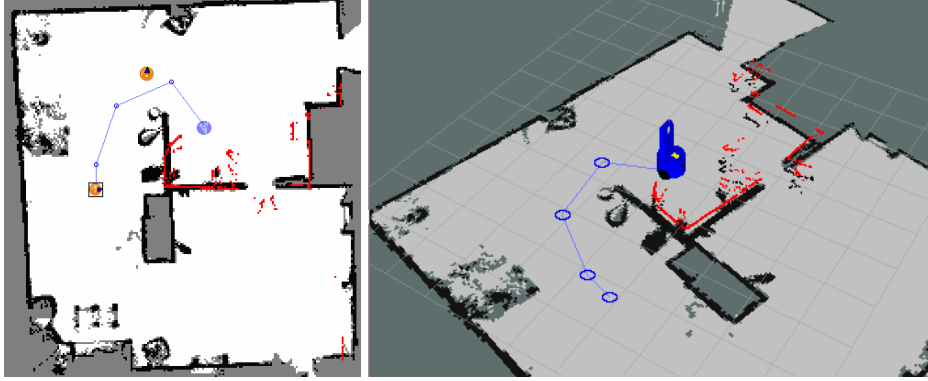


Fig. 2. 2D and 3D view of a map and a planned path (blue line). Red dots indicate the current laser scan, while orange points in the 2D map stand for navigation points.



Fig. 3. Prediction result of our home net pixel wise semantic segmentation approach. In this case the net is classifying different kinds of beers.

Object Recognition We developed an approach for pixel wise semantic segmentation based on SegNet [1]. The approach has proven to be very fast on limited hardware resources. On a Nvidia Jetson TX2 we reached 8Hz. Over currently spread object detection methods like [5] which estimate bounding boxes a pixel wise semantic segmentation approach. One of the benefits of using semantic segmentation approaches are the ability to estimate finer grasp poses. This is particularly of interest for objects that are complex to manipulate like cutlery. A package for the use of our semantic segmentation approach can be found online³.

² Manufacturer of our robotic platform: <http://www.ulrichc.de>

³ https://gitlab.uni-koblenz.de/robbie/homer_home_net

5 Technology and Scientific Contribution

5.1 General Purpose System Architecture

In the past years we have migrated step by step from our self developed architecture to ROS. Since 2014, our complete software is ROS compatible. To facilitate programming new behaviors, we created a architecture aiming at general purpose task executing. By encapsulating arbitrary functionalities (e.g. grasping, navigating) in self-contained state machines, we are able to start complex behaviors by calling a ROS action. The ROS action library allows for live monitoring of the behavior and reaction to different possible error cases. Additionally, a semantic knowledge base supports managing objects, locations, people, names and relations between these entities. With this design, new combined behaviors (as needed e.g. for the RoboCup@Home tests) are created easily and even students who are new to robotics can start developing after a short introduction.

5.2 Imitation Learning

Recently we proposed an visual approach for Imitation Learning [3]. This approach was presented during the 2018 RoboCup@Home Finals in Montreal.

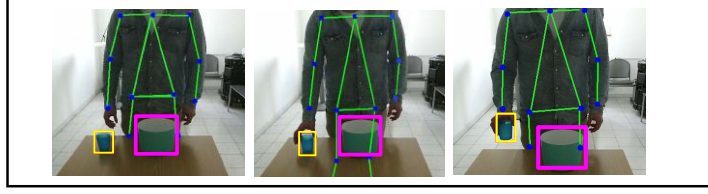
Current robotic systems that lack a certain desired behavior, commonly need an expert programmer to add the missing functionality. Contrary, we introduce an approach related to programming robots by visual demonstration that can be applied by common users. Provided a basic scene understanding, the robot observes a person demonstrating a task and is then able to reproduce the observed action sequence using its semantic knowledge base.

We presented an approach for markerless action recognition based on Convolutional Pose Machines (CPM) [6], object observations [5] and continuous spatial relations. The actions are executable on a robot that is able to execute a set of common actions. The initial scene analysis allows semantic reasoning in case the required object is not present. Further, this allows executing the same action sequence with different objects which is a major benefit over action sequencing approaches that rely on positional data only. Even so we are demonstrating our approach on 2D observations, the formulations are also adaptable in 3D. Figure 4 gives an overview of our approach.

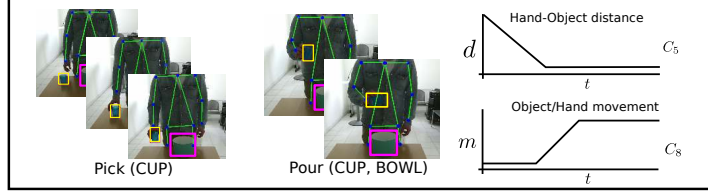
5.3 Gesture Recognition

For the restaurant and general purpose task we developed an approach for gesture recognition [2]. Human pose features in a face centric coordinate system are extracted and classifiers are trained on custom created datasets for the required gestures. Further we showed generalization of our approach onto common gestures using public available datasets. An illustration of our approach is shown in Figure 5.

1) Observation



2) Action recognition



3) Action execution

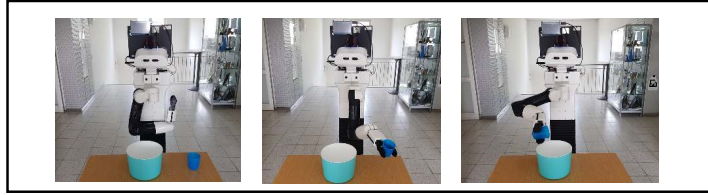


Fig. 4. Approach overview for extracting action informations from 2D image sequences in order to execute them on a mobile robot. Exemplary object detections (yellow, pink) and human pose estimates (green) are observed. Actions are recognized using a set of constraints. For replicating the observed actions we used two mobile robots equipped with an arm.

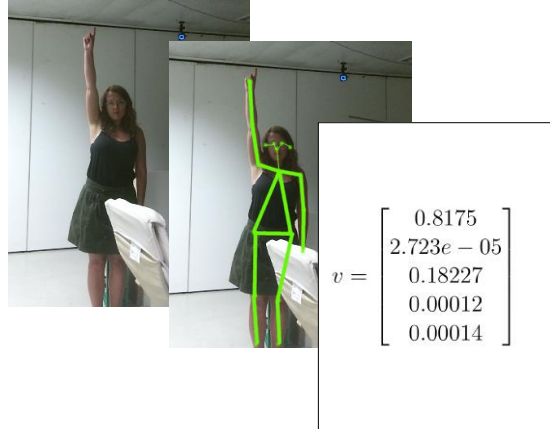


Fig. 5. Approach overview for gesture classification from single 2D image. First human pose features are extracted (green). Based on these features a gesture classifier outputs a prediction vector with estimated estimated gesture confidences.

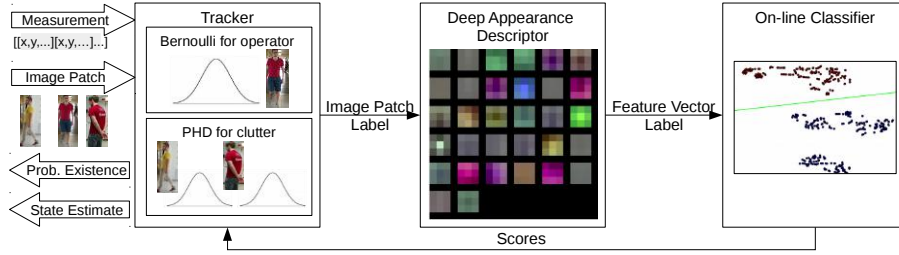


Fig. 6. Tracking Overview

A RFS Bernoulli single target tracker in cooperation with a deep appearance descriptor to re-identify and online classify the appearance of the tracked identity. Measurements, consisting of positional information and an additional image patch serve as input. The Bernoulli tracker estimates the existence probability and the likelihood of the measurement being the operator. Positive against negative appearances are contentiously trained. The online classifier returns scores of the patch being the operator.

5.4 People Detection and Tracking

We developed an integrated system to detect and track a single operator that can switch *off* and *on* when it leaves and (re-)enters the scene [7]. Our method is based on a set-valued Bayes-optimal state estimator that integrates RGB-D detections and image-based classification to improve tracking results in severe clutter and under long-term occlusion. The classifier is trained in two stages: First, we train a deep convolutional neural network to obtain a feature representation for person re-identification. Then, we bootstrap a classifier that discriminates the operator online from remaining people on the output of the state-estimator. See Figure 6 for an visual overview. The approach is applicable for following and guiding tasks.

5.5 Speech Recognition

For speech recognition we use a grammar based solution supported by a academic license for the VoCon speech recognition software by Nuance⁴. We combine continuous listening with a begin and end-of-speech detection to get good results even for complex commands. Recognition results below a certain threshold are rejected. The grammar generation is supported by the content of a semantic knowledge base that is also used for our general purpose architecture.

6 Conclusion

In this paper, we have given an overview of the approaches used by team homer@UniKoblenz for the RoboCup@Home competition. We presented a com-

⁴ <http://www.nuance.com/for-business/speech-recognition-solutions/vocon-hybrid/index.htm>

combination of out-of-the box hardware and sensors and a custom-built robot framework. Furthermore, we explained our system architecture, as well as approaches for 2D and 3D object recognition, human robot interaction and object manipulation with a 6 DOF robotic arm. This year we plan to use the TIAGo robot and blue Lisa for the main competition. Based on the existing system from last year's competition, effort was put into improving existing algorithms of our system (speech recognition, manipulation, people tracking) and adding new features (imitation learning, gesture recognition) to our robot's software framework. Finally, we explained which components of our software are currently being prepared for publication to support the RoboCup and ROS community.

References

1. Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*, 2015.
2. Raphael Memmesheimer, Ivanna Mykhalchyshyna, and Dietrich Paulus. Gesture recognition on human pose features of single images. In *Intelligent Systems (IS), 2018 9th International Conference on*, pages 1–7. IEEE, 2018.
3. Raphael Memmesheimer, Ivanna Mykhalchyshyna, Viktor Seib, Nick Theisen, and Dietrich Paulus. Markerless visual robot programming by demonstration. *arXiv preprint arXiv:1807.11541*, 2018.
4. Raphael Memmesheimer, Viktor Seib, and Dietrich Paulus. homer@unikoblenz: Winning team of the robocup@home open platform league 2017. In Hidehisa Akiyama, Oliver Obst, Claude Sammut, and Flavio Tonidandel, editors, *RoboCup 2017: Robot World Cup XXI*, pages 509–520, Cham, 2018. Springer International Publishing.
5. Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.
6. Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. *CoRR*, abs/1602.00134, 2016.
7. Nicolai Wojke, Raphael Memmesheimer, and Dietrich Paulus. Joint operator detection and tracking for person following from mobile platforms. In *Information Fusion (Fusion), 2017 20th International Conference on*, pages 1–8. IEEE, 2017.
8. Alexander Zelinsky. Robot navigation with learning. *Australian Computer Journal*, 20(2):85–93, 5 1988.
9. Alexander Zelinsky. *Environment Exploration and Path Planning Algorithms for a Mobile Robot using Sonar*. PhD thesis, Wollongong University, Australia, 1991.

Name of team homer@UniKoblenz

Member Raphael Memmesheimer, Ivanna Mykhalchyshyna, Niklas Yann Wet-
tengel, Tobias Evers, Lukas Buchhold, Patrik Schmidt, Niko Schmidt, Ida
Germann, Mark Mints, Greta Rettler, Christian Korbach, Robin Bartsch,
Isabelle Kuhlmann, Thomas Weiland

Contact information raphael@uni-koblenz.de

Website <http://homer.uni-koblenz.de>

Hardware :

- PAL Robotics TIAGo Steel Edition
- Kinova Mico
- Lenovo P50 Notebook
- Directed Perception D46-17.5 PTU
- Rode VideoMic
- Microsoft Kinect 2
- CU2WD Robot platform
- Hokuyo Laser Scanner
- SICK LMS 100
- Odroid C2

Software :

- ROS
- OpenCV
- PCL
- Mary TTS
- Ogre3D
- Caffè
- Tensorflow
- Matlab
- Nuance VoCon
- strands
- Custom software for:
 - User Interface (homer_gui https://gitlab.uni-koblenz.de/robbie/homer_gui)
 - Object recognition (homer_object_recognition https://gitlab.uni-koblenz.de/robbie/homer_object_recognition)
 - Mapping / Navigation (homer_mapnav https://gitlab.uni-koblenz.de/robbie/homer_mapnav)
 - Robot face (homer_robot_face https://gitlab.uni-koblenz.de/robbie/homer_robot_face)
 - Speech Recognition / Speech synthesis (android_speech_pkg https://gitlab.uni-koblenz.de/robbie/homer_android_speech)