



Adaptive Learning Methods for Autonomous Mobile Manipulation in RoboCup@Home

Raphael Memmesheimer^(✉), Viktor Seib, Tobias Evers, Daniel Müller,
and Dietrich Paulus

Active Vision Group, Institute for Computational Visualistics,
University of Koblenz-Landau, 56070 Koblenz, Germany
{raphael,vseib,tevers,muellerd,paulus}@uni-koblenz.de
<http://homer.uni-koblenz.de>, <http://agas.uni-koblenz.de>

Abstract. Team homer@UniKoblenz has become an integral part of the RoboCup@Home community. As such we would like to share our experience gained during the competitions with new teams. In this paper we describe our approaches with a special focus on our demonstration of this year’s finals. This includes semantic exploration, adaptive programming by demonstration and touch enforcing manipulation. We believe that these demonstrations have a potential to influence the design of future RoboCup@Home tasks. We also present our current research efforts in benchmarking imitation learning tasks, gesture recognition and a low cost autonomous robot platform. Our software can be found on GitHub at <https://github.com/homer-robotics>.

1 Introduction

This year, team homer@UniKoblenz became RoboCup@Home world champion for the fourth time. This makes our team the most successful team in the Open Platform League of the RoboCup@Home!

Team homer@UniKoblenz started in 2008 by borrowing rescue robot “Robbie” from our prior existing RoboCup Rescue team for their own participation in a so far unknown at home environment. The next year homer@UniKoblenz participated for the first time with their own robot “Lisa”. Thanks to the endeavors of the team leaders and the commitment of all team members over time we managed to be among the top 5 teams frequently. As our team that arises from new students every year, going beyond the 5th place was not easy. We had to develop strategies to teach and mentor new students and allow them to start being productive after a short introductory period. Most importantly, we had to come up with a software architecture and organization structure, which enable programming on different abstraction levels and, despite their complexity, are easy to comprehend and to learn for new team members. In this paper we first present our strategies and solutions to overcome common challenges. We provide several software package used by our team during the competition and give an

insight on how we collect data for object recognition. We hope all of these will serve as guidelines for new teams. Finally, we discuss our demonstration in the finals of 2019 and describe the results of the competition.

2 Team homer@UniKoblenz

The key to success in a competition is a good team. Here we describe our selection process for new team members and give an overview over our robotics hardware.

2.1 Members

In our case only one academic employee is working with Lisa. That employee is the team leader and conducts research for his PhD thesis on Lisa. Since the major part of the team is formed by students and there is only a limited amount of persons, which can be effectively supervised, these students have to be selected carefully. Thus, students have to write a formal application and will be invited to an interview. The application process is necessary in order to identify especially those students, which are most motivated and willing to put extensive work into the project. A student does not have to be exceptionally talented in any field to be accepted. The main requirement is to show motivation and willingness to learn. Two weeks before the start of the course all selected students participate in a daily training to learn the basics of version control, operating the robots and using ROS. The goal is to enable the students to work on their own with the robots. This is achieved by using a modular software architecture with different abstraction levels. Please consult [1,2] for details on the organization of our software. During the course focus is put on preparations for the competition. In two weekly meetings the team analyzes the requirements of the different tests and develops strategies to solve them. Over the years we reduced the amount of formal requirements engineering and software specifications to a minimum in favor of agile and permanent development. We integrate Jenkins as a build server for continuous, multi-architecture integration and deployment. Further, at all times students are encouraged to test their code using the robots at place with real sensor data or recorded ones, e.g. from ROS bagfiles. On a regular basis we organize simulated RoboCup events at which each RoboCup task is tested and scored according to the current rules. This allows to honestly assess the progress made during the course.

2.2 Robots

We use a custom built robot called Lisa and a PAL Robotics TIAGo [3] depicted in Fig. 1. Each robot is equipped with a workstation notebook with an 8 core processor and 16 GB RAM. We use Ubuntu Linux 16.04 and ROS Kinetic. Each robot is equipped with a laser range finder and wheel encoders for navigation and mapping. Further, both robots have a movable sensor head with an RGB-D camera and a directional microphone. Lisa features a 6-DOF robotic arm

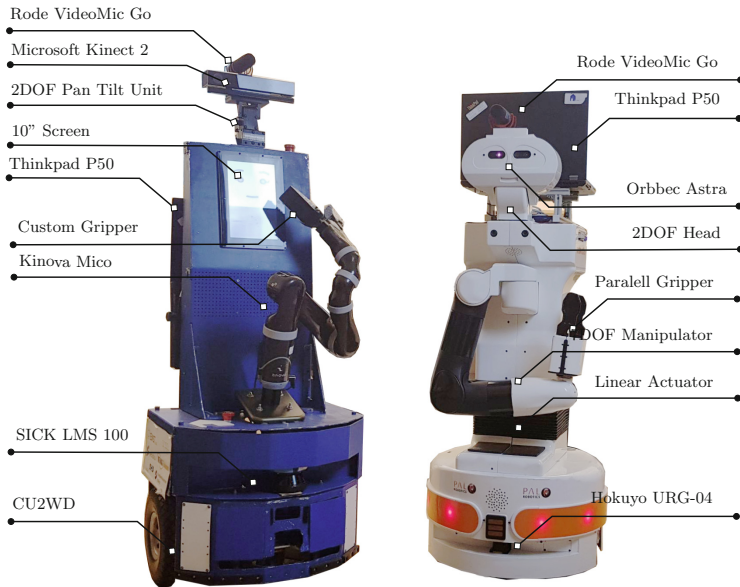


Fig. 1. The robots Lisa (left) and TIAGo (right). We use Lisa first and foremost for human robot interaction related tasks. The 10 in. screen supports guided interactions. TIAGo is used mainly as a mobile manipulation platform. The work range is far higher and the arm controllers are more reliable.

(Kinova Mico) for mobile manipulation. The end effector is a custom setup and consists of 4 Festo Finray-fingers. An Odroid C2 mini computer inside the casing of Lisa handles the robot face and speech synthesis. In contrast to Lisa, TIAGo is able to move its torso up and down and has a wider working range. This allows TIAGo to use his 7-DOF arm to reach the floor, as well as high shelves.

3 Contributing Software and Dataset

We omit a detailed description of the current approaches used by our team as these can be found in recent team description papers. Instead, in this section we focus on our contributions in the form of released software and dataset. We hope, this will help new teams to catch up with the evolved RoboCup@Home community and allow them to participate and reach a similar performance level more easily.

3.1 Software

We started releasing our packages for ROS in our old repository¹. Now we decided to make our contributions more accessible through our GitHub profile². This

¹ Previous homer repositories: <http://wiki.ros.org/agas-ros-pkg>.

² Current homer GitHub profile: <https://github.com/homer-robotics>.

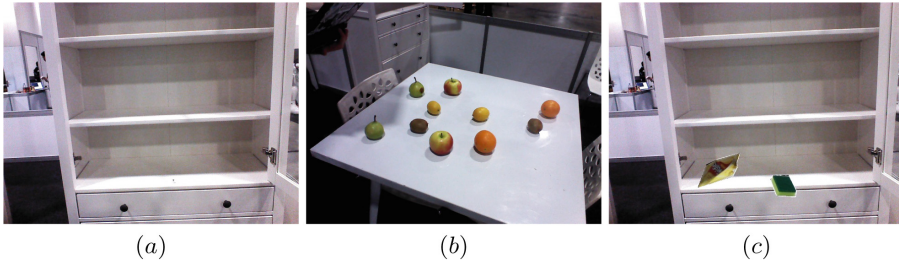


Fig. 2. Examples from the RoboCup 2019 dataset. We recorded backgrounds like in (a) and labeled images containing objects like e.g. objects of the class *fruits* in (b). The labeled objects are then augmented under a variety of transformations and then mapped onto the recorded background images (c).

includes custom mapping and navigation packages, a robotic face, integration to different text to speech systems and gesture recognition. In [2, 4] we describe in detail how new teams can create an autonomous robot capable of mapping, navigation and object recognition using our released software. Currently, as deep learning has become the state of the art in vision, we are also adopting these methods. Meanwhile, our point cloud recognition software with recent enhancements is described in [5] and can also be found on GitHub³. As development continues, we plan to release more packages and instructions on how they can be used.

3.2 RoboCup@Home 2019 Dataset

During the RoboCup 2019 competition we created a dataset that gives an insight in what data we gather for the competition attendance. In total, we recorded 189 images containing objects of different classes and 12 background images. The pixel-wisely labeled images in conjunction with the labeled object images were used to project the extracted objects to a variety of backgrounds. The backgrounds represent manipulation locations from inside the arena. In total 60572 labels were generated on 33539 augmented images. Figure 2 shows examples of backgrounds, object images and augmented images. The dataset is available online⁴.

4 Final Demonstration

During the final demonstration we showed three approaches that could potentially influence the way how RoboCup tasks will be designed in the future. The first two approaches are addressing the challenge of how robots can perform

³ Point cloud recognition: <https://github.com/vseib/PointCloudDonkey>.

⁴ RoboCup@Home 2019 dataset: https://agas.uni-koblenz.de/datasets/robocup_2019_sydney/.

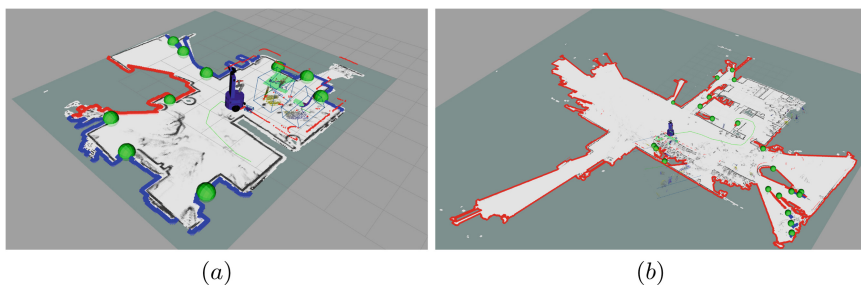


Fig. 3. LISA exploring the apartment during the RoboCup@Home finals. (a) shows a map created shortly after entering the apartment. Tables and chairs are added to the semantic representation. (b) shows the explored map at the end of the final demonstration. Once finished with the apartment the robot started to explore the outside area.

complex tasks without much prior knowledge. We demonstrated an approach for autonomous semantic exploration on our robot Lisa and an approach for adaptive learning of complex manipulation tasks. In addition, we proposed an approach for touch enforcing manipulation. A video showing the final demonstration during the RoboCup@Home 2019 is available on YouTube⁵.

4.1 Autonomous Semantic Exploration

Currently, at RoboCup@Home semantic information about the arena and the objects are provided during the setup days. After publication the majority of teams records maps, defines locations and trains objects. All this data is translated to a semantic knowledge representation which can be understood by the robot. Maps are commonly stored as occupancy grid maps with additional layers for rooms and points of interest. Object recognition tends to be trained using recent object detectors [6, 7]. This strategy works fine in a competition environment as long as the state of the arena remains unchanged or undergoes only minor modifications. In the finals, we demonstrated a more general idea addressing the question “How can we make robots behave more autonomously in previously unknown environments?”. There is a variety of pre-trained models nowadays which contains a wide variety of classes found in daily life. The COCO [8] dataset contains classes like *Chairs*, *Dining Table*, *Refrigerators*, *Couch*. The wide variety of training examples allows for creation of a generalized object detector which yields good results in completely different scenarios. We make use of the strength of those object detectors and combined them with a traditional method for exploration based on occupancy grid-maps. An example, as shown during the finals is shown in Fig. 3. Based on the autonomously created semantic representation, robots could execute manipulation tasks in the future

⁵ Final Video 2019: <https://www.youtube.com/watch?v=PZmKzngDegk>.

in previously unknown environments without any human intervention. Multi-modal sensor data extraction that is fused on different levels can increase the quality and level of the semantic representation. For our approach we plan to improve the pose estimates of the detected objects.

4.2 Adaptive Learning by Demonstration

With TIAGo we presented an approach for learning complex manipulation tasks without any prior knowledge about the task. The robot estimates the initial scene state by detecting objects. The robot’s arm is then put into gravity compensation mode, which allows for seamless human guidance. A human demonstrated relevant arm trajectories for a given task. During the demonstration trajectories are recorded relatively to the estimated objects and a static local reference.

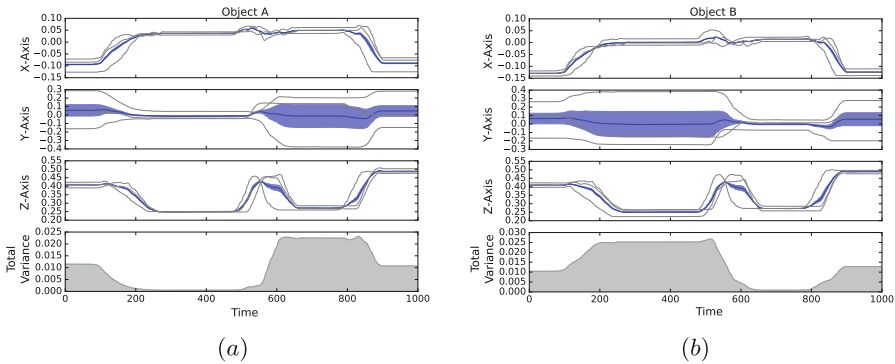


Fig. 4. Average relative position and variance of Object A and Object B. Gray lines in X-, Y- and Z-Axis represent single demonstrations. Blue lines represent average positions and blue areas variances. The total variances are shown in gray in the bottom graphics. (Color figure online)

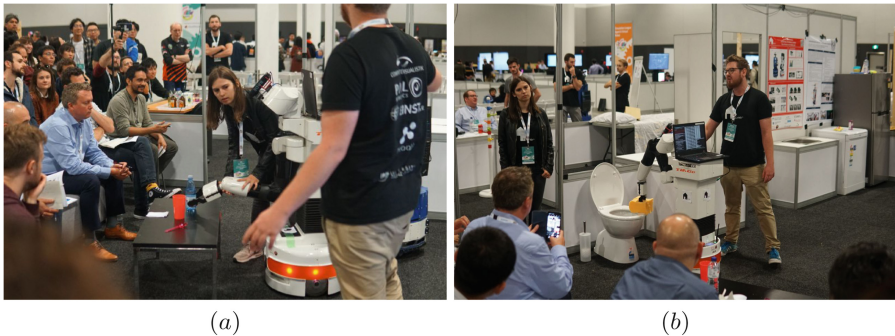


Fig. 5. TIAGo learning by demonstration how to pour a drink (a) and cleaning a toilet using a touch enforcing approach and a sponge end-effector (b).

The reference allows demonstrations with just one or no object at all as reference. With multiple demonstrations, it can be predicted if a specific object is involved in the task. More precisely, the variance in relative position of an object to the end effector over time can be calculated. A low variance indicates high importance of the object within the task at that moment and vice versa. Examples of variances are shown in Fig. 4, where an object A has been placed in an object B. The total variance of object A is low between 200–500 (Fig. 4(a)), where it has been grasped and thus is involved in the task during that time period. On the other hand the total variance of object B is low between 600–800 Fig. 4(b). The end effector moved over object B and let go in that time period. Knowing the average position of the robot’s end effector relative to the appropriate objects and estimating its importance allows to calculate a position for every time step and finally build a trajectory for any scene. Our approach is based on the ideas of Reiner et al. [9]. Yet, we use deviations of the trajectories to align the multiple trajectories. Aligning them is necessary in order to compute precise variances, since demonstrations by humans are never executed in the exact same speed. The Levenberg-Marquardt algorithm is used, which shall grant a more stable handling of singularities [10]. Figure 5(a) shows the approach as presented during the finals. Future work on this approach will include integration of joint efforts. Currently only joint angles are taken into account during the recording and reproduction. This has side effects like that if an object is grasped by closing the gripper it will be grasped with the same end effector joint states as in the demonstration. This means objects are held only loosely or smaller objects will not be stably grasped. Therefore, we suggest to also take the motor effort into account during the demonstration and reproduction.

4.3 Touch Enforcing Manipulation

In the final demonstration we also built on top of our previously presented effort based manipulation approach [11]. We propose to enforce contact during a manipulation action by continuously observing the joint efforts. Our previous approach was used in an open loop control, where a movement is interrupted once a certain effort has been reached. This is suitable for grasping where you want to determine when the effector touches the surface supporting the object. Our novel approach proposes to hold a certain effort in closed loop control. This was demonstrated during the finals as a case sample of cleaning a toilet (Fig. 5(b)). This approach can be used for a variety of tasks like cleaning tables or windows and slightly moving objects.

5 Results

This year’s tasks were categorized in two scenarios: Housekeeper and Party Host. Both scenarios contained five tasks each and teams could choose to perform three of those per scenario. For detailed descriptions of the tasks we refer to the RoboCup@Home rulebook [12]. Results from the OPL are shown in Fig. 6.

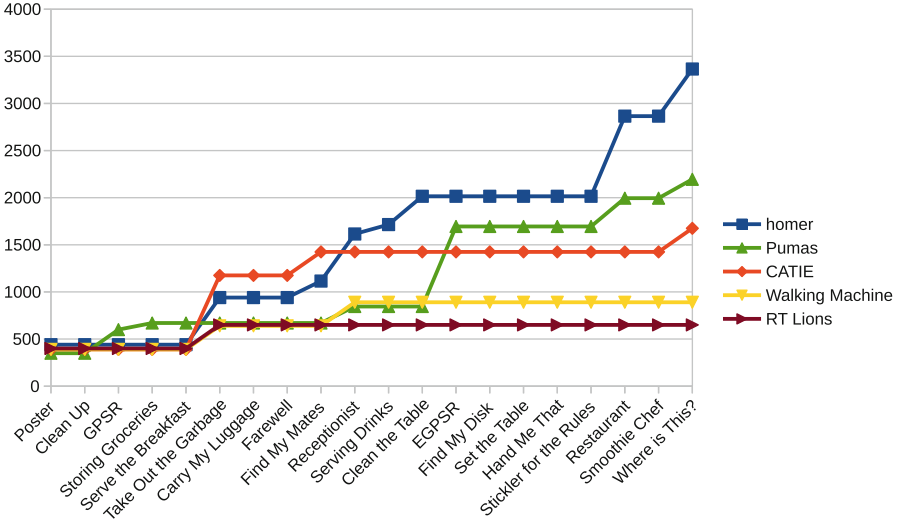


Fig. 6. Scores after Stage 1 and Stage 2 for the top five ranked teams. Stage 1 ended with *Serving Drinks*. Note: Tasks not attended by teams are still listed with no increase in points.

In Stage 1 of the Housekeeper scenario, we participated first in *Take Out The Garbage* (Fig. 7a) where we gathered 500 points. The robot grasped two garbage bags with our previously introduced closed loop touch enforcing approach. The bins were detected by finding circular patterns in the laser scan. For the *Storing Groceries* task we trained a semantic segmentation approach [6] with augmented images. Semantic annotations contained grasping trajectories. Due to a robot damage during transportation, the calibration was off and we scored 0 points. We further participated in the *Serving Breakfast* task. In *Find My Mates* we scored 175 points and additional 500 in the *Receptionist* task. We used convolutional pose machines [13] and projected them using the RGB-D camera in order to calculate distances of the persons in relation to locations of the map. This way we could estimate free seat locations and introduce the new guest to the others. In *Serving Drinks* we scored 100 points (Fig. 7b).

Table 1. The finals results

Team	Stage 1	Stage 2	Final
homer	1715	3365	100
Pumas	845	2195	78.05
CATIE	1425	1675	–



Fig. 7. TIAGo grasping a trash bag during the *Take Out The Garbage* task using an effort based grasping approach to ensure that the bag was grasped successfully (a) and Lisa handing over a drink to a guest during the *Serving Drinks* task.

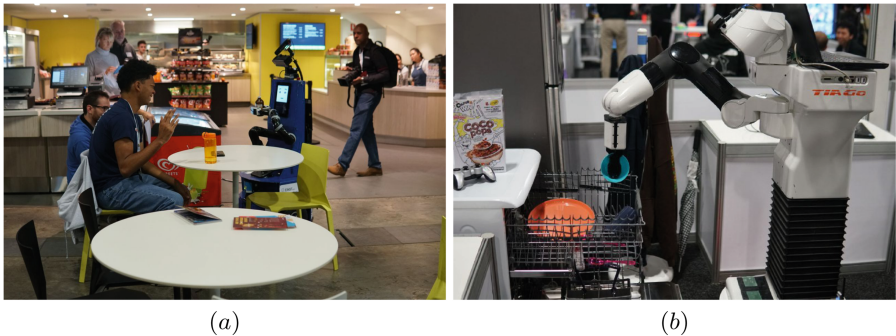


Fig. 8. LISA learning an unknown restaurant and then serving customers in the *Restaurant* task (a) and TIAGo gently placing a in the *Clean the Table* task (b) during a test run.

Stage 2 consisted of *Clean the Table*, *Enhanced General Purpose Service Robot (EGPSR)*, *Hand Me That*, *Restaurant*, *Stickler for the Rules Where Is This?*. No team attended the tasks *Find My Disk*, *Smoothie Chef* or *Set The Table*. In *Clean the Table* (Fig. 8b) we achieved 300 points. Points were given only if the whole task was completed. We used the effort based gripping approach and successfully verified our grasping approaches. When having failed to grasp the object, we asked for a handover by describing the object we had intended to take. Our approach for *Hand me That* was based on finding the operator by estimating human poses [13]. As multiple persons could be in the environment, the robot asked the operator to take a distinguishable pose. We gathered 0 points, however we demonstrated robust human robot interaction. For *Restaurant* (Fig. 8a) we used again a gesture recognition based on estimated human poses [13]. Maps were created on-line and as tables and chair's in unknown environments are hard to see obstacles we fused the laser range finder and RGB-D camera in an obstacle layer of the mapping. We managed to serve two customers and deliver

three orders with a fourth order on the way. In total we gathered 850 points. For the *Where Is This?* task we decided to use naive operators. We managed to guide two randomly chosen children from the audience to the object they were looking for, another 500 points for us. Table 1 shows the final results of the first three teams.

6 Current Research

We now briefly highlight our current research and ongoing developments related to RoboCup@Home.

Simitate: A Hybrid Imitation Learning Benchmark. Imitation learning approaches for manipulation tasks lack comparability and reproducibility. Therefore we developed a benchmark [14] which integrates a dataset containing RGB-D camera data as it is available on most robots. The RGB-D camera is calibrated against a motion capturing system. The calibration furthermore allows the reconstruction in a virtual environment which allows quantitative evaluation in a simulated environment. For that we propose two metrics that assess the quality on a trajectory and a effect level.

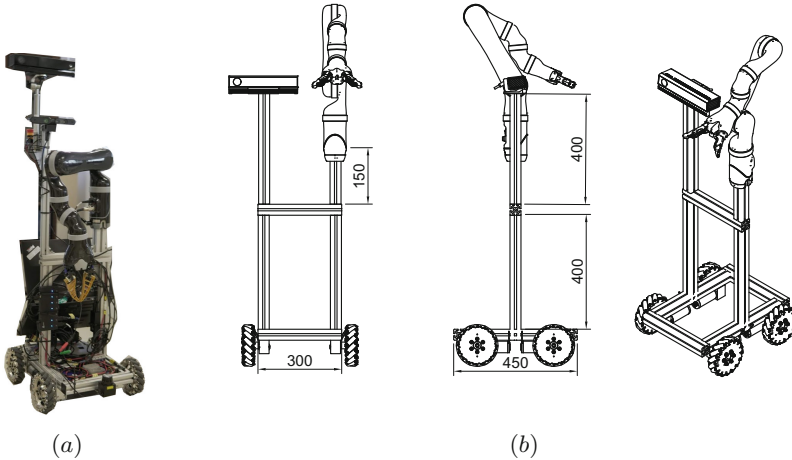


Fig. 9. Assembled *Scratchy* (a) and exemplary dimensions in (b).

A Lightweight Modular Autonomous Robot for Robotic Competitions. We proposed a novel modular robotic platform [15] for competition attendance and research. The aim was to create a minimal platform, which is easy to build, reconfigurable, offers basic autonomy and is low in cost to rebuild. Thus

researchers can concentrate on their respective topics and adapt sensors and appearance depending on their needs. *Scratchy* allows also low cost participation in robot competitions as the weight is low and the robot can be compressed up to a hand luggage size.

Gesture Recognition. Gesture recognition in RoboCup@Home becomes more and more mandatory. First we propose a gesture recognition approach on single RGB-D images. We extract human pose features and train them using a set of classifiers [16]. We then extended the approach to image sequences [17] using Dynamic Time Warping [18] and a one nearest neighbour classifier. This allows gesture classification with just one reference example.

Updating Algorithms. In the past we successfully used classic approaches for 3D point cloud processing [5] and affordance estimation [19]. We believe that transferring these ideas to the powerful domain of deep neural networks will further enhance the performance of these algorithms. Currently, we investigate methods that allow for a lightweight network structure while still retaining the performance of the classic approaches.

7 Conclusion

With our success this year we demonstrated a robust robotic architecture that is adaptable to various kinds of robots and setups. This architecture was constantly developed and improved over the years - and we will continue into the same direction. This paper gave a brief insight into the internal organisation of our teams and presented our robots. We further published some important software components on our GitHub profile. We hope that all of these will be helpful to new team who join the RoboCup@Home league. Our final demonstration of this year focused on the topic of autonomy in unknown environments. We want to motivate the league to follow this track and encourage autonomous robots that operate in previously unseen environments and execute loosely specified tasks. This will increase the flexibility of the robots and allow more research related activities in RoboCup@Home. Our methods on imitation learning, programming by demonstration, autonomous mobile exploration can be seen as potential directions. We believe that recent research results like on benchmarking imitation learning approaches, gesture recognition and a minimal autonomous research platform can support current and new teams.

Acknowledgement. We want to thank the participating students that supported in the preparation, namely Ida Germann, Mark Mints, Patrik Schmidt, Isabelle Kuhlmann, Robin Bartsch, Lukas Buchhold, Christian Korbach, Thomas Weiland, Niko Schmidt, Ivanna Kramer. Further we want to thank our sponsors (Univeristy of Koblenz-Landau, Student parliament of the University of Koblenz-Landau Campus Koblenz, PAL Robotics, Einst e.V., CV e.V., Neoalto and KEVAG Telekom GmbH).

References

1. Seib, V., Manthe, S., Memmesheimer, R., Polster, F., Paulus, D.: Team homer@UniKoblenz — approaches and contributions to the RoboCup@Home competition. In: Almeida, L., Ji, J., Steinbauer, G., Luke, S. (eds.) RoboCup 2015. LNCS (LNAI), vol. 9513, pp. 83–94. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-29339-4_7
2. Memmesheimer, R., Seib, V., Paulus, D.: homer@UniKoblenz: winning team of the RoboCup@Home open platform league 2017. In: Akiyama, H., Obst, O., Sammut, C., Tonidandel, F. (eds.) RoboCup 2017. LNCS (LNAI), vol. 11175, pp. 509–520. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00308-1_42
3. Pages, J., Marchionni, L., Ferro, F.: Tiago: the modular robot that adapts to different research needs. In: International Workshop on Robot Modularity, IROS (2016)
4. Seib, V., Memmesheimer, R., Paulus, D.: A ROS-based system for an autonomous service robot. In: Koubaa, A. (ed.) Robot Operating System (ROS). SCI, vol. 625, pp. 215–252. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26054-9_9
5. Seib, V., Theisen, N., Paulus, D.: Boosting 3D shape classification with global verification and redundancy-free codebooks. In: Tremeau, A., Farinella, G.M., Braz, J. (eds.) Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 5, pp. 257–264. SciTePress (2019)
6. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
7. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
8. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
9. Reiner, B., Ertel, W., Posenauer, H., Schneider, M.: LAT: a simple learning from demonstration method. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4436–4441. IEEE (2014)
10. Buss, S.R., Kim, J.-S.: Selectively damped least squares for inverse kinematics. *J. Graph. Tools* **10**(3), 37–49 (2005)
11. Memmesheimer, R., Mykhalchyshyna, I., Seib, V., Evers, T., Paulus, D.: homer@UniKoblenz: winning team of the RoboCup@Home open platform league 2018. In: Holz, D., Genter, K., Saad, M., von Stryk, O. (eds.) RoboCup 2018. LNCS (LNAI), vol. 11374, pp. 512–523. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27544-0_42
12. Matamoros, M., et al.: Robocup@home 2019: Rules and regulations (draft) (2019). http://www.robocupathome.org/rules/2019_rulebook.pdf
13. Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291–7299 (2017)
14. Memmesheimer, R., Mykhalchyshyna, I., Seib, V., Paulus, D.: Simitate: a hybrid imitation learning benchmark. arXiv preprint [arXiv:1905.06002](https://arxiv.org/abs/1905.06002) (2019)
15. Memmesheimer, R., et al.: Scratchy: a lightweight modular autonomous robot for robotic competitions. In: 2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), pp. 1–6. IEEE (2019)

16. Memmesheimer, R., Mykhalchyshyna, I., Paulus, D.: Gesture recognition on human pose features of single images. In: 2018 International Conference on Intelligent Systems (IS), pp. 813–819. IEEE (2018)
17. Schneider, P., Memmesheimer, R., Kramer, I., Paulus, D.: Gesture recognition in RGB videos using human body keypoints and dynamic time warping. arXiv preprint [arXiv:1906.12171](https://arxiv.org/abs/1906.12171) (2019)
18. Salvador, S., Chan, P.: Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **11**(5), 561–580 (2007)
19. Seib, V., Knauf, M., Paulus, D.: Affordance Origami: unfolding agent models for hierarchical affordance prediction. In: Braz, J., et al. (eds.) VISIGRAPP 2016. CCIS, vol. 693, pp. 555–574. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64870-5_27