Austin Villa@Home 2018
DSPL Team Description Paper

Justin W. Hart[1], Peter Stone[1], Andrea Thomaz[2],
Scott Niekum[1], Luis Sentis[3], and Raymond J. Mooney[1]

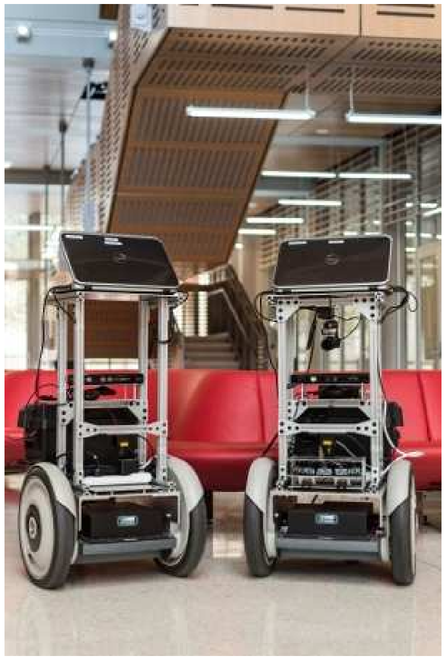
[1]Department of Computer Science
[2]Department of Electrical and Computer Engineering
[3]Department of Aerospace Engineering
University of Texas at Austin
http://www.cs.utexas.edu/~AustinVilla/?p=athome

Abstract. UT Austin Villa@Home is composed of faculty, post-docs and students across multiple laboratories at the University of Texas at Austin. Collectively, our research spans artificial intelligence, machine learning, natural language, control, and human-robot interaction. Austin Villa@Home has taken part in RoboCup@Home twice, in 2007 and 2017, where we were awarded second and third place, respectively. This Team Description Paper describes our system to date, as well as documenting our high-level approach to the 2018 competition.

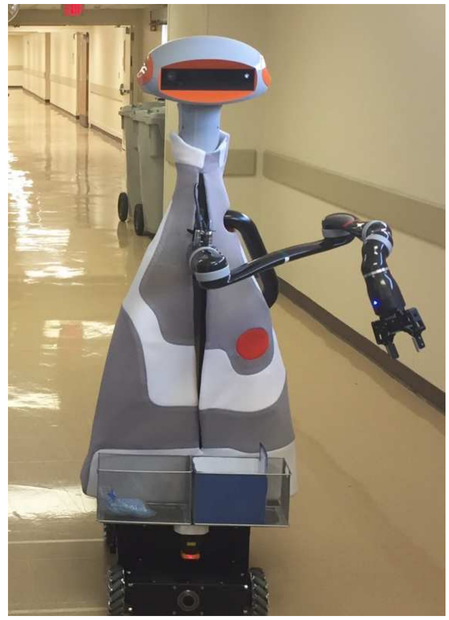
1 Introduction

The UT Austin Villa @ Home RoboCup@Home team brings together an interdisciplinary group of researchers from laboratories spanning the Computer Science, Electrical and Computer Engineering, and Aerospace Engineering departments at UT Austin. Prof. Peter Stone, the founder of the UT Austin Villa RoboCup soccer team, leads the Building-Wide Intelligence (BWI) Project. BWI's intention is to deploy a fleet of autonomous robots in the Gates-Dell Complex at UT Austin which become an integral part of the building's culture, providing services to the inhabitants of the computer science department. Prof. Andrea Thomaz, an expert in interactive robotics, leads a laboratory focused on human-robot interaction for social robots. Prof. Scott Niekum, the leader of UT's Personal Robotics Robotics Laboratory, researches learning from demonstration and safety for robot learning. Together, Thomaz and Neikum share a laboratory space which simulates a typical domestic environment in which they test and develop these technologies. Prof. Luis Sentis, an expert in robotic control, leads a laboratory focused on enhancing control for robots operating near humans. Prof. Raymond Mooney leads a group with broad interests within natural-language processing and natural-language learning.

RoboCup 2017 marked the first time Austin Villa@Home competed in the Domestic Standard Platform League (DSPL), in which we took 3rd place. Previously, in 2007, our group took 2nd place in RoboCup@Home. This TDP describes our system and research, as well as our high-level approach going into 2018's competition.



(a) Two BWIBots from the Building-Wide Intelligence Project, run by Peter Stone's laboratory. This platform is also used by Raymond Mooney's language group.



(b) Custom robot, Poli, from Andrea Thomaz's laboratory.

(c) Custom robot used in Scott Niekum's laboratory.

(d) Dreamer is a compliant humanoid robot used in Luis Sentis's laboratory.

Fig. 1: Robot platforms in use by the the respective laboratories that comprise the by the UT Austin Villa RoboCup@Home Team.

2 Software Overview

The RoboCup@Home competition closely resembles research perused by our collective group, and the Toyota Human Support Robot (HSR) is similar to the robots used in our groups (see Figure 1). Our RoboCup@Home infrastructure is based on adaptations of the BWI infrastructure, which is build upon Robot Operating System (ROS).

2.1 Object Recognition / Computer Vision

Our object recognition system handles object detection on tables and cupboards.

1. Cameras are first calibrated using a custom camera calibration suite which implements Zhang's method [1], followed by bundle adjustment.
2. Sample consensus segmentation is used to extract planes from point cloud data. Segmented planes are used to identify surfaces such as those of the table and shelves in the cupboard.
3. Objects are extracted using Euclidean clustering.
4. The 3D locations of extracted objects are projected into 2D map locations with respect to extracted planes as bounding boxes.
5. YOLO [2] classifies objects from input images. The YOLO network is fine-tuned based on the known objects during the competition.

2.2 Manipulation

Manipulation is based on code provided by TRI. Grasps are based on simple heuristics to grasp objects near their midpoints. The TRI manipulation code is closed-source, however, we wrote wrapper functions to access the robot's motion functionality which optimized the robot's motion producing faster trajectories.

2.3 Audio Processing

Our system performs speech recognition, denoising, and sound localization.

Speech Recognition Speech recognition is handled by the Google Speech API, which we access through a cloud-based interface. We bias the recognition with a corpus of words used often in RoboCup@Home in order to decrease the word error rate (WER) in speech tasks.

Denoising Prior to speech recognition, audio is passed through a nonlinear beamformer used by Google Chrome's WebRTC implementation. This helps increase the signal to noise ratio by using a form of spatial filtering where the microphone array is digitally ”steered” towards the speech source and the sensitivity of the array is decreased at all directions away from the speaker.

Sound Localization Sound localization is done using GEVD-MUSIC [3], which improves upon the classic MUSIC algorithm by using a pre-measured noise correlation matrix that our noise calibration module generates. The result of this is used directly (e.g. in the speech recognition task) and is also passed on to the beamformer. Our implementation is based off of HARK.[1]

2.4 Natural Language

Our natural language processing technology is mainly used in the speech related tasks of SPR and GPSR. Once transcripts are received from speech recognition, template matching is done to find which question the speaker is asking. This works since all questions in the tasks follow a pre-specified format. Moreover, errors in template matching are corrected by using Levenshtein distance to find the closest template. Once this processing is done, questions are then able to be answered by referencing a knowledge base.

2.5 2017 Recap

Our performance in 2017 was largely due to several well-engineered components which were re-used repeatedly in task-specific scripts which were engineered for each stage. This process led to a significant time spent on-site hand-tuning performance for each task, but ultimately allowed our team to take third place in the DSPL.

3 Research Focus

Our research emphases include reinforcement learning, learning from demonstration, controls and human-robot interaction. We believe that building integrated, autonomous systems is a worthwhile research endeavor in itself, so we have an established record of developing platforms, several of which are depicted in Figure 1.

4 Innovative Technology

Our group has made significant contributions spanning multiple areas of robotics, artificial intelligence, and human-robot interaction. Here, we provide a brief overview of published innovative technologies which are likely to impact our RoboCup@Home efforts.

[1] http://www.hark.jp/

4.1 Natural Language Understanding

Intrinsic to the RoboCup@Home competition is understanding and responding to natural language requests. The BWIBot uses a dialog agent to communicate with users through natural language. The system uses a semantic parser which is able to resolve lambdas such as possessives. This allows our robot to process verbal commands by directly interfacing with our planning technology. Thus the robot to reason about commands which it has not heard previously uttered such as “bring the pot of coffee to Justin’s office.” Our approach can recognize more complex language than keyword-based approaches. Moreover, current work on clarification dialogues allows for the robot to clarify the meaning of commands when it has only a partial understanding of what has been said.

4.2 Grounded multimodal language learning

The language grounding problem is one of pairing percepts to language symbols. For instance, “bring me the red bottle” would require the words “me,” “red,” and “bottle” to resolve to real-world interpretations. Our group has pursued significant work in the form of robots playing language games such as “I spy,” as a child might, in order to learn object groundings. The robot performs object explorations such as picking up or dropping objects in order to learn object properties paired with words [4].

4.3 Learning multi-step tasks from demonstration

We developed a series of LfD algorithms that, for the first time, allowed a robot to learn the structure of complex tasks such as IKEA furniture assembly from a small number of demonstrations [5]. This research led to state-of-the-art Bayesian nonparametric and control techniques that were able to automatically identify an appropriate number of skills from task demonstrations [6], infer the goals of each skill [7], construct controllers to accomplish these goals, and intelligently sequence these controllers based on perceptual feedback [5].

4.4 Robot-centric human activity recognition

Our robot uses its autonomous navigation capability in a large, unstructured, and human-inhabited environment. The activities learned by our robot were performed spontaneously by many different people who interacted with (or were observed by) the robot, as opposed to the standard methodology of asking study participants to perform certain actions. In contrast to classic computer vision approaches, our system uses both visual and non-visual cues when recognizing the activities of humans that it interacts with [8].

4.5 Planning using action language $\mathcal{BC}$

We have demonstrated how action language $\mathcal{BC}$ can be used for robot task planning in domains requiring planning in the presence of missing information and indirect/recursive action effects [9]. While we demonstrate using $\mathcal{BC}$ to express a mail collection task, the overall methodology is applicable to any other planning domains that require: recursive and indirect action effects, defeasible reasoning, and acquiring previously unknown knowledge through human-robot interaction. In addition, we also demonstrated how answer set planning under action costs [10] can be applied to robot task planning in conjunction with $\mathcal{BC}$ [9].

A much more extensive list of relevant research contributions, as well as links to our papers, is available from the application webpage.

5 Re-usability

We will continue our tradition of releasing components of our system to support their broad use within the research community.

Development for BWI is done "in the open" on GitHub. We maintain public repositories and release software that comprises a fully integrated system for autonomous service robots for the every-day, human-inhabited office environment. Our system has been built on top of the Robot Operating System (ROS) middleware framework [11], and is available open-source at `https://github.com/utexas-bwi/` under the BSD license. Included capabilities: Multi-robot control, navigation, floor-switching when the robot takes an elevator, symbolic/logical navigation integrated with a symbolic planner, planning and reasoning for high-level tasks.

The research platforms of Prof. Thomaz and Prof. Niekum are both a part of the HLP-R software system and share much of their code and development through this project. Included capabilities: kinesthetic teaching pipeline, primitives and tools for manipulation, and extensive, modular perception library. This software also supports research in Prof. Sonia Chernova laboratory at Georgia Tech. The HLP-R packages are available at `https://github.com/HLP-R` under the BSD license.

Our TEXPLORE code provides an open-source package for reinforcement learning on real robots.

Our ar_track_alvar ROS package has become a community standard for tag-based perception

Our ROS implementation of Dynamic Movement Primitives has become a popular tool for learning from demonstration.

We have also made research code available for: Bayesian changepoint detection, active articulated model estimation, and Bayesian nonparametric skill learning from demonstration.

6 Applicability to the Real World

Beginning this competition season, we will establish a permanent testing facility mirroring a standard competition arena. This will be hosted in a new laboratory space, and will become a home to our RoboCup@Home team. This laboratory space will supplement labs which simulate real-world scenarios in Thomaz and Niekum's laboratories. Additionally, the Building-Wide Intelligence project hosts robots in the Gates-Dell Complex at UT, and continually strives to deploy its fleet of autonomous robots in a longitudinal study in the real-world scenario or our actual computer science department. RoboCup@Home mirrors these efforts, and our team acts with the goal of unifying our code bases through our participation in RoboCup@Home. This is to say, our competition code is also code that is intended for longitudinal real-world deployment and is tuned for each of our experiments in our laboratories. Prof. Thomaz's also serves as the CEO of Diligent Robotics, a start-up aiming to bring service robotics to hospitals. This venture is seeing the direct application of her research, robotics platform, Poli, and its supporting software.

7 Plans for 2018 Competition

For 2018, our primary goal is to step up our efforts on two fronts. The first is to create a comprehensive system which enters into this year's competition, rather than a series of task-specific systems which compete in each of the competition's tasks. The second is to identify important research goals to our group which we are able to explore through our participation in RoboCup@Home, to bring to bear truly novel technologies which our group is not yet pursuing, which RoboCup@Home inspires us to investigate.

To the first end, our goal is to develop all of Stage 1 as an implementation of General Purpose Service Robot. The BWI project is based around a planner-based architecture which is provided with verbal commands through a semantic parser. Special action executors are paired to actions in the planner. To this end, we will implement the tasks in Stage 1 through our planner-based architecture. If successful, all of this functionality should be available to GPSR, and we will be able to implement the Stage 1 tasks as plans to be executed.

To the second end, it is our belief that once we have implemented Stage 1 in this fashion, that improvement on Stage 1 tasks will be a matter of enhancement of these action executors and of the representations used in our system. As such, relevant research goals will identify themselves in the forms of the technological gaps revealed in our attempt to implement the system in this fashion. This will have a real-world impact on our organization in the form of RoboCup@Home tasking informing our group's primary research goals. Technologies developed for RoboCup@Home will thus find their ways into BWI and HLP-R.

8 Conclusion

Austin Villa @ Home is positioning itself not only for victory in RoboCup@Home 2018, but also to use RoboCup@Home as a driver for the research across our organization. We look forward to seeing you all in Montreal.

Bibliography

References

1. Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.
2. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.
3. Keisuke Nakamura, Kazuhiro Nakadai, Futoshi Asano, Yuji Hasegawa, and Hiroshi Tsujino. Intelligent sound source localization for dynamic environments. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 664–669. IEEE, 2009.
4. Jesse Thomason, Aishwarya Padmakumar, Jivko Sinapov, Justin Hart, Peter Stone, and Raymond J. Mooney. Opportunistic active learning for grounding natural language descriptions. In *Proceedings of the 1st Annual Conference on Robot Learning (CoRL-17)*, 2017.
5. Scott Niekum, Sarah Osentoski, George Konidaris, Sachin Chitta, Bhaskara Marthi, and Andrew G Barto. Learning grounded finite-state representations from unstructured demonstrations. *The International Journal of Robotics Research*, 34(2):131–157, 2015.
6. Scott Niekum, Sarah Osentoski, George Konidaris, and Andrew G Barto. Learning and generalization of complex tasks from unstructured demonstrations. In *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*. IEEE, 2012.
7. Scott Niekum, Sarah Osentoski, Christopher G Atkeson, and Andrew G Barto. Online bayesian changepoint detection for articulated motion models. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
8. Ilaria Gori, Jivko Sinapov, Priyanka Khante, Peter Stone, and JK Aggarwal. Robot-centric activity recognition 'in the wild'. In *Social Robotics*, pages 224–234. Springer International Publishing, 2015.
9. Piyush Khandelwal, Fangkai Yang, Matteo Leonetti, Vladimir Lifschitz, and Peter Stone. Planning in action language $\lfloor\rfloor$ while learning action costs for mobile robots. In *International Conference on Automated Planning and Scheduling (ICAPS)*, June 2014.
10. Thomas Eiter, Wolfgang Faber, Nicola Leone, Gerald Pfeifer, and Axel Polleres. Answer set planning under action costs. *Journal of Artificial Intelligence Research (JAIR)*, 2003.
11. Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5, 2009.

Addendum

Team Name UT Austin Villa
Team Members
Professors
- Raymond Mooney, Scott Niekum, Luis Sentis, Peter Stone, Andrea Thomaz, Justin Hart

Students
- Nicolas Brissonneau, Daniel Brown, Yuchen Cui, Taylor Kessler Faulkner, Rolando Fernandez, Reymundo Alex Gutierrez, Ajinkya Jain, Yuqian Jiang, Steven Jens Jorgensen, Priyanka Khante, Minkyu Kim, Akanksha Saran, Kathryn Baldauf, Ashay Lokhande, Rishi Shah, Benjamin Singer, Nick Walker, Sam Gunn, Neil Patil, Sean Kirmani, Jamin Goo, Mayuri Raja, Jeffrey Huang

Contact Information
Prof. Justin Hart (*Team Leader*) *hart@cs.utexas.edu*
Prof. Raymond Mooney *mooney@cs.utexas.edu*
Prof. Scott Niekum *sniekum@cs.utexas.edu*
Prof. Luis Sentis *lsentis@austin.utexas.edu*
Prof. Peter Stone *pstone@cs.utexas.edu*
Prof. Andrea Thomaz *athomaz@ece.utexas.edu*
Website URL `http://www.cs.utexas.edu/~AustinVilla/?p=athome`
Hardware For Domestic Standard Platform League we will be using a Toyota Human Support Robot (HSR) to be furnished by Toyota.
External Devices
- Google Speech API
- Alienware Alpha R2 (Computer)

Third-Party Software
- Clingo
- MoveIt
- Robot Operating System (ROS)

Robot HSR Hardware Description

Specifications are as follows:

- Name: Human Support Robot (HSR)
- Footprint: 43 cm
- Robot dimensions: height: 1.35m (max)
- Robot weight: 37 kg.

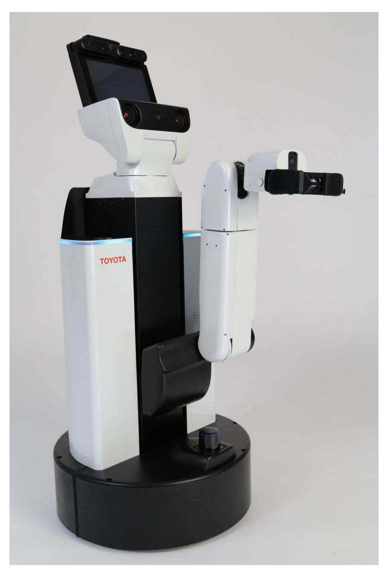


Fig. 2: Toyota HSR