CMPepperBot: Collaborative Task Robot

Manuela Veloso, Michiel de Jong, Anahita Mohseni-Kabir, Vittorio Perera, Travers Rhodes, Aaron Roth, Robin Schmucker, Kevin Zhang, Chenghui Zhou

[1]**Carnegie Mellon University**, Machine Learning Department, Computer Science Department, Robotics Institute

Abstract. We propose to participate in the new RoboCup@Home Standard Platform League with the Pepper robot. Our proposed work is composed of three main components: (i) autonomous navigation and localization in diverse environments; (ii) bi-directional natural human-robot interaction; and (iii) scene understanding, description, and reasoning. We build upon our extensive work in autonomous robots, including our mobile CoBot robots, which have been active in our university buildings for the last six years performing item-delivery and escort service tasks. We also propose three new RoboCup@Home tasks: (i) Replicating a Configuration of Objects, (ii) Balloon Sorting, and (iii) Pepper Self Knowledge and Human Understandable Communication. Although we have not recently participated in the RoboCup@Home competition, we have extensive experience at RoboCup SSL and RoboCup SPL.

1 Introduction

We have researched autonomous robots for decades and have been pursuing seamless integration of perception, cognition, and action. We have developed a variety of autonomous robots, including robot soccer teams [9], service robots [18], and human-robot interaction and learning robots [11]. This proposal builds upon this extensive past work of our team. We have gathered a team with expertise in autonomy, language, vision, and human-robot interaction, all integral to the new RoboCup@Home Standard Platform with the Pepper robot.

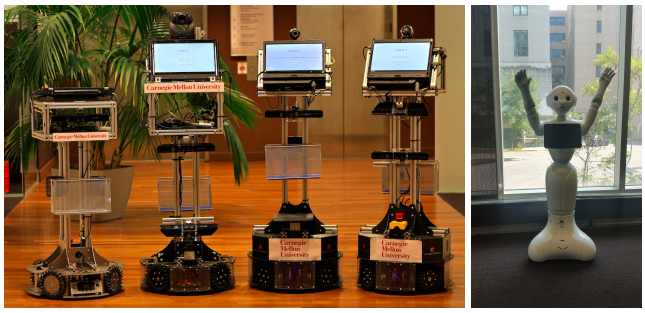
We have been working with the Pepper robot for the past year. We were able to develop full autonomous map learning, navigation, and localization through SLAM, and natural interaction through face learning. We were further able to install ROS within NAO-Qi. Our team has also implemented improved object and posture recognition, and investigated extensions to further on-board and off-board speech recognition.

In addition to discussing our past and future research, we also propose three new tasks for the competition. These tasks will inspire teams to push the Pepper robots to accomplish new feats, as well as make the tournament even more exciting for spectators and the public at large.

Finally, our proposed work complies to our pursuit of a safe, beneficial, and robust co-existence of humans and autonomous robots. We point the readers to an interview with Manuela Veloso, http://www.theverge.com/a/verge-2021/humanity-and-ai-will-be-inseparable which captures the underlying philosophy we aim with our proposed research.

2 Previous Work

We research a variety of robot platforms. Figure 1 illustrates our robots. We very briefly present the corresponding underlying research.



CoBots Pepper

Fig. 1. Our Service and Interaction Robots

2.1 Work on CoBots

We research, develop, and deploy multiple autonomous mobile robots. Our CoBots (Figure 1) are capable of performing tasks requested by users in our multi-floor office building [18, 19]. To successfuly perform service tasks, our robots have several core capabilities. The CoBots autonomously localize and navigate in the diverse types of indoor space, including corridors, elevators, and open areas with movable furniture and people, distinguishing map-known long-term features (walls) or map-missing short-term (furniture) and dynamic (people) features, supporting an effective novel overall episodic non-Markovian localization [4, 1–3]. The task scheduler for CoBots schedules conflict-free plans for multiple robots to satisfy constrained tasks specified and requested by users, capable of transferring items among them to optimize traveled time [7]. The CoBots have sensing, cognition, and actuation limitations. To overcome the robots' own limitations through symbiotic autonomy, namely by proactively asking for help from humans, and accessing and learning from the web [8, 14, 16, 17].

2.2 Work on Human-Robot Interaction

Understanding and communicating with humans is integral to the development of a mobile service robot. In order to optimize Cobot's communication capabilities, we have developed a probabilistic model, coupled with a knowledge base and a dialog system that not only enables our CoBot robots to understand and execute spoken commands, but also to learn about the environment they are deployed in [8, 12].

More recently, we have introduced novel contributions enabling our CoBot robot to respond to requests for information by explaining Cobot's past actions and experiences in natural language. [13] We define the concept of *verbalization*

as "the process by which an autonomous robot converts its own experience into language" and the idea of a *verbalization space* that represents the variations in possible explanations for the same robot experience [15]. We showed how it is possible to learn a mapping from the language used to ask robots for a verbalization task to specific point in the verbalization space [10]. Learning this mapping allows the robot to generate explanation that better matches the user expectations.

2.3 Work on Pepper

Since acquiring Pepper, the team has been experimenting with and extending Pepper's functionality. We have developed programs that allow Pepper to follow behind a human walking, move according to a human's gesture-based directives, imitate a human's arm posture, recognize and tell apart human faces, and detect and identify objects.

Succesfully using these skills requires a combination of robust vision and speech processing. To that end, we have augmented Pepper's inbuilt vision and speech processing software with external resources, using OpenPose and Yolo9000 to implement posture detection and object recognition, and using Google Cloud Speech to improve the accuracy of speech recognition.

In order to improve the robustness of Pepper's communication capabilities, we have also experimented with fall-back means of communication, such as text and button input on Pepper's tablet, in the event that voice or visual commands are not successfully perceived or processed.

Additional software used: Robot Operating System, spaCy, and LinearCRF.

3 Proposed Research Approach

Our proposed research directions are composed of three main components: (i) autonomous navigation and localization in diverse environments; (ii) bi-directional natural human-robot interaction; and (iii) scene understanding, description, and reasoning.

3.1 Autonomous Navigation and Localization in Diverse Environments

We have developed robust autonomous navigation and probabilistic localization in varying environments [5], strongly relying on the perception of long-term features, such as walls, and short-term static features, such as tables and chairs. We will focus on these capabilities in diverse environments, including crowded areas, large open areas, and corridor-bounded areas. We will equip the robot with a suite of different localization and navigation algorithms suitable for different environments. Our goal is to enable the robot to learn to detect the different types of environments and automatically adjust its algorithms and their parameters to seamlessly and robustly navigate while transitioning among different environments.

We envision this focus to be of great relevance to the RoboCup@Home competition as we expect the robots to require the capability to face home-like rich-featured corridor-bounded environments, as well as crowded and open-space environments. We expect our research to enable the Pepper robot to combine vision-detected RGB features, vision-detected depth features, as well as sound and WiFi features. We will also research on enabling the robot to learn from instruction from humans, as well as to be corrected by humans in specific conditions.

3.2 Bi-Directional Natural Human-Robot Interaction

In our research on Human-Robot Interaction we embrace a Bi-Directional approach, in the sense that natural-language-based communication will flow *from a human to the robot* (e.g., when users assign tasks to the robot or query its state), as well as *from the robot to a human* (e.g., when the robot pro-actively requests additional information to perform tasks or when the robot offers a description of its own experience). As discussed earlier, we have developed a system for communication on CoBot that allows it to respond to complex commands, learn about its environment and verbalize its past experiences. We will implement the same system on our Pepper service robot, and propose to strongly build upon this research on understanding and verbalization and enable the Pepper robot to be able to communicate and explain its autonomous experience.

3.3 Scene Understanding, Description, and Reasoning

We assume that robots, in particular also in RoboCup@Home, will need to navigate through complicated scenes that involve humans performing various kinds of motions. In our current work, our navigation algoritms in CoBot are able to avoid the obstacles in the path of CoBot, including humans, but we would like to go one step further and have the autonomous mobile robot, Pepper and also CoBot,proactively interact with the humans in the scene.

Based on our previous work on human detection using depth images [6], we will mostly focus on making use of such data from the humans. Before any interactions, the robot needs to understand the humans' intents in the scene. We propose to introduce several dimensions and features to capture human intent, such as the human motion direction, speed, and acceleration. We will use machine learning to associate these clues with intent on whether the human is walking towards the robot or not, and if so, is he/she intended to stop at the robot. If the human has the intention to stop at the robot, then we propose to develop a robot behavior to appropriately respond to the human intention, such that the interactive robot should go towards the human and stop accordingly; or if the human is simply minding their own business, the robot should continue moving on its own path.

We will assume three types of motions – constant position, constant velocity and constant acceleration – as a starting point and develop algorithms to learn and classify the human motions. Learning the velocity and acceleration provides

a more accurate description of the human trajectory, such that the robot can reason about the appropriate time to interact based on past experience. For example, the robot should predict the trajectory of the human, such that it is not too far or too late to greet him/her. After we achieve the interaction with a single human, we generalize to more complex scenarios with multiple humans.

4 Proposed New RoboCup Tasks

In this section we propose additional tasks for the Robocup@Home competition that will require Pepper to utilize a range of abilities, combining mobility, vision, speech, and social interaction. The suggested activities go beyond simply issuing commands to Pepper, and involve human-robot collaboration at the task level. Pepper and a human will work together towards a purpose, just as we hope for humans and robots to partner and collaborate in real-world environments.

The proposed tasks will also increase the entertainment value of the competition. The faster-paced nature of the task, the transparent scoring, and the ability of a spectator to get a sense of the general success or failure forward this goal. The design of these tasks was driven both by a desire to encourage new research and by an urge to engage the public. A public that is excited about the possibilities of collaborative social robots will be a boon to the field as a whole.

We further propose that each task described below would be performed multiple times (for example, 3 times), and the average or maximum score from the runs is taken as the final score for that task.

4.1 Replicating a Configuration of Objects

In this task, Pepper is allowed to view a configuration of objects. A human in a different room (or separated by a screen) has access to all of these object arranged arbitrarily. First, Pepper must describe the scene in front of it and instruct the human how to arrange their objects in an identical manner. Second, Pepper moves to the human's area and views their setup. If the setup is wrong in some fashion, Pepper will instruct the human to make the appropriate changes.

Task Details When the task begins, Pepper can view objects in front of it. Objects can include furniture such as tables, chairs, and stools, as well as smaller objects such as food containers and blocks. Objects could differ in size, color, and shape. Pepper will describe what is sees, using relational language. For example, "Place the green can on top of the yellow plate that is in the center of the table."

This first part of the task is limited to 5 minutes. Afterwards, a tournament official will go to the human's configuration and introduce at least one additional error that Pepper will have to correct. At this time, Pepper is led to the human's area for the second part of the task. After viewing the scene, Pepper should assess the errors present in the configuration and give additional instructions to fix them, such as "There should not be a coke bottle in this scene" or "The water bottle should be on the red plate, not the yellow plate."

Successive rounds might have increasing number of objects.

Scoring Pepper will have 5 minutes for the first phase (description) and 5 minutes for the second phase (critiquing the human's results). Teams will be scored according to the number of objects correctly placed in the first phase, both in terms of absolute position and in relationship to each other, and the number of successful corrections in the second phase.

4.2 Balloon Sorting

Balloons are in room 1. Pepper must gather the balloons from room 1 and put them in room 2, while following instructions. A human participant might instruct it saying, "bring at least 5 red balloons to the living room" or "leave 2 blue balloons in the other room." This is a competitive, timed task.

Task Details There are two rooms in this task. They may be labeled, such as the "bedroom" and the "living room". At the beginning of the task there are 10 balloons total in one of the rooms. There is a ceiling fence in both rooms so that the balloons do not rise above 8 feet high. Balloons are yellow, red, green, and blue, in any number of each color. Each balloon has a string with a loop on the bottom at the height of Pepper's head that can serve as a handle that Pepper is capable of grabbing.

A tournament official gives Pepper instructions regarding bringing balloons to the living room. These instructions could be piecemeal ("Please bring 3 red balloons. Now, please bring 2 blue balloons.") or more expansive ("Please bring all balloons to the living room."). Instructions could change over time, as a human changes their mind. For example, "I want 4 red balloons, not 3." Pepper could be instructed to bring balloons back from room 2 to room 1, too, as in, "Remove the yellow balloons from the living room!" If Pepper receives "conflicting" instructions, it should follow the most recent instruction (as well as previous non-conflicting instructions). Humans collectively will issue no more than 5 instructions total over the course of the task.

Scoring Pepper will have 10 minutes to complete the task. The score will take into account what instructions Pepper was able to complete within the allotted time (judged by number of balloons moved out of total balloons to be moved). Balloons moved in error will count against the score.

4.3 Pepper Self Knowledge and Human-Understandable Communication

Tournament officials prepare a series of commands (hidden from competing teams) which they issue to Pepper during this task. After performing them, Pepper must describe what they just did in a high-level, human-understandable fashion.

Task Details Tournament officials prepare a series of low-level commands that are not known by teams beforehand. These commands could include standard verbal commands, and could also be anything in the standard naoqi API, that will be fed to Pepper as a script.

At the beginning of the task, tournament officials instruct Pepper to start making note of its own actions. Then, the tournament officials issue commands verbally or electronically for Pepper to execute. After carrying out the commands, Pepper is asked to describe what it has done in a high-level human understandable way. For example, if a command was issued to rotate Pepper's shoulder a certain number of degrees, Pepper might say, "I raised my arm" or, "I wave hello in greeting."

Scoring Pepper has 5 minutes for the description part of the task. Points are awarded based on how much of the actions is encompassed by the description (maximum points for a description that touches upon each action performed) as well as quality of a description. Saying "I wave hello" is higher quality than "I raised my arm," however if such a higher-level meaning doesn't make sense in context points may be deducted instead.

4.4 Benefits of Proposed Tasks

Any of the above noted tasks will offer opportunities for technical innovation among the participating teams, as well as generate a more competitive and enthralling tournament atmosphere for all humans in attendance.

Upon selection of any of these tasks for inclusion in Robocup@Home, we commit to being engaged in specifying these tasks further, to the level of detail and specificity required for the official rulebook.

5 Conclusion and Expected Contributions

Our proposed work will advance our research in autonomous navigation and localization in diverse environments; new bi-directional natural-language-based human-robot interaction, capable of describing experience and informing humans of abnormal experience; and scene understanding, description, and reasoning to enable robots to best respond to any situation.

We hope our proposed tasks could heighten the fun nature of the RoboCup@Home event. Teams' attempts to fulfill these tasks would result in new research that will bring closer a day when autonomous humanoid robots work together with people in their own homes.

We believe we have a strong starting point to contribute to the 2018 RoboCup@Home competition, given our extensive experience with multiple autonomous robots, as well as concrete experience with the Pepper robot. Given our NaoQi and ROS development environments, we will be committed to sharing our developments and code to other teams in the competition. We have a long experience of such sharing since our pioneering AIBO robot and simulation soccer teams, as well

as crucial modules of SSL. We will also make sure that all our developments and advances with Pepper build upon and are usable by other robots, in particular the CoBot robots.

References

1. J. Biswas, B. Coltin, and M. Veloso. Corrective gradient refinement for mobile robot localization. In *Proceedings of IROS*, pages 73–78, 2011.
2. J. Biswas and M. Veloso. Depth camera based indoor mobile robot localization and navigation. In *Proc. of ICRA*, pages 1697–1702, 2012.
3. J. Biswas and M. Veloso. Localization of the CoBots over Long-term Deployments. *Int. Journal of Robotics Research*, 32(14):1679–1694, 2013.
4. Joydeep Biswas. *Vector Map-Based, Non-Markov Localization for Long-Term Deployment of Autonomous Mobile Robot.* PhD thesis, CMU-RI-14-25, 2014.
5. Joydeep Biswas and Manuela M. Veloso. Episodic non-markov localization. *Robotics and Autonomous Systems*, 87:162 – 176, 2017.
6. B. Choi, C. Mericli, J. Biswas, and M. Veloso. Fast Human Detection for Indoor Mobile Robots Using Depth Images. In *Proceedings of ICRA'13*, 2013.
7. Brian Coltin. *Multi-agent Pickup and Delivery Planning with Transfers.* PhD thesis, Carnegie Mellon University, 2014. CMU-RI-14-05.
8. T. Kollar, V. Perera, D. Nardi, and M. Veloso. Learning Environmental Knowledge from Task-Based Human-Robot Dialog. In *Proc. of ICRA*, 2013.
9. J.P. Mendoza, J. Biswas, D. Zhu, P. Cooksey, R. Wang, S. Klee, and M. Veloso. Selectively Reactive Coordination for a Team of Robot Soccer Champions. In *Proceedings of AAAI*, 2016.
10. V. Perera, S. Selvera, S. Rosenthal, and M. Veloso. Dynamic generation and refinement of robot verbalization. In *Proceedings of Ro-Man*, 2016.
11. Vittorio Perera, Sai P Selveraj, Stephanie Rosenthal, and Manuela Veloso. Dynamic generation and refinement of robot verbalization. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 212–218. IEEE, 2016.
12. Vittorio Perera, Robin Soetens, Thomas Kollar, Mehdi Samadi, Yichao Sun, Daniele Nardi, René van de Molengraft, and Manuela Veloso. Learning task knowledge from dialog and web access. *Robotics*, 4(2):223–252, 2015.
13. Vittorio Perera and Manuela Veloso. Learning to understand questions on the task history of a service robot. In *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*. IEEE, 2017.
14. S. Rosenthal, J. Biswas, and M. Veloso. An Effective Personal Mobile Robot Agent Through Symbiotic Human-Robot Interaction. In *Proc. of AAMAS*, 2010.
15. S. Rosenthal, S. Selvaraj, and M. Manuela. Verbalization: Narration of autonomous mobile robot experience. In *Proceedings of IJCAI*, 2016.
16. S. Rosenthal, M. Veloso, and A. Dey. Learning Accuracy and Availability of Humans who Help Mobile Robots. In *Proceedings of AAAI*, 2011.
17. Stephanie Rosenthal. *Human-Centered Planning for Effective Task Autonomy.* PhD thesis, Carnegie Mellon University, 2012. CMU-CS-12-110.
18. M. Veloso, J. Biswas, B. Coltin, and S. Rosenthal. CoBots: Robust Symbiotic Autonomous Mobile Service Robots. In *Proceedings of IJCAI*, 2015.
19. R. Ventura, B. Coltin, and M. Veloso. Web-Based Remote Assistance to Overcome Robot Perceptual Limitations. In *AAAI Workshop on "Intelligent Robotic Systems"*, 2013.