Web-Based Remote Assistance to Overcome Robot Perceptual Limitations*

Rodrigo Ventura
Institute for Systems and Robotics
Instituto Superior Técnico
1049-001 Lisbon, Portugal

Brian Coltin and **Manuela Veloso**
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA

Abstract

This paper addresses the problem of overcoming visual perception limitations in service robots with remote assistance from human users. In particular, consider a scenario where a user requests the robot to perform some task that requires a perceptual ability, e.g., check if a specific mug, "my mug," is in the lab or in an office, but the robot may not know how to recognize that object. We propose to equip the robots with the abilities to: (i) identify their own perceptual limitations, (ii) autonomously and remotely query human users for assistance, and (iii) learn new object descriptors from the interaction with humans. We successfully developed a complete initial version of our approach on our CoBot service mobile robot. The interaction with the user builds upon our previously developed semi-autonomous telepresence image sharing and control. The user can now further identify the object and the robot can save the descriptor and use it in future situations. We illustrate our work with the task of learning to identify an object in the environment, and to report its presence to a user. Our ongoing work includes addressing a dynamic interaction between the robot and the remote user for visual focus of attention and different object viewing, as well as the effective storage, labeling, accessing, and sharing of multiple learned object descriptors, in particular among robots. Our goal is also to contribute the learned knowledge to crowd-robotics efforts.

Introduction

One of the challenges of having mobile robots autonomously perform tasks is dealing with the robots' limitations. For instance, when a task specification is ambiguous, as often happens when natural language is used, e.g., "get me my mug," "where is my backpack," or when the robot is not able to deal with a novel situation, e.g., unable to fit a new kind of kitchenware into the dishwasher machine, or when the robot faces an unexpected situation that prevents the execution of a task such as a closed door or misplaced trash bin.

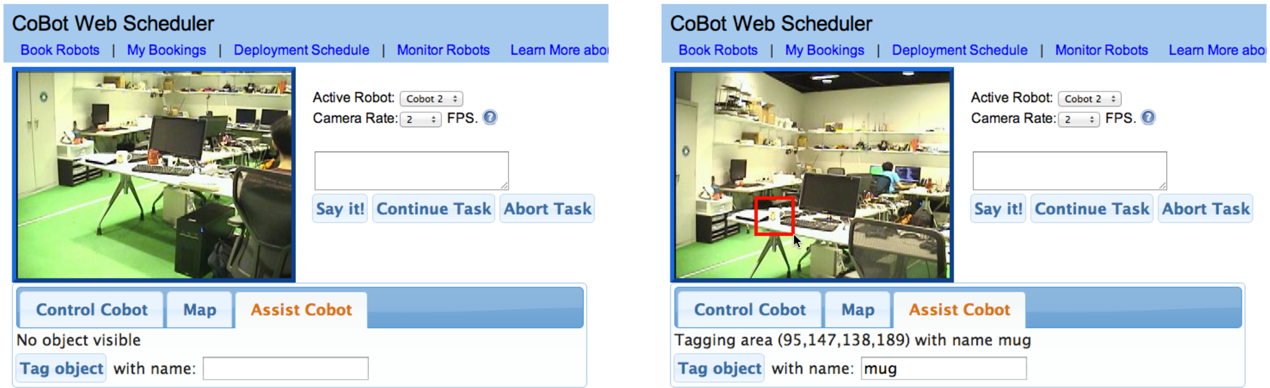
A possible approach to this problem is to endow robots with increasing physical or computational complexity. However, this complexity seems unbounded in reality and the robots will inevitably encounter unexpected unknown situations, especially once robots get out of controlled lab environments.

Instead, *symbiotic autonomy* has been actively pursued (Rosenthal and Veloso 2010; Rosenthal, Veloso, and Dey 2011; Samadi, Kollar, and Veloso 2012; Veloso et al. 2012), where robots proactively and autonomously ask for help. This approach has been deployed successfully in office environments. However, interaction has been focused on asking for actuation and localization help from humans collocated with the robot, and accessing the web for information of location of an object described from a speech-based interaction. This paper explores now the challenge of extending the approach so that the robot autonomously seeks assistance by summoning and interacting with a *remote* human about, and to overcome, its *visual perception* limitations.

Symbiotic autonomy has been implemented in the CoBot robots (Veloso et al. 2012) and deployed successfully in office environments. Tasks are requested and scheduled using a web-based interface (Veloso et al. 2012). In specific situations, the execution of CoBot's tasks require assistance from users collocated with the robot. In particular, traveling between floors requires a human to press the elevator button, to hold the elevator door open while CoBot enters the elevator, and to inform the robot of whether it is on the right destination floor. However, interaction during task execution has been limited humans collocated at the robot's location.

Several efforts have used the web to access robots, with early examples including the teleoperation of a robotic arm (Goldberg et al. 1995; Taylor and Trevelyan 1995), and interfacing with a mobile robot (e.g, (Simmons et al. 1997; Siegwart and Saucy 1999; Saucy and Mondada 2000; Schulz et al. 2000)) to which users placed requests over the web, namely for the robot to go to specific places. Notably the RoboCup@Home initiative (Visser and Burkhard 2007) provides competition setups for indoor service autonomous robots, with a wide scope of challenges increasing yearly, focusing mainly on robot autonomy and verbal interaction with users. However, most of these efforts assume that robots embody all capabilities necessary to per-

*This work was supported by project FCT [PEst-OE/EEI/LA0009/2011]



(a) CoBot Telepresence interface allowing the user to tag an object in the robot camera image stream.

(b) The user selects a region in the camera image containing the object

Figure 1: Telepresence interface of CoBot, used to assist it to visually learn and identify objects. This interface shows the camera image stream, together with the GUI elements contained in the "Assist Cobot" tab at the bottom of the screenshots.

form their tasks autonomously. A system for remote assistance of a semi-autonomous field robot was presented, focusing on remotely assisting the robot in exception-handling scenarios (Murphy and Rogers 1996).

Assisting the Robot to Locate Objects

Here we consider the problem of assisting a user to locate a specific object in the environment, e.g., in a multi-floor building. We assume a conventional pan-and-tilt camera as the main vision sensor onboard the robot. Objects are identified using feature descriptors detected in the camera image stream.

User assistance is provided by means of the CoBot telepresence interface (Coltin et al. 2011), which allows the user to move the robot and orient the camera freely. When the user intends to teach the robot a new object, he employs the web interface to tag a rectangular region in the camera image around the object location, as illustrated in Figure 1(b). Then, a tuple $\langle D, l, t\rangle$ containing the set of descriptors D extracted from features detected inside the tagged region, where each descriptor $d_i \in D$ is a 128 dimensional vector, the robot location l, and the tag string t is stored in a global database.

Object detection is performed with a nearest neighbor descriptor matcher. Once a match is found, a bounding box is drawn around the matched features on the image, together with the associated tag. This information can then be used to trigger specific actions in the robot task planner, such as informing the user that the object was found, or moving to another location to continue searching.

The successful detection of previously tagged objects in the robot camera opens the door to some interesting possibilities. For instance, looking for an object in a building can be performed autonomously by,

1. navigating towards the most likely locations where the object can be (Samadi, Kollar, and Veloso 2012), and then
2. using the stored feature descriptors to determine whether it is there.

Once the system database is populated with tagged objects, the robot can map (as a background process) their locations across the whole environment, e.g., a multiple floor building, as they are encountered. Maintaining a shared database of previously encountered objects allows the robots to know tagged object locations autonomously.

Illustrative example

In the initial version we used the telepresence interface equipping the remote web interface of CoBot. The robot camera used for telepresence tasks was used to acquire visual features. We used off-the-shelf object detection routines found in the OpenCV library[1]. The SIFT feature descriptors were used (Lowe 2004), together with a brute-force nearest neighbor descriptor matcher, both implemented in OpenCV.

Figure 1 shows the telepresence interface (left) and the operation of tagging an image region and associating it with the tag string "mug" (right). Then, whenever the robot identifies in the image features whose descriptors match an object in the database, a rectangle is drawn surrounding the features, together with the associated tag. This is shown in Figure 2.

Figure 2: Detection of a previously identified object (mug).

[1] `http://opencv.org/`

Additional objects can be tagged by the user. In this version, all objects with matching features are identified in the image. Since the database is shared among CoBots, whenever an object is tagged by an user, all other robots become readily capable of identified that object in their own camera images. In other words, robots learning objects from user assistance is a form of crowd-sourcing object recognition capabilities among them.

Conclusions and Future Work

This paper presented an approach to overcome visual perception limitations of service robots. In particular, we explored the assistance a remote user can provide to robots autonomously performing tasks. In particular, we described an initial version of the system where users assist the robot identifying objects in the camera image. We also reported on the initial version of the implementation on the CoBot robots, where a remote user teaches the robot to visually recognize objects by tagging them in the camera image. Then, the robot is shown to identify the tagged object location in the image. Future work includes the integration of this approach with the task planner, and the extension of these ideas to other situations where remote user assistant is fundamental for the robot to complete its tasks.

References

Coltin, B.; Biswas, J.; Pomerleau, D.; and Veloso, M. 2011. Effective semi-autonomous telepresence. In *Proceedings of the RoboCup Symposium*.

Goldberg, K.; Mascha, M.; Gentner, S.; Rothenberg, N.; Sutter, C.; and Wiegley, J. 1995. Desktop teleoperation via the world wide web. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 1, 654–659.

Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2):91–110.

Murphy, R. R., and Rogers, E. 1996. Cooperative assistance for remote robot supervision. *Presence* 5(2):224–240.

Rosenthal, S., and Veloso, M. 2010. Using symbiotic relationships with humans to help robots overcome limitations. In *Proceedings of the AAMAS'10 Workshop on Collaborative Human/AI Control for Interactive Experiences*.

Rosenthal, S.; Veloso, M.; and Dey, A. 2011. Task behavior and interaction planning for a mobile service robot that occasionally requires help. In *Proceedings of the AAAI Workshop on Automated Action Planning for Autonomous Mobile Robots*.

Samadi, M.; Kollar, T.; and Veloso, M. 2012. Using the web to interactively learn to find objects. In *Proceedings of AAAI'12, the Twenty-Sixth Conference on Artificial Intelligence*.

Saucy, P., and Mondada, F. 2000. KhepOnTheWeb: open access to a mobile robot on the internet. *IEEE Robotics Automation Magazine* 7(1):41–47.

Schulz, D.; Burgard, W.; Fox, D.; Thrun, S.; and Cremers, A. 2000. Web interfaces for mobile robots in public places. *IEEE Robotics Automation Magazine* 7(1):48–56.

Siegwart, R., and Saucy, P. 1999. Interacting mobile robots on the web. In *Proceedings of the IEEE International Conference on Robotics and Automation*.

Simmons, R.; Goodwin, R.; Haigh, K. Z.; Koenig, S.; and O'Sullivan, J. 1997. A layered architecture for office delivery robots. In *Proceedings of the first international conference on Autonomous agents (AGENTS'97)*, 245–252.

Taylor, K., and Trevelyan, J. 1995. Australia's telerobot on the web. In *Proceedings of the International Symposium on Industrial Robots*, volume 26, 39–44.

Veloso, M.; Biswas, J.; Coltin, B.; Rosenthal, S.; Brandao, S.; Mericli, T.; and Ventura, R. 2012. Symbiotic-autonomous service robots for user-requested tasks in a multi-floor building. In *Workshop on Cognitive Assistive Systems (CAS), IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-12)*.

Visser, U., and Burkhard, H.-D. 2007. RoboCup: 10 years of achievements and future challenges. *AI Magazine* 28(2):115–132.