Towards Enabling a Robot to Effectively Assist People in Human-Occupied Environments

Sachithra Hemachandra and **Ross Finman** and **Sudeep Pillai** and
Seth Teller and **Matthew R. Walter**
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA
{sachih,rfinman,spillai,teller,mwalter}@csail.mit.edu

Abstract

Over the last decade, we have seen an increasing demand for robots capable of coexisting with people. Enabling robots to operate safely and effectively alongside human partners within unstructured environments poses interesting research challenges that broadly span the field of artificial intelligence. This paper previews some of the challenges that we faced in developing a robot "envoy" that operates for extended periods of time throughout an office-like environment, assisting human occupants with everyday activities that include greeting and escorting guests as well as retrieving and delivering objects. We see three skill areas as critical for the robot to effectively perform these tasks. The first is shared situational awareness—the robot must interpret its environment through a world model that is shared with its human partners. Secondly, the robot should act in a safe, predictable manner and be capable of intuitive interaction with people, through such means as natural language speech. Thirdly, the robot, which we initially treat as a rookie, should efficiently utilize information provided by human partners, requesting assistance when necessary and *learning* from such assistance to become more competent with time.

Introduction

Traditionally seen as tools used in isolation for industrial automation, robots are becoming more pervasive within unstructured human-occupied environments. Our lab is currently developing a robot that is able to maintain a persistent presence alongside people in a workplace setting. The Envoy Project is an effort to enable a mobile robot to serve as an office assistant that works alongside people in a collaborative and socially acceptable manner. One task that the robot would perform in this role is that of interacting with lab visitors. The robot is currently capable of detecting and interacting with visitors and can interpret and execute spoken commands that refer to escorting the user across floors to their destination. Additionally, we would like the robot to be able to retrieve known objects based upon a simple set of directions, such as fulfilling the spoken request to "please bring the blue notebook that I left on my desk."

In this paper, we address three research areas that we feel are important to ensuring that the Envoy robot works effectively with and is accepted by people. First, the robot must

Figure 1: The Envoy robot platform.

exhibit shared situational awareness, which means that it interprets its environment, objects, and events in a manner that is similar to our own world model and that the robot makes its model transparent to the user. Secondly, the robot should be capable of intuitive interaction with people that includes the ability to understand natural language speech. Its actions should be transparent and it should operate in ways that are predictable and socially acceptable. Thirdly, the robot should be able to request assistance when necessary so as to avoid failure and utilize information that is provided as efficiently as possible. We treat the robot as initially being a "rookie" that can not perform certain tasks on its own (e.g., pushing elevator call buttons), but that can request assistance and, in the case of current work, learn to become more competent with time. In developing these capabilities, we take advantage of recent advancements in perception, navigation, motion planning, speech interpretation, and learning. Effectively integrating these potentially competing methods into a common framework is one of the challenges we face.

Our platform (Fig. 1) consists of an electric wheelchair with three LIDARs, a Microsoft Kinect camera and boom microphone mounted to a panning head, an IMU, a barometric pressure sensor, a pair of speakers, and a laptop.

Research Challenges

We discuss our recent progress towards developing the aforementioned capabilities for the Envoy platform and outline directions for future research.

Situational Awareness

Shared situational awareness on the part of the robot implies the capacity to reason about its surroundings and actions using a model that is transparent to its human partners.

The robot maintains an occupancy map that it uses for metric localization and a multi-floor semantic map of the environment. The semantic map consists of a coarse topology that includes elevator locations, passageways, and rooms bound to their colloquial names. The robot generates this map through a guided tour scenario (Hemachandra et al. 2011), in which it autonomously follows a person throughout the building, learning the representation as they narrate. We are currently working on a framework that allows people to convey richer semantic maps that model locations, objects, and their affordances through an online tour that emphasizes natural language speech.

In addition to recognizing modeled objects by name, we believe the robot also needs to identify and track people whether they are users or bystanders. Through color and depth segmentation (Felzenszwalb and Huttenlocher 2004) and learned geometric features, the robot detects people using the panning Kinect camera, and tracks them as they move around the robot. The system also uses its LIDARs to independently track users and bystanders.

Human-Robot Interaction

In order for the robot to interact effectively with visitors, we believe that it should be able to interpret task-constrained natural language speech and synthesize responses. The ability to ground speech to the corresponding objects, locations, and actions affords richer interactions (Walter et al. 2010; Tellex et al. 2011). Situational awareness is a critical factor in enabling this rich interaction.

We believe it is equally important that the robot operates safely and that its actions be transparent to bystanders. People are more willing to accept the robot in their presence if its actions are predictable and adhere to social conventions. This means, for example, that the robot should follow predictable paths through the environment, a challenge for sample-based motion planning algorithms that, while being very good at identifying feasible paths, often result in non-obvious solutions. To address this, we make the assumption that routes that are obvious to people are also efficient, and implement an RRT*-based planner (Karaman et al. 2011) to identify optimal paths.

It is also important that the robot adheres to social conventions that dictate how people interact, such as navigating down the right side of a corridor or letting people out of an elevator before entering. Some challenges we've faced include defining or learning these conventions and recognizing when they should govern the robot's actions. Hard-coded policies provide a partial solution in some cases (e.g., waiting to enter an elevator until people stop moving), but the robot should also be capable of predicting the behavior of people in its surroundings. Existing work in motion pattern modeling (Joseph et al. 2011) can be incorporated within a planning framework (Aoude et al. 2010) to support this behavior.

Uncertainty Mitigation and Learning

A fundamental challenge to operating in unstructured environments is dealing with uncertainty. We treat the robot much like a rookie, recognizing that there are actions that it may not yet have the competency to perform, such as summoning elevators, or tasks that can be made simpler with additional information, such as identifying a particular object among clutter. In an effort to improve performance, we enable the robot to request help from and judiciously use information provided by the human. For example, until the robot is able to push elevator buttons, it asks the user to summon the elevator, hold the door, and push the relevant floor at the appropriate time. The challenge we face when generalizing this capability is deciding when assistance is required and what assistance is most helpful.

While we initially allow the robot to ask the user to open doors and operate elevators, our goal is for it to become more competent and independent as it develops. Learning from demonstration (Argall et al. 2009) provides an approach to acquiring new skills by observing a human repeatedly perform a series of tasks. One focus of our current research is to extend the guided-tour scenario to manipulation, whereby the robot opportunistically observes and learns from people manipulating objects in the world, e.g., picking things up and pushing elevator buttons. Of the many challenges is that of being able to generalize the state/action sequences in the context of the human to those that are consistent with the robot's perceptual and mobile manipulation capabilities. In order to make this goal more tractable, we propose a strategy that builds on work in natural language interpretation (Tellex et al. 2011) to allow the user to provide supervision by using speech to ground objects, actions, and causal relationships.

Conclusion

We have outlined some areas of research that we believe are important if our Envoy robot is to work safely and effectively alongside humans. We briefly described our current research and outlined directions for future research.

References

Aoude, G.; Luders, B.; D.S., L.; and How, J. 2010. Threat-aware path planning in uncertain urban environments. In *IROS*.

Argall, B.; Chernova, S.; Veloso, M.; and Browning, B. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57(5).

Felzenszwalb, P., and Huttenlocher, D. 2004. Efficient graph-based image segmentation. *IJCV* 59(2).

Hemachandra, S.; Kollar, T.; Roy, N.; and Teller, S. 2011. Following and interpretting narrated guided tours. In *ICRA*.

Joseph, J.; Doshi-Velez, F.; Huang, A.; and Roy, N. 2011. A Bayesian nonparametric approach to modeling motion patterns. *Autonomous Robots*.

Karaman, S.; Walter, M.; Perez, A.; Frazzoli, E.; and Teller, S. 2011. Anytime motion planning using the RRT*. In *ICRA*.

Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M.; Banerjee, A.; Teller, S.; and Roy, N. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. In *AAAI*.

Walter, M.; Friedman, Y.; Antone, M.; and Teller, S. 2010. Vision-based reacquisition for task-level control. In *ISER*.