Communicating with Mobile Agents

David P. Miller
MITRE Corporation
dmiller@starbase.mitre.org

Introduction

Artificial Intelligence researchers in language have spent the past thirty years creating systems that concentrate on communications among disembodied agents. This has led to a preoccupation with grammar and syntax, huge annotated dictionaries and the problem of converting text to speech and speech to text. Such systems have been produced under the belief that the complete representation of the world tends to lie in the language representation, and everything the agent knows about the universe is either put into its memory a-priori, or comes in through the single channel of natural language communications. This may true for some special cases, but is certainly not the case when one is trying to understand or decode the meaning of speech as it is used when directing mobile agents in real-time.

In the past several years, there has been some attempt to ground some of the symbols generated by the communications system to the physical world. These have concentrated primarily on standard AI type problems of recognizing colored blocks in various positions. Even more recently have been attempts to have an embodied agent interact with the world using communications of this sort.

Unfortunately, all these attempts rely on either the world being heavily constrained, or the agent having sensing capabilities far beyond what is commonly available these days for aartificial agent. In addition, the sensory data that is assumed, or even used, is spatially vacuous, at best providing information in a single dimension.

This paper is concerned with guiding a mechanical agent through structured terrain (an indoor environment. The agent is outfitted with typical robotic sensors (reliable contact sensors, reasonably reliable proximity sensors, and unreliable range and vision sensors). The questions to be answered are: What types of communications between human and agent are necessary to successfully navigate through the world in a goal-directed way? Unfortunately, those questions will not be satisfactorily answered in this paper, but I hope to provide some further insight into why this is an important question, and some possible directions to look for the answers.

The wheelchair problem

The problem of particular interest to me, is that of a person trying to communicate to a robot in which they are riding. More specifically, a person being carried about by a robot wheelchair. Traditionally, this type of communications is carried out by a 2-DoF joystick. The communication is very low-level and consists of giving direction and velocity commands. While this is an effective way of communicating motion desires, it requires continuous and semi-skilled eye-hand coordination. This type of coordination is sometimes beyond the capabilities of people who need the mobility assistance the most.

A joystick communicates specific movements, but says nothing about where (symbolically, or functionally) you are trying to go. The difficulty in adding a more traditional AI communications system is that of getting the mapping from the real-world to the internal representation of the world held by the robot, and the coordination of that with the operator's representation. It's basically the pixels to predicates problem, where you want all the intermediate steps to be operator accessible.

Problem description

A wheelchair user needs to have different relationships to the physical world depending on the exact situation in which they find themselves. While going down the hallway they may want to stay well clear of all objects and other people. If they are talking to someone as they go down the hall, they will want to stay clear of all objects and people with the exception of the person they are accompanying; that person they will want to pace, and maintain their position to the side. When going onto an elevator, they will want to stay clear of people, but go as close as possible to the far side of the elevator, so that they have room to turn around, and won't block the doorway for others. When approaching a desk or table, they will want to actually go inside the vertically defined envelope of the desk, but only to a certain point (or else they risk injury). While these situations are all clearly distinct to you and me, to an agent outfitted strictly with range and proximity sensors, they may appear quite difficult to distinguish.

Navigation through doors and down hallways has been demonstrated by these types of systems, but switching between the contexts has not. One obvious solution would be to have buttons or commands for every different situation; to go through a doorway you press the doorway button,. Obviously, such a system would be cumbersome.

Desired interaction

A more desirable situation would be for the operator to talk to their chair as if it were a chauffeur. An example dialogue[1] might be:

"Go forward and turn a little to the left. Let's go through the door and turn right. Keep going down the hall. Turn left into the next doorway. Stop here. Turn to the left and go to the desk... no the one to the right... OK."

I believe that this level of interaction is possible without the robot having to be able to distinguish a wall from a desk from an elevator. Current robots can distinguish open space from an obstacle. If you know you are in a hall, then the two flat obstacles to either side are walls. If you know you are in a room, facing somewhere near the door, then the open space is where the door is. Similarly, if you know you are facing somewhere near a desk, then the indentation is probably the chair space for the desk. Sufficient contextual knowledge can go a long way towards adding structural clues to simple occupancy data.

Language generation, by the chair, would be largely unnecessary. Its actions would speak more clearly than words, and a simple intentional heading indicator could be used when telegraphing intentions before action was critical.

This application has advantages for using this type of technology over more traditional or independent robot applications. The fact that an operator is constantly with the robot is a huge advantage. The operator can make sure that the robot does not get too deeply into trouble. The operator also can feed commands and information the robot in a timely manner. By delaying commands until they need to be executed, they can be much more context sensitive than could a plan that was presented in its entirety at the beginning of the run. Finally, this is an application area where it is not at all detrimental to have an operator "baby-sit" the robot continuously.

[1]The alert reader will note that this is not really a dialogue, but a monologue. Chauffeurs usually have little to say to their employer, other than to assent. Wheelchairs usually have even less to talk about.

Conclusions

Task context can greatly simplify the communications necessary for a robot understand crude sensor data. Interfaces for assistive robotics would be a prime application area for this type of interface. I hope to have some sort of demonstration system running on a robot wheelchair by the time of the workshop.