Multimodal User Interface for Mobile Robots

Gregory K. Myers
Luc E. Julia
Adam J. Cheyer
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025 USA
1-650-859-4091
myers@erg.sri.com

Samir Gherbi
Swiss Federal Institute of Technology (EPFL)
VRAI Group, DMT-ISR
1015 Lausanne
Switzerland
+41-21-693-5850
gherbi@ai.sri.com

rnABSTRACT

SRI has developed a prototype of a multimodal user interface for small mobile robots that will help dismounted warfighters to conduct reconnaissance and safely enter urban areas occupied by hostile forces. By providing input via voice, pointing, and drawing, the multimodal user interface takes advantage of the natural ways in which humans communicate. The novel aspect of our user interface is that the operator can directly interact with the video imagery and map data so as to easily and efficiently specify spatial locations and relationships. The user interface also enables the operator to task the robots with semiautonomous behaviors that combine both operator-supplied and robot-supplied information. SRI implemented the user interface on a notebook computer and demonstrated it controlling a wheeled robot.

KEYWORDS: multimodal user interface, mobile robot control and tasking, video interaction, gesture recognition

1. INTRODUCTION

SRI is developing a prototype of a small mobile robot that will help dismounted warfighters to conduct reconnaissance and safely enter urban areas occupied by hostile forces. An operator in the field must be able to control and task one or more robots wirelessly via a small, lightweight user interface module that demands a minimum of attention. To investigate some of the issues involved in the design of the user interface module, SRI developed a prototype user interface that presents video data and robot status to an operator, and permits the operator to control and task the robot and video camera in both direct (teleoperation) and semiautonomous (supervised) modes. We describe below our design concepts and some of the prototype's demonstrated capabilities.

2. DESIGN CONSIDERATIONS

Given the dynamic warfighter operating environment for the robot and the operator, the design of the user interface must address several key issues:

- How should the operator command the robot to go to a specific location? Even if the robot has its own localization capability, it will not be practical to specify destinations in terms of absolute world coordinates. Therefore, other ways of specifying destinations and landmarks must be developed.
- How can the operator easily understand where the robot is and where the imagery is coming from? In nonthreatening environments, the operator may have the luxury of viewing a rich set of visual data to create a more realistic telepresence (such as was utilized by Hine et al. [1]). However, due to the payload and bandwidth limitations of our application, the operator may have only a limited awareness of the robot's surroundings, especially when the robot is out of sight.
- How can localization information be best presented, and how accurate does a map have to be? A map may help the operator remember where the robot has been in relation to its environment, but it is crucial to display only the most useful information on the map at any given

time during the mission, since the operator should not be overloaded with unimportant data.

Several points underlie our overall approach to the design of the user interface:

1. Multimodal user input—The user interface should be multimodal, that is, take advantage of the natural ways in which humans communicate. Tasking robots will be easier by providing input via voice, pointing, and drawing. For example, it is desirable to be able to command the robot to “Go there,” which involves simultaneous speaking and gesturing. The user interface should also provide multiple and redundant ways in which the operator can issue commands and interact with the robot and the sensor data, so that, for example, if the battle situation demands that the operator maintain silence, commands can be initiated entirely through pointing at the display.
2. Direct interaction—The operator must use his or her own intelligence to interpret the sensor data and to make decisions in order to task the robot. For example, the operator must look at the video imagery and decide where the robot should go next. Spatial locations and directions will be part of many commands to the robot. The design should permit the operator to interact directly with the video and maps of the robot’s surroundings so as to perform immediate tasking, without requiring the operator to first translate the robot’s perceived position in the video image into an absolute frame of reference.
3. Dynamically reconfigurable display interface—The screen area of a fieldable version of the user interface display will be limited. However, the information that the operator needs to view varies with the task and the situation. The display interface should accommodate this variation in need, to minimize the cognitive load on the operator. This will confer an added benefit during development and testing, since the design is likely to evolve through experimentation.

3. USER INTERFACE DESCRIPTION

3.1 Equipment

SRI explored some of the advanced concepts for the design of the user interface in a demonstration prototype implemented on a notebook computer, and demonstrated the prototype’s ability to control a wheeled robot. The prototype user interface is designed to handle multiple robots. The user interface computer consists of a Toshiba* Tecra 750CDT computer with a Margi Zoom Video Card. Its software is structured as a set of agents using the Open Agent Architecture [2, 3]. The display interface itself was developed with Borland’s Delphi, a visual programming environment that uses an object-oriented extension of the programming language Pascal. The robot was a Pioneer AT wheeled vehicle equipped with ultrasonic range sensors and a 486 processor running Saphira [4], an architecture and a set of software modules that coordinate sensing and control on the robot. Communication between the robot and the notebook computer was via an RS-232 signal over a wireless modem. A Sony EVI-D30 video camera, which has pan, tilt, and zoom controllable through an RS-232 port, was mounted on the robot. Its analog signal was transmitted to the notebook computer via a wireless transmitter.

3.2. User Interface Capabilities

Figure 1 is a screen dump of the prototype user interface. The user interface display includes live video imagery received from a camera mounted on the robot, and robot status information. The operator uses a control panel on the display to control the camera’s field of view and the movement of the robot. A map shows the robot’s path and information about the area of operation. The small pop-up window near the center contains previously captured video imagery that has been retrieved and replayed. The display can be dynamically reconfigured according to the nature of the task and the preferences of the operator. In addition to resizing individual windows, the operator can quickly exchange the positions and sizes of the map and video windows to view either window within a larger screen area. The display context associated with each robot can be accessed on command.

The operator can interact with the displayed information by speaking, pointing, and drawing. The speech recognition and natural language interface was constructed with a commercial product offered by Nuance Communications, Inc. A touchscreen enables the operator to directly point at the display controls and locations in the video and the map, or draw landmarks and paths on the map. Additional operator input can be provided by other analog devices including a mouse, a joystick, and a thumbwheel.

Figure 1. Prototype User Interface

*All product or company names mentioned in this document are the trademarks of their respective holders.

The robot's movement can be controlled in a variety of ways. Picking a point in the inner circle of the display's control panel (Figure 2) instantly selects the heading and speed of the robot. The heading can be adjusted by selecting points in the outer ring. The same controls can also be overlaid on the live video image (Figure 3), so the operator can look continuously at the video image while issuing commands to the robot. The robot can also be controlled via speech commands, such as "stop," "turn left," or "speed 200." Commands can be directed to a specific robot or set of robots ("Robot 2, go to room 1," "Red team, stop"). In addition, the direction of the robot's movement can be specified by pointing directly in the video image. Eventually, the robot will be designed to approach a destination by visually servoing on an area pointed to in the image (similar to the capability of the Ames Marsokhod Rover [5]).

The field of view of the camera can be controlled via speech commands and manipulation of the control panel icons, in a manner similar to the control of the robot's movement. In addition, the camera can be controlled through the recognition of gestures hand-drawn within the video image window. The recognition algorithms underlying these capabilities are based on pen-based stroke analysis [6]. For example, drawing a cross over the point around which the desired field of view should be centered causes the camera to pan and tilt to achieve the desired field of view.

The map helps the operator understand the robot's location and the source of the imagery. The robot's location relative to an initialization point is continuously sensed on board the robot and transmitted to the user interface computer. Information about the robot's surroundings relative to its path is generated in two forms. Still images and video sequences can be saved at various reference points during the mission. The location of each reference point along the robot's path is indicated on the map as an icon (see Figure 4). The imagery can be recalled by clicking on the icon or by a spoken command (e.g., "Show snapshot 3").

The map can also be used to command the robot to go to a specific location semiautonomously. Such behavior-based locomotion control relieves the operator from the attention burden associated with teleoperation. Constrained locomotive behaviors that have been implemented include wall following, lane following (e.g., down a sidewalk or corridor), and turning into a doorway. The operator can draw on the map to indicate the intended destination and various constraints for the robot to use as navigational aids. For example, in Figure 5 the operator draws a rough approximation of corridor walls and a landmark, then an arrow to indicate the destination and desired path. These map features, which need not be drawn to scale, become topological constraints that are matched with the robot's localization and range sensor data by the Saphira robot control software. These capabilities will be especially handy for operating a robot inside a building where it will have to negotiate many features and obstacles. These features can be verbally annotated for later reference, so that landmarks and destinations can be referred to by semantically meaningful names ("enemy 1," "top of stairs," etc.). When the robot confirms the successful execution of the behavior, the positions and lengths of the hand-drawn features are adjusted to coincide with the robot's sensed position.

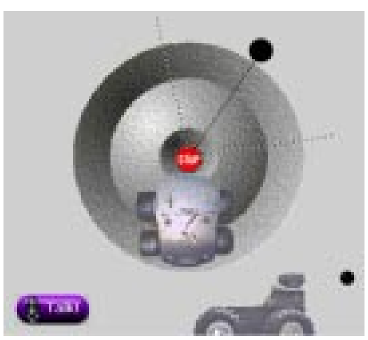

Figure 2. Control Panel for Robot and Camera

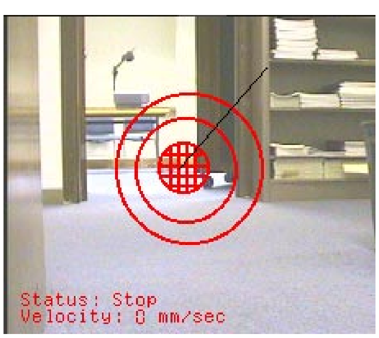

Figure 3. Control Panel Overlaid on Video Image

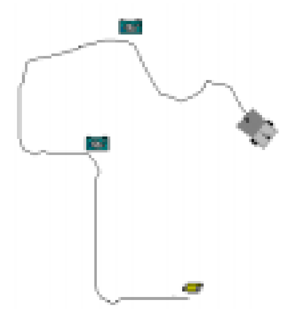

Figure 4. Video Reference Points on a Map

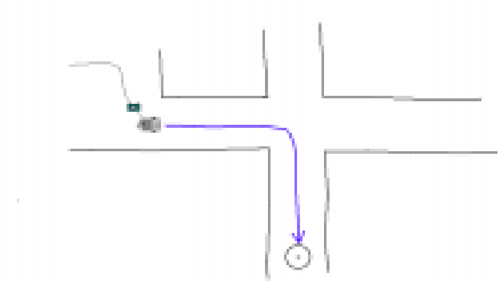

Figure 5. Supervised Autonomy with Hand Drawn Map

4. SUMMARY

SRI has developed a prototype of a multimodal user interface for a small mobile robot; this interface will help small units of warfighters conduct reconnaissance and safely enter urban areas occupied by hostile forces. By providing input via voice, pointing, and drawing, the multimodal user interface takes advantage of the natural ways in which humans communicate to task the robot. In contrast to previous work in which robots are controlled via speech and human gestures [7], with our user interface the operator can interact directly with the video imagery and map data to easily and efficiently specify spatial locations and relationships. The user interface also enables the operator to task the robot with semiautonomous behaviors that combine operator-supplied and robot-supplied information. We will continue to enhance the user interface with additional capabilities, such as the annotation of saved video, still image, and map information, with drawn and spoken comments to facilitate the preparation of multimedia surveillance and status reports.

5. ACKNOWLEDGMENTS

This work was sponsored by the Defense Advanced Research Projects Agency under contract N00014-96-C-0026.

6. REFERENCES

1. Hine, B., Hontalas, P., Fong, T., Piguet, L., Nygren, E., and Kline, A. VEVI: A Virtual Environment Teleoperations Interface for Planetary Exploration. In *SAE 25th International Conference on Environmental Systems* (July, San Diego, CA), 1995.
2. Cohen, P., Cheyer, A., Wang, M., and Baeg, S. An open agent architecture. In *AAAI Spring Symposium*, (March, Stanford University, Stanford, CA), 1994, pp. 1–8.
3. Moran, D., Cheyer, A., Julia, L., and Park, S. Multimodal user interfaces in the Open Agent Architecture. In *Proceedings of IUI-97*, (January, Orlando, FL), 1994, pp. 61–68.
4. Konolige K., Myers, K., Ruspini, E., Saffiotti, A. *The Saphira Architecture: A Design for Autonomy*, MIT Press, Cambridge, MA, 1997.
5. Wettergreen, D., Thomas, H., Bualat, M. Initial Results from Vision-based Control of the Ames Marsokhod Rover. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, Control of Wheeled Robots* (7–12 September, Grenoble, France), 1997, pp. 1377–1382.
6. Julia, L., and Faure, C. Pattern recognition and beautification for a pen-based interface. In *Proceedings of ICDAR '95*, (Montreal, Canada), 1995, pp. 58–63.
7. Perzanowski, D. *Interbot: Natural Language and Gesturing in a Robotics Environment*. Navy Center for Applied Research in Artificial Intelligence, (January 1997). Available on the World Wide Web at http://www.aic.nrl.navy.mil/~dennisp/robotgesture-work.html.