Advanced Interfaces for Vehicle Teleoperation: Collaborative Control, Sensor Fusion Displays, and Web-based Tools

Terrence Fong[1], Charles Thorpe[1] and Charles Baur[2]

[1]The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213 USA

[2]Institut de Systèmes Robotiques
L'Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne EPFL, Switzerland

Abstract

Our goal is to make vehicle teleoperation accessible to all users, novices and experts alike. In our research, we are developing a new system model for teleoperation, sensor-fusion displays and Web-based tools. Our long-term objective is to develop systems in which humans and robots engage in constructive dialogue, not merely simple interaction, to exchange ideas and to resolve differences. In short, to build a framework in which humans and robots can work together and can jointly solve problems.

1 Introduction

Sophisticated interfaces for teleoperation have become increasingly important. For some applications, of course, teleoperation is merely a temporary expedient until autonomous capabilities improve. In other applications, however, the major purpose of the robot is exploration and human-robot interaction is the main feature driving the application. Thus, it is critical that we learn how to design better interfaces, so we can build truly integrated and efficient human-robot systems.

We previously developed a number of vehicle teleoperation systems for field applications such as reconnaissance and remote science[6][10][11]. One of the lessons learned is that vehicle teleoperation is often problematic, especially for novice or untrained operators. Loss of situational awareness, poor attitude and depth judgement, and failure to detect obstacles are common occurrences. Moreover, even if a vehicle has autonomous capabilities (e.g., route following) and is supervised by experts, factors such as poor communications, malicious hazards and operator workload may still compromise task performance.

To address these problems, and to make vehicle teleoperation more effective and more productive, we need interfaces which make it easier to understand the remote environment, to assess the situation, to make decisions, and to effect control. Thus, we are developing a set of tools to facilitate efficient and robust remote driving in unknown, unstructured and dynamic environments.

2 Related Research

During the past twenty years, the majority of work in vehicle teleoperation has centered on rate-controlled systems for hazardous environments. In these systems, a trained operator controls the vehicle's rotation and translation rates via hand-controllers and receives feedback from video cameras. McGovern reported on work with a fleet of wheeled ground vehicles: small indoor robots to large outdoor military automobiles[13]. More recently, vehicle teleoperation systems have emphasized the use of multi-modal operator interfaces and supervisory control[2][5].

Our research draws on work in sensor fusion displays, supervisory control, multi-operator and cooperative teleoperation, and human-robot control architectures. Sensor fusion displays combine information from multiple sensors or data sources for display[8]. Under supervisory control, an operator divides a problem into a sequence of tasks which a system can achieve on its own[17]. In multi-operator teleoperation, humans share or trade control[4]. Cooperative teleoperation tries to improve teleoperation by supplying expert assistance[16]. Several robot control architectures have addressed the problem of mixing humans with robots[1][12].

3 Approach

Our research is driven by the following approach:

- investigate peer-to-peer human-robot interaction and adjustable autonomy through a new teleoperation system model
- develop sensor fusion displays suitable for vehicle teleoperation
- create Web-based tools to enable teleoperation by novices without instruction or training

Although our work is intended primarily to support vehicle teleoperation in field environments, we believe our approach and results are germane to applications in other domains, particularly those which involve high levels of human-robot interaction.

3.1 Collaborative control

Telerobotic systems have traditionally been designed for humans. While sufficient for some domains, it is clearly sub-optimal for multiple vehicles or planetary rovers. Thus, we propose a new approach: *collaborative control*. In this model, a human and a robot collaborate to perform tasks and to achieve goals. Instead of a supervisor dictating to a subordinate, the human and the robot engage in *dialogue* to exchange ideas and resolve differences. Hence, the robot is more equal and can treat the human as an imprecise, limited source of planning and information[7].

An important consequence of collaborative control is that the robot can decide how to use human advice: to follow it when available and relevant; to modify it when inappropriate or unsafe. This is not to say that the robot becomes "master": it still follows higher-level strategy set by the human. However, with collaborative control, the robot has more freedom in execution and can better function when the operator is distracted or unavailable. As a result, teleoperation is more robust and better able to accommodate varying levels of autonomy and interaction.

To examine the numerous human-machine interaction and design issues raised by this new approach, we are building a collaborative control system. In particular, we are investigating how to support human-robot dialogue, how to make the robot more aware, how to design the user interface, and how to handle dynamic control and data flow.

3.2 Sensor fusion displays

To improve vehicle teleoperation, we need to make it easier for the operator to understand the remote environment and to make decisions. In other words, we need to design the human-robot interface so that it maximizes information transfer while minimizing cognitive loading.

Our approach is to enhance the quality of information available to the operator. Specifically, we are developing new sensor fusion techniques using 3D sensors (lidar, stereo vision, etc.) to create a user interface which efficiently and effectively displays multisensor data[14]. In this way, we provide the operator with rich information feedback, facilitating understanding of the remote environment and improving situational awareness[3][19].

Sensor fusion has traditionally been used to support autonomous processes such as localization. To date, however, scant attention has been given to sensor fusion for teleoperation. Although many problems are common to both (sensor selection, data representation, fusion), sensor fusion for teleoperation differs from classic sensor fusion because it has to consider human needs and capabilities.

3.3 Web-based tools

Vehicle teleoperation interfaces are often cumbersome, need significant infrastructure, and require extensive training. Many systems overwhelm the user with multiple displays of multiple sensors while simultaneously demanding high levels of cognition and motor skill. As a result, only experts can achieve acceptable performance.

In order to make vehicle teleoperation accessible to all users, we need to make operator interfaces that are easy to deploy, easy to understand and easy to use. One approach is to build these interfaces using the WorldWideWeb. A Web interface is attractive because it can be accessed world-wide, requires little infrastructure, and is highly cost-effective. At the same time, Web interfaces use familiar interaction models, thus requiring little (or no) training.

Web-based teleoperation, however, raises many issues and prohibits use of traditional approaches. Specifically, we find we must develop methods which minimize bandwidth usage, which provide sensor fusion displays, and which optimize human-computer interaction.[9].

4 Results

4.1 Collaborative control

Our current collaborative control system uses a message-based architecture (shown in Figure 1) to connect task-achieving system modules which we call a *behavior*. We consider the user, connected to the system via the user interface, to be one of these modules.

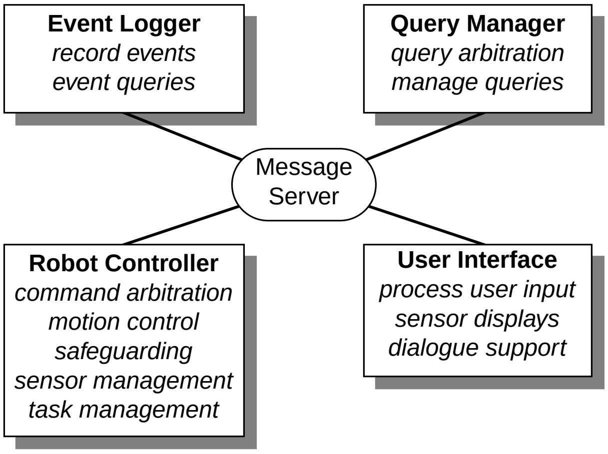


Figure 1. Collaborative control architecture

Dialogue between human and robot arises from an exchange of messages. We believe that effective dialogue does not require a full language, merely one which is pertinent to the task at hand and which efficiently conveys information. Thus, we do not use natural language and we limit message content to vehicle mobility (e.g., positioning).

We classify messages as shown in Table 1. *Robot commands* and *user statements* are uni-directional. A query is expected to elicit a response (though the response is not guaranteed and may be delayed). At present, we are using approximately thirty messages to support vehicle teleoperation. A selection of these messages is given in Table 2

Table 1. Dialogue message classes

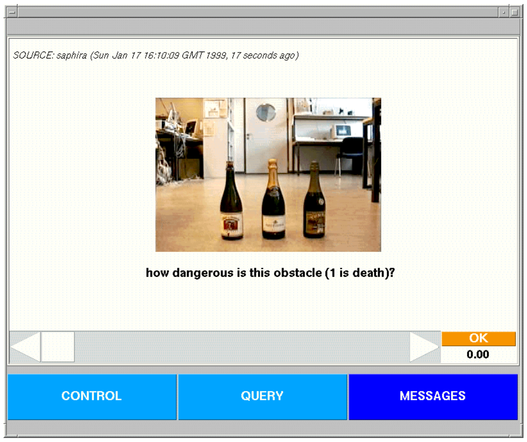
User → Robot	Robot → User
robot command (command for the robot)	user statement (information for the user)
query-to-robot (question from the user)	query-to-user (question from the robot)
response-from-user (query-to-user response)	response-from-robot (query-to-robot response)

Table 2. Example vehicle mobility dialogue messages

Category	Message
query-to-robot	How are you? Where are you?
response-from-robot	bar graphs (How are you?) map (Where are you?)
user query	How dangerous is it this (image)? Where do you think I am (map)?
response-from-user	"8" (How dangerous is this?) position (Where do you think I am?)
robot command	rotate to X (deg), translate at Y (m/s) execute this path (set of waypoints)
user statement	I think I'm stuck because my wheels spin Could not complete task N due to M

In our system, the operator sends and receives messages via a user interface. Our current interface has three modes, each of which supports two dialogue message classes[7]. This partitioning clarifies human-robot interaction, allowing focus on a specific dialogue aspect. Each mode is designed to convey messages as efficiently as possible. For example, the *query-to-user* "How dangerous is this object?" is shown as in Figure 2. The image allows the user to perform visual analysis and the slider provides a rapid, yet precise response mechanism.

We have begun studying how collaborative control influences performance of "A to B". In this scenario, the robot is commanded to make a change of pose in an unknown environment. The question we would like to answer is: how does performance (completion, execution speed, situational awareness, etc.) change as the dialogue is varied? Specifically, we would like to ascertain what effects are observable as the level of autonomy is varied.



SOURCE: saphira (Sun Jan 17 16:10:09 GMT 1999, 17 seconds ago)

how dangerous is this obstacle (1 is death)?

Figure 2. Message mode.

4.2 Sensor fusion displays

Our initial sensor fusion display incorporated coarse range data and omnidirectional camera images[3]. In this system, we displayed sonar ranges as a filled, colored circle (representing the beam cone) image overlay. We found, however, that users had difficulty interpreting the resulting images due to the poor angular resolution of our sonar sensors (i.e., large range readings resulted in large overlay circles which made obstacle identification difficult).

More recently, we have been using a multisensor system with monochrome video, stereo vision, ultrasonic sonar, and vehicle odometry[14][19]. The stereo vision system and ultrasonic sonars are co-located on a sensor platform (see Figure 3) which is mounted on a vehicle.

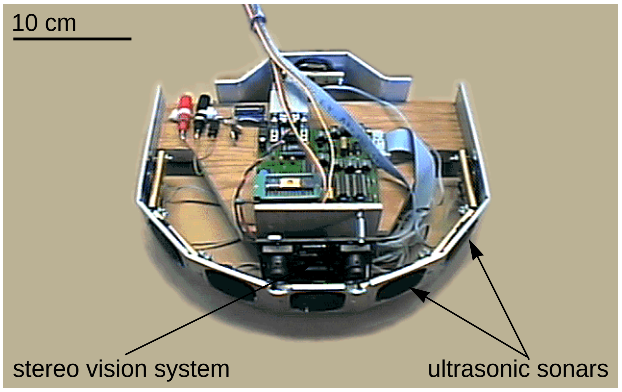


Figure 3. Multisensor platform

We chose these sensors based on their complementary characteristics. The stereo vision system provides monochrome and range (disparity) images. Ultrasonic sonars provide discrete (time-of-flight) ranges. Table 3 lists situations encountered in vehicle teleoperation. Though none of the sensors works in all situations, the group as a whole provides complete coverage.

Table 3. Sensor characteristics

Situation	2D images	Stereo vision	Sonar
smooth surfaces (with visual texture)	OK	OK	Fails[a]
rough surfaces (without visual texture)	OK	Fails[b]	OK
close obstacles (<0.6 m)	OK[c]	Fails[d]	OK[e]
far obstacles (>10 m)	OK	Fails[f]	Fails[g]
no external light source	Fails	Fails	OK

a. specular reflection
b. no correlation
c. limited by focal length
d. high disparity
e. limited by transceiver
f. poor resolution
g. echo not received

We fuse 2D and stereo images, sonar and odometry data using a *cross-filter* algorithm (Figure 4). A *Texture Filter* is applied to the 2D images to identify areas with inadequate texture for stereo matching. A *Close Range FIlter* is applied to the sonar data to identify regions containing objects too close for stereo matching. The selected regions are then processed using a Kalman filter and vehicle odometry information. Finally, the fused data is used to construct the interface displays.

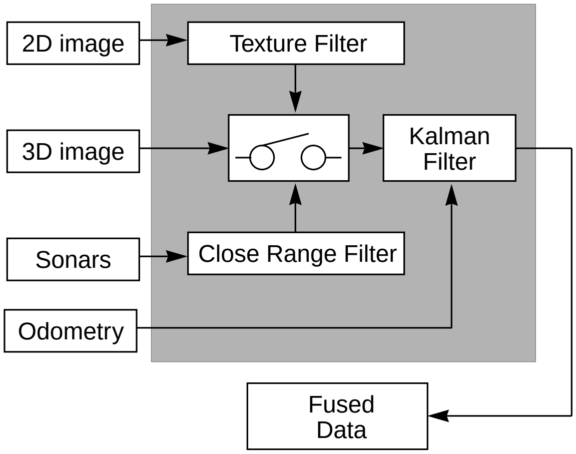
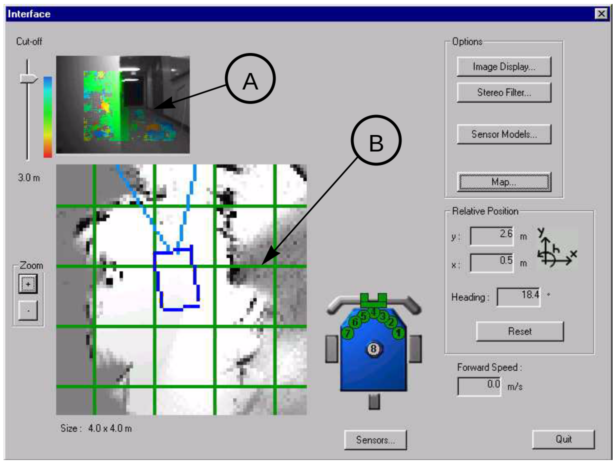


Figure 4. Cross-filter algorithm

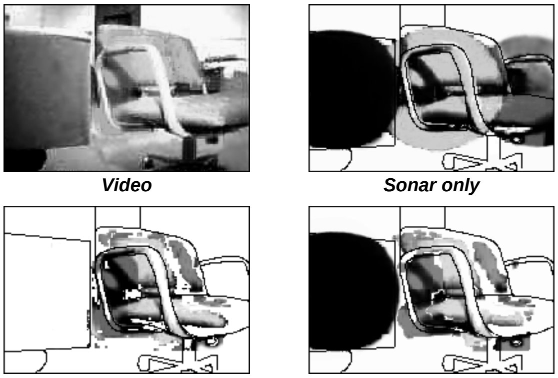
Figure 5 shows the main window of our sensor fusion based user interface. The interface contains two primary displays: (A) a 2D image with color overlay and (B) a local map constructed with sensor data. The 2D image facilitates scene interpretation and understanding by directing attention to obstacles and by aiding distance estimation. The local map displays an occupancy grid and improves situational awareness (especially monitoring of vehicle orienta-



y: 2.6 m
x: 0.5 m
Heading: 18.4 °
Forward Speed: 0.0 m/s
Size: 4.0 x 4.0 m

Figure 5. Sensor fusion user interface

Figure 6 demonstrates how sensor fusion improves the display. The top left image contains video only: from this view it is difficult to judge relative depth. In the top right image (sonar only), the obstacles are detected, but the scene remains difficult to interpret. In the bottom left image (stereo only), the chair is mapped correctly, but the box on the left is not seen because it lacks texture. Fusing data from both sensors yields the bottom right image: the chair is mapped with good resolution (stereo) and the box is clearly visible (sonar).



Video

Sonar only

Stereo only

Stereo and sonar

Figure 6. Improvement by fusing stereo and sonar

4.3 Web-based user interfaces

To date, we have created two Web-based systems. We first developed the *WebPioneer*[1] in collaboration with ActivMedia, Inc. The WebPioneer enables novices to explore an indoor environment. The WebPioneer, however, requires significant network resources and restricts expert users (i.e., it only provides a limited command set).

---

[1] http://webpion.mobilerobots.com

Our second system, *WebDriver*, is designed to minimize network bandwidth usage, to provide an active user interface, and to optimize human-computer interaction. It supports a wide range of users and enables safe and reliable Web-based vehicle teleoperation. The WebDriver differs from other systems because it enables teleoperation in unknown, unstructured and dynamic environments[9].

The WebDriver architecture is shown in Figure 7. The *User Interface* is a Java applet which runs in a Web browser. It is connected to the system via a persistent network link, accepts user commands and provides continuous feedback from the robot's sensors. The *Base Station* performs communication with the user interface, image processing, and high-level robot control. The *Robot* is equipped with on-board sensors for autonomous safeguarding and a motion controller. It is connected to the base station via a radio modem and analog video transmitter.

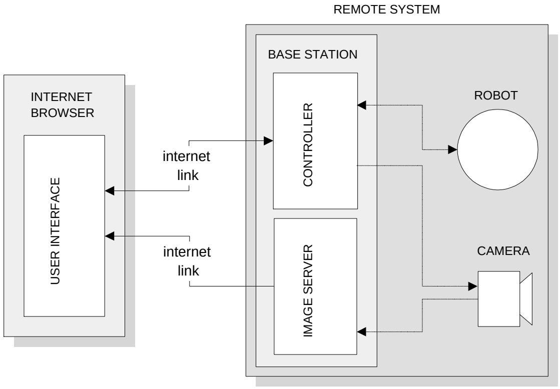
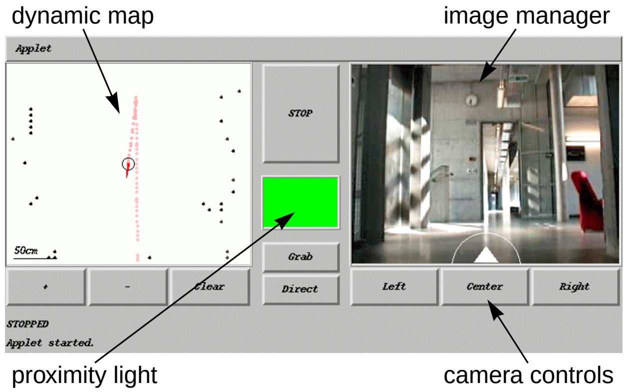


Figure 7. WebDriver system architecture

The WebDriver user interface is shown in Figure 8 and contains two primary tools, the *dynamic map* and the *image manager*, which allow the user to send commands to the robot and to receive feedback. We designed the interface so that the user is always able to see complete system status at a glance and can specify robot commands in multiple ways.



STOPPED
Applet started.

Figure 8. Web interface for vehicle teleoperation

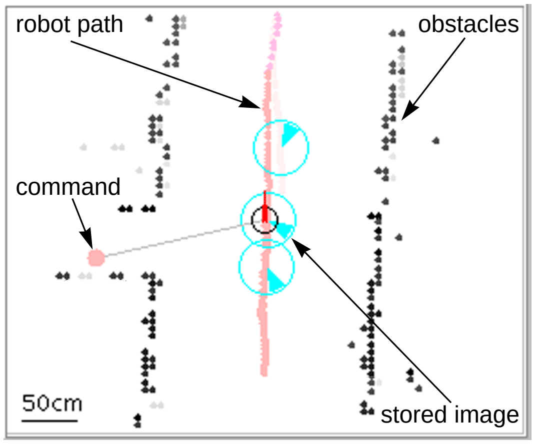


Figure 9. Dynamic map

The dynamic map (Figure 9) is constructed using ultrasonic sonar readings and robot position. The map displays sensor data as colored points; light colors indicate low confidence, dark colors indicate high confidence. The map also displays locations (blue circles) at which images were stored with the image manager. Clicking on the map designates commands the robot to move to an absolute position

The image manager (Figure 10) displays and stores images from the robot's camera. Unlike other Web teleoperation systems, such as [15] or [18], we do not use server-push video because it excessively consumes bandwidth. Instead, we use an event-driven client-server model to retrieve images when certain events (user command, obstacle detected, etc.) occur. On each image, the camera orientation and obstacles indicators are overlaid. When a stored image is shown, a "replay" symbol is displayed. Clicking on the image commands the robot to turn or translate.

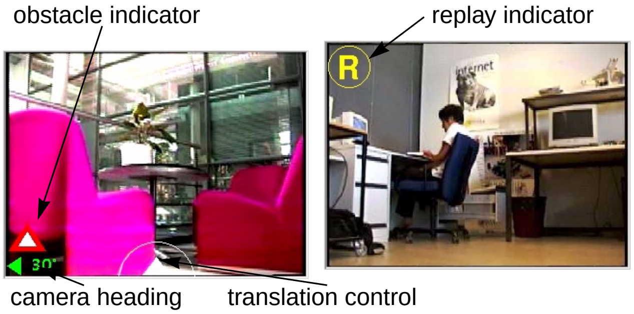


Figure 10. Image manager
current image (left), stored image (right)

The WebDriver effectively frees the system from bandwidth limitations and transmission delay imposed by the Web, thus enabling effective control of the robot. Anecdotal evidence from a range of users suggests that the system is quite reliable and robust. We found that novices are able to safely explore unfamiliar environments and that experts can efficiently navigate difficult terrain.

5 Future Work

The majority of conventional vehicle teleoperation systems require expensive infrastructure and extensive operator training. For example, the American military has begun performing remote operations (e.g., reconnaissance) with unmanned air and ground vehicles. To do this, a highly trained soldier teleoperates using multiple controls, video and data screens. These systems are expensive, time consuming to deploy, and have low productivity.

As an alternative, we plan to develop a palm-size computer system which incorporates collaborative control, sensor fusion displays, and Web-based tools. Our goal is to be able to remotely drive a mobile robot, at any time and any location, with minimal infrastructure. We believe this system will significantly advance vehicle teleoperation while providing an ideal platform for studying peer-to-peer human-computer interaction and adjustable autonomy. Moreover, such a system will be well suited for applications ranging from facility security to reconnaissance.

6 Conclusion

By treating the operator as an limited, imprecise, and noisy source of information, collaborative control enables use of human perception and cognition without requiring continuous or time-critical response. Collaborative control helps balance the roles of operator and robot, giving the robot more freedom in execution and allowing it to better function if the operator is inattentive or making errors.

By combining data from multiple, complementary sensors, sensor fusion displays allow us to increase the quality and richness of information available to the operator. With sensor fusion displays, human-machine interaction becomes more efficient, facilitating understanding of the remote environment and improving situational awareness.

By employing wide-area networks and well-known interaction models, Web-based tools can be used worldwide, require little infrastructure and are highly cost-effective. As a result, Web-based tools offer significant potential for making vehicle teleoperation accessible to all users.

Acknowledgments

We thank Gilbert Bouzeid, Sébastien Grange, Roger Meier, and Grégoire Terrien for their innumerable contributions and hard work. We also thank the Institut de Systèmes Robotiques (DMT-ISR / EPFL) for providing research facilities and infrastructure. This work is partially supported by the DARPA TTO “Tactical Mobile Robots” program (NASA JPL 1200008) and by SAIC, Inc.

References

[1] Albus, J., et. al., “NASA/NBS Standard Reference Model for Telerobot Control System Architecture (NASREM)”, Technical Note 1235, NIST, Gaithersburg, Maryland, 1987.

[2] Bapna, D., et. al. “The Atacama Desert Trek: Outcomes”, IEEE International Conference on Robotics and Automation, Leuven, Belgium, 1998.

[3] Bouzeid, G., “Acquisition and Visualization of Ultrasound Data and Sensor Fusion with a 2D Camera”, EPFL Microengineering Department, Lausanne, Switzerland, March 1998.

[4] Cannon, D. and Thomas, G. “Virtual Tools for Supervisory and Collaborative Control of Robots”, *Presence* 6(1), 1997.

[5] Cooper, B., “Driving on the Surface of Mars Using the Rover Control Workstation”, SpaceOps ‘98, Tokyo, Japan, 1998.

[6] Fong., T., et. al. “Operator Interfaces and Network Based Participation for Dante II”, SAE 25th ICES, San Diego, California, July 1995.

[7] Fong, T., et. al., “Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation”, AAAI 1999 Spring Symposium, Stanford, California, March 1999.

[8] Foyle, D., “Proposed Evaluation Framework for Assessing Operator Performance with Multisensor Displays”, SPIE Vol. 1666, 1992.

[9] Grange, S., et. al., “Effective Vehicle Teleoperation on the World Wide Web”, IEEE International Conference on Robotics and Automation, San Francisco, California, April 2000.

[10] Hine, B., et. al., “VEVI: A Virtual Environment Teleoperations Interface for Planetary Exploration”, SAE 25th ICES, San Diego, California, July 1995.

[11] Kay, J. and Thorpe, C., “Operator Interface Design Issues in a Low-Bandwidth and High-Latency Vehicle Teleoperation System”, SAE 25th ICES, San Diego, California, July 1995.

[12] Krotkov, E., et. al., “Safeguarded Teleoperation for Lunar Rovers”, SAE 26th ICES, Monterey, California, 1996.

[13] McGovern, D., “Human Interfaces in Remote Driving”, Technical Report SAND88-0562, Sandia National Laboratory, Albuquerque, New Mexico, 1988

[14] Meier, R., et. al., “A Sensor Fusion Based User Interface for Vehicle Teleoperation”, IEEE FSR ‘99, Pittsburgh, Pennsylvania, August 1999.

[15] Michel, O. et al., “KhepOnTheWeb: An Experimental Demonstrator in Telerobotics and Virtual Reality”, *VSMM’97*, Geneva, Switzerland, September 1997.

[16] Murphy, R. and Rogers, E. “Cooperative Assistance for Remote Robot Supervision”, *Presence* 5(2), 1996.

[17] Sheridan, T. *Telerobotics, Automation, and Human Supervisory Control*, MIT Press, Cambridge, Massachusetts, 1992.

[18] Siegwart, R., and Saucy, P., “Interacting Mobile Robots on the Web”, workshop, IEEE Conference on Robotics and Automation, Detroit, Michigan, April 1998.

[19] Terrien, G., and Fong, T. “Remote Driving with a Multisensor User Interface”, SAE 30th ICES, Toulouse, France, July 2000.