Autonomous Robots 11, 77–85, 2001

Advanced Interfaces for Vehicle Teleoperation: Collaborative Control, Sensor Fusion Displays, and Remote Driving Tools

TERRENCE FONG
The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; Institut de Systèmes Robotiques Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

CHARLES THORPE
The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

CHARLES BAUR
Institut de Systèmes Robotiques Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Abstract. We are working to make vehicle teleoperation accessible to all users, novices and experts alike. In our research, we are developing a new control model for teleoperation, sensor-fusion displays and a suite of remote driving tools. Our goal is to build a framework which enables humans and robots to communicate, to exchange ideas and to resolve differences. In short, to develop systems in which humans and robots work together and jointly solve problems.

Keywords: human robot interaction, mobile robots, multisensor displays, remote driving, vehicle teleoperation

1. Introduction

In our previous work, we built a number of vehicle teleoperation systems for field applications such as reconnaissance and remote science (Fong et al., 1995; Hine et al., 1995; Kay and Thorpe, 1995). One of the lessons learned is that vehicle teleoperation is often problematic, especially for novices. Loss of situational awareness, poor 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 and operator workload may still compromise task performance.

To address these problems, we are developing tools and techniques to improve human-robot interaction in vehicle teleoperation. In particular, we are investigating a new model for teleoperation, collaborative control, which facilitates adjustable autonomy. Additionally, we are creating displays to make it easier for operators to understand the remote environment and to make decisions. Finally, we are building interfaces which are easy to deploy, understand, and use.

2. Related Research

During the past twenty years, the majority of research in vehicle teleoperation has centered on rate-controlled systems for hazardous environments. For example, McGovern (1988) reported on work with a fleet of wheeled ground vehicles: small indoor robots to large outdoor military automobiles. More recently, vehicle teleoperation systems have emphasized the use of multi-modal operator interfaces and supervisory control (Fong and Thorpe, 2001).

Our research draws on work from numerous domains. Sensor fusion displays combine information from multiple sensors or data sources into a single, integrated view (Foyle, 1992). Under supervisory control, an operator divides a problem into a sequence of tasks which the robot must achieve on its own (Sheridan, 1992). Cooperative teleoperation tries to improve teleoperation by supplying expert assistance (Murphy and Rogers, 1996). Several robot control architectures, such as (Albus et al., 1987), have addressed the problem of mixing humans with robots.

3. Approach

Collaborative Control

To improve human-robot interaction in vehicle teleoperation, we are developing a new control model called *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 (Fong et al., 1999).

An important consequence of collaborative control is that the robot can decide how to use human advice: to follow it when available; to modify it when inappropriate. 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. As a result, teleoperation is more robust and better able to accommodate varying levels of autonomy and interaction.

Sensor Fusion Displays

To make it easier for the operator to understand the remote environment, we need to enhance the quality of information available to the operator. Thus, we are developing multisensor displays which fuse data from a variety of 3D sensors (ladar, sonar, stereo vision) (Meier et al., 1999). In this way, we provide the operator with rich information feedback, facilitating understanding of the remote environment and improving situational awareness (Terrien et al., 2000).

Sensor fusion has traditionally been used to support autonomous processes (e.g., localization) with scant attention given to display. Although many problems are common to both (sensor selection, data representation, fusion), sensor fusion for display differs from classic sensor fusion because it has to consider human needs and sensory capabilities.

Novel Interface 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. To make vehicle teleoperation accessible to all users, therefore, we need interfaces which are easy to deploy, understand and use.

Our approach is to develop a suite of interface tools using computer vision, Personal Digital Assistants (PDA), and the WorldWideWeb. With computer vision, we can provide flexible, user-adaptable interaction. With PDA's, we can construct portable interfaces for use anywhere and anytime. With the WorldWideWeb, we can build cost-effective interfaces which require little (or no) training.

4. Results

4.1. Collaborative Control

Our current collaborative control system is implemented as a distributed set of modules in a message-based architecture (Fig. 1). Human-robot interaction is handled by the user interface working in conjunction with the event logger, query manager and user adapter. A safeguarded teleoperation controller provides localization, map building, motion control, sensor management and speech synthesis.

Dialogue between human and robot arises from an exchange of messages. At present, we are using approximately thirty messages to support vehicle teleoperation. A selection of these messages is given in Table 1. *Robot commands* and *user statements* are unidirectional. A query (from the human or the robot) is expected to elicit a response. In our system, however, responses are not guaranteed and may be delayed. Since the robot may ask simultaneous queries (i.e., multiple modules may need human advice), we perform *query arbitration* to select which ones are given to the user (Fong et al., 1999).

We have found that collaborative control provides significant benefits to vehicle teleoperation. First, it improves performance by enabling joint problem solving. This generally produces better results than either the human or robot can achieve alone. Second, dialogue serves as an effective coordinating mechanism,

Table 1. Example vehicle mobility dialogue messages.

Category	Direction	Message
Robot command	User → robot (command for the robot)	Rotate to X (deg), translate at Y (m/s) Execute this path (set of waypoints)
User statement	Robot → user (information for the user)	I think I'm stuck because my wheels spin Could not complete task N due to M
Query-to-robot	User → robot (question from the user)	How are you? Where are you?
Response-from-robot	Robot → user (query-to-robot response)	Bar graphs (How are you?) Map (Where are you?)
Query-to-user	Robot → user (question from the robot)	How dangerous is this (image)? Where do you think I am (map)?
Response-from-user	User → robot (query-to-user response)	"8" (How dangerous is this?) Position (Where do you think I am?)

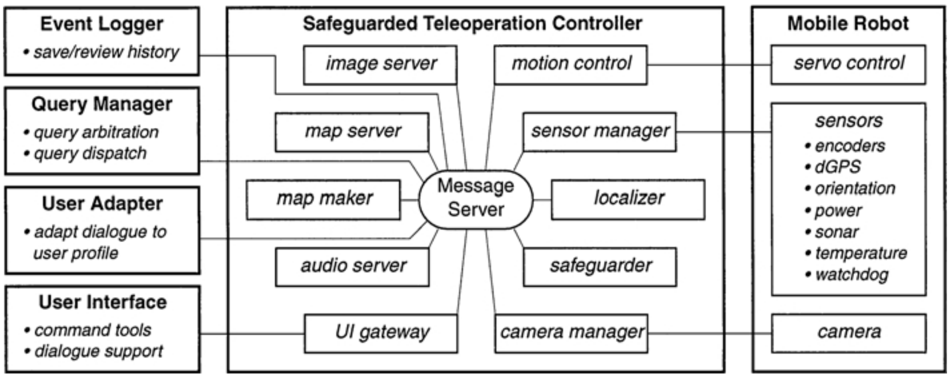


Figure 1. Collaborative control architecture.

particularly when an operator is controlling multiple vehicles. Since robot queries are prioritized (via arbitration), the operator's attention is efficiently directed to the robot most in need of assistance. Finally, because we can adapt dialogue (based on the user's availability, knowledge, and expertise), collaborative control allows us to better support non-specialists.

4.2. Sensor Fusion Displays

In teleoperation, having good depth information is essential for judging the positions of objects (obstacles, targets, etc.) in the remote environment. Our approach is to provide visual depth cues by displaying data from a heterogeneous set of range sensors. We are currently using a multisensor system equipped with a laser scanner (ladar), monochrome video, stereo vision, ultrasonic sonar, and vehicle odometry (Meier et al., 1999; Terrien et al., 2000) as shown in Fig. 2.

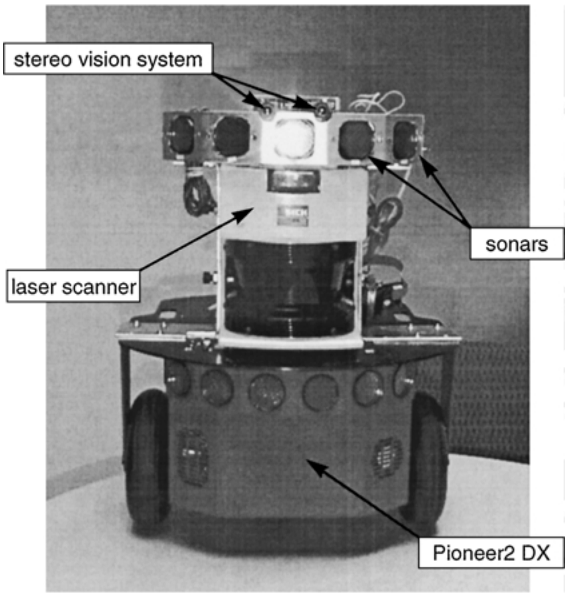
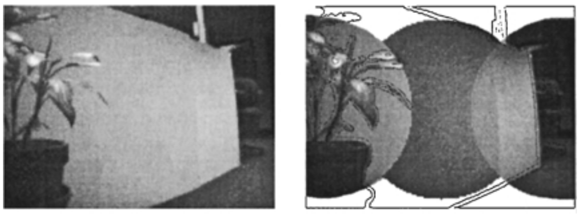


Figure 2. Multisensor platform.

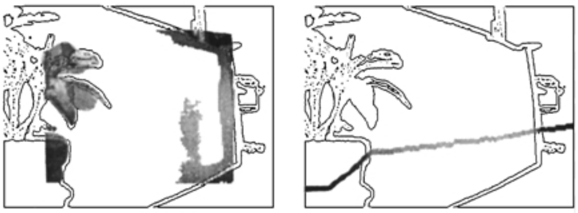
Table 2. Sensor performance in teleoperation situations.

Situation	2D Image (intensity)	3D Image (disparity)	Sonar (TOF)	Ladar (laser)
Smooth surfaces (no visual texture)	OK	Fails[a]	Fails[b]	OK
Rough surface (little/no texture)	OK	Fails[a]	OK	OK
Far obstacle (>10 m)	Fails[c]	Fails[d]	Fails[e]	OK
Close obstacle (<0.5 m)	OK[f]	Fails[g]	OK[h]	OK[i]
Small obstacle (on the ground)	Fails[c]	OK	OK	Fails[j]
Dark environment (no ambient light)	Fails	Fails	OK	OK

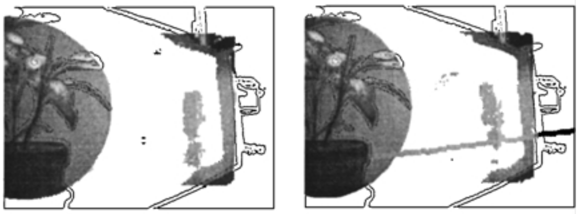
[a] No correlation.
[b] Specular reflection.
[c] No depth measurement.
[d] Poor resolution.
[e] Echo not received.
[f] Limited by focal length.
[g] High disparity.
[h] Limited by transceiver.
[i] Limited by receiver.
[j] Outside of scan plane.



Video only *Sonar only*



Stereo only *Ladar only*



Sonar and stereo *Ladar, sonar, and stereo*

Figure 3. Improvement by fusing ladar, sonar, and stereo.

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. The ladar provides precise range measurement with very high angular resolution and is a good complement to the stereo vision and sonar (both of which are less accurate but have broader field-of-view). Table 2 lists situations encountered in vehicle teleoperation. Though none of the sensors works in all situations, the group as a whole provides complete coverage.

Figure 3 demonstrates how sensor fusion improves the display of a scene with difficult sensing characteristics: in front of the vehicle is a smooth, untextured wall and close by is a large plant (shown in the top left image). In the top right image (sonar only), the plant is detected well, but the wall is shown at incorrect depths due to specular reflection. In the middle left image (stereo only), the wall edges are clearly detected and the plant partially detected (the left side is too close for stereo correlation). However, the center of the wall (untextured) is completely missed. In the middle right image (ladar only), we see that the wall is well defined, but that the planar scan fails to see the plant. In the bottom left image (fused sonar and stereo), both the wall edge and plant are detected, but the center remains undetected. In the bottom right image (all sensors), we see that all features are properly detected. The sonars detect the plant, the ladar follows the wall, and stereo finds the wall edge.

4.3. Remote Driving Tools

Visual Gesturing. *GestureDriver* is a remote driving interface based on visual gesturing (Fong et al., 2000). Visual gesturing offers two distinct advantages over traditional input methods. First, the interface is easy to deploy and can be used anywhere in the field of view of the visual tracker. More significantly, since the mapping from gesture to action is entirely software based, it is possible to adapt the interpretation to the current task and to the operator in real-time.

GestureDriver uses normalized color filtering and stereo vision for robust feature (hand and body) tracking. Color filtering provides fast 2D localization, while stereo provides 3D measurements (shape and range). GestureDriver provides several interpretations for mapping gestures to commands. For example, the *virtual*

Figure 4. Virtual joystick mode. The right hand position indicates (left to right) right, left, forward, reverse, stop.

Figure 5. Visual gesturing for vehicle teleoperation.

joystick interprets operator hand motion as a two-axis joystick (see Fig. 4). To start, the operator raises his left hand to activate the gesture system. The operator then uses his right hand to specify direction and command magnitude.

We found that GestureDriver works well almost anywhere within the vision system's field of view. Figure 5 shows an operator using the virtual joystick to directly teleoperate a mobile robot. In this mode, hand gestures are mapped directly to robot motion. Distance from a reference point (as defined by the user) sets the vehicle's speed, while orientation controls the vehicle's heading.

We also found that remote driving with visual gestures is not as easy as one might believe. Although humans routinely use hand gestures to give commands, gestures may be semantically identical but have tremendous variation in spatial structure. Additionally, several users reported that visual gesturing can be fatiguing, especially when the robot is operating in a cluttered environment. Thus, to improve the GestureDriver's usability we are considering adding additional interface modalities (e.g., speech) to help classify and disambiguate visual gestures.

PDA. *PdaDriver* is a Personal Digital Assistant (PDA) interface for vehicle teleoperation (Fig. 6). We designed it to be easy-to-use, easy-to-deploy and to function even when communication links are low-bandwidth and high-latency. PdaDriver uses multiple control modes, sensor fusion displays, and safeguarded teleoperation to enable efficient remote driving anywhere and anytime (Fong et al., 2000).

We implemented the PdaDriver using a WindowsCE Palm-size PC and Personal Java. The PdaDriver provides relative position, rate, and waypoint (image and

Figure 6. PdaDriver: user interface (left), remote driving a mobile robot (right).

map) control modes. Image-based driving is well suited for unstructured or unknown terrain as well as for cluttered environments. Our method was inspired by Kay and Thorpe (1995), but uses a planar world model. Map-based driving helps maintain situational awareness and is useful for long-distance movements.

We have conducted field trials with the PdaDriver in a variety of environments, both indoor and outdoor. Since remote driving is performed in a safeguarded, semi-autonomous manner, continuous operator attention is not required and the robot moves as fast as it deems safe. Anecdotal evidence from both novice and expert users suggests that the PdaDriver has high usability, robustness, and performance. Furthermore, users reported that the interface enabled them to maintain situational awareness, to quickly generate commands, and to understand at a glance what the robot was doing.

WorldWideWeb. We developed our first Web-based system, the *WebPioneer*, in collaboration with ActivMedia, Inc. The WebPioneer enables novices to explore a structured, indoor environment and has been in continuous operation[1] since April 1998. The WebPioneer, however, consumes significant network resources (due primarily to the use of live video) and restricts expert users (i.e., it only provides a limited command set).

We designed our second system, *WebDriver*, to address these problems as well as to support teleoperation in unknown, unstructured and dynamic environments (Grange et al., 2000). The WebDriver is implemented as a Java applet and runs in a Web browser (Fig. 7). The interface 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.

The dynamic map displays sensor data as colored points: light colors indicate low confidence, dark colors indicate high confidence. Clicking on the map commands the robot to move to an absolute position. The image manager displays and stores images from the robot's camera. Unlike other Web-based vehicle teleoperation systems, such as Michel et al. (1997), we do not use server-push video because it excessively consumes bandwidth. Instead, we use an event-driven client-server model to display images when certain events (e.g., obstacle detected) occur. Clicking on the image commands relative turn or translation.

We have found that the WebDriver's design effectively frees the system from bandwidth limitations and transmission delay imposed by the Web (Grange et al., 2000). Informal testing with a range of users suggests that the system is quite reliable and robust. In practice, we have seen that novices are able to safely explore unfamiliar environments and that experts can efficiently navigate difficult terrain.

5. Discussion

Although all our interfaces support vehicle teleoperation in unknown environments, each interface has

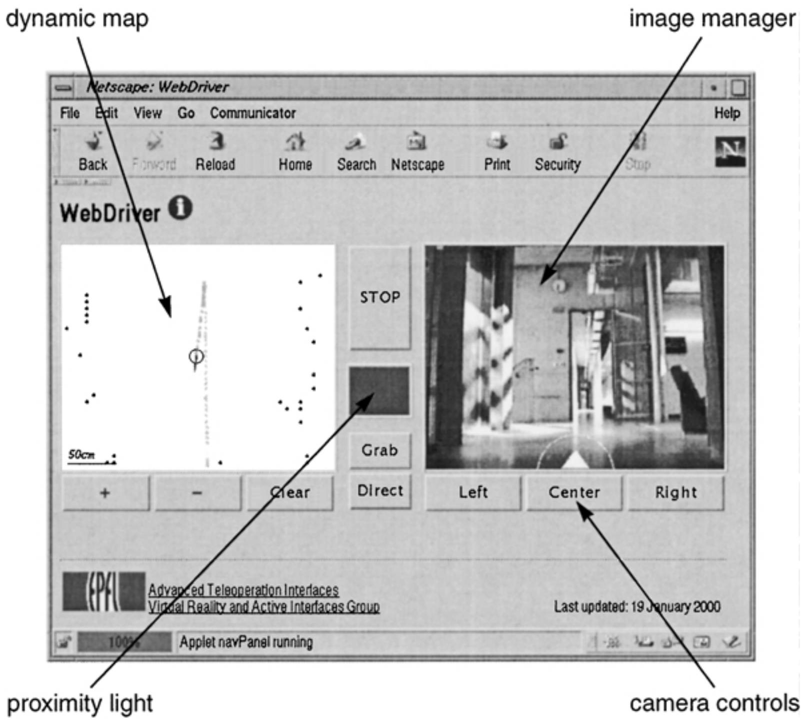


Figure 7. Web interface for vehicle teleoperation.

unique characteristics and is intended for use under different conditions. Collaborative control, for example, was designed to encourage peer interaction between a human and a robot. As such, it is most suitable for operators who have some level of expertise and can provide useful answers to robot questions. Conversely, the WebDriver interface is geared primarily towards the novice, who does not need (or may not want) the command capabilities used by experts. Table 3 provides a comparison of our interfaces.

Almost all modern computer interfaces are designed with user-centered methods. A variety of human performance or usability metrics (speed of performance, error rate, etc.) are typically used to guide the design process (Newman and Lamming, 1995). Yet, in spite of the success of these methods at increasing performance and reducing error, there has been little application of these methods to teleoperation interface design. One hypothesis is that mainstream HCI techniques are ill-suited for teleoperation (Graves, 1998). Cognitive walkthrough, for example, is generally performed for multi-dialogue interfaces and from the viewpoint of novice users, both of which are rare in teleoperation systems.

This is not to say, however, that teleoperation interfaces cannot be constructed or analyzed in a structured fashion. Rather, it is our firm belief that HCI methods *should* be applied to the greatest extent possible, especially during design. Thus, we used the guidelines presented in Graves (1998) when designing all our interfaces. In particular, all our interfaces strongly emphasize consistency, simplicity of design, and consideration for context of use. Most recently, we developed the PdaDriver interface using a combination of heuristic evaluation and cognitive walkthrough.

Our long-term objective is to develop systems in which humans and robots work together to solve problems. One area in which human-robotic systems can have a significant impact is planetary surface exploration. Thus, we intend to develop interfaces which enable EVA crew members (e.g., suited geologists) and mobile robots to jointly perform tasks such as sampling, site characterization, and survey. To do this, we plan to combine elements of our research in collaborative control, sensor fusion displays, and PDA interfaces. The challenge will be to create a portable interface for field science and to quantify how human-robot collaboration impacts task performance.

Table 3. Vehicle teleoperation interface comparison.

Interface	Design goals	Application	Control variables	Vehicle autonomy	User training
Collaborative control	Peer interaction Semi-autonomous operation Human as resource	Exploration Reconnaissance Surveillance	Rate Position (abs/rel) Waypoint (map/image)	High	Medium
Sensor Fusion	Facilitate environment assessment Improve situational awareness	Exploration	Rate Position (abs/rel)	Low	Medium
GestureDriver	Flexible, user-adaptable Physical human-robot interaction	Line-of-site operations Scientific field assistant	Rate (translate) Heading (abs)	Low	High
PdaDriver	Lightweight, portable hardware Operate anywhere & anytime	Exploration Field operations Reconnaissance	Rate Position (abs/rel) Waypoint (map/image)	Medium	Low
WebDriver	Minimal infrastructure Minimal training Novice operators	Education Public demonstrations	Position (rel) Waypoint (map/image)	Medium	Low

6. Conclusion

We are working to make vehicle teleoperation accessible to all users, novices and experts alike. To do this, we have developed interfaces which improve human-robot interaction and enable joint problem solving. Collaborative control enables use of human expertise without requiring continuous or time-critical response. Sensor fusion displays increase the quality of information available to the operator, making it easier to perceive the remote environment and improving situational awareness. Finally, by employing computer vision, PDA's, and the WorldWideWeb, we have created remote driving tools which are user-adaptive, can be used anywhere, and which require little training.

Acknowledgments

We would like to thank Gilbert Bouzeid, Sébastien Grange, Roger Meier, and Grégoire Terrien for their contributions and tireless work. This work was partially supported by grants from SAIC, Inc., the DARPA TTO TMR program and the DARPA ITO MARS program.

Note

1. http://webpion.mobilerobots.com

References

Albus, J. et al. 1987. NASREM. NIST, Gaithersburg, MD, Technical Note 1235.

Fong, T., Pangels, H., Wettergreen, D., Nygren, E., Hine, B., Hontalas, P., and Fedor, C. 1995. Operator interfaces and network based participation for Dante II. In *Proceedings of the SAE ICES*, San Diego, CA.

Fong, T., Thorpe, C., and Baur, C. 1999. Collaborative control: A robot-centric model for vehicle teleoperation. In *Proceedings of the AAAI Spring Symposium: Agents with Adjustable Autonomy*, Stanford, CA.

Fong, T., Conti, F., Grange, S., and Baur, C. 2000. Novel interfaces for remote driving: Gesture, haptic and PDA. In *Proceedings of the SPIE Telemanipulator and Telepresence Technologies Conference*, Boston, MA.

Fong, T. and Thorpe, C. 2001. Vehicle teleoperation interfaces. *Autonomous Robots* 11(1):514–525.

Foyle, D. 1992. Proposed evaluation framework for assessing operator performance with multisensor displays. *SPIE*, 1666:514–525.

Grange, S., Fong, T., and Baur, C. 2000. Effective vehicle teleoperation on the World Wide Web. In *Proceedings of the IEEE ICRA*, San Francisco, CA.

Graves, A. 1998. User interface issues in teleoperation. De Montfort University, Leicester, United Kingdom.

Hine, B., Hontalas, P., Fong, T., Piguet, L., Nygren, E., and Kline, A. 1995. VEVI: A virtual environment teleoperations interface for planetary exploration. In *Proceedings of the SAE ICES*, San Diego, CA.

Kay, J. and Thorpe, C. 1995. Operator interface design issues in a low-bandwidth and high-latency vehicle teleoperation system. In *Proceedings of the SAE ICES*, San Diego, CA.

McGovern, D. 1988. Human Interfaces in Remote Driving. Sandia National Laboratory, Albuquerque, NM, Technical Report SAND88-0562.

Meier, R., Fong, T., Thorpe, C., and Baur, C. 1999. A sensor fusion based user interface for vehicle teleoperation. In *Proceedings of the IEEE FSR*, Pittsburgh, PA.

Michel, O., Saucy, P., and Mondada, F. 1997. KhepOnTheWeb: An experimental demonstrator in telerobotics and virtual reality. In *Proceedings of the IEEE VSMM*, Geneva, Switzerland.

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

Newman, W. and Lamming, M. 1995. *Interactive System Design*, Addison-Wesley: Boston, MA.

Sheridan, T. 1992. *Telerobotics, Automation, and Human Supervisory Control*, MIT Press: Cambridge, MA.

Terrien, G., Fong, T., Thorpe, C., and Baur, C. 2000. Remote driving with a multisensor user interface. In *Proceedings of the SAE ICES*, Toulouse, France.

Terry Fong received his B.S. (1988) and M.S. (1990) in Aeronautics and Astronautics from the Massachusetts Institute of Technology. From 1990 to 1994, he was a computer scientist in the NASA Ames Intelligent Mechanisms Group and was co-investigator for virtual environment teleoperation experiments involving wheeled, free-flying and walking mobile robots. He is currently pursuing a Robotics Ph.D. at CMU and is performing his thesis research on advanced teleoperation interfaces at EPFL. His research interests include human-robot interaction, Web-based interfaces, and field mobile robots.

Charles Thorpe is Principal Research Scientist at the Robotics Institute of Carnegie Mellon University. He received his Ph.D. in Computer Science from Carnegie Mellon University (1984) under the guidance of Raj Reddy. He has published over 120 peer-reviewed papers in mobile robotics, computer vision, perception, teleoperation, man-machine interfaces, and intelligent highway systems. He is the leader of the Navlab group, which is building computer-controlled cars and trucks. His research interests include computer vision, planning, and control of robot vehicles operating in unstructured outdoor environments.

Charles Baur received his Ph.D. in Microengineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland in 1992. He is currently Adjoint Scientifique at EPFL and director of the Virtual Reality and Active Interfaces Group, which he created in 1993. In addition, he is founder and CEO of 2C3D, a start-up company specializing in real-time, 3D visualization for medical imaging and endoscopic applications.