A Sensor Fusion Based User Interface for Vehicle Teleoperation

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

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

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

Abstract

Sensor fusion is commonly used to reduce uncertainty in localization, obstacle detection, and world modeling. However, sensor fusion can also be used to improve teleoperation. In particular, we can use sensor fusion to create user interfaces which efficiently convey information, facilitate understanding of remote environments and improve situational awareness. We do this by selecting complementary sensors, combining information appropriately and designing effective representations. In this paper, we discuss sensor fusion for teleoperation, describe a vehicle teleoperation interface, and present our results.

1 Introduction

Vehicle teleoperation consists of three basic problems: figuring out where the vehicle is, determining where it should go, and getting it there. These problems can be difficult to solve, particularly if the vehicle operates in an unknown environment[5]. Furthermore, humans in continuous control may limit vehicle teleoperation. In particular, poor performance (e.g., imprecise control) and vehicle failures (e.g., roll over) are often caused by operator error[8].

Thus, to improve vehicle teleoperation, we need to make it easier for the operator to understand the remote environment, to assess the situation and to make decisions. In other words, we need to design the human-machine interface so that it maximizes information transfer while minimizing cognitive load. Numerous methods have been proposed to do this including supervisory control [11], teleassistance [10] and virtual reality [6].

Our approach is to enhance the quality of information available to the operator. Specifically, we use sensor fusion to create a user interface which efficiently and effectively displays multisensor data. In this way, we provide the operator with rich information feedback, facilitating understanding of the remote environment and improving situational awareness.

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

2 Related Research

2.1 Sensor fusion displays

VEVI

The Virtual Environment Vehicle Interface (VEVI) is an operator interface for direct teleoperation and supervisory control of robotic vehicles [6]. VEVI uses interactive 3D graphics to provide desktop and head-mounted/head-tracked stereo displays. Data from multiple, on-board vehicle sensors are used to dynamically update graphical vehicle and world models. VEVI has been used for numerous robotic exploration missions including the 1994 Dante II descent and terrain mapping of the Mt. Spurr volcano[4].

Nomad Driving Interfaces

Nomad, a mobile robot designed for planetary exploration, completed a 200-kilometer traverse of the rugged Atacama Desert (Chile) in 1997. Nomad was teleoperated by operators in North America using two primary interfaces: the *Virtual Dashboard* and the *Telepresence Interface*. The Virtual Dashboard provided a real-time visualization of Nomad's state including position on aerial images. The Telepresence Interface used panospheric camera images to create an immersive forward-looking display[13].

Situation Awareness Virtual Environment

The Situation Awareness Virtual Environment (SAVE) project is investigating applications of simulation, sensor fusion, and automation technologies for Air Traffic Control (ATC). Sensor fusion is being used to developing three-dimensional displays for surface traffic management in decreased-visibility situations[3]. It is interesting to note that sensor fusion issues for ATC are very closely related to those for teleoperation tasks discussed in this paper.

2.2 Telepresence and Augmented reality

Telepresence means that a display is sufficient and natural to create an illusion of physical presence at the remote site. Telepresence is commonly claimed to be important for direct manual teleoperation, but the optimal degree of immersion required to accomplish a task is still a topic for discussion [11]. Some researchers claim that high-fidelity telepresence requires feedback using multiple modalities (visual, auditory, haptic).

Augmented reality is a variation of *Virtual Environments (VE)*, otherwise known as Virtual Reality. Augmented reality allows users to see the real world (often with a head-mounted, see-through display) with virtual information (e.g., graphic overlays) superimposed or composited on the display[1]. To date, augmented reality has been used for a wide range of applications including medical, manufacturing, design, and entertainment.

3 Sensor Fusion for Teleoperation

In robotics, sensor fusion has been used primarily for improving the performance of autonomous processes such as localization and world modeling. It is our contention, however, that sensor fusion can (and should) also be applied to non-autonomous (i.e., human-centered) tasks. Specifically, we believe that sensor fusion can be used to create an efficient, multisensor display which provides rich information feedback and facilitates vehicle teleoperation.

3.1 Humans and Sensor Fusion

To apply sensor fusion to teleoperation, however, we need to consider not only conventional sensor fusion issues (sensor selection, sensor characteristics, data representation, fusion level, etc.) but also human needs and limitations. In particular, we need to identify what information is needed by a human, how it should be communicated, and how it will be interpreted.

Additionally, we must choose appropriate methods to combine information: the way we fuse data from a set of sensors will differ if the result is to be used by an autonomous process or by a human. For example, a world modeling process may need multiple-sensor range data to be fused globally, but a human may only require local fusion.

Finally, we need to design effective representations so that the data is accessible and understandable. As with all user interfaces, we must create displays which simplify human-machine interaction. Fused sensor data alone will not compensate for a poorly crafted display.

3.2 Integrating Multiple Sensors

For traditional teleoperation user interfaces, each part of the display is updated with data from a single sensor. Thus, the operator is forced to scan many display areas, interpret the information, and combine the (hopefully consistent) results to obtain spatial awareness. For complex situations or a multisensor system, the resulting cognitive workload can be extremely high and leads directly to fatigue, stress and inability to perform other tasks[11].

We can solve this problem by fusing the data from multiple sensors and presenting the result in a way that enables the operator to perceive quickly what is important for a specific task. This will reduce cognitive workload for the operator, leaving his mental resources to concentrate on the task itself. A particularly effective approach would be to dynamically select sensors and the fusion method based on the task being performed.

Multiple sensors provide information which can be considered as either redundant or complementary. We can use redundant information to reduce the uncertainty of measurements or (in case of sensor failures) to increase the reliability of the system. The major problem in fusing redundant information is that of registration: determining that the information from each sensor refers to the same features (spacial and temporal) in the environment. We can use complementary information to improve the coverage and effectiveness of sensing. For example, we can use a set of heterogeneous sensors to compensate (mask) the failure modes or limitations of each individual sensor.

3.3 Teleoperation Display Considerations

Representing depth

For teleoperation, good depth information is essential for making judgement about the relative position of objects in the remote world. In fact, many teleoperation errors can be directly attributed to inaccurate distance estimation[8]. Thus, when we build a teleoperation system, we need to provide ways for operators to accurately view depth.

The fundamental problem is that to do so, we must represent multi-dimensional data on a flat screen[12]. Artists have long relied on visual cues (see Table 1) for depicting three-dimensional scenes on paper.

Table 1. Visual depth cues

Visual Cue	Examples
Color and Brightness	Aerial perspective, shadows, relative brightness, texture gradient
Size	Retinal or familiar
Position	Occlusions, linear perspective, height in plane, stereopsis, motion parallax
Physiological	Depth by focus, eye convergence

User interfaces can also provide a sense of depth by rendering one or more of these depth cues. However, not all of these cues can be simulated on a flat screen. Stereopsis and motion parallax, for example, can only be created using special hardware (e.g. head mounted devices).

We must point out, however, that even under ideal conditions (i.e., direct natural viewing) humans are not accurate or consistent at making judgments of absolute distance. This means that even if a perfect illusion of depth can be created, spatially precise teleoperation requires that absolute information needs to be added to the display.

Use of color

Color provides a natural and efficient means for encoding multi-dimensional information. We can use color to provide specific display functions, e.g., red shading to indicate danger or to provide warning. However, we must avoid overusing color to prevent clutter and confusion.

Conventional computer displays encode colors with the RGB color space model. Unfortunately, RGB differs greatly from the way humans perceive color. A more appropriate model is HSV (Hue-Saturation-Value), which closely mimics humans color perception. HSV provides us with three distinct parameters for encoding information.

4 System Configuration

To investigate the use of sensor fusion for teleoperation, we have developed a vehicle teleoperation user interface which combines information from multiple sensors and displays the fused data to the operator in real-time[9].

4.1 Hardware

Sensors

We process data from a stereo vision system, a ring of ultrasonic sonars and vehicle odometry (wheel encoders). The stereo vision system and ultrasonic sonars are co-located on a sensor platform (see Figure 1) which may be mounted on a vehicle.

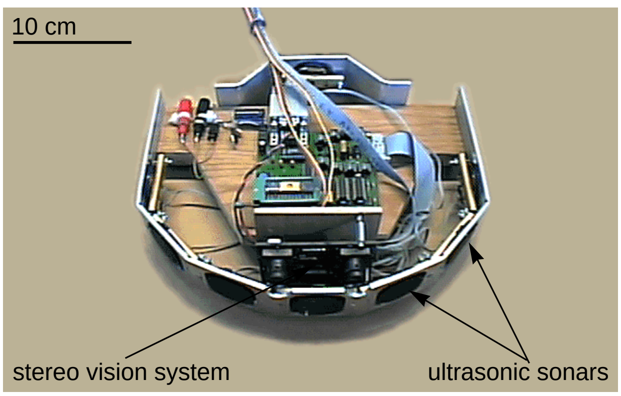

Figure 1. Multisensor platform

We chose these sensors based on their complementary characteristics (see Table 2) and their wide range of applications in mobile robotics. The stereo vision system is a Small Vision Module (SVM)[7]. The SVM provides 2D intensity (monochrome) images and 3D range (disparity) images at 5Hz frame rate. The ultrasonic sonars provide time-of-flight range at 25Hz. The beam cones of the three front sonars overlap with the SVM's stereo field-of-view. The remaining sonars are placed to optimize obstacle detection.

The primary advantage of stereo vision is its good angular resolution. Additionally, stereo vision can be done at relatively low cost and high speed. We do not consider the non-linear depth resolution of stereo vision to be a problem for teleoperation. This is because, in almost all cases, we are only concerned with areas close to the vehicle (where depth resolution is high) and not with distant areas (where depth resolution is low).

There are two primary problems associated with stereo vision. First, if there is not sufficient texture in the image to make a correlation, the output becomes noisy. This occurs when object surfaces are smooth or in low contrast scenes. Second, if objects are close to the cameras, the disparity becomes too large. Thus, there is a minimal distance (maximum disparity) for which range values can be computed.

Table 2. Characteristics of stereo vision and sonar

Criteria	Stereo Vision	Sonar
ranging	stereo correlation	time of flight
measurement	passive	active
range	0.6 to 6 m	0.2 to 10 m
angular resolution	high	low
depth resolution	non-linear	linear
data rate	5×10^5 bps	250 bps
update	5 Hz	25 Hz
field of view	40˚ horizontal / 35˚ vertical	30˚ beam cone
failure modes	low texture low/high intensity low bandwidth	specular reflection cross-talk noise

The advantage of using sonars is that they can detect obstacles with high confidence. Since sonars make active measurements, they are independent from the energy (and its associated noise) of the environment. Thus, if an object is well defined (i.e., located perpendicular to sonar axis and has good ultrasonic reflectivity) a very precise range measurement can be obtained.

Sonar, however, suffers from a number of drawbacks. Most significantly, sonar ranging is highly susceptible to error caused by non-perpendicular and/or off-axis targets. Additionally, range errors may arise due to multiple or specular reflections. Lastly, sonar transducers almost always have an inherently wide beam cone, which results in poor angular resolution.

The complementarity of 2D intensity images, stereo vision, and sonar is readily apparent if we examine failure situations. Table 3 lists several situations frequently encountered in vehicle teleoperation. As the table shows, none of the sensors works in all situations. However, the sensors as a group do provide complete coverage.

Table 3. Sensor failure situations

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

Vehicle

We initially placed the multisensor platform on an electric wheelchair equipped with wheel encoders (Figure 2). Although we were unable to teleoperate this system, we were able to design and verify concepts for the sensor fusion interface.

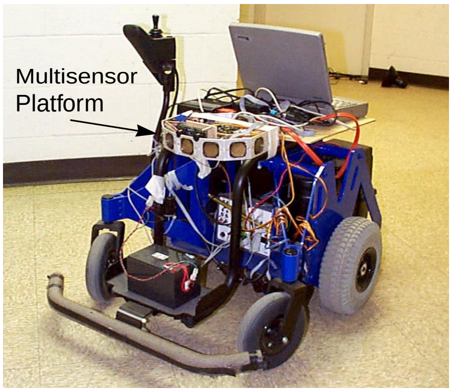

Figure 2. Multisensor platform on a wheelchair

Later, we mounted the multisensor platform on top of a PioneerAT mobile robot (Figure 3).The PioneerAT is a skid-steered, wheeled vehicle which is capable of traversing moderately rough natural terrain. We equipped the robot with an analog video transmitter and a RF modem for wireless communications. We teleoperated the robot using a combination of position and rate commands.

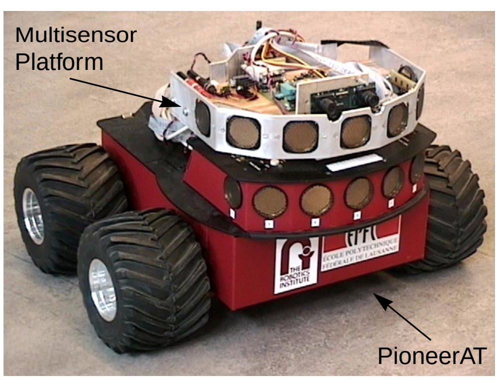

Figure 3. PioneerAT with multisensor platform

4.2 User Interface

Figure 4 shows the main window of our sensor fusion based user interface. The interface contains two primary display areas: (A) a 2D image with color overlay and (B) a local map constructed with sensor data. The 2D image is designed to facilitate scene interpretation and understanding. The color overlay directs the operator's attention to obstacles located near the vehicle and also aids distance estimation. The local map displays an occupancy grid which is updated in real-time from sensor data. The map is designed to improve situational awareness (especially monitoring of vehicle orientation) and maneuvering in cluttered environments.

The interface allows the operator to select from a number of sensor noise models (gaussian, uniform distribution,

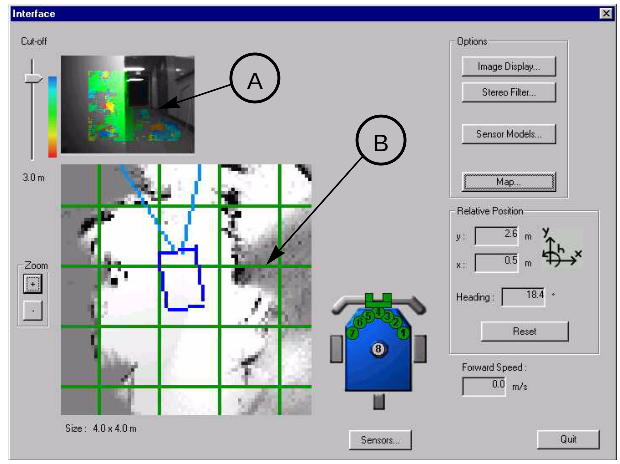

Figure 4. Sensor fusion user interface

etc.), to specify sensor filters (e.g., stereo texture detection), and to directly control each sensors's function. Additionally, the interface allows the operator to customize each display (color mapping, map scroll mode, display area, display priority, etc.).

4.3 Architecture

Fusing stereo and sonar

We fuse 2D image, stereo vision, sonar and odometry data using a *cross-filtering* algorithm. The flow of data through the cross-filter algorithm is shown in Figure 5.

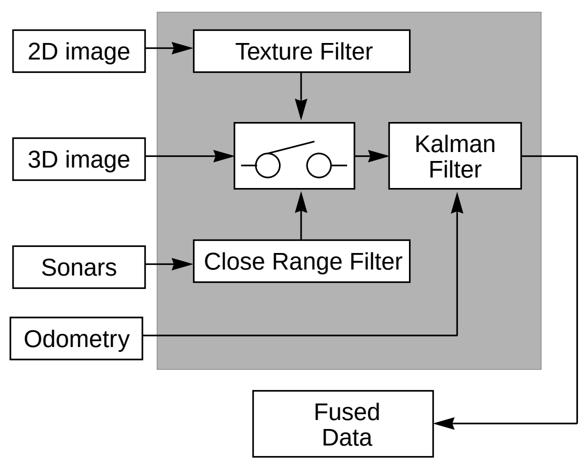

Figure 5. Cross-filter algorithm

This cross-filter algorithm produces fused data by first filtering the raw 2D image and sonar data, then using the filtered data and a Kalman filter to process the stereo information:

Texture Filter. Measures the amount of texture in the 2D image. This is used to filter regions with low textures (e.g. a white wall) where the stereo output would be noisy.

Close Range Filter. Filters regions where objects are too close (based on sonar range) for computing a correlation and the stereo output alone would not allow the operator to recognize dangerous obstacles.

Stereo switch. Declares regions in the 3D image "valid" or "invalid". Invalid data will not be used for further processing or displaying.

Kalman Filtering. Estimates the next stereo frame based on vehicle speed and the time between frames. We combine this estimate with the actual measurement to reduce noise and to improve stability.

Processing

When the system is running, we continually process the sensor data from the stereo vision system, the sonars and on-board odometry to generate the two user interface displays. An "event generator" produces messages for the operator when certain events occur. For example, if a sensor fails or gives suspicious data, a message warns the operator that the sensor is faulty.

Image Display

We create the image display by overlaying range information as colors on a 2D intensity image taken from one of the cameras. This method does not provide an absolute indication of range because humans cannot accurately identify color hue. However, it focuses the operator's attention on near objects, warns the operator if an object is very close, and enhances estimation of relative distances. In addition, the image display also contains a projected grid which is obstructed (hidden) by above-ground obstacles. This grid also improves distance estimation (e.g. the size of a grid cell corresponds to the size of the vehicle and helps the operator to identify free and occupied space).

We rely primarily on stereo vision for range data because it has good angular resolution. This information is filtered according to Figure 6.The concept is to use the other sensors to filter or replace stereo ranges. For example, if we detect from the 2D image that the scene has low image texture, then the stereo range data is not mapped. Similarly, if we detect nearby obstacles from the sonar, the stereo information is replaced by the sonar information.

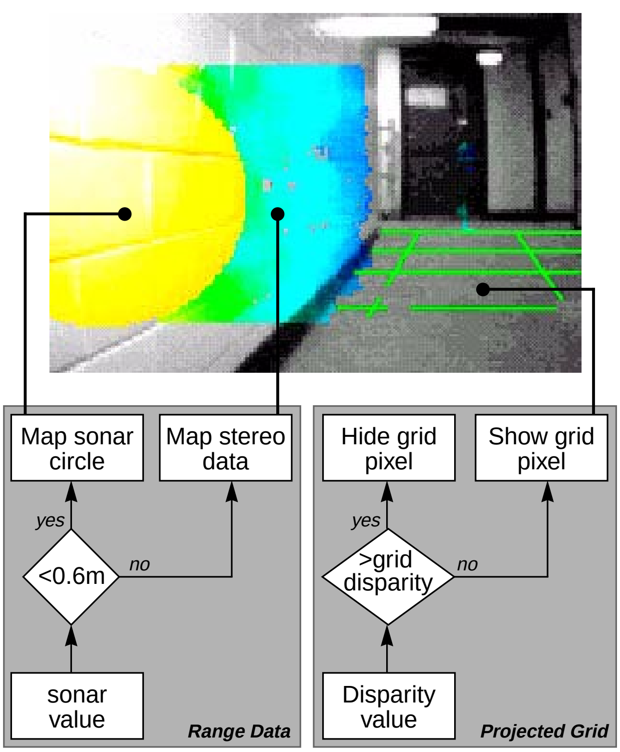

Figure 6. Image display processing

Local Map Display

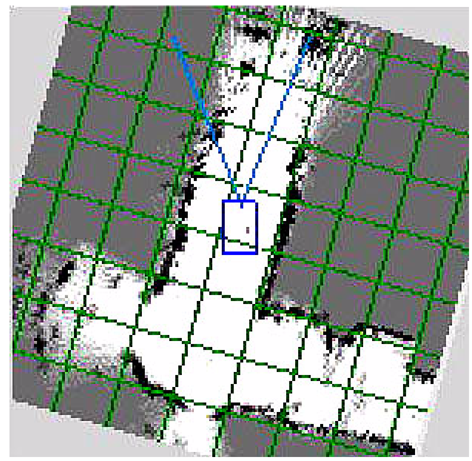

Figure 7. Local map display

We build the map display by combining vehicle odometry with stereo and sonar ranges onto an occupancy grid using Histogramic In-Motion Mapping[2]. Occupancy grids are a probabilistic method for fusing multiple sensor readings into a surface map. The advantage of this framework is that sensor fusion is done very straightforward by updating a single, centralized map using each range sensor. We visualize the occupancy grid by encoding the certainty of a cell being occupied as a gray level (see Figure 7).

5 Results

Image Display

Figure 8 shows an example where we first map only the stereo information (top left), then only the sonar information (top right) and then the fused stereo and sonar information (bottom).

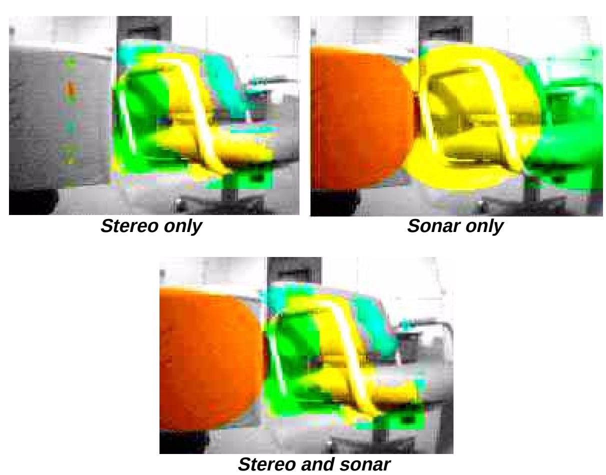

Figure 8. Improvement by fusing stereo and sonar

In the top left image, the chair is mapped correctly, but the obstacle on the left cannot be seen because it does not have enough texture and is too close for stereo. In the top right image, the objects are detected by sonar, but the resolution is very low and the image is difficult to interpret. Fusing data from both sensors yields the bottom image: the chair is mapped with good resolution (stereo) and the obstacle on the left side is now clearly visible (sonars)

Local Map Display

A significant problem with sonars is poor angular resolution which may result in considerable uncertainty about object locations. Nevertheless, if we take numerous sonar readings from a vehicle in motion, the contours of objects become visible and false measurements (e.g. due to specular reflections) tend to be eliminated. Figure 7 shows the map of an indoor corridor produced purely by sonar data (note that corridor walls are somewhat rough).

By fusing stereo with the sonar data, we can improve the map. At each update, we extract a single (horizontal) line from the disparity image and apply it to the grid. With the high angular resolution from stereo, object contours in front of the vehicle (in the stereo field of view) are mapped more clearly. Figure 8 shows the vehicle approaching some stairs. The stairway walls appear clearly with the fused data. With sonars alone, they are not seen at all.

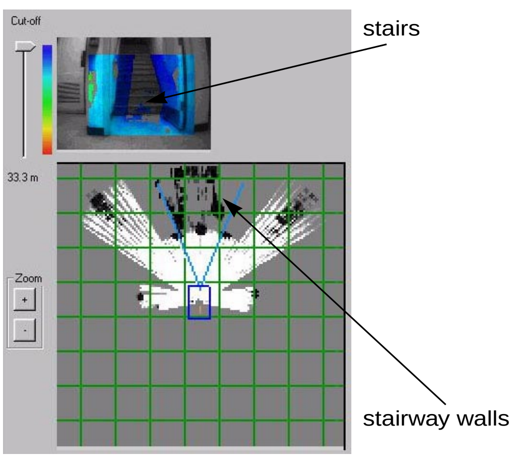

Figure 9. Local map created with stereo and sonar

6 Conclusion

In our work, we have implemented a user interface for vehicle teleoperation which demonstrates the utility of fusing multiple sensor data. We have used stereo vision, sonar information, and odometry to create a 2D image overlay which improves estimation of relative distance and spotting of nearby obstacles. Similarly, we use the fused data to improve occupancy grid-based map building.

By using sensor fusion, we believe we can build better user interfaces. Combining data from multiple, complementary sensors allows us to increase the quality of the information available to the operator and to make human-machine interaction more efficient. In short, sensor fusion offers us the potential to greatly improve teleoperation.

Acknowledgments

We would like to thank Nicolas Chauvin (CMU & EPFL) for providing his assistance and feedback during the development of the multisensor platform. We would also like to thank Illah Nourbakhsh for use of the electric wheelchair from the CMU Mobile Robot Programming Lab.

References

[1] Azuma, R., “A Survey of Augmented Reality”, *Presence: Teleoperators and Virtual Env.*, (6), 1997.

[2] Borenstein, J, and Koren, Y., “Histogramic In-motion Mapping for Mobile Robot Obstacle Avoidance”, *IEEE Journal of Rob. and Automation*, 7(4), 1991.

[3] Foyle, D.C. Proposed evaluation framework for assessing operator performance with multisensor displays. SPIE Volume 1666, 514-525, 1992.

[4] Fong, T., et. al., “Operator Interfaces and Network-Based Participation for Dante II”, *SAE Intl. Conf. on Environmental Systems*, 1995.

[5] Fong T., Thorpe, C., and Baur, C.,. “Collaborative Control: A Robot-Centered Model for Vehicle Teleoperation”, *AAAI Spring Symposium on Agents with Adjustable Autonomy*, Stanford, CA, 1999.

[6] Hine, B., et al., “VEVI: A Virtual Environment Teleoperations Interface for Planetary Exploration”, *SAE Intl. Conf. on Environmental Systems*, 1995.

[7] Konolige, K., “Small vision system: Hardware and Implementation”, *Eighth International Symposium on Robotics Research*, Hayama, Japan, 1997.

[8] McGovern, D. “Human Interfaces in Remote Driving”, Technical report SAND88-0562, Sandia National Laboratory, Albuquerque, NM, 1988

[9] Meier, R. “Sensor Fusion for Teleoperation of a Mobile Robot”, Diplome Thesis, Swiss Fed. Inst. of Technology Lausanne, Switzerland, March 1999.

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

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

[12] Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Princeton, NJ, 1990.

[13] Wettergreen, D., et. al., “Operating Nomad During the Atacama Desert Trek”, *Field and Service Robotics Conference*, Canberra, Australia, 1997.