

# Mobile Robotic Teleguide Based on Video Images

## Comparison Between Monoscopic and Stereoscopic Visualization

Robot teleoperation is inherently related to sensor data transmission. Sensor data can be interpreted by the robotic system before being transmitted and presented to a user. This typically happens with sonar and laser sensors. In other cases, sensor data can directly be presented to a user to let him or her interpret the transmitted information. This happens, for example, with visual data.

Robot teleoperation systems typically rely on two-dimensional (2-D) displays. These systems suffer from many limitations, e.g., misjudgment of self-motion and spatial localization; limited comprehension of remote ambient layout, object size, and shape; etc. This leads to unwanted collisions during navigation and long training periods for an operator.

An advantageous alternative to traditional 2-D (monoscopic) visualization systems is represented by the use of stereoscopic viewing. In the literature, we can find works demonstrating that stereoscopic visualization may provide a user with a higher sense of presence in remote environments because of higher depth perception, leading to higher comprehension of distance as well as aspects related to it, e.g., ambient layout, obstacle perception, and maneuver accuracy [2]–[6], [10].

The earlier conclusions can in principle be extended to teleguided robot navigation, where the use of stereo vision is expected to improve the navigation performance and driver capabilities [3]–[6]. However, it is hard to find works in the literature addressing stereoscopic mobile robot teleguide. In addition, it is not straightforward how stereo viewing would be an advantage for indoor workspaces where the ambient layout, typically man-made, would be simple and emphasizing monocular depth cues such as perspective, texture gradient, etc., hence diminishing the advantage of a binocular stereo.

The goal of the proposed work is to analyze the characteristics and advantages of a telerobotic system based on video transmission and stereoscopic viewing.

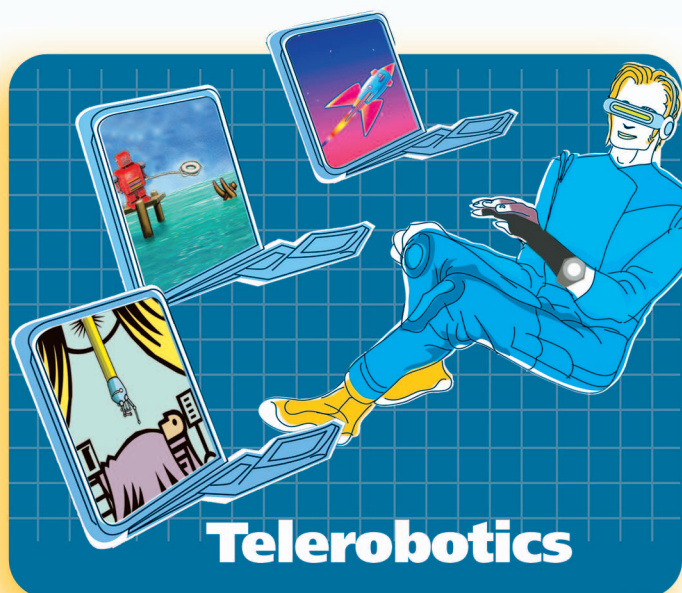
The proposed investigation follows a systematic approach based on the identification of main factors and a usability evaluation designed according them. Two different three-dimensional (3-D) visualization facilities are considered to evaluate its performance on systems with different characteristics, cost, and application context. The aim is to gain an insight into the problem and to understand on what system, and to what extent, is the stereo viewing beneficial.

In the next section, we introduce visual sensors and stereo viewing. Then, the proposed investigation strategy is presented, followed by experimental design, test setup, and result analysis. Some final remarks conclude the article.

### Video-Based Teleoperation and Stereo Viewing

When operating in unknown or hazardous environments, accurate robot navigation is of paramount importance. Errors and collisions must be minimized. Performance in robot teleoperation can

**BY SALVATORE LIVATINO,  
GIOVANNI MUSCATO,  
SALVATORE SESSA,  
CHRISTINA KÖFFEL,  
CARMELO ARENA,  
ALBA PENNISI,  
DANIELE DI MAURO,  
AND ERINC MALKONDU**



© CREATAS, ARTVILLE, & DIGITAL VISION

be improved by enhancing the user's sense of presence in remote environments (telepresence). Vision being the dominant human-sensor modality, much attention has been paid to the visualization aspect.

The use of visual sensors in telerobotics has become very common, because video images provide rich and highly contrasting information. Therefore, they are largely used in many tasks that need accurate observation and intervention.

The rich information provided by a camera may require a large bandwidth that needs to be transmitted at interactive rates. This often presents a challenge in video-based robot teleoperation in case of transmission to distant locations or when the employed medium has limited communication capabilities.

Several video-compression techniques have been developed to reduce or solve the transmission-delay problem. In case of stereo images, the information to be transmitted is larger (double, in principle). However, this can be greatly reduced, e.g., based on redundant information in stereo images, while specific networks for streaming video have been proposed [4].

The bandwidth constraint may lead to transmission delays, and this may affect the interaction performance, e.g., response speed and accuracy. Corde et al. [7] claim that a delay of more than 1 s leads to a significant decrease in performance.

Different approaches and display technologies have been developed for generating 3-D stereo-visualization systems [4]. The basic idea supporting stereoscopic visualization is that this is closer to the way we naturally see the world, which tells us about its great potential in teleoperation.

Stereoscopic visualization can play a major role toward increasing the user's involvement and immersion because of the increased level of depth awareness. This is expected to lead to a more accurate performance of the task and comprehension of the environment. There are several works in the literature that focus on stereoscopic visualization, typically assessing stereoscopic versus monoscopic visualization, which can be classified as either application specific (i.e., application-oriented user studies) or abstract test (i.e., abstract tasks and content with general performance criteria) [2].

Test trials often deal with assessing the role of the most-dominant depth cues, e.g., interposition, binocular disparity, movement parallax, and their consequence to user adaptation to new context (e.g., user's learning capabilities) [9].

The parameters that help to assess stereoscopy benefits are item difficulty and user experience, accuracy, and performance speed [9].

Stereoscopic visualization is believed to present the necessary information in a more natural way that facilitates all human-machine interactions [3], and stereoscopy improves comprehension and appreciation of the presented visual input, perception of structure in visually complex scenes, spatial localization, motion judgment, concentration on different depth planes, and perception of surface materials.

Most of the benefits of stereo viewing may affect the robot teleguide, because stereopsis enhances the perception of relative locations of objects in the observed remote worlds [3], impression of telepresence and of 3-D layout [1], ability to skillfully manipulate a remote environment [4], response time and accuracy when operating in dangerous environments, etc.

The main drawback that has yet prevented many applications is that the users are called to make some sacrifices [10]. A stereo view may be hard to get right at the first attempt; hardware may cause cross talk, misalignment, and image distortion (due to lens, displays, projectors); and all this may cause eye strain, double-image perception, depth distortion, and look-around distortion (typical for head-tracked displays).

## Proposed Investigation

We have identified a set of factors that affect the user's performance in mobile robot teleguide. These are related to the capability of a user to estimate the following factors:

- ◆ *Spatial localization*: Robot position in relation to surrounding objects
- ◆ *Spatial configuration*: Ambient layout and 3-D structures
- ◆ *Depth relationships*: Egocentric and relative distances
- ◆ *Motion perception*: Robot motion and environment dynamics
- ◆ *Action control*: Response to provide commands and robot feedback.

These factors are not independent. An improvement to one of them typically has a consequence on the others. The user's ability to estimate those factors can be increased by a training activity and by improving the system design.

The possibility for stereoscopic visualization may affect some of the aforementioned factors to a different extent. This depends on the following factors:

- ◆ *Space and budget*: The available space for user interface has an effect on the choice for system dimension, structure, and projection modality. Different approaches and displays have very different costs.
- ◆ *Robot and sensors*: This concerns mobile platform kinematic and control modality, processing speed, and sensor data. The sensor data contribute to image type and quality of the visual output.
- ◆ *Stereo approach and display*: This affects the performance in terms of sense of presence, depth impression, level of realism, viewing comfort, and adequacy to application. The display size plays an important role [2], [10].

In this work, we consider only the use of the visual sensor (video camera) to analyze its potential and limitation in mobile robot teleguide. The use of a different sensor and collaboration among different sensor modalities are for future work. We have limited our investigation to two different aspects.

- ◆ *Stereoscopic approach (colored anaglyph and polarized filters)*: These approaches are very different in terms of cost and performance. Colored anaglyph is cheap, easy to produce, and very portable. However, it has poor color reproduction, and it often generates cross talk, which affects precision and viewing comfort. On the other hand, polarized filters nicely reproduce colors, has nearly no cross talk, and it is very comfortable to the viewer. However, it requires a more complex and expensive setup, and it is less portable.
- ◆ *Visual display (laptop and wall)*: These displays are different in terms of size and technology. A laptop display uses liquid crystal display (LCD) technology, and it has a relatively small display size, typically up to 19 in with

high resolution. A wall display is instead composed by a projector and a screen with a size up to several meters.

The visualization systems used in our tests are as follows (Figure 1).

- ◆ *Anaglyph laptop*: We believe that the anaglyph approach can still be effective despite the disadvantages. Therefore, we wanted to test it. The colored anaglyph approach is proposed on a 15-in laptop display. This results in having a stereo on a portable system, which is suitable for tasks requiring a user to be close to a mobile robot operational environment and a low-cost hardware.
- ◆ *Polarized wall*: The polarized filters approach is proposed on a  $2 \times 2$  m wall display. This results in higher user involvement, 3-D impression, and comfort suitable for training purposes or for tasks requiring accurate maneuvering and long operational sessions.

We have restricted our experimental conditions to indoor environments and a factory-like scenario. The stereo camera system is set based on the following objectives:

- ◆ *Realistic observation*: This is intended as an observation that appears close to that obtained when looking at the environment with our own eyes.
- ◆ *Expected distances*: The average distance to objects of interest is about 2 m.
- ◆ *Comparable evaluation*: All different experimental trials run under the same conditions.

Based on the earlier objectives, a compromise setting is estimated for the camera parameters. The camera system sits on top of the robot, with a height of 95 cm, and it looks  $25^\circ$  downward (tilt angle) and has a baseline of 7 cm. Our stereo camera system has been designed based on directives from the literature. To investigate the role of camera parameters in a stereo setting is out of the focus of this article. Our aim is the usability evaluation with a realistic camera setting.

## Experimental Design

The proposed evaluation aims at detecting the overall usability of the proposed system. The purpose is to obtain a tangible proof of user's navigation skills and remote environment comprehension under different circumstances.

The research question involves the following two aspects:

- ◆ *Mono versus stereo*: What are the main characteristics and advantages of using stereoscopic visualization in

mobile robot teleguide in terms of navigation skills and remote environment comprehension?

- ◆ *Anaglyph laptop versus polarized wall*: How may the characteristics and advantages associated to stereoscopic viewing vary for different approaches of stereo and display systems?

The usability study is a within-subject evaluation, and it is designed according to recommendations gathered from the literature and authors' experience and previous works on the evaluation of virtual-reality (VR) applications [8]. Twelve participants were tested.

Each participant is asked to teledrive a remotely located mobile robot on both the proposed facilities (laptop and wall system), using both stereo and mono visualization modalities. This results in four navigation trials per participant.

We conform to the traditional approaches in terms of forms and questionnaires, with few additions [8]. We use a seven-scale semantic differential for answering the questions.

The schedule for participant activities includes the timing of the single tasks, the overall completion time (with breaks, form filling, debriefing, etc.), and the task sequence per participant. It is very important to counterbalance the sequence of tasks to avoid fatigue and learning effects.

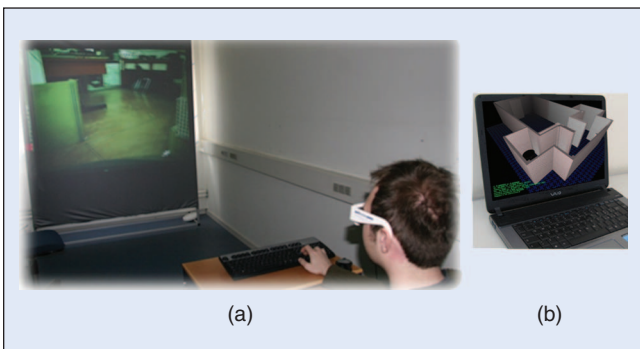
An initial practice session is administered to get acquainted with the task and the system. A pilot study is also performed before executing the formal study so as to debug and refine the experimental design.

The study includes qualitative and quantitative evaluations. The data related to qualitative evaluation are gathered through questionnaires that are designed after the following subjective parameters:

- ◆ *Depth impression*: The extent of perceived depth when observing different objects
- ◆ *Suitability to application*: The adequacy of the system and stereo approach to the specific task
- ◆ *Viewing comfort*: The eye strain and general body reaction
- ◆ *Level of Realism*: The realism of the visual feedback including object dimension and general appearance
- ◆ *Sense of presence*: The perceived sense of presence and isolation from surrounding space.

The data related to quantitative evaluations are gathered through robot sensors that are processed to obtain the following evaluation measurements.

- ◆ *Collision rate*: The collision number divided by the completion time. It provides information about obstacle detection and avoidance that is independent from user speed. This is the most relevant measurement as it provides explicit information about driving accuracy.
- ◆ *Collision number*: The number of collisions registered during a trial. It may provide information about obstacle detection and avoidance.
- ◆ *Obstacle distance*: The mean of minimum distance to obstacles along the path followed during a trial. It provides information about obstacle detection and avoidance.
- ◆ *Completion time*: The time employed to complete the navigation trial. It provides information about the user's environment comprehension. This parameter may also show the user's confidence (sometimes, a false confidence).



**Figure 1.** The visualization systems used in our tests. (a) The wall. (b) The laptop.

## Robot teleoperation is inherently related to sensor data transmission.

The knowledge of the completion time is needed to estimate the collision rate.

- ◆ *Path length*: The length of the robot journey. It may provide information about drive efficiency and obstacle detection.
- ◆ *Mean speed*: The mean speed of each trial. It may show the user's confidence.

All the acquired data go through statistical and graphical evaluations to identify potential tendencies (based on mean, standard deviation, and specific observations) and precise trends [based on the analysis of variance (ANOVA)].

### Test Setup and Procedure

The experiment involved facilities on different sites: local and remote. The remote site is the location where the robot operated. This was the Robotics Laboratory at the Department of Engineering Electrical Electronics and Systems (DIEES), University of Catania, Italy. The local site is the location where the user (tele-) operated. This was the Medialogy Lab at the Aalborg University in Copenhagen, Denmark.

The experiment required a close coordination between the remote and local sites.

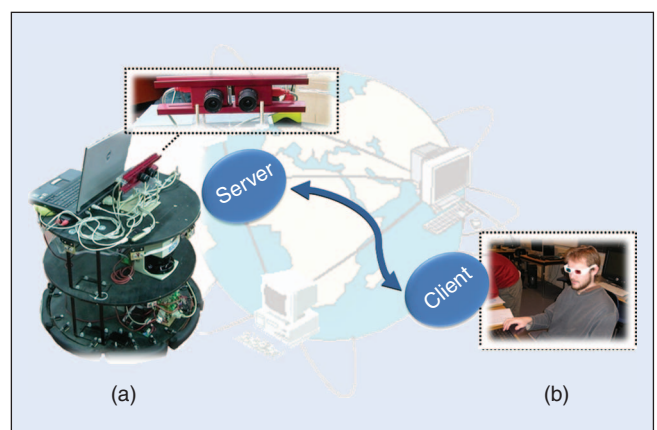
- 1) *Robotics system*: Our robot is the 3rd Version of MOBILE Robot DIEES University of Catania (Italy) (3MOR-DUC), a mobile robot platform with a cylindrical shape (diameter = 75 cm × height = 85 cm) (Figure 2). The platform is equipped with bumpers, encoders, laser, and a stereo camera on top. The platform carries two car batteries and a laptop system on board.
- 2) *Visualization systems*: We use two visualization systems. The first one is an ordinary laptop PC with 1 GB RAM and a 15-in wide screen, displaying stereo using the colored anaglyph approach. The second system is a polarized wall composed of a 2 × 2 m silver screen, located approximately 5 m away, and two powerful projectors equipped with linear-polarized filters. Figure 1 shows the two systems.
- 3) *Network connection*: The teleoperation system is implemented as a client-server architecture based on a standard network protocol [third International Standard Organization's Open System Interconnect (ISO/OSI) level]. The server sits onboard of the robotic platform. The client runs on the user system. The client and server are connected through an Internet link. We considered the hypertext transfer protocol (HTTP) to send packages related to our teleoperation. We choose the HTTP because of the presence of firewalls and proxy systems in our local site. A faster, but less reliable alternative could be the user datagram protocol (UDP), designed to handle real-time control.

The system used in our experimentation has a nearly constant delay of 1 s, which allowed for interactive teleoperation. Despite the delay, the robot can be sufficiently well managed at the price of a relative slower drive. We consider the aforementioned delay as a realistic setting for many applications requiring visual feedback in teleoperation. There is no difference in transmission delay between

mono- and stereo-viewing conditions, because the streamed video is always sent in stereo (and it is up to the local operator to set for monoscopic or stereoscopic viewing). This assured us that the performed tests were not biased in principle by a delay difference.

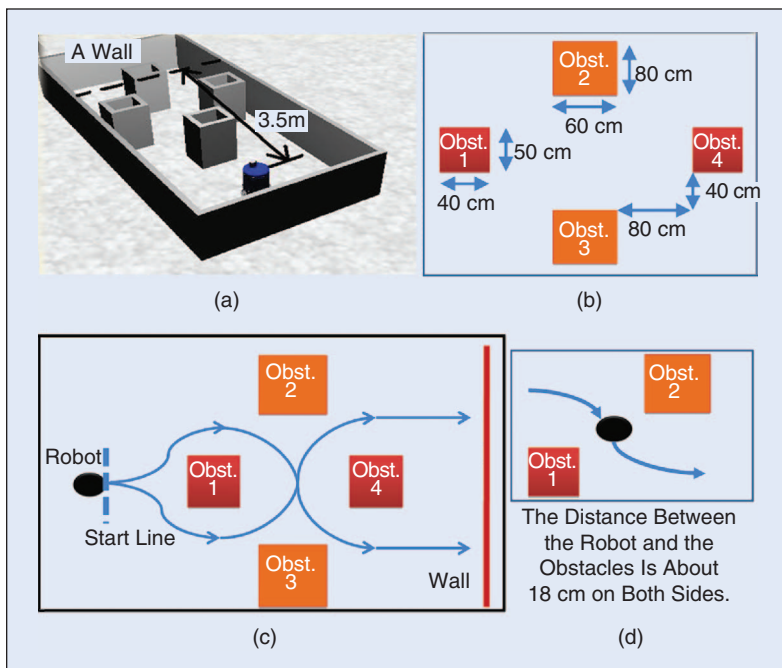
The images of the two cameras, each with resolution 640 × 480 pixels, are linked together side by side. They are then jpeg compressed and sent by the server through the HTTP package. The client decompresses the image that is then resized according to the visualization resolution and stereo approach.

- 4) *Participants*: Twelve subjects are tested among students or staff members of the university. The target population is composed of participants with varying background and have none or medium experience with VR devices. The age of the participants ranged between 23 and 40, with an average of 25.8. This is done to guarantee a great internal variance for unbiased and reliable results.
- 5) *Organization*: The test trials are conducted during several days. The average execution time per participant is about 40 min. Each participant executes the same number of tasks under the same conditions. Participants are assisted by a test monitor and a technician during the entire test session. The participant task and facility order is given according to a predetermined schedule to counterbalance the sequence of tasks.
- 6) *Procedure*: Four steps were performed. First, an introductory phase included a practice drive. Then, the

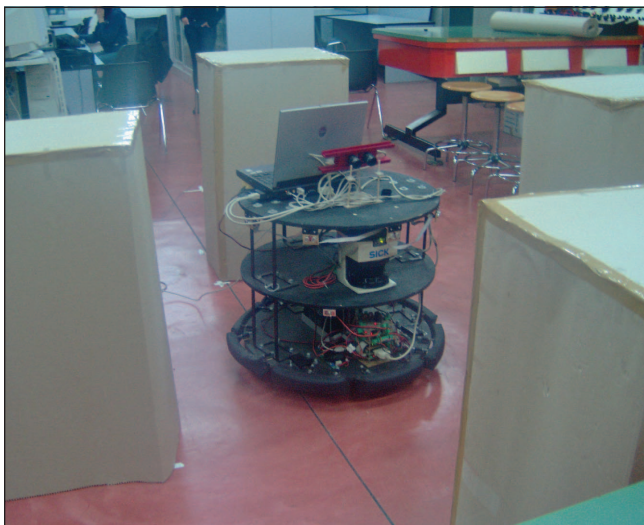


**Figure 2.** A representation of the local-remote system interaction. (a) Mobile robot in Robotics Lab, University of Catania, Italy, equipped with a stereo camera and sitting on the platform, responsible for capturing stereo images or mono images. (b) User in Medialogy Lab, Aalborg University, Denmark, in front of a laptop (or wall) system. He wears goggles to obtain 3-D visual feedback of the remote environment.





**Figure 3.** The robot workspace. (a) The 3-D representations of the environment given to the user. (b) Obstacle sizes. (c) A 2-D floor map with expected trajectories, start line, and the final position (on a wall line). (d) Distances to robot.



**Figure 4.** The robot 3MORDUC operating during a test trial.

user teledrives the robot toward a final location while avoiding collisions. The participants were eventually asked to complete predesigned questionnaires. Figure 3 shows the robot workspace.

Figure 4 shows our mobile robot during experimental trials.

## Result Analysis

The results of the experimentation are shown in Figures 5 and 6 (the descriptive statistics) and Table 1 (the inference).

We measured statistical significance of results by estimating the ANOVA. In particular, a two-way ANOVA was applied to measure the effect of stereo–mono and laptop–wall on each of

the dependent variables. We set the  $P$  value to 0.05 to determine whether the result is judged statistically significant.

The results are commented according to the proposed research questions.

### Mono Versus Stereo

1) *Collision*: Under stereoscopic visualization, the users perform significantly better in terms of collision rate. The ANOVA shows the main effect of stereo viewing on the number of collisions per time unit:  $F = 5.83$  and  $P = 0.0204$ . The improvement when comparing mean values is 20.3%. Both collision rate and collision number are higher in case of monoscopic visualization in most of the users' trials. The diagram in Figure 7 shows the collision number for a typical user in both the facilities. This supports the expectation, based on the literature, that the higher sense of depth provided by stereo viewing may improve the driving accuracy.

2) *Obstacle distance*: There is no relevant difference in the mean of minimum distance to obstacles between mono- and stereo driving.

The result from the ANOVA is not significant, and the improvement when comparing mean values is only 3.3%.

- 3) *Completion time*: There is no significant difference in completion time. Nevertheless, we have observed that the time spent for a trial is greater in stereo visualization in 77% of the trials. The test participants have commented that the greater depth impression and sense of presence provided by stereoscopic viewing make a user spending a longer time in looking around the environment and avoid collisions.
- 4) *Path length*: There is no significant difference in path length. Nevertheless, the user shows different behaviors under mono- and stereo conditions. Under stereo-viewing conditions, the path is typically more accurate and well balanced.
- 5) *Mean speed*: The results for the mean speed show a clear tendency in reducing speed in case of stereo viewing. The ANOVA shows a tendency to be significant ( $F = 3.04, P = 0.0891$ ). In general, a slower mean speed is the result of a longer time spent to drive through the environment.
- 6) *Depth impression*: All users had no doubts that depth impression was higher in case of stereo visualization. The result from ANOVA shows the main effect of stereo viewing:  $F = 51.86$  and  $P = 0.0$ . This result is expected and agrees with the literature.
- 7) *Suitability to application*: There is no significant difference in terms of adequacy of the stereo approach and display to the specific task. Nevertheless, we notice an improvement of 74% on mean values in the case of polarized stereo (anaglyph stereo penalizes the final result).
- 8) *Viewing comfort*: There is no significant difference in viewing comfort between stereo and mono visualization,

which contradicts the general assumption of stereo viewing being painful compared with mono. Stereo viewing is considered even more comfortable than mono in the polarized wall. The higher sense of comfort of the wall system is claimed to be gained by a stronger depth impression obtained in stereo. Our conclusion is that the low discomfort of polarized filters is underestimated as an effect of the strong depth enhancement provided in the polarized wall.

9) *Level of realism*: All users find stereo visualization closer to how we naturally see the real world. The result from the ANOVA shows the main effect of stereo viewing:  $F = 23.79$  and  $P = 0.0$ . The mean values show an improvement of 84%.

10) *Sense of presence*: All users believe that stereo visualization enhances the presence in the observed remote environment. The ANOVA has  $F = 51.86$  and  $P = 0.0$ . The improvement in mean values is 97%.

### Anaglyph Laptop Versus Polarized Wall

1) *Collision*: Users perform significantly better in the laptop system in terms of collision rate. The ANOVA has  $F = 8.65$  and  $P = 0.0054$ , and the improvement when comparing mean values is 10.3%. The collision number ANOVA shows a tendency to be significant ( $F = 3.32$ ,  $P = 0.0757$ ). The effect of stereoscopic visualization compared with the monoscopic one is analogous on both facilities.

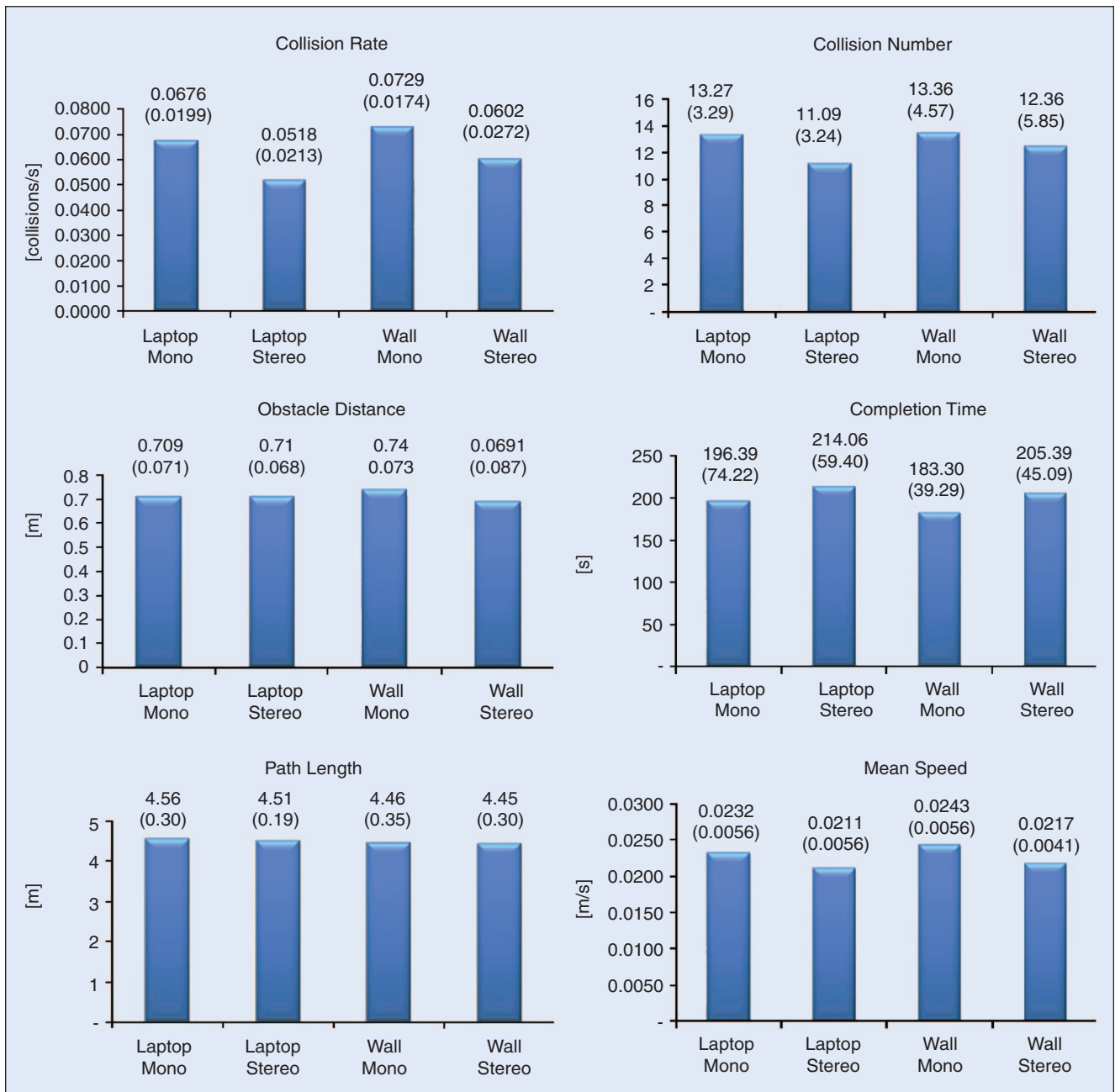
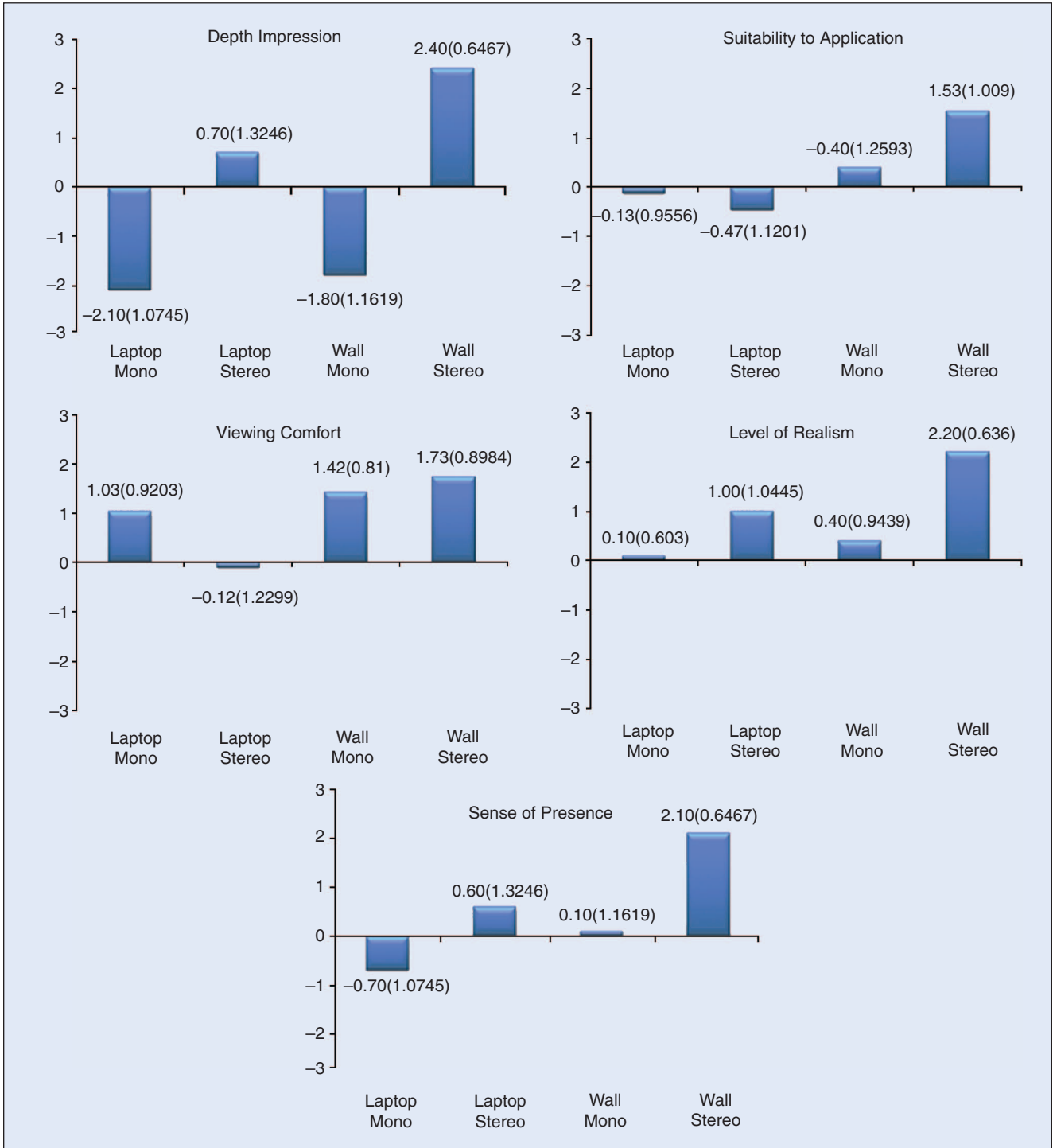


Figure 5. Bar graphs illustrating mean values and standard deviation (in brackets) for the quantitative variables.

- 2) *Obstacle distance*: When sitting in front of the laptop system, users perform significantly better compared with the wall in terms of mean of minimum distance to obstacles. The ANOVA has  $F = 7.63$  and  $P = 0.0086$ .
- 3) *Completion time*: There is no significant difference between the two systems. Nevertheless, a faster performance is noted in larger screens. Most of the participants argued that the faster performance is due to the higher

sense of presence given by the larger screen. The higher presence enhances driver's confidence. Therefore, smaller time is employed to complete a trial.

- 4) *Path length*: There is almost no difference between the two systems in terms of path length.
- 5) *Mean speed*: There is no significant difference in mean speed between the two systems. The higher mean speed is typically detected on the wall. The large screen requires



**Figure 6.** Bar graphs illustrating mean values and standard deviation (in brackets) for the qualitative variables. The qualitative data were gathered through questionnaires, where the participants provided their opinions by assigning values that ranged between +3 (best performance) and -3 (worst performance).

users to employ their peripheral vision, which allows for spending less time looking around and explains the wall better performance. The mean values show the same patterns on both facilities.

- 6) *Depth impression*: There is no significant difference between the two facilities. This confirms that the role played by the stereoscopic visualization is more relevant than the change of facilities. The improvement when driving in stereo is 76% on the laptop and 78% on the wall.

It may surprise the reader that most users claim a very high 3-D impression with laptop stereo. Confirmation that perceived depth impression can be high in small

screens is found in the work of Jones et al. [6], which shows how the range of depth tolerated before the loss of stereo fusion can be quite large on a desktop. In our case, the range of perceived depth in the laptop stereo typically corresponds a larger workspace portion than in large screens systems (in other words, the same workspace portion corresponds to a wider range of perceived depth for large screens), but we typically lose stereo after 5–7 m.

- 7) *Suitability to application*: There is no significant difference between the two systems; however, we can observe that users believe that a large visualization screen is more suitable to the mobile robot teleguide. This goes along with

**Table 1. The results of two-way ANOVA for the quantitative and qualitative measurements.**

<b>Collision Rate</b>										<b>Collision Number</b>									
	<b>SS</b>	<b>DoF</b>	<b>F</b>	<b>P</b>		<b>SS</b>	<b>DoF</b>	<b>F</b>	<b>P</b>		<b>SS</b>	<b>DoF</b>	<b>F</b>	<b>P</b>		<b>SS</b>	<b>DoF</b>	<b>F</b>	<b>P</b>
Mono–Stereo	0.00228	1	5.83	0.0204	Mono–Stereo	27.841	1	1.57	0.2181	Mono–Stereo	4348.3	1	1.4	0.2435	Mono–Stereo	0.0001	1	3.04	0.0891
Laptop–Wall	0.00338	1	8.65	0.0054	Laptop–Wall	59.114	1	3.32	0.0757	Laptop–Wall	2992.9	1	0.96	0.332	Laptop–Wall	0.00007	1	2.18	0.1473
Interaction	0.00017	1	0.45	0.5076	Interaction	2.75	1	0.15	0.6962	Interaction	373.2	1	0.12	0.7306	Interaction	0.00001	1	0.35	0.5553
Error	0.01561	40			Error	711.273	40			Error	124120.4	40			Error	0.00154	40		
<b>Obstacle Distance</b>										<b>Completion Time</b>									
Mono–Stereo	6359	1	1.28	0.2638	Mono–Stereo	4348.3	1	1.4	0.2435	Mono–Stereo	4348.3	1	1.4	0.2435	Mono–Stereo	0.0001	1	3.04	0.0891
Laptop–Wall	37757.9	1	7.63	0.0086	Laptop–Wall	2992.9	1	0.96	0.332	Laptop–Wall	2992.9	1	0.96	0.332	Laptop–Wall	0.00007	1	2.18	0.1473
Interaction	124.9	1	0.03	0.8746	Interaction	373.2	1	0.12	0.7306	Interaction	373.2	1	0.12	0.7306	Interaction	0.00001	1	0.35	0.5553
Error	198013	40			Error	124120.4	40			Error	124120.4	40			Error	0.00154	40		
<b>Path Length</b>										<b>Mean Speed</b>									
Mono–Stereo	0.00445	1	0.05	0.8164	Mono–Stereo	0.0001	1	3.04	0.0891	Mono–Stereo	0.0001	1	3.04	0.0891	Mono–Stereo	0.0001	1	3.04	0.0891
Laptop–Wall	0.14136	1	1.73	0.1954	Laptop–Wall	0.00007	1	2.18	0.1473	Laptop–Wall	0.00007	1	2.18	0.1473	Laptop–Wall	0.00007	1	2.18	0.1473
Interaction	0.00123	1	0.02	0.9029	Interaction	0.00001	1	0.35	0.5553	Interaction	0.00001	1	0.35	0.5553	Interaction	0.00001	1	0.35	0.5553
Error	3.26154	40			Error	0.00154	40			Error	0.00154	40			Error	0.00154	40		
<b>Depth Impression</b>										<b>Suitability to Application</b>									
Mono–Stereo	75.142	1	51.86	0	Mono–Stereo	1.3359	1	0.78	0.3824	Mono–Stereo	1.3359	1	0.78	0.3824	Mono–Stereo	1.3359	1	0.78	0.3824
Laptop–Wall	2.506	1	1.73	0.196	Laptop–Wall	0.1237	1	0.07	0.7895	Laptop–Wall	0.1237	1	0.07	0.7895	Laptop–Wall	0.1237	1	0.07	0.7895
Interaction	0.96	1	0.66	0.4204	Interaction	0.1237	1	0.07	0.7895	Interaction	0.1237	1	0.07	0.7895	Interaction	0.1237	1	0.07	0.7895
Error	57.955	40			Error	68.5051	40			Error	68.5051	40			Error	68.5051	40		
<b>Viewing Comfort</b>										<b>Level of Realism</b>									
Mono–Stereo	2.1976	1	1.63	0.2091	Mono–Stereo	19.1136	1	23.79	0	Mono–Stereo	19.1136	1	23.79	0	Mono–Stereo	19.1136	1	23.79	0
Laptop–Wall	3.1824	1	2.36	0.1323	Laptop–Wall	1.4545	1	1.81	0.186	Laptop–Wall	1.4545	1	1.81	0.186	Laptop–Wall	1.4545	1	1.81	0.186
Interaction	0.1067	1	0.08	0.7799	Interaction	0.2045	1	0.25	0.6166	Interaction	0.2045	1	0.25	0.6166	Interaction	0.2045	1	0.25	0.6166
Error	53.9293	40			Error	32.1364	40			Error	32.1364	40			Error	32.1364	40		
<b>Sense of Presence</b>																			
		<b>SS</b>	<b>DoF</b>	<b>F</b>	<b>P</b>														
Mono–Stereo			75.142	1	51.86														
Laptop–Wall			2.506	1	1.73														
Interaction			0.96	1	0.66														
Error			57.955	40															

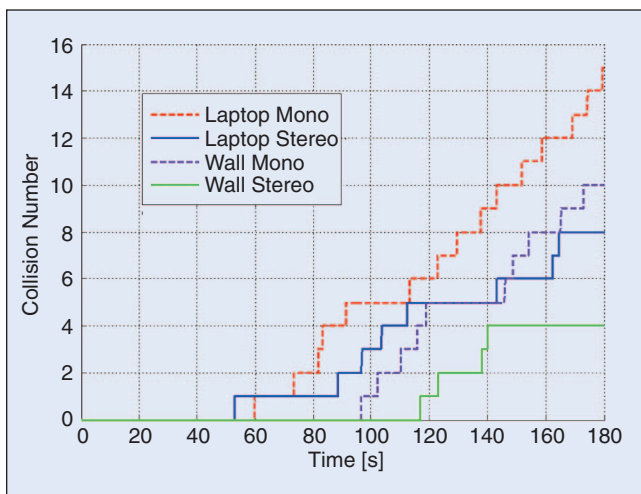
Rows show values for the independent variables (stereo–mono, laptop–wall), their interaction, and error. Columns show the sum of squares (SS), the degrees of freedom (DoF), the *F* statistic, and the *P* value.



Demiralp et al. considerations [2], telling that looking-out tasks (i.e., where the user views the world from inside-out as in our case), require users to use their peripheral vision more than in looking-in tasks (e.g., small-object manipulation). A large screen presents the environment characteristics closer to their real dimension, which enforces adequacy of this display to the application.

The polarized wall in stereo is considered the most suitable for teledriving tasks, which makes this facility very suitable for training activities. On the other side, the laptop stereo is considered inadequate for long tele-driving tasks because of the fatigue an operator is exposed to. The laptop system remains nevertheless most suitable as a low-cost and portable facility.

- 8) *Viewing comfort*: There is no significant difference between the two systems; however, the mean bar graph and typical users' comments show that a higher comfort is perceived in case of a polarized wall. This result is expected, and it confirms the benefit of front projection and polarized filters that provide limited eye strain and cross talk, and great color reproduction. The passive anaglyph technology (laptop stereo) strongly affects viewing comfort, and it calls for high brightness to mitigate viewer discomfort. The mean values show an opposite tendency between the two facilities in terms of stereo versus mono.
- 9) *Level of realism*: The mean level of realism is higher in case of the wall system, with a mean improvement of 58%. This is claimed due to the possibility given by large screens to represent objects with a scale close to real. The realism is higher under stereo viewing on both facilities.
- 10) *Sense of presence*: The mean sense of presence is higher in case of the wall system, with a mean improvement of 40%. The large screen involves user's peripheral vision more than the small screen, which strongly affects sense of presence. The presence is higher under stereo visualization on both facilities.



**Figure 7.** Collision number for a typical user in laptop and wall facilities.

## Conclusions

In this work, a comparison between monoscopic and stereoscopic visualization in mobile robot teleguide was performed. Two different visualization systems were considered. The main aim was to experimentally demonstrate the performance enhancement in mobile robot teleoperation when using video-based stereoscopic visualization. This took place despite the increased complexity of the system and, in some cases, it decreased the level of viewing comfort. Furthermore, the experimentation was performed in an indoor man-made environment where the advantage of binocular stereo is challenged by the presence of strong monocular depth cues.

A usability evaluation was proposed that involved several users and two different working sites located approximately 3,000 km apart.

The results were evaluated according to the proposed research question. This involved two factors: monoscopic versus stereoscopic visualization and laptop system versus wall system. The two factors were evaluated against different quantitative variables (collision rate, collision number, obstacle distance, completion time, path length, mean speed) and qualitative variables (depth impression, suitability to application, viewing comfort, level of realism, sense of presence).

The result of the evaluation on the stereo-mono factor indicated that 3-D visual feedback leads to fewer collisions than 2-D feedback and is therefore recommended for future applications. The number of collisions per time unit was significantly smaller when driving in stereo on both the proposed visualization systems. A statistically significant improvement of performance of 3-D visual feedback was also detected for the variables such as depth impression, level of realism, and sense of presence. The other variable did not lead to significant results on this factor.

The results of the evaluation on the laptop-wall factor indicated significantly better performance on the laptop in terms of the mean of minimum distance to obstacles. No statistically significant results were obtained for the other variables. The interaction between the two factors was not statistically significant.

Further studies are under development, with the aim of decreasing the requirements for communication bandwidth by using laser sensor signals and graphical reconstruction of the remote environment. Further visualization systems are also being considered.

We expect that the 3-D visualization will largely be adopted in future on different application contexts, e.g., interactive television and computer games, and their use in telerobotics will certainly become popular.

## Acknowledgments

Special thanks go to Prof. G. Gallo and Dr. M. Fanciullo (Dipartimento di Matematica e Informatica, University of Catania, Italy) for their valuable support. Thanks also to the anonymous reviewers for the interesting suggestions and corrections given for the final article.

## Keywords

Telerobotics, stereo vision, 3-D displays, virtual reality, mobile robotics.

## References

- [1] M. Bocker, D. Runde, and L. Muhlback, "On the reproduction of motion parallax in videocommunications," in *Proc. 39th Human Factors Society*, 1995, pp. 198–202.
- [2] C. Demiralp, C. Jackson, D. Karelitz, S. Zhang, and D. Laidlaw, "CAVE and fishtank virtual-reality displays: A qualitative and quantitative comparison," *IEEE Trans. Visual. Comput. Graph.*, vol. 12, no. 3, pp. 323–330, 2006.
- [3] D. Drascic, "Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing," in *Proc. 35th Human Factors Society*, 1991, pp. 1367–1371.
- [4] M. Ferre, R. Aracil, and M. Navas, "Stereoscopic video images for tele-robotic applications," *J. Robot. Syst.*, vol. 22, no. 3, pp. 131–146, 2005.
- [5] G. Hubona, G. Shirah, and D. Fout, "The effects of motion and stereopsis on three-dimensional visualization," *Int. J. Hum. Comput. Stud.*, vol. 47, no. 5, pp. 609–627, 1997.
- [6] G. Jones, D. Lee, N. Holliman, and D. Ezra, "Controlling perceived depth in stereoscopic images," in *Proc. SPIE*, 2001, vol. 4297, pp. 422–436.
- [7] L. J. Corde, C. R. Caringnan, B. R. Sullivan, D. L. Akin, T. Hunt, and R. Cohen, "Effects of time delay on telerobotic control of neural buoyancy," in *Proc. IEEE. Int. Conf. Robotics and Automation*, Washington, USA, 2002, pp. 2874–2879.
- [8] S. Livatino and C. Koeffel, "Handbook for evaluation studies in virtual reality," presented at the IEEE Int. Conf. Virtual Environments, Human-Computer Interface and Measurement Systems, Ostuni, Italy, 2007.
- [9] U. Naepflin and M. Menozzi, "Can movement parallax compensate lacking stereopsis in spatial explorative tasks?" *Displays*, vol. 22, no. 5, pp. 157–164, 2006.
- [10] I. Sexton and P. Surman, "Stereoscopic and autostereoscopic display systems," *IEEE Signal Process. Mag.*, vol. 16, no. 3, pp. 85–89, 1999.

**Salvatore Livatino** received his M.Sc. degree in computer science at the University of Pisa, Italy, in 1993, with a specialization undertaken at the Scuola Superiore S. Anna, Pisa, where he pursued his research activity in the coming years at the Advanced Robotics Technology and Systems Laboratory (ARTS Lab) and Perceptual Robotics Laboratory (PERCRO) in 1992–1995. He was a visiting researcher at the University of Leeds, United Kingdom, in 1995, The French National Institute of Research in Computer Science and Control (INRIA) Grenoble, France, in 1996, and University of Edinburgh, United Kingdom, in 2001. He worked for 11 years at the Aalborg University, Denmark, where he obtained his Ph.D. degree and became an assistant and associate professor at the Department of Media Technology (medialogy studies in Copenhagen and VR Media Lab in Aalborg). He is currently with the School of Electronic, Communication and Electrical Engineering at the University of Hertfordshire, United Kingdom. His main research areas are VR, computer graphics, computer vision, and mobile robotics, with works combining those fields, e.g., photorealistic image synthesis and vision-based robot navigation. He has published several journal and conference papers. His most recent works involve the use of 3-D visualization and VR facilities in telerobotics, cultural heritage, and computer games. His teaching experience has mostly focused on problem-based learning and multidisciplinary education.

**Giovanni Muscato** received his electrical engineering degree from the University of Catania in 1988. Following graduation, he worked with Centro di Studi sui Sistemi in Turin. In 1990, he joined the Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi of the University of Catania, where he is currently a full-time professor of robotics and automatic control. His research interests include model reduction, service robotics, and the use of soft-computing techniques in the modeling and control of dynamical systems. He was the coordinator of the EC project, Robovolc: A Robot for Volcano Explorations, and is a local coordinator of several national and European projects in robotics. He is a Senior Member of the IEEE Control Systems Society and Robotics and Automation Society and is cochair of the service robotics technical committee. He is on the board of trustees of the Climbing and Walking Robots (CLAWAR) association. He has published more than 200 papers in scientific journals and conference proceedings and three books on control and robotics.

**Salvatore Sessa** is with the Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi, University of Catania, Italy, where he got his Ph.D. degree in 2008 and his M.Eng. degree in 2004. He was visiting the Aalborg University in 2007 and is currently in Waseda University, Tokio, with a Japan Society for the Promotion of Science (JSPS) postdoctoral fellowship. His research interests are focused on mobile robotic navigation and localization.

**Christina Köffel** is with the Center of Usability Research and Engineering (CURE), Vienna, Austria. She got her M.Sc. degree in medialogy at the Aalborg University in 2008 and her M.Sc. degree in digital design at the Upper Austria University at Hagenberg in 2007.

**Carmelo Arena** is with the Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi, University of Catania, Italy, where he got his M.Eng. degree in 2008. His master's thesis was in collaboration with the Aalborg University.

**Alba Pennisi** is with the Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi, University of Catania, Italy, where she got her M.Eng. degree in 2008. Her master's thesis was in collaboration with the Aalborg University.

**Daniele Di Mauro** is an M.Sc. student at Dipartimento di Matematica e Informatica, University of Catania, Italy. He was an exchange student at the Aalborg University.

**Erinc Malkondu** got his M.Sc. degree in medialogy at the Aalborg University in 2008.

**Address for Correspondence:** Salvatore Livatino, School of Electronic, Communication, and Electrical Engineering, University of Hertfordshire, Hatfield, AL 10 9AB, United Kingdom. E-mail: s.livatino@herts.ac.uk.