

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**

Open access books available

**122,000**

International authors and editors

**135M**

Downloads

Our authors are among the

**154**

Countries delivered to

**TOP 1%**

most cited scientists

**12.2%**

Contributors from top 500 universities



**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



# Testing Stereoscopic Vision in Robot Teleguide

Salvatore Livatino<sup>1</sup>, Giovanni Muscato<sup>2</sup> and Christina Koeffel<sup>3</sup>

<sup>1</sup>*Engineering and Technology, University of Hertfordshire*

<sup>2</sup>*Ingegneria Elettrica Elettronica e Sistemi, University of Catania*

<sup>3</sup>*Center for Usability Research and Engineering*

<sup>1</sup>*United Kingdom, <sup>2</sup> Italy, <sup>3</sup> Austria*

## 1. Abstract

The aim of this chapter is to provide a focused guideline on how to test virtual reality (VR) systems used in robot teleoperation. The guideline is demonstrated based on a real experiment. The goal of this experiment is to test the characteristics and the advantages of a telerobotic system based on video transmission and stereoscopic viewing. The experimentation design follows a systematic approach that relies on identification of a number of key parameters and a usability evaluation designed according them. Two different 3D visualization facilities are considered for evaluating performance on systems with different characteristics, cost and application context. The results of the experiments are expected to provide insight into stereoscopic robot teleguide, and to understand on what system, and to what extent, is stereo vision beneficial.

## 2. Introduction

Robot telerobotion is typically related to survey and intervention in inaccessible, unknown, or hazardous environments. Despite of the latest generation of robotic systems possessing a high level of autonomy, remotely directed robot intervention is still typically human-driven. Humans are irreplaceable in tasks that require high-accuracy or deep environment cognition. The latter is typically needed to resolve situations with high unpredictability, where fast decision making and comprehension is required.

Robot teleoperation systems typically rely on 2D displays. These systems suffer from many limitations. Among them are misjudgement of self-motion and spatial localization, limited comprehension of remote ambient layout, object size and shape, etc. The above limitations may lead to unwanted collisions during navigation and long training periods for an operator.

An advantageous alternative to traditional 2D (monoscopic) visualization systems is represented by the use of a stereoscopic viewing. In the literature we can find works demonstrating that stereoscopic visualization may provide a user with 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, obstacles perception, and manoeuvre accuracy.

The aim of the proposed testing activity is to test stereoscopic vision on mobile robot teleguide. The experimentation is run in an indoor workspace. This represents a challenging setup for our testing because indoor ambient layouts, typically man-made, are simple and emphasize monocular depth cues such as perspective, texture gradient, etc., so they diminish the advantage of binocular stereo.

This chapter provides a brief introduction to stereo vision in video-based robot teleoperation. The next section (Section 3) presents the design of the proposed experiment. Then, the guideline together with its application to the experiment is presented (Section 4), followed by an analysis of the results (Section 5). Some final remarks conclude the chapter (Section 6).

## 2.1 Video Images in Robot Teleoperation

Performance in robot teleoperation can be improved by enhancing the user's sense of presence in remote environments (telepresence). Vision being the dominant human sensor modality, large attention has been paid by researchers and developers to the visualization aspect.

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

The rich information provided by a camera may require a large bandwidth to be transmitted at interactive rates. This often represents a challenge in transmission to distant locations or when the employed medium has limited communication capabilities.

Several video compression techniques have been developed which may 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 greatly be reduced, e.g. based on redundant information in stereo images, while specific networks for streaming video have been proposed, (Ferre' et al., 2005).

The bandwidth constraint may lead to transmission delays and this may affect interaction performance, e.g. response speed and accuracy. (Corde et al., 2002) claims that a delay of more than 1 sec. leads to eminent decrease of performance.

## 2.2 Stereoscopic Viewing

Stereoscopic visualization can play a major role towards increasing the user's involvement and immersion, because of the increased level of depth awareness. This is expected to give more accurate action performance and environment comprehension.

Stereoscopy improves: comprehension and appreciation of presented visual input, perception of structure in visually complex scenes, spatial localization, motion judgement, concentration on different depth planes, and perception of surface materials.

Most of the benefits of stereo viewing may affect robot teleguide because stereopsis enhances: perception of relative locations of objects in the observed remote worlds [3], impression of telepresence and of 3D layout (Bocker et al., 1995) , ability to skilfully manipulate a remote environment (Ferre' et al., 2005), response time and accuracy when operating in dangerous environments, etc.

The main drawback of stereoscopic viewing, which has yet prevented its large application, is that users may have to make some sacrifices, (Sexton et al., 1999). A stereo image may be hard to “get right” at first attempt, hardware may cause crosstalk, misalignment, image distortion (due to lens, displays, projectors), and all this may cause eye strain, double images perception, depth distortion, look around distortion (typical for head-tracked displays).

### 3. Testing Stereoscopic Teleguide

The testing of the proposed stereoscopic teleguide is organized by having a robot operating on a factory like scenario created in the Robotics lab at the DIEES department of the University of Catania, Italy, and a user driving it sitting at the Medialogy lab at the Aalborg University in Copenhagen, Denmark. The two sites are approximately 3,000 km apart. Figure 1 shows a representation of the local-remote system interaction.

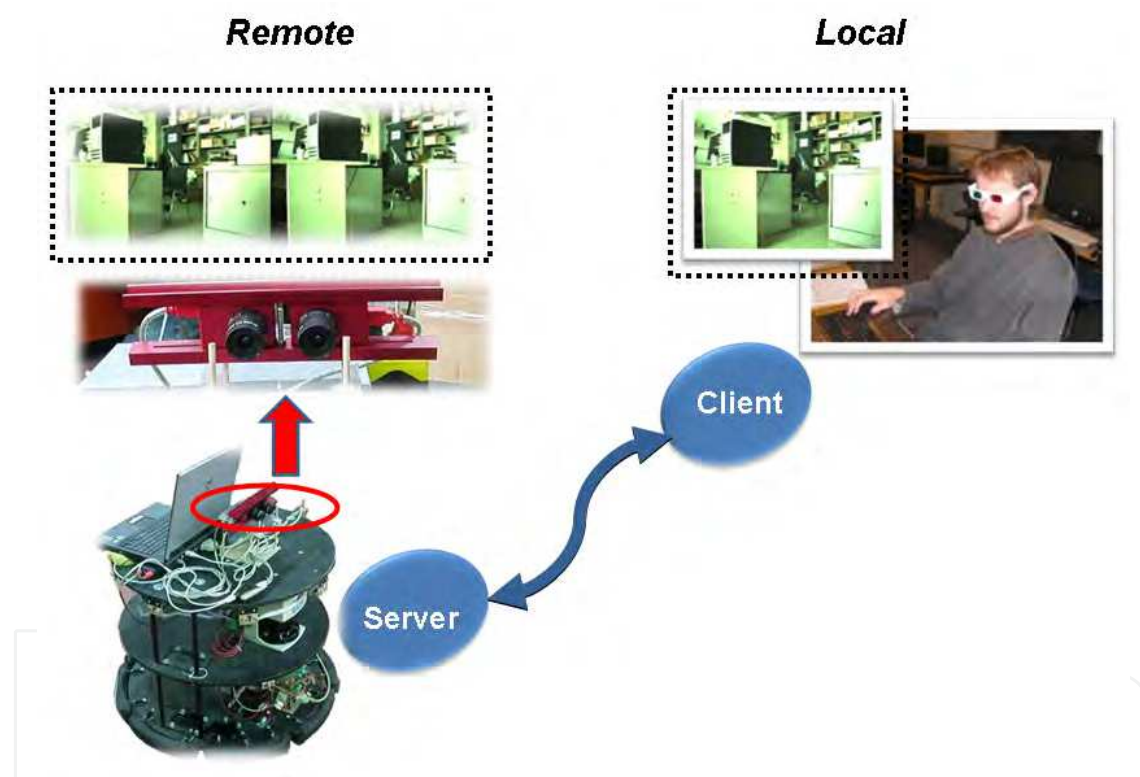


Fig. 1. A representation of the local-remote system interaction. The mobile robot on the figure left-hand side is located at the Robotics lab at the DIEES department of the University of Catania, Italy. The user (tele-) driving the mobile robot, shown in the right-hand side, is sitting at the Medialogy lab of the Aalborg University, Copenhagen, Denmark. The two sites are approximately 3,000 km apart.

The stereoscopic testing is delimited to two different visual displays and stereo approaches.

The visual displays are very different in size and type of technology. They are presented in the following points:

- **Wall.** A Wall display is typically composed by a projector and a screen with a size up to several meters.
- **Laptop.** A Laptop display uses LCD technology and it has a relatively small display size, typically up to 19 inches with high resolution.

The proposed stereoscopic approaches are very different in cost and performance. They are presented in the following points:

- **Polarized Filters.** The Polarized Filters nicely reproduce colours, have nearly no crosstalk, and they are very comfortable to a viewer. However, they require a complex and expensive setup and the system is less portable.
- **Coloured Anaglyph.** The Colored Anaglyph is cheap, easy to produce and very portable. However, it has poor colour reproduction and it often generates crosstalk which affects precision and viewing comfort.

The wall display uses Polarized Filters, therefore we call it Polarized Wall. The display's dimension is 2x2 meters. It is capable of providing high user involvement, 3D impression and comfort, suitable for training purposes or for tasks requiring accurate manoeuvring and long operational sessions.

For the Laptop system we use Coloured Anaglyph, therefore we call this setup Anaglyph Laptop. The display diagonal is 15 inches. This results on having stereo on a portable system, which is suitable for tasks requiring a user to be close to mobile robot operational environments. Furthermore the hardware is available at low-cost.

The figure 2 shows the visualization systems used in our investigation.



Fig. 2. The visualization systems used in our tests. The Wall (left) and the Laptop (right).



We have restricted our experimental conditions to indoor environments and a "factory-like" scenario. The stereo camera system follows the objectives of having a realistic observation. It is intended to appear close to the observation obtained when looking at the environment with our own eyes. Therefore, the stereo cameras pair should have a baseline close to the one typical for human eye distance. The same baseline should also satisfy the condition of showing effective left-right image separation for an expected average distance to visualized objects of about 2 meters.

A compromise setting is estimated for the camera parameters. The camera system sits on the top of the robot with a height of 95 cm, it looks 25 deg. downwards (tilt angle) and it has a baseline of 7 cm. Our stereo camera system has been designed based on the directions given in the literature.

It is important for the experiment that all trials are run under the same conditions.

#### 4. Evaluation Guidelines

This section describes the proposed set of directives for the usability evaluation of stereoscopic vision in robot teleoperation. Usability describes how effectively and efficiently a user is able to fulfil tasks using a system. Especially in the field of VR the user's experience should also be taken into account. The following paragraphs will be addressing main issues when conducting usability evaluations of VR systems. Nevertheless also user experience related problems will be briefly described.

The following paragraphs are divided into sub-sections addressing specific aspects. The content of the description is based on selected literature in the field of VR and stereo viewing that the authors have investigated.

The test designer is left with some freedom of choice depending on:

- The guideline specific aspects
- Application context
- Available time
- Pre-determined objective

To support the designer's decision in making a choice, the guideline often directly refers to the results of our investigation in specific aspects in terms of percentage of literature works.

##### 4.1 Test Plan

When forming the idea of conducting an evaluation, a test plan should be created. This document contains in principle every kind of knowledge necessary for the usability study. It serves as the basic document for communication to other people that might be involved in the user study (e.g. second test monitor).

Using the test plan, every involved person knows the main principles and ideas behind the evaluation. Therefore open questions and misunderstandings can be clarified. Furthermore, the test plan describes the resources needed and gives an overview of the milestones already accomplished. A properly formulated test plan for user studies in the field of VR should contain the following items:

- **Purpose:** The purpose describes the research question and the main problems treated as well as the current state of the art of the project.

- **Problem statement/test objectives:** The problem statement treats the main issues and questions connected to the evaluation, e.g. the questions derived from the hypothesis.
- **User profile:** The user profile describes the target group and the participants to be acquired for the study.
- **Test design:** The test design includes decisions about the entire session of the usability study, such as the evaluation method, e.g. if doing a between or within subjects evaluation. Furthermore the test design specifically describes each single step during the user study, starting from the arrival of the participants until the time they leave.
- **Task list:** The task list describes every task and sub-task that the participants will be asked to accomplish and on which VR device tasks are accomplished.
- **Test environment/equipment:** This section elaborates the test environment and equipment used in the test, e.g. VR devices and rooms needed.
- **Test monitor role:** The description of the test monitor role includes information about the test monitor and possible spectators.
- **Evaluation measures:** The evaluation measures should be described on a list enumerating all data collected during the user study (data logging, questionnaires, etc.).
- **Report contents and presentation:** This section gives a short preview on the data contained in the final test report and the presentation of the results obtained during the user study.

#### 4.2 Research Question

Before starting to build a setup for an evaluation, the research question for the usability study needs to be formulated. A general research question defining the purpose of the entire project should already exist; nevertheless a specific research question should be formulated for the special purpose of the evaluation. This defines the main subject of the study.

It is very important to create a strong and valid research question that summarizes the goal of the evaluation in only one sentence/paragraph.

It is essential that the purpose of the entire project as well as the evaluation is clear to everybody on the project/evaluation team. Additionally, the research question should help to formulate the hypothesis we want the project to be tested against.

To formulate the research question we start from the problem statement, which in our case study has two hypotheses. They are:

- Users performing tasks employing stereo visualization perform better than users performing the same tasks employing mono visualization.
- The same task is not performed with the same efficiency and accuracy on different VR facilities.

The research question can be synthesized as in the following points.

- **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?

### 4.3 Ethics

Since user tests are conducted with humans, it is essential to assure that there will be no harm to the participants and that their personal rights are maintained, (Burdea & Coiffet, 2003). Users' mental and physical health must not be at risk and they need to be informed about potential hazards. Furthermore, users have to be able to stop whenever they feel uncomfortable and desire the test to end.

Certain universities or research centers dispose of an ethical department that administrates all studies and evaluations conducted involving humans. In this case, the researchers have to apply to this committee and do have to obey certain rules. If the institution where the evaluation is supposed to take place does not dispose of such a department, ethical considerations have to be taken into account as well. Especially when there is no board reviewing the studies, one has to make sure that all ethical concerns are respected. Furthermore also legal considerations of the country where the study is planned should be reviewed.

Virtual reality applications offer many possible risks to the participants of a user study, e.g. in cases when new devices are invented and tested or when existing devices have not entirely been tested for health risks. Additional hazards can appear through the use of e.g. head mounted displays, laser diodes, etc. Different mechanical devices in use, such as haptic tools can endanger the participants' health when applied incorrectly, this also includes stereoscopic viewers. Side-effects such as the occurrence of cybersickness need attention when using VR systems depending, e.g. on the type of stereoscopic approach and display size. They might even require a participant to stop the test.

### 4.4 Evaluation Method

At the very beginning of each user study it is important to choose and define the appropriate evaluation methods applicable to the setup to be tested. According to J. Nielsen these are: performance measures, thinking aloud, questionnaires, interviews, logging actual use and user feedback. These evaluation methods can also be applied in a combined version. Depending on the time when the evaluation takes place and the kind of data collected, one can distinguish between formative and summative user studies. Formative usability evaluations usually take place several times during the development cycle of a product to collect data of prototypes. Typically summative evaluations are applied at the end of a project, for example, to compare different products. Formative user studies are rare in VR. When comparing two or more different VR devices/applications (summative evaluation), one can decide whether to use a within or between subjects design. Between subjects studies are more common in VR. A statistical analysis conducted in (Koeffel, 2008) has shown that a total of 61% of user studies in VR were designed as between subjects studies.

In our experiments different types of VR devices were compared against each other, therefore the study is designed as summative evaluation. Because of the limited number of participants and the difficulty of finding equally skilled participants, a within subjects design was preferred over a between subjects design. Therefore each participant fulfilled the same amount of tasks on all available VR devices.

The study includes quantitative and qualitative evaluations. The following evaluation measures were collected through robot sensors and calculated for the following quantitative evaluation measures:



- **Collision Rate.** The Collision Number divided by the Completion Time. It provides information about obstacle detection and avoidance which 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 user's environment comprehension. This parameter may also show user's confidence, (sometime a false confidence). 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 user's confidence.

The following evaluation measures were collected through questionnaires and calculated for the following qualitative evaluation measures:

- **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 objects dimension and general appearance.
- **Sense of Presence.** The perceived sense of presence and isolation from surrounding space.

During the evaluation of the data, the questions were grouped into five categories corresponding to the five qualitative judgement categories, in order to be able to compare the results in each area. The 7 scale semantic differentials were used for the answer of questionnaires.

#### 4.5 Setup

In our recommendations the setup is distinguished into the testing environment and the technological setup.

- **Testing environment**  
Evaluations conducted by students and academic researchers usually take place in the facilities of universities or research centers. In some cases these institutions dispose of their own usability labs for conducting evaluations, but in most of the cases the evaluations occur in computer labs or classrooms. Since classrooms are not always comfortable (and hard to find relaxing), while it is required that the participants feel at ease, it is very important to create a comfortable environment. It has to be avoided the presence of people that are not involved in the project, the presence of those running

around hectically and preparing the evaluation, and other distractions such as loud noises.

It is generally important not to give the participants unnecessary information about the project or to bias the results by telling the users some weaknesses or previous results. If the user study requires a test monitor logging data while the participants perform the testing, it is fundamental that he/she respects the participants' privacy by not sitting too close to them. Furthermore, any kind of stress and emotional pressure has to be kept away from the participants in order not to influence the results.

- **Technological setup**

Student and research evaluations often base on an already finished project (summative evaluations), referring to the hardware and/or the software. Therefore the technological setup might already be given. Considering the different VR setups, it is very important to assure that all needed devices are at the test monitors' disposal on the day(s) of the usability study. Furthermore it is very important to test if the application, the software, and the data logging, are well functioning. Since VR devices and applications are still considered to be "new technology" they are sometimes unstable and tend not to work all the time. Hence, it is crucial to organize and reserve technical facilities and rooms, and to inspect the functionalities of the project to be tested.

#### 4.6 Participants

Several fundamental elements of evaluations are related to participants. Before recruiting volunteers, it is very important to investigate the target population of the user study. Therefore users with the desired attributes such as age, gender, education, experience with VR, computer experience, gaming experience, visual abilities, etc. can be selected. Generally, it is advisable to test user groups with a great internal variance. Users should be recruited from different age groups, gender and experience. A careful selection of participants should also be according to expected system users. In any case, main users' characteristics such as average age, gender, experience, etc., should clearly be stated.

Based on authors' experience the men participating in user studies in the field of VR are twice as many as women. This could be acceptable in case of men being the main users of a VR product.

Concerning the number of participants, it mainly depends on the kind of user study conducted (i.e. formative or summative evaluation, between or within subjects design, etc.). Generally, usability experts (Faulkner, 2000; Nielsen, 1993; Nielsen & Mack, 1994; Rubin, 1994) hold that 2 to 4 participants suffice for conducting a representative pilot study, and 5 to 20 participants suffice for conducting a formal user study. Nevertheless, more recent approaches on evaluations in the field of VR have suggested testing a higher number of participants in order to obtain meaningful results.

A number of approximately 23 participants is suggested for within subject designs and 32 for between subject designs. In case of pilot studies, a minimum number of 6 participants is proposed. These figures are based on our literature analysis.

Participants are typically volunteers and/or they do not receive any financial compensation. Nevertheless, it is highly recommended to hand the participants a small token of appreciation after finishing the user study.

Twelve subjects take part in the described evaluation study. They 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

#### 4.7 Forms

Forms are different documents that are handed to the participants during the course of a user study. Concerning the forms given to the participants, this guideline conforms to the traditional approaches introduced in (Nielsen, 1993; Rubin, 1994) and the results of the statistical analysis of relevant literature in (Koeffel, 2008). Therefore we recommend the use of the following forms:

- **Information sheet:** The information sheet (also called test script) provides an overview of the entire testing process. This form should be handed to the participants at the very beginning before the actual testing, and it should contain information about: the title of the project, names and contact information, introduction to the project, duration of the study, tasks to be completed, and the possibility to withdraw from the study at any time. In 5 out of the 18 studies investigated, the participants have reported to have received written or displayed information before the testing process.
- **Consent form:** The consent form states that the researchers are allowed to use and publish the data collected during the user study. This may also include pictures or videos taken during experiments. It is a reassurance for the participants that their data will not be used for any other purpose than the one explained in the consent form and/or in the information sheet. For the researcher this form is a legal reassurance that he/she is allowed to use and publish the obtained data.
- **Questionnaires:** Generally questionnaires should contain the information required by the research question which is not possible to be collected automatically through data logging and performance measures. Therefore, mostly subjective qualitative data is collected using questionnaires. Special issues should be treated in questionnaires in order to emphasize the conclusion and the results of the data collection. Questionnaires can provide answers about personal feelings or preferences. We distinguish among: screening, pre-test, device, post-test, background, presence, simulator sickness, and the EVE-experience questionnaire.
- **Task scenarios:** It might be necessary to provide participants with a task scenario (describing each step in detail) for each task he/she should complete. This allows every participant to gain the same amount of information. Furthermore it clarifies the knowledge necessary to complete a given task.
- **Data collection forms:** Experience has shown that it is not always sufficient to auto-log data using software. Sometimes it is necessary that the test monitor writes down notes or information during a task session. This can be additional information such as time or estimates expressed by participants.
- **Thank you form:** In addition to the possible personal gratification that participants may receive by taking part in a user study, a thank you letter should also be handed to them. This is important in order to formally thank the participants and tell them where to find further information about the progress of the project, e.g. published papers.

The forms should be adapted to the needs of the user study. Generally we suggest the employment of semantic differentials as answering options.

In our experiments we provide a page as information sheet and consent form. We use questionnaires in our qualitative evaluation. We have a number of questions for gather users' background information, including their experience and gaming abilities (e.g. hours per week). We have then a questionnaire for each of the five proposed qualitative evaluation measures, (depth impression, suitability to application, viewing comfort, level of realism, sense of presence), and users' overall impression after the user study. The questions are designed to get answers for the research questions.

As example, the questionnaire for the level-of-realism parameter included the following questions: "How realistic is the environment layout?", "How realistic are the visualized objects size and shape?", "How natural was the driving?", "What mark would you give as general level of realism?". The questionnaire also included user's suggestion for improvement and general remarks. A conclusive comparative questionnaire was provided at the end of each experiment.

We conform to the traditional approaches in terms of forms and questionnaires, with few additions (Livatino et al., 2007). We use a seven scale semantic differential for answering the questions. In particular the possible values range between -3 and +3, with -3 is for the worst and +3 for the best result.

We provide our users with a task scenario page that contains information about the workspace they tele-explore. We do not use data collection forms in our quantitative evaluation. A thank you letter is provided at the end of user trials together with sweets and drinks.

#### **4.8 Procedure**

The test procedure is part of the test design and it very much depends on the specific subject and application context. The procedure should be carefully designed in order to provide meaningful data.

In our experiment four steps are performed. First, an introductory phase that includes a practice drive. Then, each participant is asked to teledrive the robot (remotely located) toward a final location while avoiding collisions. The drive is performed on both the proposed facilities (wall and laptop), using both stereo and mono viewing conditions. This results in four navigation trials per participant.

The participants are eventually asked to complete predesigned questionnaires. Practice sessions are administrated before testing. A debriefing phase ends the test session.

The figure 3 right-hand side shows images from our test sessions: the top-right image shows our forms and stereoscopic goggles ready before testing; the middle image shows test users filling in the questionnaires.

#### 4.9 Schedule

It is essential to estimate the overall completion time per participant and to prepare a schedule showing the sequence of participants and their assigned tasks. In particular, the schedule should include: timing of the single tasks, overall completion time, the sequence of the tasks per participant, possible breaks, time needed for introduction and debriefing, room locations, etc.

The studies analyzed indicate an overall completion time per participant that ranges from 23 to 240 minutes with an average completion time of 45 minutes. This time includes the time from when a participant arrived at the testing facility until the time he/she left.

In the field of VR it is very important to keep the single task sessions as well as the overall completion time as short as possible. A maximum of 30 minutes per task is recommended by Bowman et al. (Bowman et al., 2002). Too long sessions might cause exhaustion of the participants and side effects such as cyber-sickness, which could negatively affect the results. It is important to counterbalance the sequence of the single tasks in order to avoid learning effects and biasing of the results.

In our experiments the test trials runs during several days. The average execution time is per participant is 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. We turn special attention on the counterbalancing of the tasks, therefore the participant tasks and facilities are given according to a predetermined schedule. The sequence during the entire user study is designed to avoid fatigue and learning effects.

#### 4.10 Test Monitor and Spectators

The role of each person present during the user study has to be predefined. Especially the test monitor should be well instructed and capable to serve his/her purpose.

The test monitor is present during all parts of the usability study and interacts with the participants. If possible somebody who has ground knowledge in usability (especially evaluations) should be employed as test monitor. In case that there is no expert in usability available, the person in the role of test monitor should acquire basic knowledge in this area. The test monitor should be able to comfortably interact with the participants, which requires an open and friendly personality (i.e. a "people-person"). It is also important that the test monitor does not get too close to the participants physically as well as mentally, to give them some privacy.

In case other people than the test monitor and the participant, are present during a test session, e.g. technical staff, VR project development team, spectators, etc., they should be introduced to participants at the beginning and the roles of the spectators need to be defined clearly. Generally, the number of spectators during a testing session should be kept small since they tend to make the users nervous. If not part of the research question, spectators should avoid talking during task sessions. This is especially crucial for VR applications, since distractions such as loud noises might disturb the sense of presence.

Since VR systems are still considered new technology and unstable, it might happen that the participant gets frustrated because something is not working properly or it is very difficult to accomplish. In such a case, the test monitor should not judge the participant or the system by expressing that e.g. "this error always occurs" or otherwise by negatively influencing the user. The test monitor should encourage the participant to go on as long as possible.



The figure 3 images bottom-right and top-left show test monitors interacting with users during the introduction to the test and forms (bottom-right) and assisting the user during robot teleguide (top-left).

#### **4.11 Pilot Study**

It is generally recommended to perform a pilot study before testing a project in a formal user study. The pilot study should be conducted in the same way as the formal study and each participant should be treated as if he/she were in the formal study (including the forms to be used).

The pilot study is useful for removing errors from the project/setup, debug the test design, debug the experimental design, detect biased questions in the questionnaires, refine the questionnaires and detect the overall time necessary per participant. Furthermore, rooms and technical facilities should be tested of their functionality.

A minimum number of 6 participants is suggested. In general, the more participants are tested, the more indicative the results are.

The pilot study is essential in case of problems that may not be predicted and only occur during the study.

#### **4.12 Formal Study**

In an ideal case, a pilot study has been conducted before the formal study and the results of the pilot study have been taken into account when planning and conducting the formal study. If required, additional tests could be conducted at the very beginning of the study in order to categorize the participants. Furthermore, a practice session should be administrated for all testing activities which need a test-user to become acquainted with system commands and behavior. In the literature that we have reviewed, an average of 4.1 tasks is accomplished per participant in practice sessions.

In order to avoid negative side effects (such as motion sickness) and fatigue, long enough breaks should be held between the single task sessions.

The figure 3 left-hand side (middle and bottom image) shows some moments of our formal study with test-users teleguiding the robot on different facilities and working on the questionnaires. An assistant is also monitoring the robot platform at the remote site.



Fig. 3. Some moment of our usability evaluation.

The right-hand side, top and middle images, show our forms ready before the testing together with stereoscopic goggles, and test-users filling in the questionnaires.

The bottom-right and top-left images show test monitors interacting with users during the introduction to the test and forms, or assisting the users during robot teleguide.

The left-hand side, middle and bottom images, show some moments of our usability study with test-users teleguiding the robot on the different facilities.

The image in the center of the figure shows our robot at the remote site together with an assistant that monitors its actions.

## 5. Results and Presentation

Another important part of conducting usability studies is the processing and evaluation of the collected data. The processing of the results can be very complex and time consuming since most of the time a lot of data is collected. Therefore it is recommended to employ statistical tools. The most frequently used are: mean, median, frequency distribution, Bonferroni, standard deviation, t-test, and ANOVA (Analysis of Variance).

For the graphical display of the gained data, frequency distributions (in form of histograms) are very popular (83% of the cases in our investigation). Their main purpose is to display error rates and time.

As inferential statistics the analysis of variance (ANOVA) is used the most to detect the statistical significance of test results. The ANOVA is a common method of separating the effects of multiple investigation factors (independent variables) on evaluation measures (dependent variables). The ANOVA examines which factors have a significant influence on a dependent variable by comparing the variance within a factor to the variance between factors, (Wanger et al. 1992).

A one-way ANOVA is to be used to estimate the effect of one factor (independent variable) on one of the evaluation measure. A two-way ANOVA is to be used to estimate the effect of two factors (independent variables) on one evaluation measures. According to the literature it is hard to analyze more than two factors using an ANOVA.

In case the effect of a factor is to be estimated on more than one evaluation measure, a multivariate ANOVA (MANOVA) should be applied. A MANOVA is an extension of the ANOVA that reports for multiple dependent variables.

The results of ANOVAs should be displayed in tables, while bar graphs are mostly used to display descriptive statistics.

The researcher may decide to expose statistically significant results only, as well as display results of the descriptive statistics only when those show meaningful trends. A different approach could be to present all data regardless their meaning, to give a reader a complete overview of all the experimental findings. A middle-ground popular approach is to expose all statistical output synthesized on tables and then only comment on text most meaningful findings. This gives readers the opportunity to discover specific data relations and trends on their own, while keeping a concise and to-the-point text description. A result analysis and conclusions may be included along the exposition of results if these are of brief content. It is instead advisable to include an extra section dedicated to the result analysis if the authors wish to elaborate and thoroughly discuss the experimental output.

In our experiments the collected evaluation measures were analyzed through inferential and descriptive statistics and the results were graphically represented by diagrams.

We measure statistical significance of results by estimating a two-way ANOVA. This is applied to measure the effect of the two dependent variables: stereo-mono and laptop-wall. We set the P value to 0.05 to determine whether the result is judged statistically significant.

We additional measure mean, standard deviation, and percentage of improvement, to observe general trends and specific tendencies.

We present all the results in tables (excluding the percentage of improvement), and report and comment on text only most meaningful findings and specific trends observation. On text we also add some conclusions based on our result analysis.

The figures 5 and 6 show the descriptive statistics and the table 1 the inferential statistics. The results are presented on text commented according to the proposed research questions.

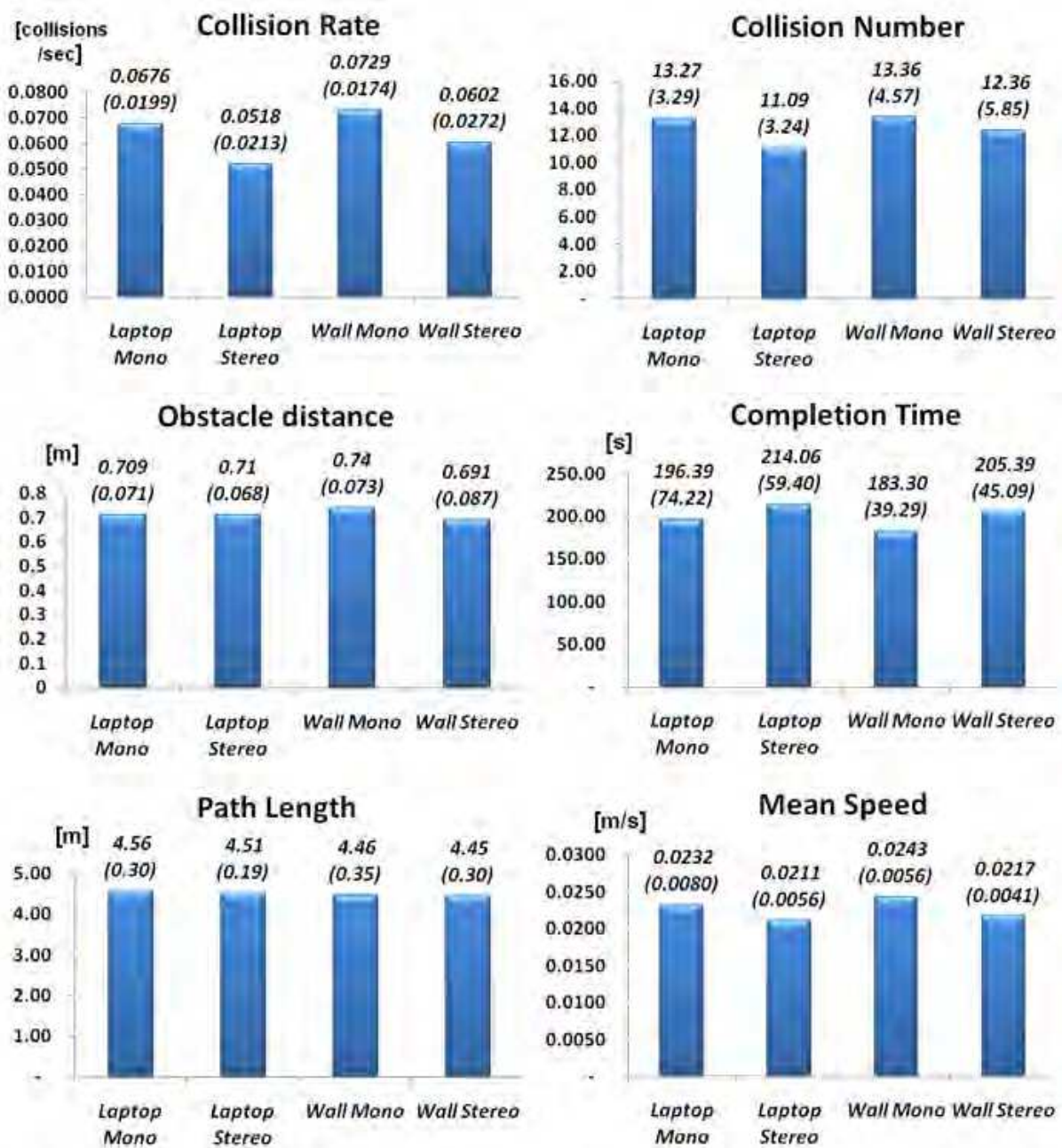


Fig. 5. Bar graphs illustrating mean values and standard deviation (in brackets) for the quantitative variables.



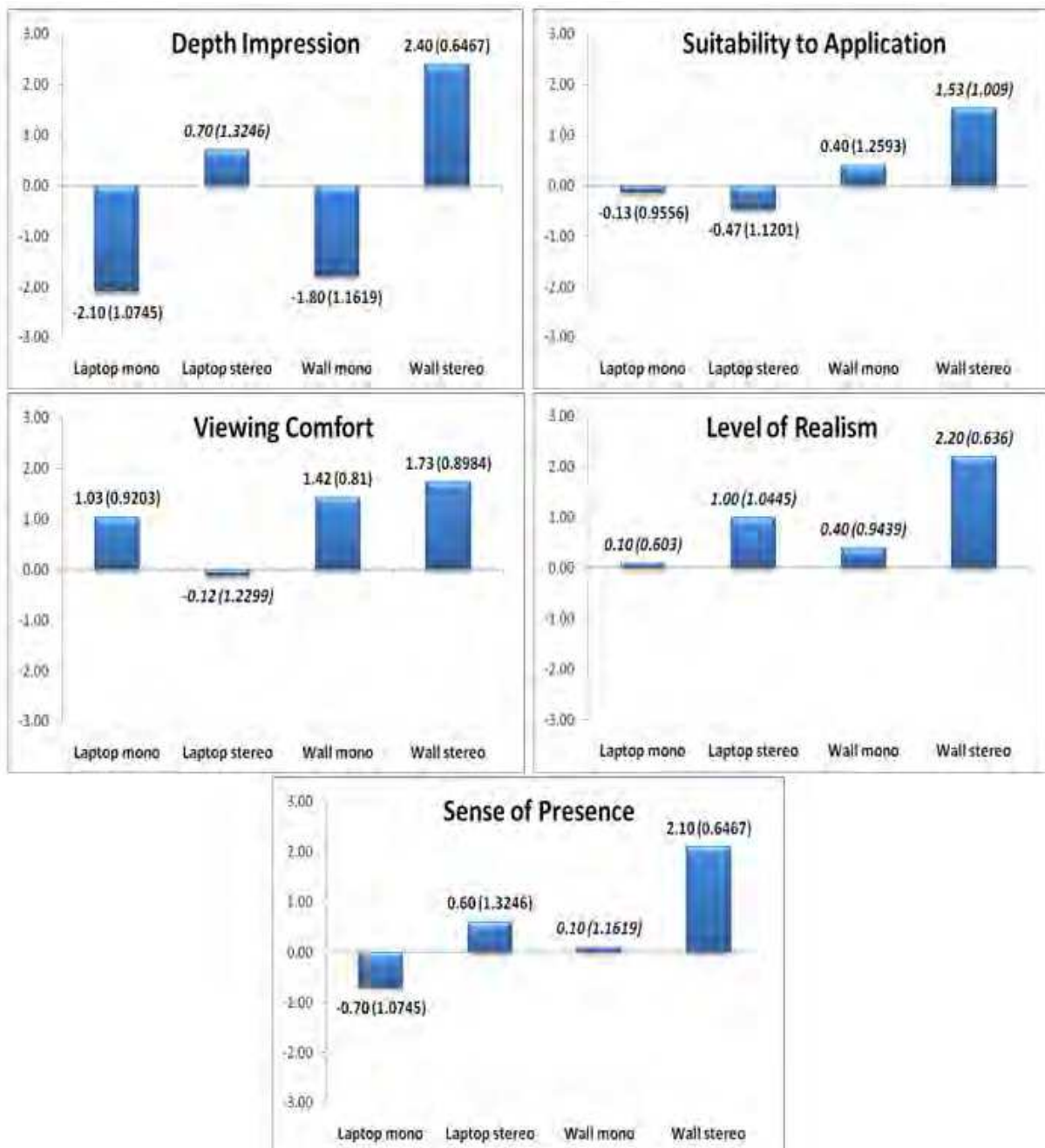


Fig. 6. Bar graphs illustrating mean values and standard deviation (in brackets) for the qualitative variables. The qualitative data were gathered through questionnaires, where the participants provided their opinions by assigning values that ranged between +3 (best performance) and -3 (worst performance).



Collision Rate					Collision Number				
	SS	df	F	p		SS	df	F	p
Mono-Stereo	0.00228	1	5.83	0.0204	Mono-Stereo	27.841	1	1.57	0.2181
Laptop-Wall	0.00338	1	8.65	0.0054	Laptop-Wall	59.114	1	3.32	0.0757
Interaction	0.00017	1	0.45	0.5076	Interaction	2.75	1	0.15	0.6962
Error	0.01561	40			Error	711.273	40		
Obstacle distance					Completion Time				
	SS	df	F	p		SS	df	F	p
Mono-Stereo	6359	1	1.28	0.2638	Mono-Stereo	4348.3	1	1.4	0.2435
Laptop-Wall	37757.9	1	7.63	0.0086	Laptop-Wall	2992.9	1	0.96	0.332
Interaction	124.9	1	0.03	0.8746	Interaction	373.2	1	0.12	0.7306
Error	198013	40			Error	124120.4	40		
Path Length					Mean Speed				
	SS	df	F	p		SS	df	F	p
Mono-Stereo	0.00445	1	0.05	0.8164	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
Interaction	0.00123	1	0.02	0.9029	Interaction	0.00001	1	0.35	0.5553
Error	3.26154	40			Error	0.00154	40		
Depth Impression					Suitability to Application				
	SS	df	F	p		SS	df	F	p
Mono-Stereo	75.142	1	51.86	0	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
Interaction	0.96	1	0.66	0.4204	Interaction	0.1237	1	0.07	0.7895
Error	57.955	40			Error	68.5051	40		
Viewing Comfort					Level of Realism				
	SS	df	F	p		SS	df	F	p
Mono-Stereo	2.1976	1	1.63	0.2091	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
Interaction	0.1067	1	0.08	0.7799	Interaction	0.2045	1	0.25	0.6166
Error	53.9293	40			Error	32.1364	40		
Sense of Presence									
	SS	df	F	p					
Mono-Stereo	75.142	1	51.86	0					
Laptop-Wall	2.506	1	1.73	0.196					
Interaction	0.96	1	0.66	0.4204					
Error	57.955	40							

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

### 5.1 Mono-Stereo

- **Collision Rate and Number:** 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 driving accuracy.
- **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%.
- **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.
- **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.
- **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.
- **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.
- **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).
- **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.
- **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%.
- **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%.

## 5.2 Laptop versus Wall

- **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.
- **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  $P=0.0086$ .
- **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.
- **Path length:** There is almost no difference between the two systems in terms of path length.
- **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 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.
- **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. (Jones et al., 2001), 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.
- **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 Demiralp et al. considerations (Demiralp et al. 2006), 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 teledriving tasks because of the fatigue an operator is exposed to. The laptop system remains nevertheless most suitable as a low-cost and portable facility.

- **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.
- **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.
- **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.

## 6. Conclusion

The present chapter introduced a guideline for usability evaluation of VR applications with focus on robot teleoperation. The need for an effort in this direction was underlined in many literature works and was believed relevant by the authors being human-computer interaction a subject area in great expansion with an increasing need for user studies and usability evaluations. The proposed work targets researchers and students who are not experts in the field of evaluation and usability in general. The guideline is therefore designed to represent a simple set of directives (a handbook) which would assist users drawing up plans and conducting pilot and formal studies.

The guideline was applied to a real experiment while it was introduced. The goal was to facilitate the reader's understanding and the guideline actual use. The experiment involved mobile robot teleguide based on visual sensor and stereoscopic visualization. The test involved two different 3D visualization facilities to evaluate performance on systems with different characteristics, cost and application context.

The results of the experiments were illustrated in tables and described after key parameters proposed in the usability study.

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.

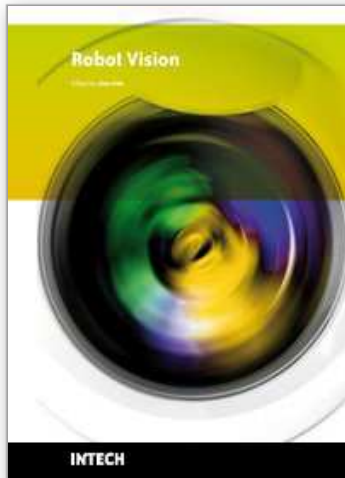
The results of the evaluation on the laptop-wall factor indicated significantly better performance on the laptop in terms of the mean of minimum distance to obstacles. No statistically significant results were obtained for the other variables. The interaction between the two factors was not statistically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

The results therefore provide insight on the characteristics and the advantages of using stereoscopic teleguide.

## 7. References

- Bocker M., Runde D., & Muhlback L., "On the reproduction of motion parallax in videocommunications," in *Proc. 39th Human Factors Society*, 1995, pp. 198-202.
- Bowman, D.A., Gabbard, J.L. & Hix, D. (2002). A survey of usability evaluation in virtual environments: classification and comparison of methods. In *Presence: Teleoperation in Virtual Environments*, 11(4):404-424
- Burdea, G.C., & Coiffet, P. (2003). *Virtual Reality Technology*, John Wiley & Sons, Inc., 2nd edition, ISBN 978-0471360896
- Corde L. J., Caringnan C. R., Sullivan B. R., Akin D. L., Hunt T., and Cohen R., "Effects of time delay on telerobotic control of neural buoyancy," in *Proc. IEEE. Int. Conf. Robotics and Automation*, Washington, USA, 2002, pp. 2874-2879.
- Demiralp, C., Jackson, C.D., Karelitz, D.B., Zhang, S. & Laidlaw, D.H. (2006). CAVE and fishtank virtual-reality displays: A qualitative and quantitative comparison. In *proc. Of IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 3, (May/June, 2006). pp. 323-330.
- Faulkner, X. (2000). *Usability engineering*. Palgrave Macmillan, ISBN 978-0333773215
- Fink, P.W., Foo, P.S. & Warren W.H. (2007). Obstacle avoidance during walking in real and virtual environments. *ACM Transaction of Applied Perception.*, 4(1):2
- Ferre M., Aracil R., & Navas M, "Stereoscopic video images for telerobotic applications," *J. Robot. Syst.*, vol. 22, no. 3, pp. 131-146, 2005.
- Jones G., Lee D., Holliman N., & Ezra D., "Controlling perceived depth in stereoscopic images," in *Proc. SPIE*, 2001, vol. 4297, pp. 422-436.
- Koeffel, C. (2008). *Handbook for evaluation studies in vr for non-experts*, Tech.Rep. Medialogy, Aalborg University, Denmark, 2008.
- Livatino, S. & Koeffel, C. (2007), *Handbook for evaluation studies in virtual reality*. In *proc. Of VECIMS '07: IEEE Int. Conference in Virtual Environments, Human-Computer Interface and Measurement Systems*, Ostuni, Italy, 2007
- Nielsen, J. (1993). *Usability engineering*, Morgan Kaufmann, ISBN 978-0125184069
- Nielsen, J., & Mack R.L. (1994). *Usability Inspection Methods*, John Wiley & Sons, New York, USA, May 1994, ISBN 978-0471018773
- Rubin, J. (1994). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*.
- Sexton I, & Surman P., "Stereoscopic and autostereoscopic display systems," *IEEE Signal Process. Mag.*, vol. 16, no. 3, pp. 85-89, 1999.
- Wanger, L.R., Ferweda J.A., Greenberg, D.P. (1992). Perceiving spatial relationships in computer generated images. In *Proc. of IEEE Computer Graphics and Animation*.





## **Robot Vision**

Edited by Ales Ude

ISBN 978-953-307-077-3

Hard cover, 614 pages

**Publisher** InTech

**Published online** 01, March, 2010

**Published in print edition** March, 2010

The purpose of robot vision is to enable robots to perceive the external world in order to perform a large range of tasks such as navigation, visual servoing for object tracking and manipulation, object recognition and categorization, surveillance, and higher-level decision-making. Among different perceptual modalities, vision is arguably the most important one. It is therefore an essential building block of a cognitive robot. This book presents a snapshot of the wide variety of work in robot vision that is currently going on in different parts of the world.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Salvatore Livatino, Giovanni Muscato and Christina Koeffel (2010). Testing Stereoscopic Vision in Robot Teleguide, Robot Vision, Ales Ude (Ed.), ISBN: 978-953-307-077-3, InTech, Available from: <http://www.intechopen.com/books/robot-vision/testing-stereoscopic-vision-in-robot-teleguide>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen