

Augmented Reality Stereoscopic Visualization for Intuitive Robot Teleguide

Salvatore Livatino
School of Engineering and Technology
University of Hertfordshire
Hatfield, United Kingdom
Email: s.livatino@herts.ac.uk

Giovanni Muscato, Davide De Tommaso, Marco Macaluso
Dipartimento di Ingegneria Elettrica
Elettronica e dei Sistemi
University of Catania, Italy
Email: gmuscato@diees.unict.it

Abstract—This paper proposes a method to simultaneously and coherently present visual and laser sensors information through an augmented reality visualization interface further enhanced by stereoscopic viewing. The use of graphical objects is proposed to represent proximity measurements which are superimposed and suitably aligned to video information through image processing. This new methodology enables an operator to quickly comprehend scene layout and dynamics, and to respond in an accurate and timely manner. Therefore the human-robot interaction is expected to be intuitive, accurate and fast. The use of graphical elements to assist teleoperation, sometime discussed in the literature, is here proposed following a systematic approach and developed based on authors' previous works on stereoscopic teleoperation. The approach is experimented on a real telerobotic system where a user operates a robot located approximately 3,000 kilometers apart. The results of a pilot test were very encouraging. They showed simplicity and effectiveness of the approach proposed and represent a base for further investigations.

I. INTRODUCTION

Many applications in robotics are very complex and they require intervention with unknown, inaccessible, or dangerous environments where unpredictable situations may occur. E.g. in case of industrial robots operating in deep waters, in planetary or volcanos exploration, in applications of safety and prevention, for bomb finding and their deactivation, [1]-[3]. These applications and many more make impossible the use of fully autonomous robotic systems and require direct human (tele-) intervention to resolve issues. These are the cases where human-cognition is irreplaceable because of the high operational accuracy that is required, as well as deep environment understanding and fast decision-making. E.g. to escape deadlock situations, to perform skilful manoeuvring.

Robot teleoperation systems typically rely on 2D displays. These systems suffer of many limitations. Among them: misjudgement of self-motion and spatial localization, limited comprehension of remote ambient layout, object size and shape, etc. These limitations lead to unwanted collisions during navigation, as well as long training periods for an operator. An advantageous alternative to traditional 2D (monoscopic) visualization systems is represented by the use of stereoscopic viewing. The authors of this paper have proposed in a previous work a mobile robot teleguide based on video images, [4]. The use of stereoscopic viewing increased user's sense of presence

and improved understanding of the remote scene structure and distance to surrounding objects.

Stereoscopic viewing improved performance on a system relying on the visual sensor only, but there are a number of other sensors a robot can rely on, and they can well complement visual sensor output, e.g. laser, infrared, odometry, sonars, bumpers, etc. The use of these sensors can significantly improve teleoperation performance because of the additional information they provide. There is however a problem related to contextual visualization of different sensor measurements. This information is often shown to a user independently into the same interface, lacking of coherence and being visualized in a not intuitive manner. Sometime the panel size is also limited despite several information need to be shown.

This paper proposes a method to simultaneously and coherently present both video and proximity information within a stereoscopic augmented reality viewing context. The approach focuses on the use of graphical objects showing (laser-based) information superimposed and integrated with video objects. The proposed approach exploits stereoscopic visualization too.

II. MULTI-SENSOR DATA AND AUGMENTED REALITY

Vision being the dominant human sensor modality, large attention has been paid in telerobotics to the use of visual sensors. Video images provide rich and high contrasted information and this comes presented in a way that is intuitive for humans. There are limitations to the use of visual sensor in teleoperation, e.g. a possible delay in image transmission, [5]. In fact, the rich information provided by a camera may require a large bandwidth to be transmitted over a network at interactive rates. Furthermore, the captured images may come from a constrained viewpoint being a camera typically placed very close to robot.

Some of the limitations can be reduced by using additional sensors. A laser sensor can be very useful to assist robot navigation and to complement visual information. We find several contributions in the literature proposing the use of camera and laser sensors, e.g. [6]. Other sensors like sonars or infrared have also been proposed, e.g. [7], [8]. A laser rangefinder can be very effective in measuring position and orientation of walls and obstacles surrounding a robot, and their distance to the robot. This sensor can provide accurate

estimates which has made it suitable for extracting 2D floor-maps of a robot workspace. 3-D spatial information can also be obtained by combining more sensor readings or a by letting the laser device move, [9]. Furthermore, it is possible to rely on 3D spatial information extrapolated from 2D laser measurements in case of simple structured environments, [10]. The figure 1 shows our robotic system equipped with a stereo camera and laser sensors, and it includes an example of the visual output they may generate. Other sensors can also be exploited to assist robot teleoperation, e.g. infrared, odometry, bumpers, and sonars.

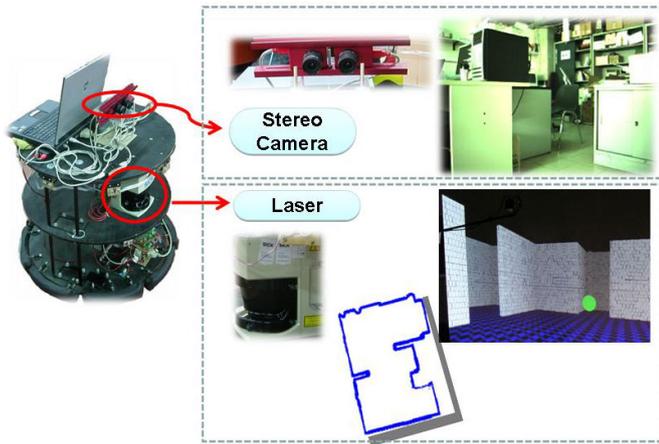


Fig. 1. The stereo camera and laser sensors with an example of the visual output they may generate. An example image of a generated 2-D floor-map reconstructed by the laser, is also shown.

The different types of information can be directly presented to a user on different displays. Alternatively, a methodology needs to be devised in order to make more effective the visualization of multi-sensor information. A relevant issue is how to present proximity measurements to an operator in a way that this can effectively and quickly be understood during tele-intervention. Furthermore, in case a visual input is also present, one wishes to combine both video and laser information in an intuitive and consistent manner. If the sensorial information is provided in a natural and intuitive manner, it paves the way to fast information processing and comprehension by a human. An intuitive visual context is a precondition for immediate and accurate reaction. The visual interface development is therefore a fundamental pre-requisite to improve human-robot interaction.

Augmented Reality (AR) represents a convenient methodology to present multi-sensor information. Our focus is on AR visual representations. From when it first appeared, AR has been proposed to robotic systems to assist tele-users, [11]. In Rastogi et al. [12], an operator is provided with visual feedback of a virtual wireframe robot superimposed over an actual physical robot. AR has also been proposed industrially to be used in the programming of painting robots to manipulate a spray gun and to teach a robot how to use it, [13], [14]. The benefits of AR include an increased task-related intuitiveness, which improves users' performance, and make

more efficient to train an operator too, [15]. AR has also been applied to maintenance, manual assembly [16], and computer-assisted surgery [17]. A number of works, e.g. [18], [19], have shown that AR can be a very efficient method for sensor fusion and for status information delivery in teleoperation. Nonetheless, an effective and consistent representation of multi-sensor information is also a challenging problem.

III. PROPOSED APPROACH

The overall aim for the proposed approach is to represent robot sensors information: visually, consistently, and three-dimensionally.

- **Visual Scenario.** To integrate visual and proximity information into the same visual scenario, so to avoid having different displays that compete for operator's attention.
- **Consistent Representation.** To simultaneously present multi-sensor information in a coherent and intuitive manner. The display is expected to show both video and graphical objects within a multi-layer AR context.
- **3-D Visualization.** To enhance our AR interface by stereoscopic viewing. An effective 3-D visualization of the environment allows a user to comprehend environment layout, objects shape, and their dimensions, in a way more natural and accurate than using 2-D displays.

There are a number of issues that need to be carefully addressed in order to achieve the above objectives. A main issue is information alignment. To obtain visual consistency of the different sensors measurements we need to calibrate the different sensors. In the literature we find proposals for off-line calibration, e.g. [20], [21]. This type of approach requires a number of specific actions to be taken before teleoperation takes place. A robot teleoperator may need to be trained for this calibration. An automatic calibration process would certainly be more desirable. However, automatic calibration is unreliable in practise for general application contexts.

In this work we propose an online semi-automatic calibration process. This is expected to take no longer than a couple of minutes. The user follows an interactive procedure to "align" different sensor information in an easy and intuitive manner. The calibration process takes place before starting a teleoperation activity and the initial calibration settings can typically be employed for an entire teleoperation session. The calibration setting can also be stored to be reused on future teleoperation sessions under the same environment setup.

The system starts by automatically processing video and graphical information and showing the result of the interaction within an AR context. The users can react to what they observe by adjusting the visualization settings, e.g. in order to make a object contour (estimated through laser measurements) superimpose the correspondent area captured by the camera. The system will react to users' adjustments and will re-process sensor measurements based on new sensors information. The new result will be presented to the users and they will so be able to provide further input to refine the result. The users are expected to be satisfied with suggested setup within very short time. The possibility for users to immediately see the results

of their actions, makes the proposed semi-automatic process very effective (human-cognition is directly involved).

An interactive semi-automatic approach to system calibration has been proposed in the past in different research areas. It has for example been responsible for successful applications in the area of vision-based 3-D reconstruction (an area where fully automatically processing had been sought for several years). A good example is the successful work of Debevec et al. [22]. We believe that a semi-automatic calibration in multi-sensors based teleoperation using AR, can make this approach popular in teleoperation interfaces.

IV. THE ALGORITHM

The method proposed for multi-sensor viewing includes the processing phases described below and illustrated in figure 2.

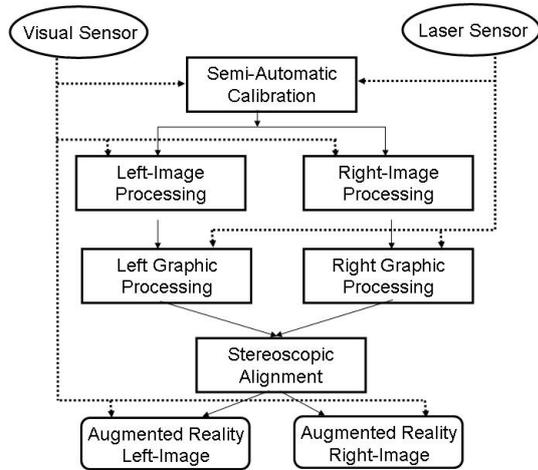


Fig. 2. The processing phases of the proposed interface.

A. Semi-Automatic Calibration

The first phase is to align visual, proximity, and positional information. The calibration process is composed by two automatic processing steps. They are:

- **Color Mapping.** The 2-D laser provides a number of distance measures (one for each degree) taken at sensor height on the horizontal plane. The biggest and smallest distance-values delimit a distance range. A color look-up table is consequently generated, where the color "red" represents the smaller range value (closest distance) and the color green represents the biggest range value (farer distance). Each distance measurement obtained through the laser sensor is mapped to a color (between red and green). This process generates a color map for each laser degree to be assigned to corresponding pixel. In our experiments we have used the following formula:

$$red = MAX * (maxdist - dist) / (maxdist - mindist)$$

$$green = MAX - red$$

The *green* and *red* values are the light intensity values for the red and green colors; *MAX* is the max range for light

intensity e.g. the constant value 255; *dist* is the current distant value; *maxdist* and *mindist* are the maximum and minimum distance values.

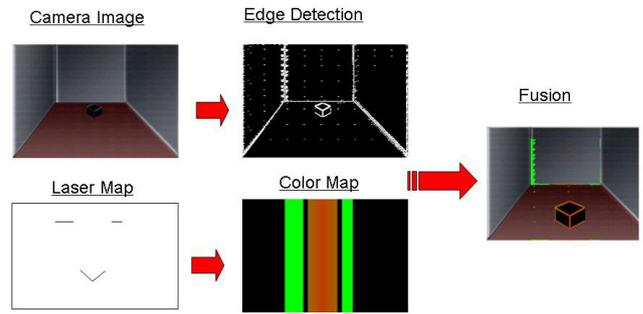


Fig. 3. The process of obtaining a transparent colored layer.

- **Transparency Layer.** The color map is added to the video images incoming from the robot camera. By performing a summation of values at pixel level we obtain a transparency effect. The figure 3 shows the process and the figure 4 shows the video image and its corresponding color map.



Fig. 4. The figure shows the video image and its corresponding color.

The proximity measurements acquired by the laser and superimposed to video image with the transparent layer technique, provide an operator with clear visual feedback about sensors alignment. In fact, any misalignment becomes visible, (though, it depends on environment scene). The operator can consequently take action and manually trigger the visualization of laser measurements to make them correspond to appropriate environment objects. Operators will get immediate feedback about consequences of their actions. They are expected to complete the calibration process within a couple of minutes. The calibration process will also provide as result the subset of laser readings corresponding to camera field of view. The figure 5 shows an image from the calibration step.

B. Image Processing

The Image Processing phase is independently applied to both left and right camera images. The objective of this step is to get reliable edges of the relevant objects present in the scene. The edges are important in order to identify image segments on top of which graphical elements should be drawn.

The edge detection is implemented through the use of the Canny's algorithm on OpenCV framework, [23], which follows a thresholding process. A relevant role is played by parameters tuning. The figure 6 shows an image before and after the Image Processing.

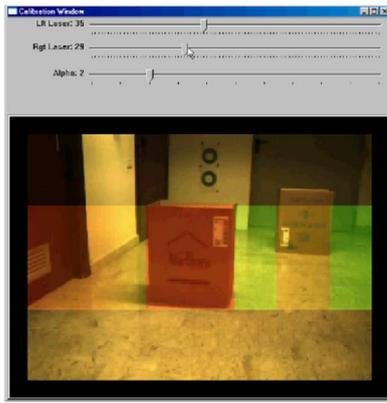


Fig. 5. An image from the calibration phase. The different transparent colored layers should superimpose the correspondent video objects.



Fig. 6. The image processing result (right) applied to a camera image (left).

C. Graphic Processing

The graphic processing phase starts with performing the Color Mapping and Transparency Layer steps as described in the Semi-Automatic Calibration. The look-up table used to build a color map can be thought to contain different levels of light intensity values, which can be tuned according to the transparency level one wishes. The graphical elements used in our work are of three types:

- *Proximity Plane.* It follows the same idea of the Transparency Layer proposed in the calibration phase, i.e. to superimpose transparent colored planes onto different objects present in the video scene. The layer is colored according to the sensor detected proximity value so that an operator can visually associate a color to a distance value and get immediate understanding of possible hazards. In this case the transparent layer can be adapted to incoming video context (illumination, distance to the object, navigation modality). The transparency of a proximity plane can be set to different values by the operator according to application or specific situations. The figure 7 shows an example of proximity planes of different colors superimposed to environment objects.
- *Ray Casting.* As alternative or complement to proximity planes, we propose to superimpose graphical "rays" to the video incoming image. The rays depart from camera position (approximately the center of the onboard stereo-camera system) to reach the closest object on a specific direction. A ray is emitted for a subset of directions, (typically those related to closest objects). Each ray is

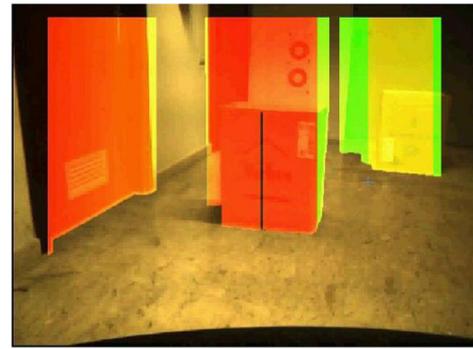


Fig. 7. Proximity planes of different colors superimposed to different environment objects.

colored based on the look-up table concept described above. The number of rays per degree can be controlled so as ray thickness.

- *Figure Value.* The estimated distance to an object as well as the robot position can also be superimposed to a video image. This type of information can be useful in case a precise estimate is required, e.g. during very slow careful motion. The displayed figures can be colored based on the look-up table concept described above, while their size and thickness can also be adjusted (manually or automatically) depending on the navigation context. e.g. figure size can be inverse proportional to robot speed.

D. Stereoscopic Alignment

The left and right images captured by the onboard stereo-camera and sent through the network, are very similar but not identical. In fact, they differ because of the stereoscopic view difference. This image difference is necessary in order for an observer to appreciate the 3-D effect, [24]. As a consequence, the image processing performed in each image may generate different results. We have two different types of image processing outcome:

- *Correspondent Features.* In this case, the detected image-features, e.g. edges, represent the same environment feature, despite they are not located in the same image-pixel on different stereo-images. In other words, the same part of the same object is detected on both left and right image.
- *Undetected Features.* In this case, it happens that for some of the environment objects a features detected in the left images does not have a correspondent one in the right image. E.g. an edge detected in the left image is not detected in the right image. This is especially the case for features appearing close to image borders, but it also happens for some other features, e.g. because they are "weakly" presented. The undetected features typically are a very limited number. Nonetheless, these features cause that some of the graphic elements (generated by the graphic processing) do not appear on both left and right images. Even if we are likely referring to a small set of graphic elements, they generate uncomfortable

viewing when the augmented scenario is observed in stereo. The figure 8 shows an example. The green circles show correspondent features while the white circles show undetected features.

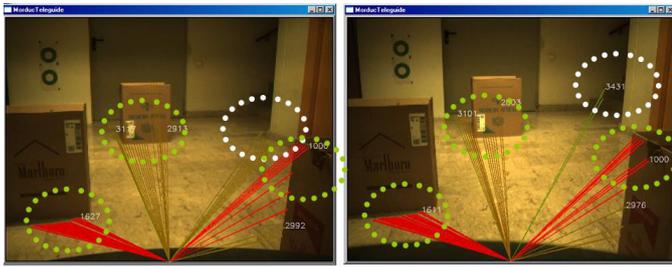


Fig. 8. An example where Stereo Alignment is needed. The green circles show correspondent features while the white circles show undetected features.

The Stereoscopic Alignment step aims at solving the problem with undetected features. A special procedure is run to check for correspondences between graphic elements in both the stereo images. If a correspondence is not found the procedure classifies the graphic element as outliers and this is removed. In case graphic elements on different stereo images do correspond but their image discrepancy is above the threshold of comfortable viewing, the element can be recomputed (e.g. based on the average value) and presented in the resulting AR view.

V. SYSTEM SETUP AND TEST TRIALS

The proposed method is tested on a real telerobotic system which allows users to operate a robot located approximately 3,000 kilometers apart from them. The robot operates in the Robotics Lab at the University of Catania, Italy, while the operator is sitting in the 3D Visualization and Robotics Lab at the University of Hertfordshire, United Kingdom. The figure 9 shows a schematization of the system setup.

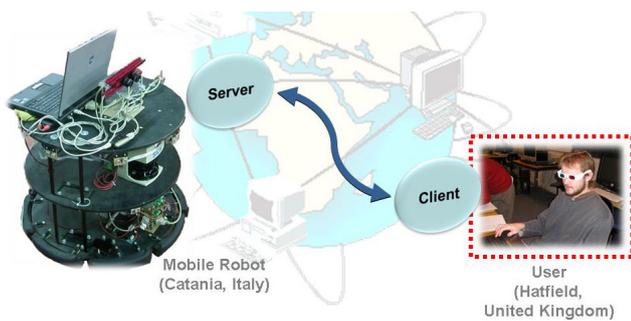


Fig. 9. A schematization of the system setup.

The robot is equipped with a stereoscopic video camera (Videre Sth-dcsg-var) and a bi-dimensional laser rangefinder (Sick LMS-200). The figure 1 shows the sensor systems and the way the information is processed.

The telerobotic system is an evolution of the system adopted in previous work by the authors, [4], but this time the system features a new operational setup and processing scheme

because of the different sensors information being simultaneously processed and displayed, (video, laser, odometry).

The processing phases illustrated in figure 2 take place at operator's site. We use C++ programming language and the OpenCV libraries both for the image and graphic processing.

The proposed pilot test aims at obtaining a first assessment of performance and reliability of the proposed AR visual interface based on the proposed image and graphic processing and information alignment strategy. The test-users were asked to perform typical robot movements such as forward translation and rotation parallel to floor. A simple working environment is proposed which is suitable for laser detection. Our laser system has nonetheless been proven reliable also on more articulated experiments and setups, [10].

We ran 4 pilot trials with the aim of testing: the simplicity and effectiveness of the semi-automatic calibration; the reliability of image processing; the support provided by the different options of the graphic processing (proximity planes, ray casting, figure values), and the stereoscopic setup. The figure 10 shows snapshots of the different processing phases as they appear to our operator during one of our pilot trial.

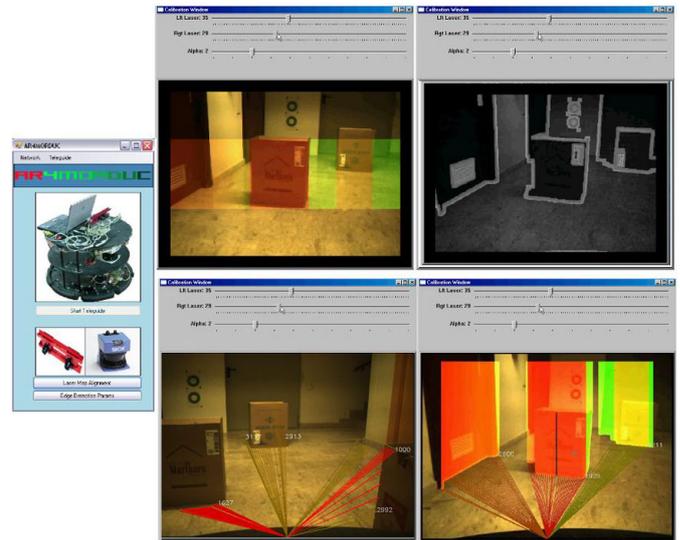


Fig. 10. Snapshots of the different processing phases as they appear to our operator during one of our pilot trial.

The results were very encouraging. The users' commands were correctly executed. The resulting visual output showed a correct alignment along the driven trajectories (few meters of translation and up to 45 degrees of rotation). Our objective to assess feasibility and reliability of the proposed methodology was accomplished. We were surprised edge detection worked very well with any robot movement we tried and the rendered graphic elements (proximity planes, ray casted and figure values) were correctly appearing. The test showed that the most critical phases are:

- *Calibration.* The semi-automatic calibration worked very well, however, the calibration needs to be done carefully because a slight misalignment, possibly not noticed at the

initial stage but that may become more noticeable later on. The figure 11 shows an example of misalignment due to unnoticed misalignment during the calibration phase.



Fig. 11. An example of misalignment.

- *Stereoscopic Alignment.* The stereoscopic visualization of augmented worlds needs careful tuning to recognize outliers and keep the sensor reading aligned. In fact, while stereoscopic video images provided great comprehension of the remote observed environment, graphical element were sometime shown with an unsuitable disparity level which affected operator's viewing comfort.

The pilot test gave us useful insight about how to tune the system before performing the foreseen extended test (a formal test) and we believe we have identified problems and solutions. For example, it would be very beneficial the possibility to recalibrate the system at run-time during navigation.

VI. CONCLUSION

This paper presented a new methodology for robot teleoperation based on AR visualization further enhanced by 3D stereoscopic viewing. The aim was a users observation accurate and timely. A method was described which included a strategy for semi-automatic calibration and a consistent way to visually represent video and graphic elements. An algorithm was proposed and implemented on a real telerobotic systems. The algorithm was tested on a simple working environment and for few typical movements. The aim was to test simplicity and effectiveness of the proposed method.

The calibration procedure worked really well when set carefully. The image processing was robust and graphical elements were correctly superimposed to video information. The stereoscopic visualization worked well on video images but it needs further attention to cope with graphic outliers.

The next step will be an extended usability evaluation after improving the system based on the lesson learned where the possibility for run-time system calibration will also be proposed. The new version of our system will also include a carefully designed "filter" to detect and correct graphical stereo outliers. A wider and more challenging workspace will also be proposed.

ACKNOWLEDGMENT

Special thanks go to Valentina Neri, Filippo Bonaccorso and Giovanni Lanteri from the Robotics Lab, at the DIEES, University of Catania, Italy, for their valuable support.

REFERENCES

- [1] R. Murphy, *Human-robot interaction in rescue robotics*, IEEE Syst., Man, Cybern., C, Appl. Rev., vol. 34, no. 2, pp. 138-153, May 2004.
- [2] Z. Zheng, M. Shugen, L. Zhenli, C. Binggang, *Communication Mechanism Study of a Multi-Robot Planetary Exploration System*. IEEE International Conference on Robotics and Biomimetics. 49-54, Dec. 2006.
- [3] G. Muscato, D. Caltabiano, S. Guccione, D. Longo, M. Coltelli, A. Cristaldi, E. Pecora, V. Sacco, P. Sim, G.S. Virk, P. Briole, A. Semerano, T. White, *ROBOVOLC: a robot for volcano exploration result of first test campaign*. Journal of Industrial Robotics, 30 (3), 231-242, 2003.
- [4] S. Livatino, G. Muscato, C. Koeffel, S. Sessa, C. Arena, A. Pennisi, D. Di Mauro, E. Malkondu. *Mobile Robotic Teleguide Based on Video Images*. IEEE Robotics and Automation Magazine, Vol. 14, No. 4, 2008
- [5] L.J. Corde, C.R. Caringnan, B.R. Sullivan, D.L. Akin, T. Hunt, R. Cohen, *Effects of Time Delay on Telerobotic Control of Neural Buoyancy*, IEEE Proceedings of Int. Conference on Robotics and Automation (ICRA), pp 2874-2879, Washington, USA, 2002
- [6] H. Baltzakis, A. Argyros, P. Trahanias, *Fusion of laser and visual data for robot motion planning and collision avoidance*, Machine Vision and Applications, Springer 2003.
- [7] R. Meier, T. Fong, C. Thorpe, C. Baur, *A Sensor Fusion Based User Interface for Vehicle Teleoperation*, IEEE Fuzzy Signals and Rules, 1999
- [8] G. Terrien, T. Fong, C. Thorpe, C. Baur *Remote Driving With a Multisensor User Interface*. Int Conf. Environmental Systems, Toulouse, France, July 2000.
- [9] Fraunhofer IAIS Institute *3DLS: 3D Laser Scanner*. <http://www.3d-scanner.net>, Dec. 2009.
- [10] S. Livatino, G. Muscato, S. Sessa, V. Neri, *Depth-Enhanced Mobile Robot Teleguide based on Laser Images*, Elsevier Journal of Mechatronics, 2010.
- [11] P. Milgram, S. Zhai, D. Drasic, J. Grodski, *Applications of augmented reality for humanrobot communication*. In Proc. of the 1993 IEEE/RSJ international conference on intelligent robots and systems (IROS 1993), p. 146772, Japan, July 1993.
- [12] A. Rastogi, P. Milgram, *Augmented telerobotic control: a visual interface for unstructured environments*. In Proc. of the KBS/robotics conference, Canada, p. 1618, Oct. 1995.
- [13] Inropa, *Intelligent Robot Painting*, <http://www.inropa.com>, Dec. 2009
- [14] T. Pettersen, J. Pretlove, C. Skourup, T. Engedal, T. Lokstad, *Augmented reality for programming industrial robots*, In Proc. of the international symposium on mixed and augmented reality (ISMAR), Tokyo, Japan, p. 31920, Oct. 2003.
- [15] J.W.S. Chong, S.K. Ong, A.Y.C. Nee, K. Youcef-Youmi, *Robot programming using augmented reality: An interactive method for planning collision-free paths*. Robotics and Computer-Integrated Manufacturing Vol. 25 pp. 689 701, 2009
- [16] J. Molineros, R. Sharma, *Computer vision for guiding manual assembly*. In Proc. IEEE Int. symposium on assembly and task planning (ISATP), p. 3628, Fukuoka, Japan, May 2001.
- [17] R.T. Azuma, *A survey of augmented reality*. Presence: Teleoperators Virtual Environ, 6(4):35585, 1997.
- [18] M. Baker, R. Casey, B. Keyes, H.A. Yanco, *Improved interfaces for human-robot interaction in urban search and rescue*. In Proc. IEEE Conf. on Systems, Man and Cybernetics, vol. 3, pp. 2960-2965, 2004.
- [19] C.W. Nielsen, M.A. Goodrich, R.W. Ricks, *Ecological interfaces for improving mobile robot teleoperation*. IEEE Transactions on Robotics, 23(5):927, 2007.
- [20] Y. Bok, Y. Hwang, I. So Kweon, *Accurate Motion Estimation and High-Precision 3D Reconstruction by Sensor Fusion*, IEEE Int. Conf. on Robotics and Automation, Roma, Italy, Apr. 2007.
- [21] Q. Zhang, R. Pless, *Extrinsic Calibration of Camera and Laser Range Finder*. In Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Sendai, Japan, Sept 2004.
- [22] P.E. Debevec, C.J. Taylor, J. Malik, *Modelling and Rendering Architectures from Photographs: A Hybrid Geometry- and Image- Based Approach*, In Computer Graphics (Siggraph), pp. 11-20, Aug. 1996.
- [23] OpenCV Software. *Open Source Computer Vision: OpenCV*, <http://opencv.willowgarage.com>, 2009
- [24] G.J. Kim, *Designing Virtual Reality Systems. The Structured Approach*. Springer, ISBN-10: 1-852339586, 2005