Robotica (2015) volume 33, pp. 2166–2185. © Cambridge University Press 2014
doi:10.1017/S026357471400126X

Design and evaluation of a head-mounted display for immersive 3D teleoperation of field robots

Henrique Martins†, Ian Oakley‡ and Rodrigo Ventura†*

†*Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal*
‡*School of Design and Human Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Republic of Korea*

(Accepted April 24, 2014. First published online: May 29, 2014)

SUMMARY

This paper describes and evaluates the use of a head-mounted display (HMD) for the teleoperation of a field robot. The HMD presents a pair of video streams to the operator (one to each eye) originating from a pair of stereo cameras located on the front of the robot, thus providing him/her with a sense of depth (stereopsis). A tracker on the HMD captures 3-DOF head orientation data which is then used for adjusting the camera orientation by moving the robot and/or the camera position accordingly, and rotating the displayed images to compensate for the operator's head rotation. This approach was implemented in a search and rescue robot (RAPOSA), and it was empirically validated in a series of short user studies. This evaluation involved four experiments covering two-dimensional perception, depth perception, scene perception, and performing a search and rescue task in a controlled scenario. The stereoscopic display and head tracking are shown to afford a number of performance benefits. However, one experiment also revealed that controlling robot orientation with yaw input from the head tracker negatively influenced task completion time. A possible explanation is a mismatch between the abilities of the robot and the human operator. This aside, the studies indicated that the use of an HMD to create a stereoscopic visualization of the camera feeds from a mobile robot enhanced the perception of cues in a static three-dimensional environment and also that such benefits transferred to simulated field scenarios in the form of enhanced task completion times.

KEYWORDS: Urban search and rescue; Head-Mounted Display; Stereopsis; Teleoperation; Human factors; User study.

1. Introduction

Human−Robot Interaction (HRI) is a broad field of research where two general categories can be identified: remote and proximate interactions, depending on whether human and robot are collocated spatially and possibly also temporally.[25] This paper focuses on remote interaction between a teleoperator and a field robot. In this context, Situation Awareness has been defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."[16,18] This is a key factor in the successful teleoperation of field robots, in particular in situations in which the operator has no direct visual contact with the robot. The typical solution in such cases is for the robot to be equipped with video cameras, the images from which are streamed to an operator console. However, empirical studies have shown that the weaknesses of such an approach at providing an effective HRI, with respect to operator situation awareness, are significant impediments to the success of missions.[8,24]

Most commonly, the images captured by field robots are presented in a console user interface alongside a large amount of other information, thus obliging the operator to divide his or her attention between the relatively small- and low-resolution video feeds and the rest of the interface. Additionally, space restrictions on the overall display size typically limit the dimensions of the camera images to a small fraction of the available display area. Altogether, these issues leave operators with a diminished

* Corresponding author. E-mail: rodrigo.ventura@isr.ist.utl.pt

awareness of the environment surrounding a robot they control. Correspondingly, they are prone to general disorientation and often make cognitive mistakes (e.g., in decision making), which negatively impact overall performance during the field missions.[56]

In this paper, a solution to improve on the level of situation awareness provided by console interfaces is presented. We propose a system capable of (1) supplying the operator with three-dimensional (3D) perception of the robot environment, and (2) automatically adjusting the robot cameras field of view according to the head movements of the operator. To provide depth perception to the operator, a head-mounted display (HMD) is employed, in which the two displays are fed with a pair of independent video streams, originating from separate video cameras mounted on the robot. The head movement is measured by a 3-DOF tracker integrated with the HMD. This design is motivated by the observation that 3D perception is particularly important for situation awareness. For instance, according to Murphy, "The key feature for comprehending the navigational context is 3D depth: not just independently of the robot (seeing if there is an obstacle or a drop off up ahead), but also in relation to the robot (e.g., if a camera or sensor is going to snag on something hanging from overhead or just to the side of the vehicle)."[46]

To validate the proposed system, we used RAPOSA,[42] a tracked robot designed for search and rescue (SAR) operations. It is composed of a main body and an articulated frontal body (see Fig. 1) and is equipped with a large number of sensors and actuators, namely two frontal video cameras, a back camera, accelerometers, and gas sensors. The original console interface of RAPOSA is similar to most teleoperated robot interfaces: a GUI displayed in a computer screen (shown in Fig. 1), together with a joy-pad input device for motion control. In order to empirically confirm the added value of the alternative interface proposed here, a systematic evaluation measuring the performance of human operators was conducted.

The contribution of this paper is twofold. First, the presentation of a SAR teleoperation system based on immersive 3D perception of the robot's environment using a binocular-calibrated HMD, and second, the evaluation of this system via a series of brief quantitative studies. This paper builds upon the system presented in ref. [43], where preliminary results were discussed. The current paper provides an extended explanation of the implemented system, as well as a thorough and improved analysis of the user studies.

The remainder of this paper is organized as follows: after an overview of related work in Section 2, a description of the system architecture is provided in Section 3. Section 4 presents a detailed explanation of the *3D vision module*, endowing the operator with 3D perception. Section 5 describes the *head-teleoperation module*, responsible for moving the point of view according to the operator head movement. The usage evaluation results are described in Section 6. Finally, Section 7 presents conclusions and a discussion of potential future work.

2. Related Work

In the last decade, there has been a sustained increase in research activity concerning field robots. In particular, their use in real disaster scenarios has highlighted the importance of their application to SAR scenarios.[3,8,11,24,47] Robots specifically designed for SAR operations have been subject to field evaluations,[10] providing crucial feedback to the scientific community. In particular, such field tests have highlighted open problems hindering the widespread adoption of SAR robots in real situations. Academic international competitions have been held, where competing platforms are tested in a common scenario with a common goal. These include Robocup Rescue[34] and ELROB[1]. In ref. [59], four teams were evaluated in a Rescue Competition at the AAAI-02 Conference, with respect to the sensors and the teleoperation interfaces used. Finally, NIST, a key player in many of these initiatives, has been developing fundamental work to standardize key aspects in SAR robotics.[29]

Although a significant area of research in SAR robotics is concerned with autonomous robots, this mode of operation is a poor fit for realistic disaster scenarios which typically depend on a strict chain of command. In such situations, all active agents (both human and robotic) must comply with centrally issued orders and centrally organized activity. Consequently, fully autonomous robots are not desirable, at least in the mid-term, and situations in which SAR robots are under close human supervision are more realistic for the foreseeable future. Field studies have identified that teleoperation

[1] European Land-Robot Trial (ELROB), `http://www.elrob.org/`.

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

of SAR robots usually requires a operator-to-robot ratio of 2:1. One possible approach to reduce this ratio, and allow for additional robots to be supervised by the same human team, is that of adjustable autonomy.[47] According to this approach, human operators are given the ability to adjust, in real-time, the level of autonomy of the robots they are controlling. For instance, an operator could choose to manoeuvre a robot through a narrow passage (low autonomy), or to command it to perform an automated task, such as stair climbing[20] or docking.[21]

The use of HMDs has been studied since the 1960s (see ref. [9] for a review), and has been progressively gaining momentum in particular application areas, such as computer-aided surgery,[4,40] video gaming, aviation[45] and robotic applications.[13,22,28,54] Depending on the type of application, HMDs may be used for different purposes, whether to simply project an enlarged image to a user, to superimpose a computer-generated image upon a real-world view (i.e., augmented reality), or even to enable the user to interact with a virtual environment. In this final case a head tracking system is typically integrated into the HMD.

Some quantitative studies of HMD applications can be found in the field of medicine,[40,41] typically focused on highly specific tasks such as colonoscopy.[36] Generally, HMD use has been found to yield performance benefits, although the literature is more inconclusive when considering two-dimensional (2D) presentations in HMDs.[36] In the area of teleoperation tasks, the usage of HMD has been evaluated in various problems, such as in simulated pick-and-place tasks[32,38] and in telemanipulation using a physical robotic arm.[57] The 3D view provided by the HMD is in general found to be beneficial, when compared with a monocular view. These studies are however limited to static robots. In ref. [26] an HMD, coupled with a pan and tilt camera system mounted on a transportation vehicle is evaluated. However, the study targets driving and loading tasks only. Recently, Labonte *et al.* reported on an extensive evaluation of a telepresence robot,[35] comparing various display modalities. However, this study does not cover HMD devices.

More recent work on the use of HMDs in robotics includes investigations of how these displays can convey a sense of a robot's embodiment to a remote operator.[2,12] HMDs have also been proposed for simultaneous teleoperation of multiple robots,[37] as well as free-viewpoint navigation in augmented reality.[49] Furthermore, automatic calibration methods for HMDs have been recently proposed.[30,31] These methods provide promising techniques to automate the determination of the rectification parameters used in our system.

Despite this growing body of work, there is relatively little research in SAR robotics which explores the full potential that HMDs offer. One exception to this is the integration of a HMD to the operation of a SAR robot presented by L. Zalud *et al.*[60–62,64] *Orpheus*[63] is a mobile robot controlled with a joystick and a HMD with an inertial head-tracker, through a user interface named *ARGOS*. This system supported the superimposition of distinct types of information on the HMD video feeds, such as thermal information or 3D proximity. Movements of the HMD were used to control a cursor on the display. Although the overlaid visualizations used in this system seem highly valuable, it makes little comment on the role of 3D information, and the use of head motions to adjust the position of a cursor seems poorly related to the operators main task of controlling the robot in its environment.

3. System Architecture

The proposed approach relies on a mobile robot equipped with a pair of stereo cameras and streaming images to an external computer for display. The HMD, together with the head tracker, is connected to this external computer, where the captured images are rectified before being displayed in the HMD. This rectification is necessary in order to provide the operator with effective and comfortable stereo vision. Moreover, the robot movement is controlled by this external computer. The head tracker signal is used to (1) rotate the images according to the head roll angle, in order to compensate for the fact that the HMD also rotates with the operator's head, and (2) to control the robot movement according to the head pitch angle and yaw motion. The image rectification algorithm is described in Section 4, while the robot control system is presented in Section 5.

The implemented system builds upon the existing operator console of RAPOSA, comprising a computer running the graphical user interface software (GUI), and a joy-pad for commanding the robot. Figure 1 depicts the resulting system architecture. This interface displays the video stream

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

Fig. 1. Implemented system architecture using the RAPOSA robot.

from three of the robot's cameras[2] on the computer screen. Specifically, the images from the two front cameras are displayed, side by side, in the GUI. The HMD is connected to the graphics board of this computer, via a VGA connector and configured as a secondary screen. The manufacturer's software drivers allow for feeding each one of the HMD twin displays independently, which is a fundamental requirement for stereopsis. To accomplish that using a single VGA connector, the HMD multiplex images based on even or odd frames. The head tracker connects using a USB cable. The manufacturer driver software provides data from the head tracker.

Two software modules implement the alternative interface proposed here: a *3D vision module*, responsible for the rectification of the images, and the *head-teleoperation module*, which commands the robot movement according to the head tracker angles. These two modules are described in the next two sections.

4. 3D Vision Module

The *3D vision module* aims at providing stereopsis to the human operator, most commonly known as depth perception, by using the HMD. Stereopsis is the process in visual perception leading to the sensation of depth from two slightly different projections of the world onto our eyes, i.e., one of the cues that humans use for perceiving the world in three dimensions.[51] Given that the robot has two stereo cameras, it is possible to provide 3D cues, starting from the captured pair of 2D images (hereafter named stereo pair), processing it, and feeding the final result to a device capable of achieving a stereoscopic effect, namely the HMD.

However, the cameras in the robot's frontal body are prone to misalignment, due to mechanical vibration. Displaying the raw images in the HMD as they are captured would cause confusion and eye strain.[44] For this reason, the images are rectified using image processing techniques, prior to displaying them in the HMD. Moreover, they are displayed in the HMD full screen, which in principle allows for an improved detailed perception.

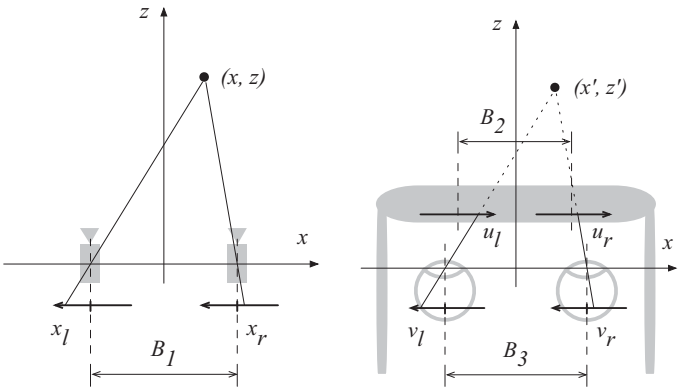
The HMD used in our implementation is a consumer model[3] offering a 32° field of view, corresponding to a 62″ (157 cm) screen watched from a distance of 9 feet (270 cm). Although other models of HMD were not tested, other similar models, including both (1) binocular view and (2) a 3-DOF head tracker, should be equally effective in this context.

4.1. Stereopsis

For the analysis of depth perception, we will consider a simple pinhole model for the cameras and for the operator eyes.[50] Furthermore, we will restrict our analysis to the geometry of the problem in the 2D plane passing through the stereo pair optical axis (assumed parallel). Figure 2 introduces the notation for the problem geometry. Consider a point (x, z) in 2D space, that gets projected to the

[2] RAPOSA is equipped with a total of four cameras: two in the front, the ones displayed in the HMD, one thermal camera also in the front, and one camera in the back for remote docking.[21,42]
[3] Vuzix iWear VR920: twin 640 × 480 LCD displays, 32° field of view, 24-bit color, 60 Hz progressive scan, weight 3.2 oz.

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press



(a) Robot side (b) Operator side

Fig. 2. Geometry of the stereopsis problem.

coordinates x_l and x_r in the left and right image planes, according to their projective geometry:

$$x_l = \frac{x + B_1/2}{z}, \qquad x_r = \frac{x - B_1/2}{z}, \tag{1}$$

where B_1 is the cameras baseline, and the cameras focal length is assumed to be unitary. The points x_l and x_r are then displayed by the HMD with a scale factor of λ, at the coordinates $u_l = \lambda x_l$ and $u_r = \lambda x_r$ in the head display coordinate frames. Following the operator's eye geometry, these points are perceived at the retina's points v_l and v_r, according to the following expressions:

$$v_l = \frac{u_l + (B_3 - B_2)/2}{d}, \qquad v_r = \frac{u_r + (-B_3 + B_2)/2}{d}, \tag{2}$$

where B_2 and B_3 are the HMD baseline and inter-pupil distance (IPD), and d is the distance of the HMD and the eyes' focal points (assuming parallel). The perception corresponding to the coordinates v_l and v_r in the retina is equivalent to the perception of a virtual point (x', y') in space:

$$v_l = \frac{x' + B_3/2}{z'}, \qquad v_r = \frac{x' - B_3/2}{z'}. \tag{3}$$

From (2) and (3) we can derive the relationship between the original point (x, z) and the operator's perception of it at (x', y'), in terms of depth perception:

$$z' = \frac{d}{\lambda \frac{B_1}{B_3} + (1 - \frac{B_2}{B_3})z} z. \tag{4}$$

This relation is nonlinear, except when $B_2 = B_3$. If B_2 and B_3 are very close, then we can approximate (4) by

$$z' \approx \frac{d\, B_3}{\lambda\, B_1} z. \tag{5}$$

Note that for $B_2 = B_3$, the IPD and the HMD baseline have to be equal. If they are not, B_2 can be effectively adjusted, by shifting horizontally both images in opposite directions, such that $B_2 = B_3$. This shift is part of the image rectification process proposed here, and described in the next section. Otherwise, if $B_2 \neq B_3$, there is a nonlinearity between z and z', resulting in a distorted depth perception.

This model assumes perfect alignment of the cameras. This is however not always the case in practice, since mechanical misalignments are hard to eliminate. Even in special purpose commercial stereo camera pairs, such as the PointGrey Bumblebee® series, image rectification in the form of

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

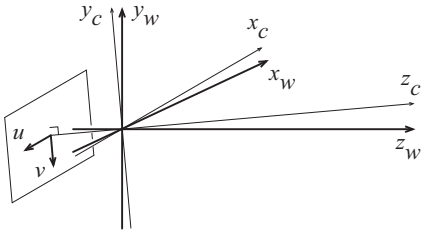


Fig. 3. Geometry for the projective transformation for a misaligned camera. A unit focal length is assumed.

software image processing is performed.[52] In the case of our experimental platform, the RAPOSA robot, this problem is worsened by the fact that the camera mounting is prone to misalignment in the long term, mainly due to mechanical vibration caused by vehicle motion. Therefore, it is necessary to rectify the image streams before displaying them in the HMD.

4.2. Image rectification

The rectification process is straightforward, comprising a rotation and a translation of the displayed images. More rigorous and sophisticated methods can be found in the literature,[23,39] however the simple method presented here proved to be effective for the application. Note that the camera misalignment is relatively small and the objects of interest in the environment and SAR scenario we are considering are reasonably distant from the cameras (typically in excess of 50 cm).

The rectification is applied to each camera individually. Let us consider the misalignment of one camera, with respect to the orientation in Fig. 2(a), by small pitch, roll, and yaw Euler angles (see Fig. 3). Consider two coordinate frames, the world frame $\mathcal{W}$ and the camera frame $\mathcal{C}$, with common origin but a small misalignment in orientation. A point $p_w = (x_w, y_w, z_w)^T$ in world frame corresponds to a point $p_c = (x_c, y_c, z_c)^T$ in camera frame, related by a rotation matrix R:

$$p_c = R\, p_w \tag{6}$$

In the camera frame, point p_c is projected to a 2D point $q = (u, v)^T$ in the image frame, according to the usual projective transformation (assuming unit focal distance)

$$q = P(p_c) = \begin{pmatrix} x_c/z_c \\ y_c/z_c \end{pmatrix}. \tag{7}$$

Assuming small rotation angles, one can approximate the resulting transformation $q = P(R\, p_w)$ by a first-order Taylor series approximation around $p_c = p_w$:

$$q \simeq P(p_w) + \frac{\partial}{\partial \Delta} P(R\, p_w)\, \Delta, \tag{8}$$

where $\Delta = (\delta\alpha, \delta\beta, \delta\gamma)$ represent the small pitch, yaw, and roll angles, expressed in R. The Jacobian matrix in this expression can be obtained from the matrix product

$$\frac{\partial}{\partial \Delta} P(R\, p_w) = \frac{\partial P}{\partial p_c} \frac{\partial p_c}{\partial \Delta}. \tag{9}$$

The first term in the right hand side follows from partial derivation of (7):

$$\frac{\partial P}{\partial p_c} = \begin{bmatrix} 1/z_c & 0 & -x_c/z_c^2 \\ 0 & 1/z_c & -y_c/z_c^2 \end{bmatrix}, \tag{10}$$

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

while the second one can be derived from the Lie algebra of the rotation group $SO(3)$,[55] generated by

$$\frac{\partial R}{\partial \Delta} = \begin{bmatrix} 0 & -\delta\gamma & \delta\beta \\ \delta\gamma & 0 & -\delta\alpha \\ -\delta\beta & \delta\alpha & 0 \end{bmatrix}, \tag{11}$$

where $\delta\alpha$, $\delta\beta$, and $\delta\gamma$ are the small misalignment angles around the x, y, and z axes, from which we can derive

$$\frac{\partial p_c}{\partial \Delta} = \frac{\partial R}{\partial \Delta} p_w = \begin{bmatrix} -y_w\delta\gamma + z_w\delta\beta \\ x_w\delta\gamma - z_w\delta\alpha \\ -x_w\delta\beta + y_w\delta\alpha \end{bmatrix}. \tag{12}$$

Substituting these two derivatives in (9) and then in (8), and taking into account that the Jacobian (9) is being computed at $p_c = p_w$, we can obtain the first-order approximate relation

$$q \simeq A\, q_o + B, \tag{13}$$

where $q_o = (u_o, v_o)^T$ is the camera projection of p_w when $p_c = p_w$, A is the matrix

$$A = \begin{bmatrix} 1 - v_o\delta\alpha & -\delta\gamma \\ \delta\gamma & 1 + u_o\delta\beta \end{bmatrix}, \tag{14}$$

and B is the vector

$$B = \left(1 + \frac{1}{z^2}\right) \begin{pmatrix} \delta\beta \\ -\delta\alpha \end{pmatrix}. \tag{15}$$

For small angles $\delta\alpha$ and $\delta\beta$, and relatively distant objects ($z_w \gg 1$), matrix A can be approximated by

$$A \simeq \begin{bmatrix} 1 & -\delta\gamma \\ \delta\gamma & 1 \end{bmatrix} \tag{16}$$

since $|v_o\delta\alpha| \ll 1$ and $|u_o\delta\beta| \ll 1$, and vector B by

$$B \simeq \begin{pmatrix} \delta\beta \\ -\delta\alpha \end{pmatrix} \tag{17}$$

since $z^2 \gg 1$. Note that the matrix A approximation in (16) corresponds to a first-order approximation of a 2D rotation matrix $R_2(\delta\theta)$ around $\theta = 0$:

$$R_2(\delta\theta) \simeq I + \delta\theta \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{18}$$

when $\delta\gamma = \delta\theta$. We can thus conclude that the misaligned camera transformation in (7) can be approximated in the camera image plane, under small misalignment angles and sufficiently distant objects, by (13): a *rotation* by $\delta\gamma$ followed by a *translation* by $(\delta\beta, -\delta\alpha)$. This is consistent with our phenomenological experience that, for sufficiently distant objects, small pitch and yaw movements of a camera produce a vertical and a horizontal shift of the image, while roll movements rotate the image by the same amount.

The proposed rectification algorithm corrects the cameras misalignment by performing a rotation transformation and a crop transformation (effectively a translation) on the original stereo pair of images. This adopted solution involves the usage of a common chessboard calibration pattern,[1] for estimating the rotation and crop parameters.

First, the pixel position of all chessboard inner corners is determined using OpenCV's `cvCalibFilter()`[5,6] (as shown in Fig. 4(b)). Given an image I, only the leftmost and rightmost

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

inner corners are considered; for each row k, these corners are denoted (x_a^k, y_a^k) and (x_b^k, y_b^k). The camera rotation is thus estimated by averaging the angles of the lines formed by these pairs of points, i.e.,

$$A(I) = \frac{1}{N}\sum_{i=1}^{N} \arctan\left(\frac{y_b^k - y_a^k}{x_b^k - x_a^k}\right). \tag{19}$$

Let $A_L = A(I_L)$ and $A_R = A(I_R)$ be the final rotation angles of the left and right images (I_L and I_R). After computing A_L and A_R, and rotating the images (as shown in Figure 4(c)), it is necessary to align them vertically and translate them according to the desired eye vergence offset. These alignments can be easily achieved by cropping the images in the opposite borders. For instance, eliminating the topmost Δy pixels from the left image and the bottom Δy pixels from the right image, the two resulting images are moved $2\Delta y$ pixels relative to each other (the same reasoning can be applied horizontally). Thus, it is necessary to recalculate the position of the pattern inner corners of the new rotated images. Consequently, let (x_L, y_L) be one arbitrary inner corner in the left rotated image, and (x_R, y_R) the corresponding corner in the right rotated image, as shown in Fig. 4(c). The crop parameters are obtained in the following way:

$$\begin{cases} \Delta y = y_L - y_R \\ \Delta x = x_L - x_R - D, \end{cases} \tag{20}$$

where Δy and Δx are the amounts of pixels to crop vertically and horizontally, respectively, and D is the baseline horizontal disparity. A positive Δx means removing the leftmost Δx pixels of the left image and the rightmost Δx pixels of the right image, and a positive Δy means removing the topmost Δy pixels of the left image and the bottommost Δy pixels of the right image. We set $D = B_2 - B_3$, in order to guarantee a linear relationship between z and z' in (5). Thus, even if $B_2 \neq B_3$, we are able to provide an undistorted depth perception to the user.

Figure 4(d) shows the rectified images from the calibration pattern, after the calibration process.

During operation, both left and right camera images are subject to the rectification algorithm: a rotation according to (19), followed by a crop according to (20). Both operations are performed by efficient OpenCV routines, thus introducing a negligible computational burden to the operator PC.

5. Head-Teleoperation Module

The proposed system uses a 3-DOF tracker, which permits the detection of the user's head attitude, e.g., in terms of yaw, pitch, and roll angles. In the case of the implemented system, a consumer HMD was used, providing absolute attitude estimation via three magnetometers and three accelerometers mounted inside the display enclosure. This tracker should be calibrated before use, setting in particular the zero yaw angle, using software provided by the manufacturer.

The robot cameras in RAPOSA are mounted in the articulated frontal body (Fig. 1), hence the camera pitch angle can be adjusted by moving this part. The proposed system uses the yaw angle of the HMD to rotate the robot body in differential drive mode, and the pitch angle to move the articulated frontal body.

We acknowledge that an improved solution can be attained by having independent pan and tilt control of the camera orientation. First, it allows, in principle, for faster camera movement thus contributing to a better immersive perception of the operator, and second, having the camera yaw angle with the robot orientation together in a single DOF, as in RAPOSA, is not natural from the teleoperation point of view, a position supported by our empirical study (see Section 6.5 below).

Given these 2 DOFs of the cameras, the pitch and yaw angles measured by the head tracker are used to directly control them. This allows for the cameras to follow the operator view direction, allowing for the operator to quickly sweep the environment and gain situation awareness faster than with traditional methods (see illustration in Fig. 5).

Since the HMD is worn by the operator's, head roll movements cause the HMD, and therefore the displayed images, to also roll. A consequence of this is that an operator who rolls his or her head would incorrectly perceive the world, as viewed by the robot, to have also rolled. Two alternative

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

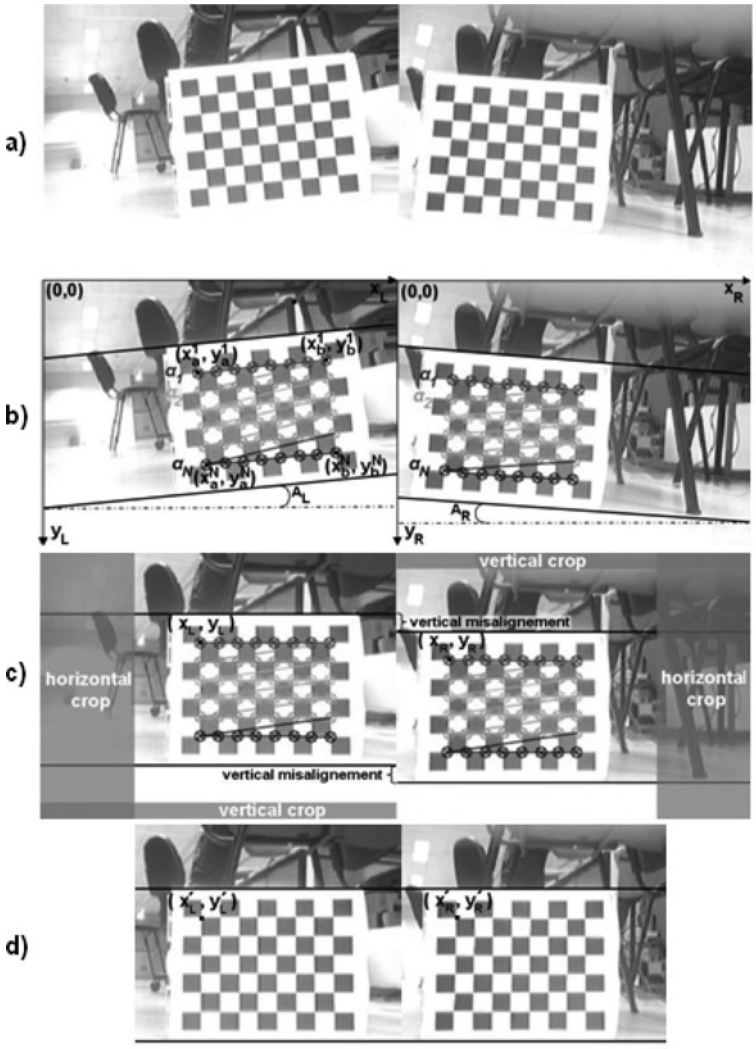


Fig. 4. Image rectification algorithm applied to a sample image containing the calibration pattern: (a) original unrectified stereo pair; (b) detection of the chessboard inner corners; (c) stereo pair after rotation transformation; (d) stereo pair after cropping transformation.

Fig. 5. Illustration of the HMD motion controlling the robot.

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

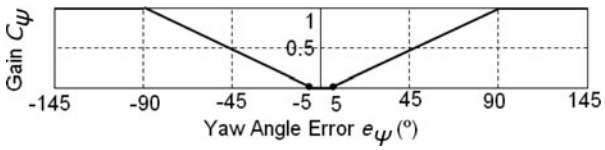


Fig. 6. Gain C_ψ as a function of the yaw angle error.

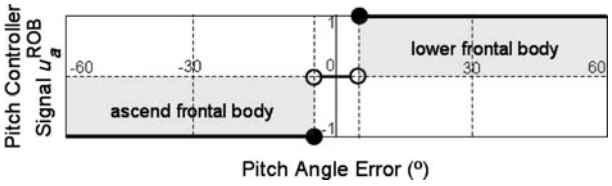


Fig. 7. Pitch controller signal as a function of the pitch angle error.

ways of compensating for this effect are: (1) to show in the HMD the images as captured by the robot cameras, or (2) to rotate the images to the inverse of the angle of the head roll. Even though option (1) allows the operator to perceive the environment in the way that it is perceived by the robot, we found it to induce discomfort. One possible reason for this is that, when a person observes a real scene, casual roll movements of the head imply a rotation of the projected image on the retina. This will not happen in the case of option (1), since the HMD, and consequently the image projected on the retina, rolls together with the head. However, with option (2), the image projected on the operator's retina will rotate as if he or she were directly looking at the scene. This ensures that operators will perceive the world as if they were looking directly at it, regardless of head roll movement. This manipulation is explained in more detail in Section 5.3.

However, it is worth noting that roll of the robot is not compensated in this manner. The reason for this is that such a compensation would impair, if not prevent, the user from accurately perceiving roll movements experienced by the robot. Detecting such events is important in a teleoperation task, as they relate to the slope of the terrain the robot is traversing. Therefore, and in line with previous systems ,[26,35] the immersion effect in the system described in this paper assumes a planar ground.

5.1. Yaw controller

The objective of the yaw controller is to control the yaw rotation of the robot, in order to minimize the yaw angle error, i.e., the difference between the yaw angle read by the HMD tracker ψ^{HMD} and the same angle given by the robot's odometry ψ^{ROB}. Because of the yaw angle bias resulting from the offset between the robot odometry and the head tracker calibration, the neutral yaw angle corresponds to the one when the system is launched. The nonlinear control law used to compute the robot's angular velocity $\omega_\psi^{\mathrm{ROB}}$, given the angular error $e_\psi = \psi^{\mathrm{HMD}} - \psi^{\mathrm{ROB}}$, is given by

$$\omega_\psi^{\mathrm{ROB}} = C_\psi(e_\psi)e_\psi. \tag{21}$$

The gain $C_\psi(e_\psi)$ depends on the angle error, according to the profile shown in Fig. 6. This gain is zero for a dead zone around zero, in order to make it stop when it is considered to be close enough to the desired goal (besides filtering out sensor noise), constant for high error values, and with a linear profile for smoothness otherwise.

The desired angular velocity is then multiplied by a conversion factor k, in order to convert it into velocities u_l^{ROB} and u_r^{ROB} that will be sent to RAPOSA's left and right track wheels. The robot rotates in differential drive mode, so these velocities are symmetric.

5.2. Pitch controller

Similarly to the yaw controller, the objective of the pitch controller $C_\theta(e_\theta)$ is to control the robot's frontal body, in order to minimize the pitch angle error e_θ, i.e., the difference between the HMD pitch angle θ^{HMD} and the robot's frontal body pitch angle θ^{ROB}. In the case of RAPOSA, the frontal body is controlled with a discrete signal u_a^{ROB}, either to maintain the frontal body current position (0),

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

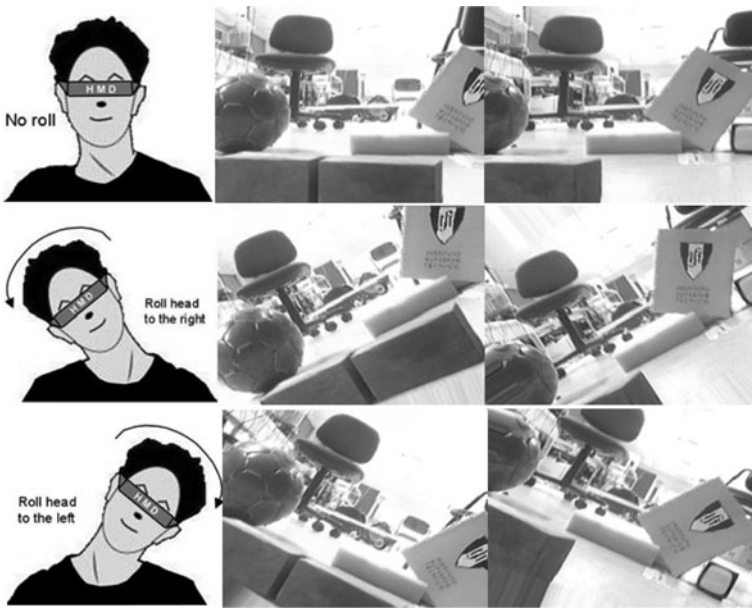


Fig. 8. HMD output when rolling head.

move it up (−1), or move it down (+1). Given an error $e_\theta = \theta^{\text{HMD}} - \theta^{\text{ROB}}$, the control signal u_a^{ROB} in function of this error is shown in the profile of Fig. 7 (bang-bang control with a dead-zone).

5.3. Roll movement

When we roll our head, the images projected on our retina rotate accordingly. However, when a HMD is worn by an operator, the displayed images rotate together with the head, creating the illusion that the video source feeding the HMD has also rotated. This paper compensates for this effect by rotating the images displayed in the HMD by the inverse angle of the user's head roll as derived from the HMD tracker. Figure 8 shows the resulting transformed images shown on the HMD, according to the roll angle measured by the HMD tracker.

It should be noted that this rotation affects stereopsis: rotating the camera images is not equivalent to rotating the robot body. As the images are rotated, the disparities caused by the camera parallax, which are horizontal, are also rotated. Since stereopsis depends upon horizontal disparity, once the disparities turn from horizontal to vertical, stereopsis becomes impaired[4]. However, for small angles this effect is not perceivable and does not cause disturbances or disorientation.

6. Usage Evaluation

Four brief empirical user studies were conducted to compare a typical screen-based display of dual camera video feeds against the head mounted stereoscopic display system introduced in this paper. The first two studies measured detailed perception (of textual information and 3D object placement) while the third study tested depth perception of 3D objects. The final study contrasted the combined input/output components of the HMD system against joy-pad input plus screen output in a simulated search and rescue task in which users had to navigate the RAPOSA robot around an arena and identify targets of interest. The four studies are limited in scope and intended to serve as a broad comparison between head-tracked HMD and on-screen displays in order to highlight performance variations, which can be studied in further depth in future work. The four studies were completed by the same group of 10 participants. Nine were male, one was female, and they were aged between 15 and 40. Two participants reported previous experience with HMD technology. All possessed normal, or corrected to normal, vision. The following sections describe the studies and their results and then close with a discussion of their implications.

[4] In the limit, a 90° image rotation is equivalent to one camera on top of the other, and the pair jointly rotated 90° accordingly. Note that in this case the horizontal disparity is independent from depth.

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

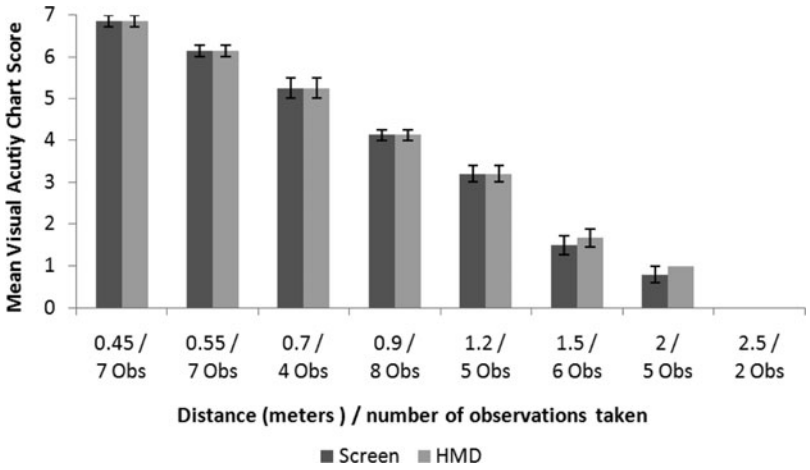


Fig. 9. Mean visual acuity (as measured by performance reading an eye chart) by distance for RAPOSA camera feeds shown on screen and HMD. Bars show standard error.

All user studies with the RAPOSA robot remotely teleoperated and participants could not directly observe the robot's environment—they were situated in an adjacent and visually isolated area. Furthermore, the processing time for the image rectification process was negligible compared with the frame-rate obtained over the wireless video transmission system. The approximate frame rate of the final system was 15 frames per second (67 ms), although this figure fluctuated substantially according to prevailing wireless conditions.

6.1. Experiment I—2D detail perception

This study evaluated the level of visual acuity attained with the screen-based and HMD displays. Participants were required to read an eye-chart featuring seven lines of letters of decreasing size placed in front of the robot's cameras at different distances (between 0.45 and 2.5 m). As in a typical optician's exam, they were scored based on the last line at which they could verbally report the characters with accuracy. Each participant performed the test at either four or five distances and in each case always used both the screen and HMD. This led to a total of 88 observations, 44 for each of the two systems. These data are summarized in Fig. 9. As expected, a one-way ANOVA showed performance decreased significantly as distance from RAPOSA increased ($F(7, 80) = 404$, $p < 0.001$). However, a repeated measures t-test showed no differences between performance with the screen and HMD ($p = 0.08$).

6.2. Experiment II—3D detail perception

The experiment investigated object perception in a 3D environment. It involved six distinctive objects placed in a complex scene in front of RAPOSA. The objects are shown in Fig. 10 (a) and a sample scene in (b). Participants were required to subjectively rate their ability to detect each of the six objects from 0 (indicating an inability to locate the object) to 10 (indicating complete confidence in their perception) using both the screen and HMD systems, leading to a total of 60 observations for each. The distance between RAPOSA and the objects varied among the subjects (from 2.5 m to 3.3 m) as did the object placement in the scene. These variations were introduced to increase the ecological validity of the study: to test performance in varying and general environments, rather than static and fixed ones. To ensure these manipulations did not confound the results, all scene setups were observed using both the screen and HMD. The mean confidence reported for all objects using the two systems is shown in Fig. 11. A matched pairs t-test revealed this difference to be significant ($p < 0.001$).

6.3. Experiment III—depth perception

This study continued the examination of object perception begun in the previous experiment. In this evaluation, 14 objects were placed in front of the RAPOSA's camera and participants were required to precisely identify the location of each using both the screen-based and HMD interfaces. The objects

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

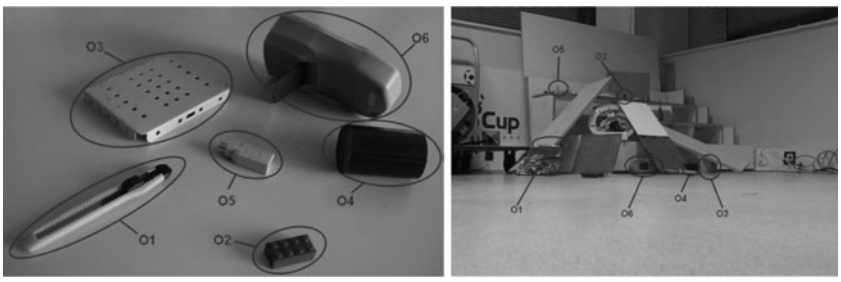


Fig. 10. Objects used (a) and an example scene (b) from experiment 2.

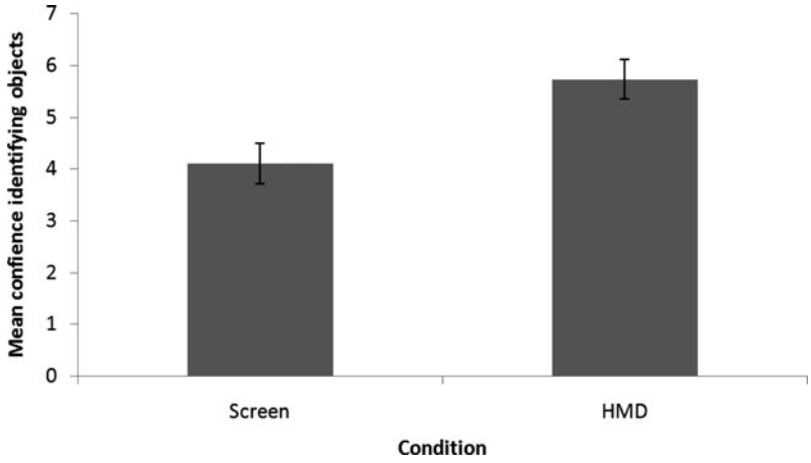


Fig. 11. Mean confidence in object recognition using screen and HMD. Bars show standard error.

were grouped into four categories: (a) a group of three objects situated close together; (b) a group of three identically shaped and sized objects; (c) a group of four identically shaped, but differently sized objects, some of which obscure parts of each other; and (d) a group of four objects arranged in a rough line so that each one partially obscures the one behind it. The layout used in this scenario can be seen in Fig. 12. Participants indicated object position by producing a plan view sketch of the environment and objects; a sample sketch is shown in Fig. 13. These sketches were rated for accuracy on a scale of 0 (inaccurate) to 3 (highly accurate position and scale) by one of the authors, leading to simple numerical data reflecting the precision with which objects could be located using the two screen and HMD systems. These data are shown in Fig. 14. An ANOVA revealed no differences between performance among the different groups of objects ($F(3, 9) = 1.26$, $p = 0.29$, figure not shown), while a repeated measures t-test showed a significant variation between the two conditions ($p < 0.001$).

6.4. Experiment IV—simulated SAR operation

In this experiment, participants performed three simulated SAR operations with the RAPOSA robot, each using a different display/control system. The systems, which served as conditions in the experiment, were:

1. A standard screen GUI display coupled with joy-pad input;
2. The HMD display plus the pitch and roll angles of the HMD head-tracker, together with the joy-pad input; and
3. The HMD display plus the pitch, roll input, and yaw angles from the HMD, together with the joy-pad input.

In all conditions the joy-pad controls the robot in differential drive (move forward/backward + rotate clock/anti-clock wise) and the frontal body angle. The HMD output is capable of controlling the robot rotation and the frontal body angle. The distinguishing factor between these conditions is the extent to which the HMD output controls the robot kinematics. Conditions 2 and 3 were deployed to explore the

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

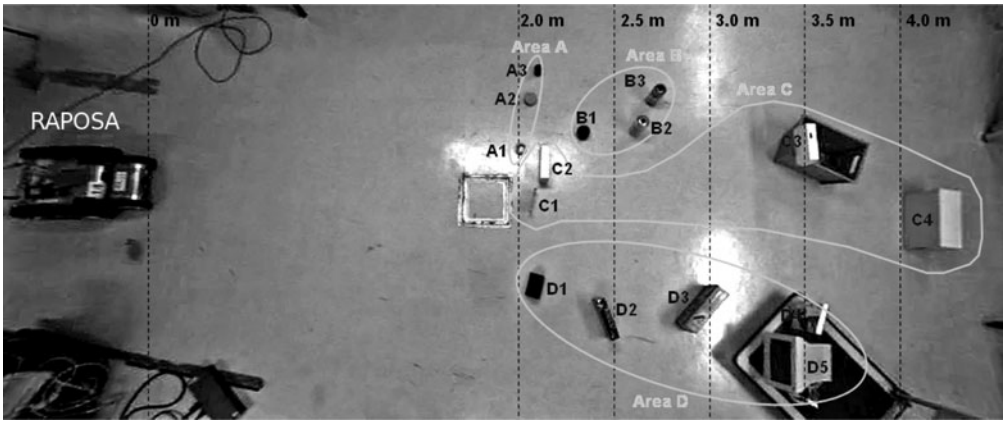


Fig. 12. Layout of the objects in experiment 3 (depth perception) with an overlay showing the object groups (A, B, C, and D). RAPOSA can be seen on the left of the image.

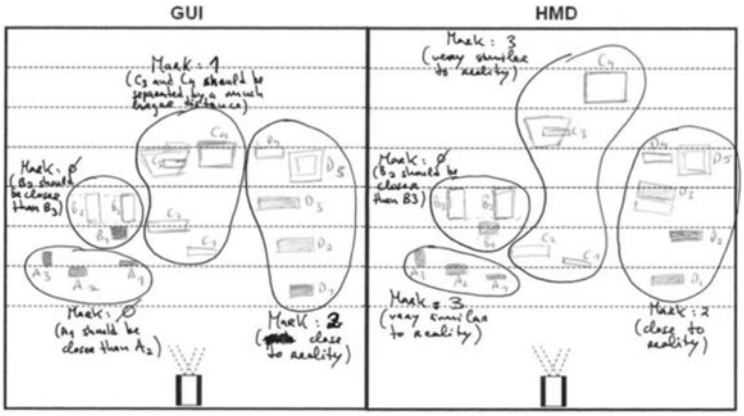


Fig. 13. A sample user sketch showing perceived object placement from both screen-based GUI and HMD conditions in experiment 3 (depth perception).

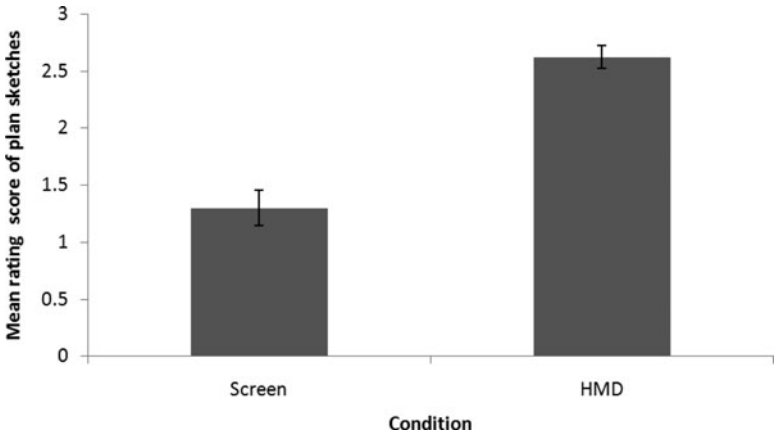


Fig. 14. Rating scores from participant sketches in experiment 3. Bars show standard error.

implications of using the yaw controller, which actually alters the spatial orientation of the robot rather than simply the angle of the camera or viewed scene. For example, with the yaw controller enabled, looking to the right or left signifies turning to the right or left. While this mapping is appealing simple, it may also engender confusion as the exploratory action of looking and the navigational action of turning are merged into one. The simulated SAR task used in the study was based on the NIST reference test arena[29] and incorporated a range of challenges of different levels of difficulty

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

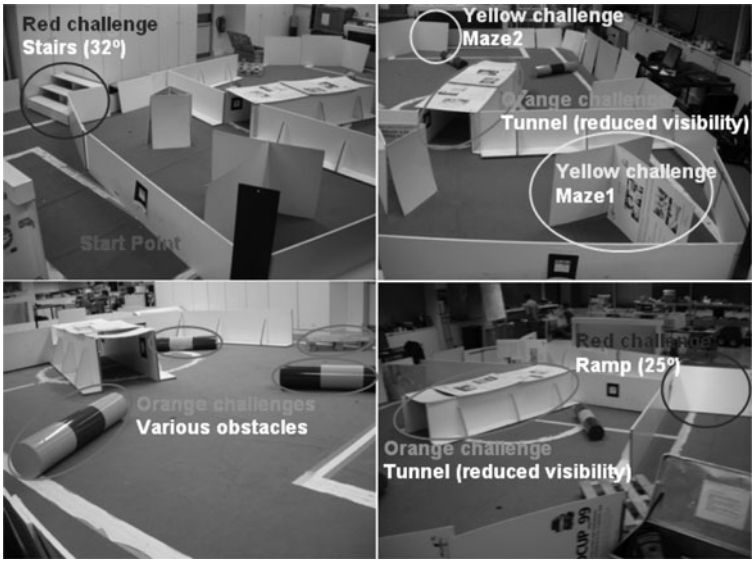


Fig. 15. Images of the simulated SAR environment used in experiment 4. Annotations indicate and define the hazards presence in this environment.

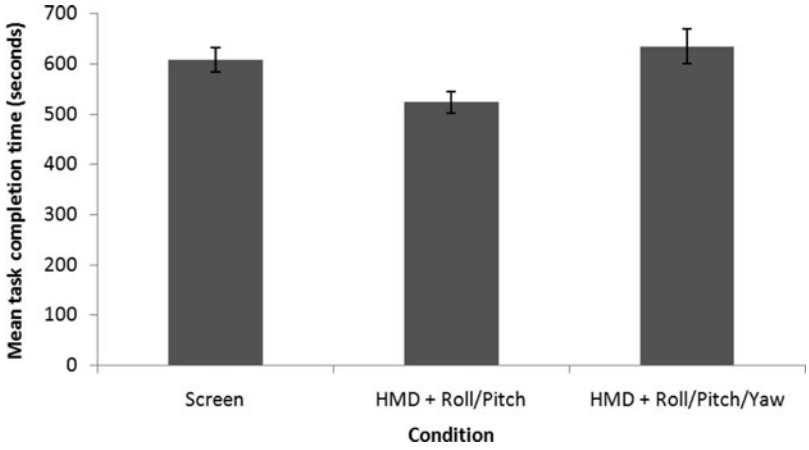


Fig. 16. Mean task completion times for the three conditions in experiment 4 (simulated SAR task). Bars show standard error.

(from yellow, orange, and red arenas) including a simple maze, obstacles on the ground, a tunnel, and a set of stairs. The whole arena was approximately 50 m^2 in size and is shown in Fig. 15. The order of the conditions was partially balanced among the participants (with two subjects performing four of the six possible order conditions and one subject in the remaining two order conditions). The task itself involved navigating RAPOSA (from a preset starting point) around the arena seeking labeled objects of interest, which were then photographed. When participants were satisfied no more objects remained unfound, they returned to the start point. The number and location of objects in the arena varied randomly from condition to condition and subject to subject, preventing participants from simply learning the location of the target objects. However, over the entire experiment, the mean number of objects hidden in each condition was maintained at 4.5. The data captured in this study were total task completion time and the percentage of objects found in each trial. These data are presented for each of the three conditions in Figs. 16 and 17. An ANOVA revealed significant differences in task completion time ($F(2, 9) = 4.359,\ p < 0.02$), but not percentage objects found ($F(2, 9) = 1.6,\ p = 0.22$). Subsequent matched pairs t-tests on the time data showed participants performed with significantly increased speed in the condition incorporating the HMD plus the roll and pitch input when compared the two others ($p < 0.015$ against the screen condition and $p < 0.001$ against the second HMD condition).

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

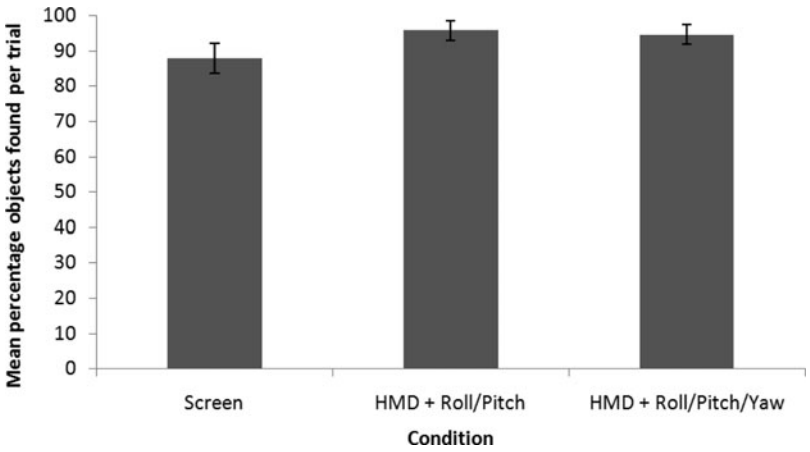


Fig. 17. Mean percentage objects found for the three conditions in experiment 4 (simulated SAR task). Bars show standard error.

6.5. Discussion

Four experiments were conducted exploring diverse aspects of perception and performance using a stereoscopic head mounted display in conjunction with a search and rescue robot. The first three studies examined aspects of visual perception while the robot remained stationary. The final study linked head motion (captured by the HMD) to the robot's movements and focused on a more realistic simulated search and rescue task. Although the studies were brief and limited in scope, when considered together, their conclusions are diverse and informative, suggesting future avenues for research and experimentation.

For instance, experiment one, which explored 2D detailed perception, showed no significant difference between the screen and HMD systems while experiment 2, which considered 3D detailed perception, revealed the HMD led to a significant improvement in the confidence with which subjects reported the location of objects in a complex visual scene. These results suggest that the benefits conferred by the use of HMD do not relate to issues of resolution or the size of the viewing area. Instead they are due solely to the increased quality of 3D perception achieved through the use of stereoscopy. Experiment 3 extends these claims by considering object perception in a more complex environment and by introducing a method (based on participants drawing plan sketches) by which this can be reliably assessed. The results indicate that a significantly improved level of perception of 3D aspects of scenes (such as depth) is achieved through the use of the stereoscopic HMD. This set of results effectively validates the design of the 3D vision module described in this paper and suggests that such systems could have a role to play in real robotic SAR operations; improved perception will confer many benefits. For example, in an influential case study of aircraft accidents, Endsley[19] reported that 72% of situation awareness (SA) errors were due to failures of perception. Interpreting this finding in the context of the studies described in this paper leads to the conclusion that HMD use may improve levels of operator SA when compared to on-screen displays when viewing 3D environments.

Experiment 4 further explored this issue in the context of a simulated SAR task. Once again, the standard screen-based condition was shown to be significantly inferior (in terms of task completion time) to the optimal performance achieved with camera feeds shown in an HMD. However, the results also suggested that using input from head motions (as sensed by the HMD) to control the movement of a robot is a potentially confusing interaction style. Specifically, although subjects performed tasks faster when they controlled the camera viewpoint independently from the robot's motion, using head movements to adjust the orientation of the robot in the world resulted in the slowest task completion times in the study. This latter finding was unexpected, and appears counter intuitive. One plausible explanation is that, perceptually, when operators rotate their heads, their torso remains fixed and correspondingly their egocentric frame of reference is unchanged. Head rotation is qualitatively different from rotating the entire body. In particular, due to experience with activities such as walking and driving, operators are likely to expect that head rotations will not alter the gross body orientation and the direction of "forward." However, as the RAPOSA camera is fixed in terms of yaw, such

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

rotations affect the robots entire body, altering the response of the system to commands requesting it move forward. This kinematic disparity breaks the sense of immersion that the HMD control aims to induce. One reason underlying this breakdown may relate to the fundamental distinction between epistemic and pragmatic action,[33] or activities intended to explore a problem space (or strictly speaking, to alter or simplify the nature of a cognitive problem) and those intended to make concrete steps to solve it. By confounding the primarily exploratory action of looking with the primarily pragmatic action of moving, the task of controlling the robot may have been made inadvertently more challenging. A potential solution to this issue would be to deploy a fully independent pan and tilt camera that can support looking around independently from gross movements of the robot, e.g., refs. [27] and [54]. However, this solution is not without its problems, particularly in SAR scenarios. Specifically, as the RAPOSA is expected to navigate in hazardous environments featuring uneven terrain composed of large obstacles and drops, a camera system capable of showing view only in the direction of motion seems highly advantageous. Recently, we implemented such a system on an improved version of RAPOSA, in which a stereo camera pair is mounted on a pan and tilt motorized base.[58] The evaluation of the system is ongoing and preliminary results indicate that decoupling camera movement from robot motion addresses the above-mentioned kinematic disparity.[53]

In summary, the experiments described in this paper effectively illustrate that novel display and control paradigms for vehicular teleoperation can yield performance improvements compared to traditional screen and keyboard interfaces. However, it also suggests that such systems need to be introduced with care. This is particularly true when novel forms of control, rather than simply display (an approach more common in the literature[15]) are considered. Exploring these issues through further design and experimentation is a clear next step for future work on this topic.

7. Conclusion and Future Work

This paper has described a visualization/control system for field robots based on an HMD with integrated sensors to detect head motions. The implementation of two system modules to achieve these new functionalities was described. The first of these modules rectifies and corrects the video feeds from the robot's stereo cameras to enable seamless and informative stereoscopic display. The second mapped sensor input to movements of the robot such that "natural mappings"[48] were maintained. The impact of these changes to a human operator was assessed through a series of four experiments. These revealed significant improvements in basic 3D perception which translated to improved task completion times in a simulated SAR task. Overall these results are strongly positive for the inclusion HMD and motion-sensing technology in field robots. We argue that they provide evidence that HMDs will lead to increased levels of situation awareness in operators. The results are also highly consistent with previous work on stereoscopic displays[14] and the main practical contribution of this work is to show the validity of this approach in a simulated field robotics scenario.

However, the studies also revealed that novel control interfaces for vehicular teleoperation need be designed with care. In particular, experimental results indicated that using head motions to control the direction of motion of a robotic vehicle can be problematic and lead to impaired performance. Further exploring this issue through additional experimentation is a key next step for future work. We anticipate tradeoffs between intuitive methods for control and those that maximize an operator's situation awareness.

Other directions for future work aiming to develop the usefulness of HMDs for field robotics will be in the design of superimposed, augmented reality style information displays shown on stereoscopic video feeds.[61,64] Such enhancements will allow operators to maintain awareness of additional data reported from the robot's sensors while using immersive HMDs. Once again, key issues with this process will be designing information to adequately support the operator's task while maintaining his or her situation awareness and minimizing cognitive overload. These are significant, often contradictory challenges and a key topic for future study is in exploring the trade-offs between the importance of presenting particular data and the impact that presentation has on human performance. A final avenue for future work is to perform experiments using measures, such as SAGAT,[17] that are specifically designed to capture levels of situation awareness.

In summary, this paper has presented research which ties technological development with empirical studies on human performance in the domain of novel control and display interfaces for vehicular teleoperation. Ultimately, we believe that work of this nature will make important steps to address

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press

the current limitations preventing effective human control of robots. In the long run, it may be able to reduce the current human-to-robot ratio of 2:1,[7] addressing an issue previously viewed as a fundamental bottleneck.

References

1. S. Arca, E. Casiraghi and G. Lombardi, "Corner Localization in Chessboards for Camera Calibration," *Proceedings of International Conference on Multimedia, Image Processing and Computer Vision*, Madrid, Spain (Mar. 30–Apr. 1, 2005).
2. C. Becker-Asano, S. Gustorff, K. O. Arras, K. Ogawa, S. Nishio, H. Ishiguro and B. Nebel, "Robot Embodiment, Operator Modality, and Social Interaction in Tele-Existence: A Project Outline," *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Tokyo, Japan (Mar. 3–6, 2013) pp. 79–80.
3. A. Birk and S. Carpin, "Rescue robotics—A crucial milestone on the road to autonomous systems," *Adv. Robot. J.* **20**(5), 595–605 (2006).
4. W. Birkfellner, M. Figl, K. Huber, F. Watzinger, F. Wanschitz, J. Hummel, R. Hanel, W. Greimel, P. Homolka, R. Ewers and H. Bergmann, "A head-mounted operating binocular for augmented reality visualization in medicine—Design and initial evaluation," *IEEE Trans. Med. Imaging* **21**(8), 991–997 (2002).
5. M. Bleyer and M. Gelautz, "Video-Based 3D Reconstruction of Moving Scenes Using Multiple Stationary Cameras," *Proceedings of the 27th Workshop of the AAPR* (2003) pp. 181–187.
6. G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library* (O'Reilly, Sebastopol, CA, USA, 2008).
7. J. Burke and R. Murphy, "Human-Robot Interaction in Usar Technical Search: Two Heads are Better than One," *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, Kurashiki, Okayama, Japan (Sep. 20–22, 2004) pp. 307–312.
8. J. L. Burke, R. R. Murphy, M. D. Coovert and D. L. Riddle, "Moonlight in Miami: A field study of human robot interaction in the context of an urban search and rescue disaster response training exercise," *Hum. Comput. Interact.* **19**, 85–116 (2004).
9. O. Cakmakci and J. Rolland, "Head-worn displays: A review," *J. Disp. Technol.* **2**(3), 199–216 (2006).
10. J. Casper and R. Murphy, "Workflow Study on Human-Robot Interaction in USAR," *Proceedings of IEEE International Conference on Robotics and Automation (ICRA-02)*, vol. 2, Washington, DC, USA (May 11–15, 2002) pp. 1997–2003.
11. J. Casper and R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center," *IEEE Trans. Syst. Man Cybern.* **33**(3), 367–385 (2003).
12. C.-T. Chang, S. Takahashi and J. Tanaka, "A remote communication system to provide "Out together feeling" ". *J. Inf. Process.* **22**(1) (2014).
13. R. Chiou, Y. (James) Kwon, T.-L. (Bill) Tseng, R. Kizirian and Y.-T. Yang, "Enhancement of online robotics learning using real-time 3D visualization technology," *J. Syst. Cybern. Inform.* **8**(3), 46–51 (2010).
14. S. Das and N. Ahuja, "Performance analysis of stereo, vergence, and focus as depth cues for active vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(12), 1213–1219 (1995).
15. D. Drascic, "Skill Acquisition and Task Performance in Teleoperation Using Monoscopic and Stereoscopic Video Remote Viewing," *Proceedings of the 35th Annual Meeting of Human Factors Society*, San Francisco, USA (Sep. 2–6, 1991) pp. 1367–1371.
16. J. Drury, J. Scholtz and H. Yanco, "Awareness in Human-Robot Interactions," *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, Washington, DC, USA (Oct. 5–8, 2003) pp. 912–918.
17. M. Endsley, "Measurement of situation awareness in dynamic systems," *Hum. Factors* **37**(1), 65–84 (1995a).
18. M. R. Endsley, "Situation Awareness: Analysis and Measurement," *Proceedings of the 32nd Annual Meeting of Human Factors Society*, Anaheim, California, USA (Oct. 24–28, 1988).
19. M. R. Endsley, "A Taxonomy of Situation Awareness Errors," **In:** Human Factors in Aviation Operations; *Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP)*, Dublin, Ireland (1995b) pp. 287–292.
20. J. Ferraz and R. Ventura, "Robust Autonomous Stair Climbing by a Tracked Robot Using Accelerometer Sensors," **In**: *Mobile Robotics: Solutions and Challenges (CLAWAR-2009)* (O. Tosun, H. L. Akin, M. O. Tokhi and G. S. Virk, eds.) (World Scientific, 2009) pp. 415–422.
21. F. Ferreira and R. Ventura, "Autonomous Docking of a Tracked Wheels Robot to its Tether Cable Using a Vision-Based Algorithm," **In**: Workshop on Robotics for Disaster Response; *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA-09)*, Istanbul, Turkey (Sep. 9–11, 2009) pp. 415–422.
22. M. Fiala, "Pano-Presence for Teleoperation," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Edmonton, Canada (Aug. 2–6, 2005) pp. 3798–3802.
23. A. Fusiello, E. Trucco and A. Verri, "A Compact Algorithm for Rectification of Stereo Pairs," *Mach. Vis. Appl.* **12**, 16–22 (2000).

24. A. Gage, R. Murphy, E. Rasmussen and B. Minten, "Shadowbowl 2003 [simulated mass-casualty exercise]," *IEEE Robot. Autom. Mag.* **11**(3), 62–69 (2004).
25. M. A. Goodrich and A. C. Schultz, "Human–robot interaction: A survey," *Found. Trends Hum.–Comput. Interact.* **1**(3), 203–275 (2007).
26. A. Halme, J. Suomela and M. Savela, "Applying telepresence and augmented reality to teleoperate field robots," *Robot. Auton. Syst.* **26**(2–3), 117–125 (1999).
27. S. Hughes, J. Manojlovich, M. Lewis and J. Gennari, "Camera Control and Decoupled Motion for Teleoperation," *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics*, Washington, DC, USA (Oct. 5–8, 2003) pp. 5–8.
28. J. Hwang, S. Lee, S. C. Ahn and H. gon Kim, "Augmented Robot Agent: Enhancing Co-presence of the Remote Participant," *Proceedings of the 7^{itth} IEEE/ACM International Symposium on Mixed and Augmented Reality*, Cambridge, UK (Sep. 15–18, 2008) pp. 161–162.
29. A. Jacoff, E. Messina, B. A. Weiss, S. Tadokoro and Y. Nakagawa, "Test Arenas and Performance Metrics for Urban Search and Rescue Robots," *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, Las Vegas, USA (Oct. 27–31, 2003) pp. 3396–3403.
30. J. A. Jones, L. C. Dukes and M. Bolas, "Automated Calibration of Display Characteristics (ACDC) for Head-Mounted Displays and Arbitrary Surfaces," *Proceedings of the IEEE Virtual Reality Conference*, Minneapolis, USA (Mar. 29–Apr. 2, 2014) pp. 85–86.
31. F. Kellner, B. Bolte, G. Bruder, U. Rautenberg, F. Steinicke, M. Lappe and R. Koch, "Geometric calibration of head-mounted displays and its effects on distance estimation," *IEEE Trans. Vis. Comput. Graphics* **18**(4), 589–596 (2012).
32. W. Kim, A. Liu, K. Matsunaga and L. Stark, "A Helmet Mounted Display for Telerobotics," **In**: Compcon Spring '88. *Proceedings of the 33^{itrd} IEEE Computer Society International Conference*, San Francisco, USA (Feb. 29–Mar. 3, 1988) pp. 543–547.
33. D. Kirsh and P. Maglio, "On distinguishing epistemic from pragmatic action," *Cogn. Sci.* **18**(4), 513–549 (1994).
34. H. Kitano and S. Tadokoro, "RoboCup rescue: A grand challenge for multiagent and intelligent systems," *AI Mag.* **22**(1), 39–52 (2001).
35. D. Labonte, P. Boissy and F. Michaud, "Comparative analysis of 3-D robot teleoperation interfaces with novice users," *IEEE Trans. Syst. Man Cybern.* **40**(5), 1331–1342 (2010).
36. L. S. Lee, D. L. Carr-Locke, R. Ookubo and J. R. Saltzman, "Randomized trial of a video headset vs. a conventional video monitor during colonoscopy," *Gastrointest Endoscopy* **61**(2), 301–306 (2005).
37. M. Lichtenstern, M. Angermann, M. Frassl, G. Berthold, B. J. Julian and D. Rus, "Pose and Paste—An Intuitive Interface for Remote Navigation of a Multi-Robot System," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan (Nov. 3–8, 2013) pp. 1632–1639.
38. A. Liu, G. Tharp, L. French, S. Lai and L. Stark, "Some of what one needs to know about using head-mounted displays to improve teleoperator performance," *IEEE Trans. Robot. Autom.* **9**(5), 638–648 (1993).
39. C. Loop and Z. Zhang, "Computing Rectifying Homographies for Stereo Vision," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, Fort Collins, Colorado, USA (Jun. 23–25, 1999) pp. 131.
40. S. K. Maithel, L. Villegas, N. Stylopoulos, S. Dawson and D. B. Jones, "The benefits of stereoscopic vision in robotic-assisted performance on bench models," *Surgical Endoscopy* **18**(4), 611–616 (2004).
41. S. K. Maithel, L. Villegas, N. Stylopoulos, S. Dawson and D. B. Jones , "Simulated laparoscopy using a head-mounted display vs. traditional video monitor," *Surgical Endoscopy* **19**(3), 406–411 (2005).
42. C. Marques, J. Cristovão, P. Alvito, P. Lima, J. Frazão, M. I. Ribeiro and R. Ventura, "A search and rescue robot with tele-operated tether docking system," *Ind. Robot* **34**(4), 332–338 (2007).
43. H. Martins and R. Ventura, "Immersive 3-D Teleoperation of a Search and Rescue Robot Using a Head-Mounted Display," *Proceedings of 14th IEEE International Conference on Emerging Technologies and Factory Automation*, Palma de Mallorca, Spain (Sep. 22–25, 2009) pp. 1–8.
44. J. S. McVeigh, M. Siegel and A. Jordan, "Algorithm for Automated Eye Strain Reduction in Real Stereoscopic Images and Sequences," *Human Vision and Electronic Imaging*, vol. 2657, San Jose, CA, USA (Jan. 29–Feb. 1, 1996) pp. 307–316.
45. S. J. Merhav and S. LifShitz, "Adaptive suppression of biodynamic interference in helmet-mounted displays and head teleoperation," *J. Guid. Control Dyn.* **14**(6), 1173–1180 (1991).
46. R. Murphy, "Human-robot interaction in rescue robotics," *IEEE Trans. Syst. Man Cybern., Part C: Applications and Reviews* **34**(2), 138–153 (2004).
47. R. Murphy, J. Casper, J. Hyams, M. Micire and B. Minten "Mobility and Sensing Demands in USAR," *Proceedings of the 26th Annual Conference of the IEEE Industrial Electronics Society*, vol. 1, Nagoya, Japan (Oct. 22–28, 2000) pp. 138–142.
48. D. A. Norman, *The Design of Everyday Things* (Basic Books, New York, USA, 2002).
49. F. Okura, Y. Ueda, T. Sato and N. Yokoya, "Teleoperation of Mobile Robots by Generating Augmented Free-Viewpoint Images," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan (Nov. 3–8, 2013) pp. 665–671.
50. F. Olivier, *Three-Dimensional Computer Vision—A Geometric Viewpoint* (MIT Press, 1996).
51. S. E. Palmer, *Vision Science: Photons to Phenomenology* (MIT Press, 1999).
52. Point Grey, *Triclops Software Development Kit (SDK), Version 3.1* (Point Grey Research Inc., 2003).

53. J. J. Reis and R. Ventura, "Immersive Robot Teleoperation Using an Hybrid Virtual and Real Stereo Camera Attitude Control," *Proceedings of RecPad 2012*, Coimbra, Portugal (Oct. 26, 2012) pp. 99–100.
54. D. Ryu, S. Kang, M. Kim and J.-B. Song, "Multi-Modal User Interface for Teleoperation of Robhaz-dt2 Field Robot System," *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, Sendai, Japan (Sep. 28–Oct. 2, 2004) pp. 168–173.
55. D. H. Sattinger and O. L. Weaver, *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics* (Springer-Verlag, Berlin, Germany, 1986).
56. J. Scholtz, J. Young, H. Yanco and J. Drury, "Evaluation of Human-Robot Interaction Awareness in Search and Rescue," *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, Barcelona, Spain (Apr. 18–22, 2005) pp. 2327–2332.
57. S. Tachi and K. Yasuda, "Evaluation Experiments of Tele-Existence Manipulation System," *Proceedings of the 3rd International Conference on Artificial Reality and Tele-Existence*, Tokyo, Japan (Jul. 6–7, 1993) pp. 17–26.
58. R. Ventura, "Two Faces of Human–Robot Interaction: Field and Service Robots," **In**: *New Trends on Medical and Service Robots: Challenges and Solutions*, vol. 20 of *MMS* (A. Rodic, P. Doina and H. Bleuler, eds.) (Springer, Berlin, Germany, 2014) pp. 177–192.
59. H. A. Yanco, J. L. Drury and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Hum.-Comput. Interact.* **19**(1), 117–149 (2004).
60. L. Zalud, "ARGOS – System for Heterogeneous Mobile Robot Teleoperation," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China (Oct. 9–15, 2006) pp. 211–216.
61. L. Zalud, "Augmented Reality User Interface for Reconnaissance Robotic Missions," *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN-2007)*, Jeju, South Korea (Aug. 26–29, 2007) pp. 974–979.
62. L. Zalud and P. Kocmanova, "Fusion of Thermal Imaging and CCD Camera-Based Data for Stereovision Visual Telepresence," *Proceedings of IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Linkoping, Sweden (Oct. 21–26, 2013) pp. 1–6.
63. L. Zalud, L. Kopecny and F. Burian, "Orpheus Reconnaissance Robots," *Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics*, Sendai, Japan (Oct. 21–24, 2008) pp. 31–34.
64. L. Zalud, L. Kopecny and T. Neuzil, "3D Proximity Scanner Integration to Rescue Robotic System," *Proceedings of the 4th WSEAS International Conference on Signal Processing, Robotics and Automation*, Salzburg, Austria (Feb. 13–15, 2005).

https://doi.org/10.1017/S026357471400126X Published online by Cambridge University Press