The 23rd IEEE International Symposium on Robot and
Human Interactive Communication
August 25-29, 2014. Edinburgh, Scotland, UK,

Be the Robot: Human Embodiment in Tele-Operation Driving Tasks

Luís Almeida[1,3] Bruno Patrão[1] Paulo Menezes[1,2] Jorge Dias[1,4]

Abstract— **This paper proposes a new interaction mechanism for tele-operating a mobile robot. The approach explores the notion of telepresence and physical embodiment to create what may be called tele-embodiment. Its principle is that the operator will see himself at the remote site and this will enable him/her to better operate the robot.**

Four interaction styles were experimentally compared, from the traditional joystick approaches to more innovative based on natural body posture intentions. The environment perception is provided by the visual feedback, according to head pose behaviour. The results show that the gesture and body based methods improves the user dexterity performing this kind of task. Moreover, the present study suggests that, when a person is focused on the task, achieving the ownership illusion towards remote body, there are autonomic responses that correspond to what would be expected in events that take place in reality (like avoiding collisions).

I. Introduction

The evolutions of communications and robotics have created the perfect opportunities for new tele-operated robots. These are expected to play an important role in maintenance or exploration tasks to be performed in remote or hazardous environments.

In these teleoperation tasks, the operator usually controls the movements of the remotely located robot by observing the evolution of its position on a map-like representation, or by using live video streams to perceive the remote environments. In the latter case, the robot has a camera that becomes the "eyes of the operator". It is common to have joystick-like interfaces to control both the robot motions and the camera orientation. Apart from being unnatural, this type of interaction styles can overload operator's handling capability and require extra attention to multiple controls, decreasing the main task performance [1][2][3][4].

This paper proposes a new strategy for the multitasking problem of having a human simultaneously driving a robot and controlling a remote camera to perceive the remote environment. To maximize the task performance and minimize the operator's physical and cognitive workload, we propose to exploit the operator induced sensation of being at the remote environment [5][6]. Moreover, the generation of remote physical embodiment feeling is explored where the user can perceive the robot's structure as his/her own body [7].

[1] Institute of Systems and Robotics - University of Coimbra, Coimbra, Portugal

[2] Department of Electrical and Computer Engineering - University of Coimbra, Coimbra, Portugal

[3] Instituto Politécnico de Tomar, Tomar, Portugal

[4] In a leave at Khalifa University of Science, Technology and Research (KUSTAR), Abu Dhabi, UAE.

{bpatrao, paulo, jorge}@isr.uc.pt, laa@ipt.pt

This new principle consists in mixing teleoperation, telepresence and physical embodiment, resulting in virtually transferring the operator to the remote robot. To better understand the intended principle we can recall James Cameron screenplay Avatar [8]. Here, the physical embodiment is well described on for following excerpt of the movie script: "Blinking, Jake slowly sits up on the gurney. He looks down at his AVATAR BODY, touching his chest with one hand." What is described in this excerpt is that our perception of ourselves can be fooled if synchronous actions and sensations are perceived, like in the case of "the rubber hand illusion" [9]. Our goal is close to this, in the sense that we want the operator to perceive the remote environment as if he/she is really there, and control the robot as if it is his/her own body.

In the remaining of the paper we present the proposed interaction mechanism, the developed interaction support software architecture and the experimental tasks for evaluating our proposed approach and compare it to traditional remote operation controls. Finally the results of the experiments are presented and discussed.

II. The human in the teleoperation loop

To better understand the possible influence of the various interaction mechanisms proposed on this paper, we start by proposing a simplified model of the teleoperation problem.

A. A simplified tele-operation model

Several authors have studied the teleoperation problem. Some propose models where the human operator is part of a control loop, sending commands though a delayed transmission channel [10], and receiving the manifestations of the robot motion through the same delayed channel. The relevance of such models comes from the fact that the transmission delays have important effects on the controllability of the overall system.

In our study we are concerned with the relationship between the human ability to control the remote robot and the interaction mechanisms in use. Concerning this, figure 1 presents a diagram that models the tele-operation process. This model is composed of an outer tele-operation control loop, that utilises an inner perception control loop.

Tele-operation Loop - As already stated, tele-operating a robot can be modelled as a control loop. In this model the human compares a given goal with the position of the robot in the remote environment. From the perceived difference, he/she develops an intention that is then translated to robot commands through some interaction mechanism. This is represented in figure 1, where block A represents

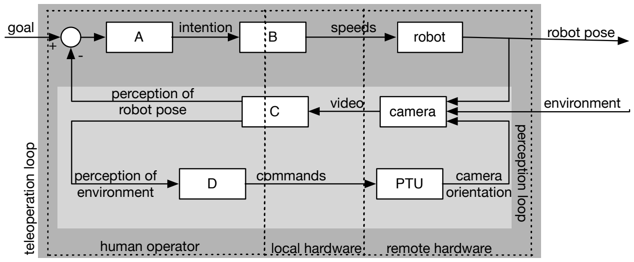

Fig. 1: Teleoperation and perception as control loops

the perception of the error and production of an intention. This intention is transformed into a command through block B which represents the human action over some interface that produces the adequate command. This is possible if the loop can be closes by having the user perceiving the pose of the robot on the remote environment.

Perception Loop - In this paper we consider teleoperation systems where the point of view of the operator is from inside the robot, i.e. using an embedded camera. Ideally the user should be able to perceive both the environment and the robot motions as if he was driving the robot from the inside. Nevertheless, due to technical limitations this is not the common case.

This perception process can also be described as a control loop. Here the human controls the robot camera orientation and uses the visual feedback to compensate his scanning actions required to pursue a goal. Controlling the camera, the user can scan the environment, track objects, etc.. Similarly to the previous case, the camera captures images that are viewed by the user. These results in the perception of the remote environment and the relative pose of the robot in it, and it is modelled by block C. Tracking, searching of objects, or snooping can be made by controlling the camera orientation. This can be done through the actuation of the user onto some interaction mechanisms represented by block D, whose commands are sent to the PTU.

III. FROM TELEOPERATION TO REMOTE EMBODIED OPERATION

Considering the presented model, we are now going to present the correspondences between different perception and control interaction mechanisms and the referred blocks A,B,C, and D.

A. The traditional teleoperation approach

In the traditional teleoperation systems, both blocks B and D represent joystick-based operation for respectively the robot motions and the PTU. In this case block C corresponds to viewing the remote environment images in a screen, and block A to the transformation of the perceived error into the intention to move the robot.

B. Viewpoint transfer using an HMD

Aiming at solving the problem of controlling the remote viewing camera our approach is to use a head mounted display (HMD) that the user can rotate left-right, up-down and using this control a pan-and-tilt unit (PTU) that supports the camera on the robot (block C and D). Using this, the camera is controlled in an egocentric way, as the user has the view centered on the camera and can control it by rotating his head in the desired direction.

This aims at being the first step for giving the user the sensation of being physically embodied on the remote robot, as ”the user will see what the robot can see”.

To accomplish this we use the HMD inertial measurement unit (IMU), which is composed of a three axial accelerometer, gyros and magnetometer. From this unit we estimate the user head pose, which are then transmitted to control the remote PTU that supports the camera. Using this system, whenever the user looks up-or-down, left-or-right, the remote PTU will replicate these movements giving the remote camera a direction equivalent to the user's gaze. By consequence, the user can ”look around”, ”track a moving object”, or just look down to see his own belly and feet, now replaced by the robot structure. This enables the user to have an egocentric perception of the remote surrounding environment, as if he/she was at the robot position and orientation. This will enable him/her to naturally explore the environment and navigate as if he/she was really there.

C. Deictic gesture-based control of the robot

Looking for natural strategies to tele-operate a mobile robot, other than the classic joystick, the present work was focused on the development of a gesture-based framework (block B). Aiming at creating a natural interaction mechanism, the initial idea was to use deictic gestures to control the robot movements.

Instead of using the motion transfer principle that is exploited by the X-BOX® games, where the movements of the user are transferred to some virtual character that appears on the screen, the idea here is give the user that he/she is controlling his/her own motion, as it he/she was at the remote location. One again, the idea is that of exploiting the physical embodiment sensation by making the user to perceive the robot structure as his own body. If accomplished the user will see him as being the robot, or inside the robot, and his pointing gestures will be used to control ”his own” motions on the remote environment.

On a typical mobile robot these gestures should be mapped to control the linear and angular velocities of the robot. Here we propose to recognize gestures that mean forward, back, turn right, turn left, which are similar to the ones used by a person to help another to navigate or maneuver a car.

For this we used a Kinect® sensor and the OpenNI™ [11] library to track the body posture. The produced output is an estimate of the coordinates of 17 points (joints) that define the 3D configuration of a skeleton-like model. On figure 2 on the left there is a representation of this model, and on right there is a frame of a tracking sequence where the skeleton is superimposed on the user silhouette. The deictic gestures can then be extracted from selected joints of this model. These gestures were defined from the basic set of pointing

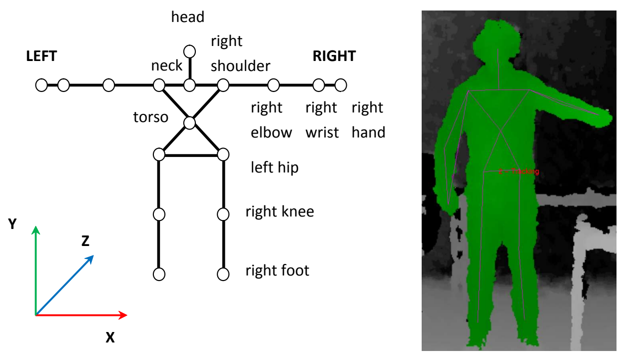

Fig. 2: Skeleton Model Joints

forward, backward, left or right and can be directly mapped into going forward, going backward, rotating left and rotating right. We can define that gesture along the sagital plane (see figure 3) of the operator (or very close to it) correspond to going straight forward or backward, and that gestures along the operator's coronal plane (or very close to it) will represent pure rotation, whose rotating direction depends to which side of the sagital plane it is done. Intermediate positions of the hand will result in combinations of forward/backward movements and rotations.

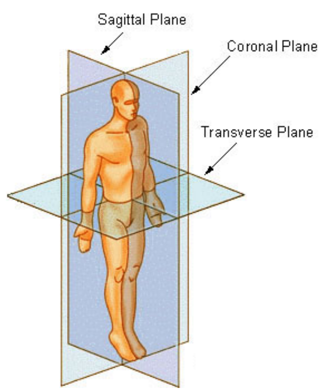

Fig. 3: Body planes representation (from http://en.wikipedia.org/wiki/Sagittal_plane).

Now we are able to define robot commands based on gestures, by extracting parameters from some joint combinations. The angular velocity Ω, can be related with the angle between the arm and the sagital plane, meaning that if the arm points to the front we get 0 degrees (no angular velocity), and if the arm points to the right we get 90 degrees (maximum positive angular velocity). Similarly, we relate the linear velocity V (forward/backward velocity) as the angle between the arm and the coronal plane. Therefore the maximum linear velocity is obtained when the arm points to the front and zero linear velocity when the arm is down (hand near the hip). To simplify the conversion of those gesture into linear and angular velocities to be applied to the robot, we consider a referential located at the intersection of the body planes referred on figure 3, where the X axle is at the intersection between the coronal and transverse planes, the Y axle the intersection between the sagital and transverse planes, and Z axle the intersection between the sagital and coronal planes. Now the linear and angular velocities are computed from the projection of the hand into the Z and X axles respectively, and then normalized by the arm length (eq. (1) and (2)) . There resultant values are the fraction of the maximum defined velocities.

$$V = \left(\frac{rightShoulder.Z - rightHand.Z}{\|rightShoulder - rightHand\|}\right) V_{max} \quad (1)$$

$$\Omega = \left(\frac{rightShoulder.X - rightHand.X}{\|rightShoulder - rightHand\|}\right) \Omega_{max} \quad (2)$$

D. Body intention-based control of the robot

We have defined another way of controlling the robot movements based in the body expressed intentions (block B). We define body intention as changes in the body posture, with respect to the rest position, that may be used to express intention to move. For example to walk forward or backward we displace one foot forward or backward respectively. To rotate left or right we rotate the shoulder in the corresponding direction (eq. (4)). To simplify this method of interacting we have marked a square on the floor with marks inside for both feet rest positions. Now it becomes simple to compute the linear velocity (eq. (3)) from the projection onto the Y axle of the displacement of the moving foot with respect to the rest position.

$$V = \left(\frac{movingFoot.Z - restingFoot.Z}{\|movingFoot - restingFoot\|}\right) V_{max} \quad (3)$$

$$\Omega = \max\left(1, \left(\frac{leftShoulder.Z - neck.Z}{leftShoulder.X - neck.X}\right)\right) \Omega_{max} \quad (4)$$

IV. Implementation

A setup composed of a mobile robot and a control station was built, both represented on figure 4.

The robot platform is a Scout II, onto which a structure was built to support a Directed Perception® PTU with a camera at about 1.60m height. Both the robot and the PTU are controlled from a module that receives the commands from the remote control station. The camera view is transmitted to the remote station via Skype®, what enables a good video quality with minimal bandwidth loss.

In what concerns the remote control station, its configuration varies according to the experiment. As will be described later, depending on the case, different combinations of joystick controller, screen monitor, RGB-D sensor and head-mounted display with IMU, will be used.

A. Architecture

The architecture shown on figure 5 was designed to enable the testing of different configurations by creating different mixes of the software modules developed for both the robot and the control station. In fact, as different interaction strategies were to be tested and evaluated, the active modules

Fig. 4: Hilario Robot (left) and Remote Control Station (right): (1) Scout, (2) Laptop, (3) PTU Camera, (4) Joystick, (5) RGB-D Camera, (6) Head-Mounted Display

that compose the setups vary from experiment to experiment, as described hereafter. To simplify the development the communication between modules at both sites is done via wireless TCP/IP based connections.

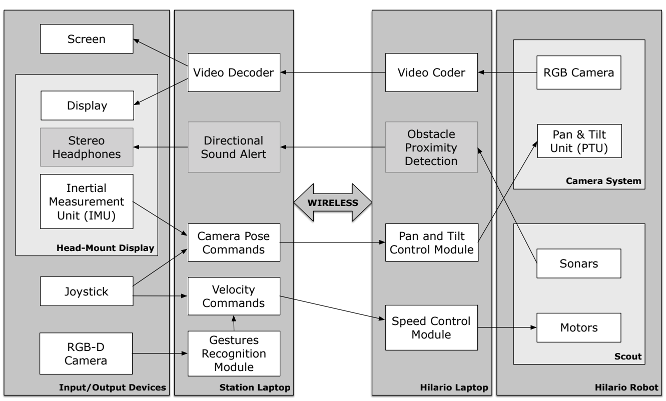

Fig. 5: Architecture Diagram

First experimental setup: a double-pad joystick is used to control both camera's PTU and robot motors. The operator observes a monitor that displays a video stream from the camera on top of the robot, a joystick to control the pan and tilt of that camera (right pad) and a second joystick to drive the robot (left pad). This is probably the most conventional tele-operation setup that can be found in the literature.

Second experimental setup: one joystick pad is used to control the robot motors, while the camera's PTU is controlled from the operator's head pose. The operator sees through the camera on the robot, using a Head-Mounted Display (HMD) and the camera's PTU mimics the operator's head movements of which are tracked by an Inertial Measurement Unit (IMU) on the HMD.

Third experimental setup: A Kinect device is used to track hand pointing gestures. The operator is standing up wearing the HMD to see through the camera on the robot.

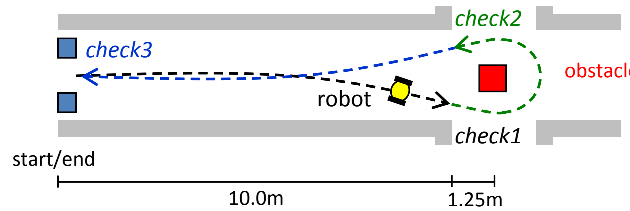

Fig. 6: Experimental task path, divided in 3 section (check 1, check 2 and check 3) with one obstacle (red box).

The IMU on the HMD is used for head tracking, similar to the previous setup. The robot motion control is performed giving explicit deictic gestures, for example, pointing to a direction make the robot move forward/backward and/or change its orientation.

Fourth experimental setup: A Kinect device is used to capture the user's body pose and from it extract natural body poses that are related with self motion intentions. As in the previous case, the HMD is used to give the user the remove view and control the PTU using IMU signals. In this case the operator can perform natural body and head movements to manipulate the robot and camera's PTU motions, respectively.

V. Experiments and Results

A. Methodology

The objective of the experiments is to understand the influence of different interaction styles on the teleoperation of a mobile robot. To assess how natural can a user interact and perceive the remote robot structure as his own body, four experiments were designed to test and compare the embodiment effect on task performance. The designed task consisted in tele-operate a robot through a path (figure 6), avoiding obstacles and finishing in the shortest period of time. The participants repeated this task using all the four experimental setups enabling comparison between standard and natural view point scan strategies, visual feedback, joystick, deictic gestures and body intention based robot control.

The four interaction setups developed are:

- **Setup 1:** Both the robot motions and camera orientation are controlled using two joysticks.
- **Setup 2:** The robot motion is controlled using a joystick, and the camera orientation is directly controlled using with the orientation of the HMD worn by the user.
- **Setup 3:** The robot motion is controlled using deictic gestures and he camera orientation is directly controlled using with the orientation of the HMD worn by the user.
- **Setup 4:** The robot motion is controlled using changes in the body posture and the camera orientation is directly controlled using with the orientation of the HMD worn by the user

In all cases the participants were standing controlling the robot and the camera as described. For the first case, the participants could view the remote camera video steam through a screen located just in front on the user. For the remaining cases the video was displayed on the HMD.

B. Procedure

Thirteen persons (2 female, 11 male), whose ages have a mean value of 26.46 and standard deviation of 5.03, participated on the experiment. These participants were informed about the purpose of this experiment, the procedures involved, that their cooperation was voluntary and that the personal data was going to be kept confidential. They had no prior knowledge about the experiment or the involved technologies. They were instructed that the main goal of this experiment was tele-operate the robot through a path, avoiding obstacles and finish in the shortest period of time. Then they tried to execute it using each of the above explained setups. The execution times were measured, and at the end they were asked to answer a short questionary.

C. Results

As stated above, in these experiments we tried to evaluate the performance changes and the subjective satisfaction of the users. The performance of the user is measured in terms of time spent for executing the task, for each of proposed interfaces. Instead of using a single time measure, the trajectory was divided in three sections (check 1, check 2, check 3). This enabled us to detect variations in performance between each of them, and to separate the initial adaptation and learning of the interaction mechanism from the rest of the experiment.

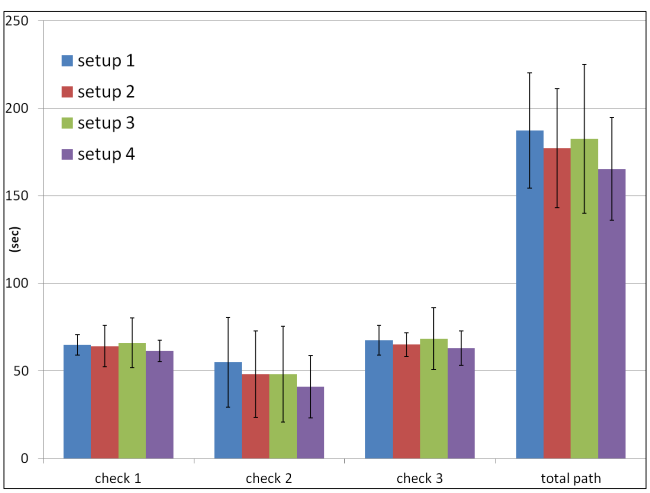

Fig. 7: Users mean time spent to perform the task on each experiment (seconds). A measurable comparison effect on task performance while using different teleoperation interaction styles designed to enhance embodiments sensations

Figure 7 depicts the performance mean times (sec) and standard deviation for the three trajectory segments of the four setups.

A subjective evaluation was developed, based on questionnaires that the users had to fill at the end of the experiment. The participant feedback questionnaires were based in 7 point Likert scale and comments. The questionnaire contents follows the IBM Computer Usability Satisfaction Questionnaire [12], in which users compared their experiences with different visual feedback models and robot control interfaces concerning subjective measures like consciousness, easiness, embodiment feeling, usability and perceived used time. The questions to be scored from 1 to 7 were:

- Q1: It was intuitive to use.
- Q2: It was easy to learn.
- Q3: I could get enough visual feedback.
- Q4: I felt embodied on the robot while performing the task.
- Q5: I completed the task quickly.

Figure 8 summarizes the obtained scores.

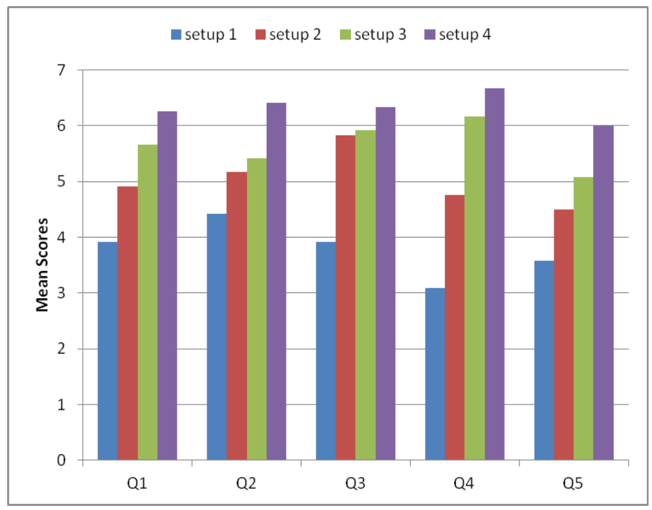

Fig. 8: Mean scores from user questionnaire feedback, scale : 1- Strong Disagree to 7-Strong Agree

D. Discussion

Objective measure results demonstrate that when the operators controls the camera point of view with his head pose (freeing their hands and minimizing the cognitive workload), obtaining a more natural visual feedback through an HMD, they improve significantly their task performance. The introduction of this visual strategy was done in setup 2 and maintained through setup 3 and setup 4. In Setup 1, while manipulating camera and robot with joysticks, operators took mean times of 65s (SD=5.9), 55s (SD=25.6), 67s (SD=8.5) to perform path sections 1, 2, 3 respectively.

The visual feedback control strategy, introduced in setup 2, enabled to outperform the setup 1 time values and obtain means of 64s (SD=11.8), 48s (SD=24.7), 65s (SD=6.7) to perform path sections 1, 2, 3 respectively). The evidence is more notorious in path section 2 (check2), where the operator had to scan the environment and plan a trajectory to turn around the obstacle. Due some visual feedback delays and limited field of view, part of path section 2 trajectory had to be mentally anticipated by the operators. Small collisions with the obstacle were registered, and the number of operators that collided with the obstacle were less in setup 2 than in setup 1 (4 in opposition to 6).

When questioned about gains in visual feedback (Q3), operators have scored high, setup 2, 3, and 4 (mean scores 5.83, 5.92 and 6.33 respectively) in opposition to setup 1 (mean score 3,92)

The introduction of natural deictic gestures based robot control (setup 3) presented gains in task performance when compared with setup 1, specially where higher skills were required (check 2 section)(setup 1 check 2: 55s (SD=25.6) and setup 3 check 2: 48s (SD=27.4)). Operators spent less time in setup 4 (setup 4 check 2: 41s (SD=17.7)). Setup 2 and setup 3 had equivalent times (setup 2 check 2 : 48s (SD=24.7) and setup 3 check 2: 48s (SD=27.3)).

Body intention-based robot control (setup 4) was the operators choice in all questions, confirmed by the time performance measures (means of 61s (SD=6.1), 41s (SD=17.1) and 63s (SD=9.7) to perform path sections 1, 2, 3 respectively). Notice that the standard deviation of setup 4, in section 2 check were smaller than in other equivalent section setup, meaning that operators were more regular performing this section.

All the operators felt comfortable operating the robots with their shoulders. In setup 4, practically there weren't operators colliding with the obstacle in check 2 section, as they were more familiar with path and had also a precise orientation control.

When questioned how easy is to use the control interfaces, operators scored high setup 3 and setup 4 (mean scores 5.6, 6.2 respectively), that is, the natural interaction styles. They felt easy to adapt to the interfaces of setup 4 (scoring 6.4). Once the controller of the visual feedback was identical in setup 2, 3, and 4 operators have scored with similar maximum values of 5.8, 5.9, and 6.3 respectively. The embodiment feeling high scores were associated to natural interaction styles implemented on setup 3 and setup 4. Without knowing exactly the chronometer time when filling the questionnaires, the operator felt that they have performed setup 3 and setup 4 faster than others experiments (mean scores of 5 and 6 respectively), even that objective measure point setup 2 as the fastest. An interpretation of this result might indicate that operators were immerse in their task, and that did not had a correct time perception.

Operators complained about limited field of view when looking down, as they could not see the ground and the robot base simultaneous, making difficult to turnaround the obstacle. Another comment was related with the lack of wider field of view and the knowledge of robot size. Corridors with white walls, without reference cues, were pointed as complexity factors. Commands execution delay was also referred as a relevant factor.

From the experiment external video records, it was possible to observe operator reactions, to critical events, similar to ones as they would behave in the remote environment. For instance, when the robot was about to collide, the user felt the moment and tend to get away from the control station.

To sum up, the use of natural teleoperation styles presents higher embodiment feeling, being demonstrated either by the analysis of self-questioners and by the task time performance. By using operator's head pose to control his visual feedback and body motion to control the robot movement, operator seems to agree that the physical and cognitive workload decrease from setup 1 to setup 4.

VI. Conclusions and Future Directions

This study results show that introducing new interaction and view control mechanisms that improve the physical embodiment sensation in tele-operation tasks can improve both the user satisfaction and performance. It is clear that viewing the remote environment as one being controlling the robot from inside of it, reduces the required mental workload to compute, the otherwise needed, view and control transformations.

In future experiments we intend to explore the introduction of auditory feedback as represented in figure 5, to map the proximity of obstacles. Another point that will be verified, is the suggested changes towards a wider field of view, by using wide angle lenses and HMD.

References

[1] J. Chen, E. Haas, and M. Barnes, "Human performance issues and user interface design for teleoperated robots," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1231–1245, Nov 2007.

[2] D. Zhu, T. Gedeon, and K. Taylor, "Exploring camera viewpoint control models for a multi-tasking setting in teleoperation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 53–62.

[3] L. Almeida, P. Menezes, L. Seneviratne, and J. Dias, "Incremental 3d body reconstruction framework for robotic telepresence applications," in *Robo 2011: The 2nd IASTED International Conference on Robotics*, Pittsburgh, USA, November 2011.

[4] S. Islam, X. Liu, A. Saddik, L. Seneviratne, and J. Dias, "Control schemes for passive teleoperation systems over wide area communication networks with time varying delay," *International Journal of Automation and Computing*, vol. 11, no. 1, pp. 100–108, 2014.

[5] M. Slater, B. Lotto, M. M. Arnold, and M. V. Sanchez-Vives, "How we experience immersive virtual environments: the concept of presence and its measurement," *Anuario de Psicologia*, vol. 40, pp. 193–210, 2009.

[6] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, "First Person Experience of Body Transfer in Virtual Reality," *PLoS ONE*, vol. 5, no. 5, pp. e10 564+, May 2010.

[7] P. Zahorik and R. L. Jenison, "Presence as being-in-the-world," *Presence: Teleoper. Virtual Environ.*, vol. 7, no. 1, pp. 78–89, Feb. 1998.

[8] J. Cameron. Avatar. [Online]. Available: http://www.imsdb.com/scripts/Avatar.html

[9] M. Costantini and P. Haggard, "The rubber hand illusion: Sensitivity and reference frame for body ownership," *Consciousness and Cognition*, vol. 16, no. 2, pp. 229 – 240, 2007.

[10] S. Islam, X. Liu, A. Saddik, L. Seneviratne, and J. Dias, "Control schemes for passive teleoperation systems over wide area communication networks with time varying delay," *International Journal of Automation and Computing*, vol. 11, no. 1, pp. 100–108, 2014.

[11] Openni. [Online]. Available: http://www.openni.org/

[12] J. R. Lewis, "Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," *Int. J. Hum.-Comput. Interact.*, vol. 7, no. 1, pp. 57–78, Jan. 1995.