Immersed Remotely: Evaluating the Use of Head Mounted Devices for Remote Collaboration in Robotic Telepresence

Sven Kratz[1] and Fred Rabelo Ferriera[2]

***Abstract*— Mobile Telepresence Robots (MTR) are an emerging technology that extend the functionality of telepresence systems by adding mobility. MTRs nowadays, however, rely on stationary imaging systems such as a single narrow-view camera for vision, which can lead to higher operator error rates due to view-related deficiencies in situational awareness. We therefore developed an improved imaging and viewing platform that allows immersive telepresence using a Head Mounted Device (HMD) with head-tracked mono and stereoscopic video. Using a remote collaboration task to ground our research, we examine the effectiveness head-tracked HMD systems in comparison to a baseline monitor-based system. We performed a user study where participants were divided into three groups: fixed camera monitor-based baseline condition (without HMD), HMD with head-tracked 2D camera and HMD with head-tracked stereo camera. Results showed the use of HMD reduces task error rates and improves perceived collaborative success and quality of view, compared to the baseline condition. No major difference was found, however, between stereo and 2D camera conditions for participants wearing an HMD.**

I. Introduction

Mobile telepresence robots (MTR) are devices that extend the functionality of telepresence systems by adding mobility that can be controlled by remote operators (or *teleoperators*). The advantage of mobility is that it adds flexibility in the location of the telepresence system, which does not need to reside at a fixed location anymore, and more personal styles of communication are possible [1] due to the teleoperator embodying the MTR platform. MTRs have long been present in research [2], and have recently become increasingly popular in consumer or business applications [3], domestic care [4], [5], education [6], [7], or healthcare [8], [9].

The design of most contemporary MTR simply *transplants* the existing stationary (i.e., desktop-based) designs onto a moving robotic base, thus bringing with them many of the affordances of stationary systems: a typical configuration is mounting a video tele conferencing screen on a mobile robotic base [10], [11].

As demonstrated in previous work, using MTRs can augment the remote users' feeling of presence in local spaces [12] and improve the sense of engagement in collaborative work scenarios [3], [13]. However, Johnson et al. [14] argue that simply replicating existing paradigms from fixed telepresence setups is not enough to be beneficial for all collaborative tasks. They state that collaborative tasks, in particular, require improved user interfaces for teleoperation.

[1]Sven Kratz is with FX Palo Alto Laboratory, Palo Alto, CA, USA kratz@fxpal.com

[2]Fred Rabelo Ferreira is with Columbia University, New York, NY, USA fredinhu@gmail.com

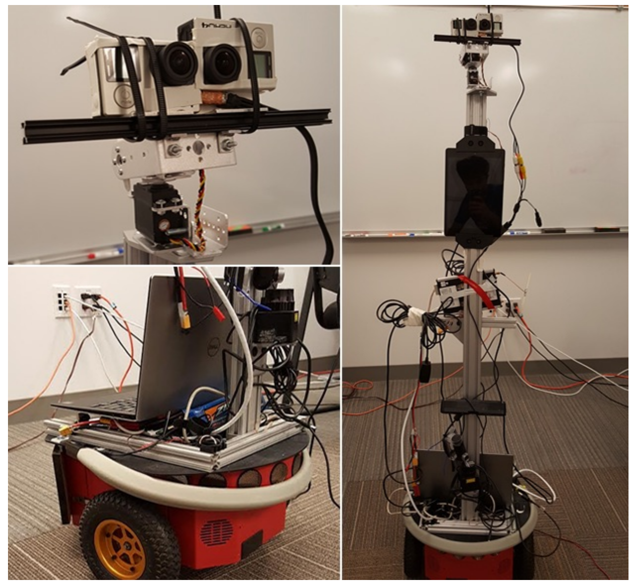

Fig. 1. Mobile telepresence robot system used in our study. *Upper left*: stereo camera pair and pan/tilt rig. *Lower left*: robot base and control laptop. *Right*: Strut assembly with mounted tablet and stereo pan/tilt rig.

In essence, they argue that teleoperation interfaces need to go beyond just showing a fixed, 2D view for the remote operator. Furthermore, Johnson et al. state that, although new video technologies (e.g., wide-angle lenses, pan-tilt setups, etc.) have been evaluated in terms of "operator efficiency and awareness", there have not been many in-depth studies of novel teleoperation interfaces in collaborative work settings.

A further issue which confronts users of MTRs is that very few of the available MTR platforms incorporate autonomous navigation capabilities. Thus, the remote operator's task of manually navigating the robot through a space (e.g., to the next meeting) can be one of the major problems in the user experience of these devices.

What makes manual teleoperation difficult is that on most MTRs the teleoperation UI shows one or more fixed camera views (e.g., forward and forward-down on the VGO [11]). In this type of configuration, the orientation of the camera view follows that of the robot base in a fixed manner. We believe that using fixed camera views can be problematic both for navigating MTRs as the fixed viewpoint allows the user only a limited field of view, and for orientation of the camera view as the whole robot needs to be reoriented to change the view direction. Furthermore, with 2D camera views it is difficult to effectively judge distances in the image, due to a lack of depth perception. The result is that the remote operator may have a reduced awareness of the position of the robot in a space, the position of obstacles relative to the robot or the best trajectory to take to navigate through a space.

To address this problem, we propose increasing the immersiveness of MTR navigation by using a head-mounted display (HMD) with head tracking. Previous work by Draper et al. [15] supports this approach. They argue that immersion is a predictor of telepresence success.

Thus, in a prototype we developed for this paper, head tracking is used to control a robot-mounted pan/tilt servo system on which a stereo camera pair is mounted to provide a stereo video feed to a user wearing an HMD. We implemented an AR overlay for the HMD that shows task-relevant information. We did this to compensate for the fact that although immersive HMDs such as Oculus Rift [16] improve the awareness of the *remote* environment, they shut the user off from the *local* environment, making context switches between the local and remote environments (e.g., to look at local paper instructions) unwieldy.

Overall, this work contributes to human-robot interaction and the design of mobile telepresence systems by examining the usability of head-tracked, head-mounted displays in the context of remote collaboration tasks. Our experimental results show that such HMD-based systems can have some key task-relevant advantages over a baseline, desktop-based teleoperation user interface. We furthermore highlight how several key UI design features for the HMD system (wider camera viewing angle, transparent overlay of task-relevant reference material, and a visual robot pose indicator) help significantly improve task performance and user ratings vs. the desktop-based baseline system.

We now proceed with a discussion of related work. We then establish our hypotheses and present the results from a user study with our system. Finally, we provide a discussion of our findings and their consequences for future research and design work.

II. Related work

Rae et al. describe several design dimensions for telepresence [17]. Within their framework, our work contributes mainly to better understanding of the aspects of *vision* and *communication* within telepresence systems: we propose the use of an immersive video system for mobile robot teleoperation and study how the use of this system affects a collaborative task. In the following, we review prior work and discuss how the present contribution builds upon and extends past efforts.

A. Robot Teleoperation

A number of past works have focused on mobile robot teleoperation interfaces, i.e., looking at techniques that allow operators to move their robots. Fong et al. provide a broad overview of teleoperation interfaces [18], [19], covering various graphical operation UIs and gesture, haptic and tablet-based interfaces for robot navigation. In addition, voice-based [20] and even indirect, fiducial-based [21] operation interfaces have been suggested. Although substantial progress has been made in the field of autonomous robot navigation [22], a study conducted by Takayama et al. [23] showed that although autonomous navigation assistance is useful to avoid local navigation errors (e.g., avoid obstacles), it can also increase the total time required for navigation tasks. We therefore believe that improving user interfaces for manual robot teleoperation is still an important and valuable field of research.

In the present work, we deliberately keep to a baseline *input* technique for robot motion commands, i.e., using keyboard arrow keys for motion commands. Because some of the most common pitfalls of robot operations are "*loss of situation awareness, poor attitude and depth judgment, inadequate perception of the remote environment, and failure to detect obstacles*" [19], we focus in this paper on drastically improving the operator's view system in order to increase overall situation awareness. With a better view system, the number of errors and the overall workload for the operator should be decreased.

B. Use of Pan-Tilt Camera Systems and Stereo Vision

A number of fixed-location telepresence systems used a pan/tilt configuration to change the remote camera view angle. Venolia et al. used a pan-tilt camera when studying telepresence communication in distributed work scenarios [3]. Biehl et al. studied the affordances of a telepresence system, that, besides the camera view, can pan/tilt its entire display [24]. They analyze the perception of telepresence of the remote operator and participants of a meeting who are collocated with the aforementioned embodied telepresence system. Polly [25] is a semi-mobile system that can be used either in a stationary configuration or worn by a local *guide* to gain mobility. Field evaluations of Polly indicated that the ability to actively control the remote viewpoint increased the sense of presence at the remote location.

Several previous projects have proposed using a user-controlled pan/tilt camera on mobile robots, e.g., Fiala et al. [26] or Zalud et al. [27]. These, however, did not use stereo imagery or head-mounted displays. Martins et al. [28] conducted a user study of teleoperation with a robot that had a stereo camera pair which was mounted rigidly on the robot, thus, in contrast to our system, the camera orientation was coupled to the robot's pose, which could cause latency problems when executing head-tracking motions. Notably, the usability study they conducted with their system yielded positive results for a search-and-rescue task. Lawson et al. [29] proposed a stereo pan-tilt robot with head tracking for remote environment exploration, very similar to the setup used in our paper. Pittman et al. [30] studied immersive control of aerial robotic vehicles using an HMD as an alternative viewing technique in order to improve the user's immersion level. Finally, the *Dora Platform* [31] is a commercial 6 DOF head tracking telepresence robot, which provides higher-fidelity head tracking than the solution used in this paper.

C. Study of Collaborative Tasks for Telepresence

Recently, Johnson et al. [14] studied the effects of different camera viewing angles for collaborative telepresence tasks. Their results favorize wider viewing angles, although extremely wide viewing angles did increase usage difficulty in some cases. In the present work, we study immersive head-tracked video using an HMD as a further method of improving the view quality for telepresence, and how this affects remote collaboration tasks.

Rae et al. [12] studied the effects of mobility on the outcome of collaborative telepresence tasks. They found that mobility contributes to an understanding of the remote space, due to the ability of generating multiple camera angles. However, they did also observe disorientation effects due to the fixed camera on their telepresence robots. The results of our studies suggest that using a head-tracked display has advantages over a fixed 2D camera for collaborative telepresence tasks and reduces the perceived difficulty of the task, which might also be due to less viewpoint-related disorientation experienced by the users.

III. User Study

The goal of this user study is to compare the use of an HMD with head-tracked stereo and mono camera systems, and a fixed monitor-based (without HMD) baseline condition for a collaborative assembly task using a mobile telepresence robotic platform.

A. Method and Choice of Conditions

We conducted a between-participants controlled laboratory experiment with three groups of three different conditions:

C1: Using a fixed 2D camera view displayed on a monitor.
C2: Using an HMD with a *Monocular* Tead-Tracked pan/tilt camera View (MHTV) system.
C3: Using an HMD with a *Stereo* Head-Tracked pan/tilt camera View (SHTV) system.

Condition **C1** (without the use of any HMD), was intended as a baseline condition, to simulate commercially available robots in the market today, such as [10], [11]. The two other conditions, with the use of an Oculus Rift DK2 [16], were intended to evaluate how this technology can improve upon robotic telecollaboration by offering new capabilities. We designed our experiment to study the effects of view type (**C1**, **C2** and **C3**) in a collaborative telepresence task. Our study is inspired by the results of a previous preliminary study [32] that suggested a higher SUS rating for the use of a standard 2D camera compared to head-tracked stereo display for the task of driving the robot. The main focus of this previous work, however, was a pure driving task. In the studies presented in the current paper, we want to analyze the full user experience: driving the robot to a remote location, and, most importantly, completing a collaborative task with a remote person.

B. Participants and Confederate

A total of 21 participants (15 male, 6 female) took part in our study, where each group condition was composed by 5 males and 2 females. Participants were aged 18–58 (M = 29.56, SD = 9.75). Participants in each condition (i.e., the telepresence users) were asked to operate the MTR, drive it to a particular remote location and then collaborate with a confederate on a furniture assembly task. The same confederate, a 24-year-old male student, participated in all trials, from a different room in the same building complex. During the task, the confederate did not initiate any action, but followed orders and engaged in discussion, fully cooperating with current participant (remote user) while trying to maintain consistency across participants by, mostly, just answering questions (in the best possible way) and asking for clarifications if there were any doubts in what actions to take and how.

C. Collaborative Assembly Task

Participants were assigned a task in which they had to remotely control a telepresence robot to help a confederate (at a remote location) assemble a piece of IKEA furniture. Our choice of remote collaboration task was grounded by similar task designs utilized by recent related works [12], [14]. Specifically, we selected a furniture assembly task because it requires a broad skill set of cognitive abilities, requiring the telepresence user to be fully engaged in the remote place, paying attention to the details of the object and its parts while interacting with the confederate and interpreting the instructions provided by the furniture assembly manual. For completion of the task, the telepresence user is required to correlate a relatively large amount of information between the local and remote contexts. For instance, participants in our study needed to identify objects and parts, pictured in the assembly manual and match those within the visual feed provided by the MTR system.

The selected furniture was the VILDAPEL Plant Stand [33], shown in Figure 2, which also had components of different shapes and sizes. We chose this particular piece because it seemed to balance our requirement of having a relatively short assembly time, but still have a nontrivial amount of parts in different shapes and sizes that could be placed in many different positions. For instance, the orientation of the crossbars on each leg, were easy to be be assembled in a mis-oriented way if the user was not paying full attention to the manual and constantly checking the confederate's assembly process. Therefore, participants were forced to genuinely try to identify the orientation of each object in order for it to be mounted correctly.

D. System design

Our mobile telepresence robotic system for the user study was based on a modified version of the Pioneer P3-DX Mobile Robot [34]. For hardware mounting, we affixed an aluminum truss structure of approximately 120 cm of height onto the robot's base, as shown in Figure 1. This height was consistent for all participants in all experiments. At the top

Fig. 2. VILDAPEL plant stand used for the collaborative assembly task in the study. Disassembled (*left*) and assembled (*right*).

of the robot, we mounted a pan/tilt servo system in which a stereo camera pair is fixed to provide a stereo video feed to the user wearing an HMD.

The Oculus Rift DK2 provides an overall resolution of 1080p (960 x 1080 per eye), 60 Hz refresh rate, and a wide field of view. It has full 6 degree of freedom rotational and positional tracking. We used the Oculus Rift SDK [35] to obtain real-time rotational tracking (head tracking) and to render images on the headset.

We used a pair of GoPro Hero 4 cameras for the stereo vision system. We used the analog video output of the GoPro cameras and digitized it using an Analog to USB video capture device, the *Hauppauge USB-Live2* [36]. Each of the camera signals captured by the USB-Live2 were broadcasted over the network using the WebcamXP [37] video streaming software, running on a dedicated Windows laptop used only for the vision system. In total, we used two robot laptops, one for the vision system, other for robot control.

We mounted the GoPro cameras on a pan/tilt servo servo rig on top of the robot to provide the stereo video feed to the user wearing the HMD. To control the pan and tilt system via PWM signals, an Arduino board [38] was used. The Pan and Tilt servomotors were connected to the analog pins of the Arduino. We used the RosSerial package [39] to send the head tracking data to the Arduino board via ROS (Robot Operating System) [40] Messages. The head tracking data provided by the Oculus Rift had to be parsed, in real time, to value ranges that made sense to the servos before sent to the Arduino board.

For the fixed-view baseline condition only, we used a HTC Nexus 9 tablet [41] fixed to the front of the robot (Figure 1). The tablet provides a 8.9 inch IPS LCD, 1536×2048 (QXGA) with a 1.6MP front (fixed) camera. However, all experiment conditions used the tablet to provide the audio input and output for the videoconferencing application (Google Hangouts [42]) that we used to transmit and receive audio between the participant and confederate.

Study participants remotely controlled the P3-DX mobile robot base using a desktop (Windows) computer which communicated with the robot control laptop (Figure 4) using ROS [40]. We implemented a custom web application for the additional user interface features used in the Study. The

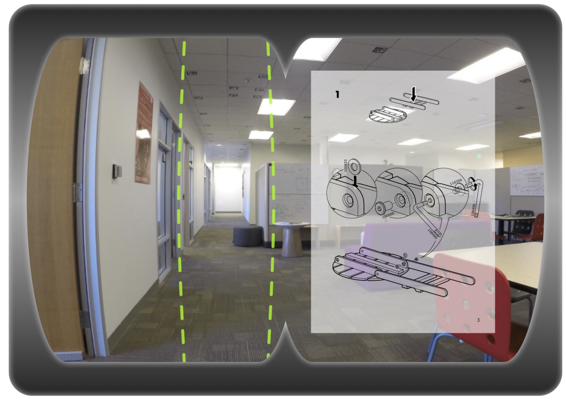

Fig. 3. Screenshots of the overlays developed for the HMD viewing conditions in the study. (a) Green dotted lines represent the robot's *forward* direction. (b) Part of the assembly manual shown as a semi-transparent overlay on the HMD.

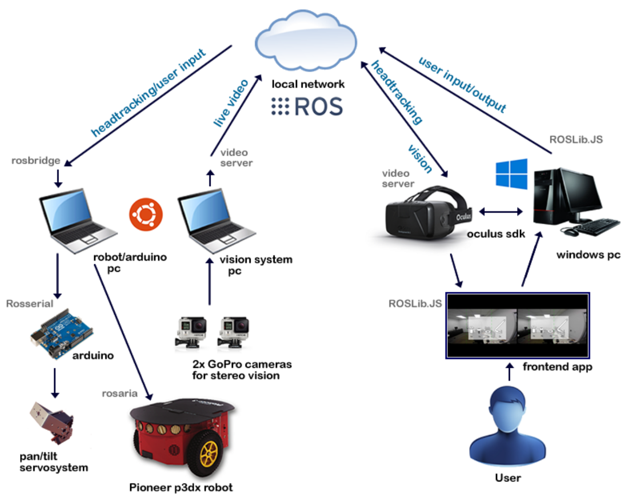

Fig. 4. System architecture used in the study.

keyboard arrow keys were used as main control keys for driving the robot, and the head-tracking data provided by the Oculus Rift was accessed via the Java Oculus SDK and sent to the robot laptop via ROSBridge. The left and right camera images, broadcasted from the vision system laptop using WebcamXP, were captured and rendered back to the Oculus Rift within the same web application.

The web application allowed the overlay of assembly manual pages on top of the camera feed with an opacity set to 0.5 (Figure 3 (b)). Users could control which page of the manual they wanted to see using the *Page Up* and *Page Down* keyboard keys. They were also able to toggle the display of the instruction manual overlay using the *Enter* key. To improve the participants' awareness of the relative orientations of the robot base front and the HMD view direction, we added an overlay of two vertical green lines (Figure 3 (a)) that showed the center orientation of the robot base (i.e., the robot's direction of forward motion).

E. Hypotheses

To study the effects of the use of HMDs and the use of stereo and mono camera systems in robotic remote collaboration, we evaluated a number of objective and subjective measures relevant to the outcome of collaborative tasks: task error rates, perceived collaboration success, perceived visual quality, and perceived task difficulty. We had the following hypotheses prior to the experiment:

H1: The use of a *monocular* head-tracked pan/tilt camera view system (MHTV) together with an HMD will improve overall user's task performance relative to the fixed 2D camera monitor-based baseline condition. By increasing the user's situational awareness, it will reduce the *number of collisions*, reduce the number of *assembly mistakes* and reduce the total *task time*.

H2: The use of a *stereo* head-tracked pan/tilt camera view system (SHTV) will improve user's overall task performance relative to fixed 2D camera monitor-based baseline condition with the same particular effects as stated in **H1**.

H3: The use of SHTV will improve user's overall task performance relative to MHTV condition with the same particular effects as stated in **H1**.

H4: The use of MHTV will increase user's feelings of engagement in the collaboration task, improving communication and improving the perceived collaborative success. In particular, it will improve the perceived *view quality*, perceived *feeling of presence* at the remote site, and reduce perceived *task difficulty* in comparison to the fixed 2D camera monitor-based baseline condition.

H5: The use of SHTV will increase user's feelings of engagement in the collaboration task, improving communication and improving the perceived collaborative success, with the same particular effects as stated in **H4**.

H6: The use of SHTV will increase user's feelings of engagement in the collaboration task, improving communication and improving the perceived collaborative success, with the same particular effects as stated in **H4**, over MHTV.

F. Procedure

To test our hypotheses, we conducted a controlled laboratory study with a between-participants design. Figure 5 shows an overview of the spatial layout of the experiment. Study participants were asked to sit in front of a desktop computer in an office room and operate the mobile telepresence robot, which, at the start of the study task, was present in the same room as the participant. Participants were given ten minutes of instruction on how to operate the system and navigate the robot, so that they were able to familiarize themselves with the controls and to practice driving the MTR while personally looking at it. Participants were then shown which path to drive the robot (Figure 5, red dashed line) in order to get to the room where the confederate person was waiting. The experimental remote room simulated a regular office room, with a big desk and a chair. The route between the participant's room and the experimental remote room was about 25 m in length, and involved three 90 degree turns. After providing instructions to the participants, the experimenter started the audio conferencing system and participants of both HMD conditions (**C2** and **C3**) were asked to put on the goggles, test the head-tracking system. They were then instructed on how to control the virtual assembly instructions overlaid on top of the camera feed. Participants of the baseline **C1** condition (without HMD) were given paper assembly instructions and also tested the video conferencing system prior to start of the activity. Each session took between 15–25 minutes in total.

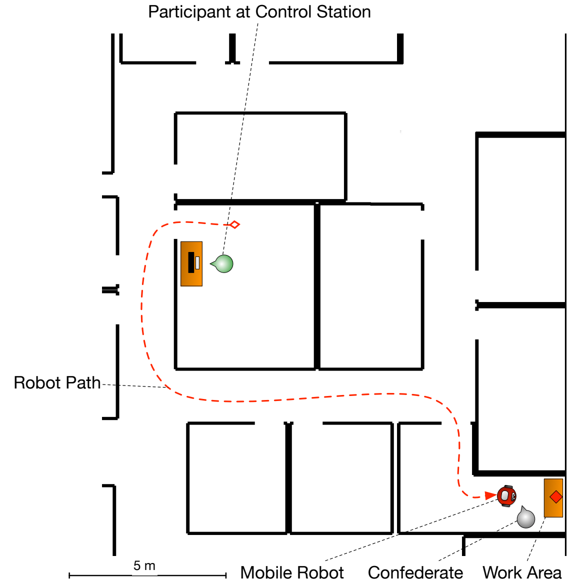

Fig. 5. A sketch of the spatial layout used in both experiments, showing the locations of the mobile robot control station and the assembly area in relation to each other. The red dashed line is the path along which the participants were asked to navigate at the start of the experiment.

G. Measures

To capture and analyze the outcomes of the collaborative task, we used several objective and subjective measures. These measurements allowed us to quantify and qualify the experience, which after an statistical analysis provided the results and insights we discuss in the sections below.

1) Objective Measures: We measured task performance and situation awareness using three variables: task completion time (beginning from when the robot left the room to when the participant indicated task completion), number of assembly mistakes/errors (incorrectly placed or misoriented parts) and the number of collisions (the participant running into something with the robot while driving it). Task completion time, number of collisions and number of assembling mistakes were measured manually by the experimenter during each session.

2) Subjective Measures: After completing the task, participants were asked to answer a questionnaire designed to assess the perceived collaborative success, quality of view/visualization, sensation of telepresence and task difficulty. Perceived collaborative success and quality of view/visualization were measured via participants' agreement on a ten-point Likert scale, e.g., "I was able to visualize all furniture parts" (1 = *Strongly Disagree* to 10 = *Strongly Agree*). Participants were also asked to measure how difficult they thought it was to complete the assembling task in the current experiment condition on a ten-point Likert scale (1 = *Very Easy* to 10 = *Very Hard*). We also asked participants open-ended questions about their thoughts and feelings on several aspects of the system.

H. Results

H1 and **H2** predicted that the use of an HMD with MHTV and SHTV conditions would improve the users' overall task performance relative to the fixed 2D camera monitor-based baseline condition. We found partial support for these hypotheses. The average task completion time, in seconds, was 581.43 (SD = 66.13), 602.57 (SD = 43.90) and 595.14 (SD = 50.85) for the "no HMD" (**C1**), MHTV (**C2**) and SHTV (**C3**) conditions, respectively. Users reported that using the HMD made it easier and faster to drive the robot, compared to the fixed monitor condition, since they had the liberty to look around and see obstacles better. After arrival at remote room, however, the assembly part of the task was done faster with the ability to look at the paper manual without HMD, compared to the virtual (overlaid) manual, wearing the HMD. This resulted in no significant difference for final average times for task completion, due to different experiences in different parts of the task, $F(2,18) = .271$, $p = .765$. The Average number of collisions and Average number of assembly mistakes, however, showed a slight advantage for the use of an HMD. Users wearing the HMD made fewer collisions: .86 (SD = .69) for **C1** condition, .57 (SD = .79) for **C2** and .43 (SD = .53) for **C3** condition and fewer mistakes .71 (SD = .76), .14 (SD = .38) and .14 (SD = .38).

H3 predicted that users in the HMD with SHTV condition group would outperform MHTV in terms of user's task performance. We did not find support for this hypothesis. In all metrics, SHTV and MHTV averages were not significantly different.

H4 and **H5** predicted that the use of an HMD with MHTV and SHTV conditions would improve the perceived collaborative success and view quality in comparison to the fixed 2D camera monitor-based baseline condition. As shown below, we did find support for these hypotheses. The average answers for the perceived collaborative success scale were 6.29 (SD = 1.50), 8.14 (SD = 0.69) and 8.43 (SD = 1.13) for the **C1**, **C2**, and **C3** conditions respectively, $F(2,18) = 7.107$, $p = .005$. Average responses to the Quality of View scale also indicated a better visualization for both HMD conditions. Average responses were 4.00 (SD = 1.41) for **C1** condition, 6.86 (SD = 1.21) **C2** and 6.86 (SD = 1.46) for **C3**, $F(2,18) = 10.169$, $p = .0011$.

H4 predicted that the users in **C2** would have a higher overall perception of collaborative success compared to **C1**. We found support for this hypothesis, $F(1,12) = 8.894$, $p = .011$. It also predicted that users would perceive a better quality of view. We found support for this hypothesis, $F(1,12) = 16.438$, $p = .001$.

H5 predicted that the users in **C3** condition would have a higher overall perception of collaborative success compared to **C1**. We found support for this hypothesis, $F(1,12) = 9.121$, $p = .010$. It also predicted that users would have better quality of view. We found support for this hypothesis, $F(1,12) = 13.793$, $p = .002$.

H6 predicted the users in **C3** would perceive an improved view quality relative to **C2**, given they should have a better visualization by the addition of depth. We did not find support for this hypothesis, $F(1,12) = .00$, $p = .99$. In the post task questionnaire, users reported perception of 3D, but were not sure of how useful it was to complete the task. Responses to the perceived collaborative success scale also showed no difference between the two groups, $F(1,12) = .324$, $p = .579$.

The average responses on task difficulty showed no big difference between any of the three groups. Responses were 6.57 (SD = 1.51), 5.29 (SD = 1.50) and 5.14 (SD = 2.12) for **C1**, **C2**, and **C3** conditions respectively. We can see a small variance between the perceived difficulty by users in **C1** and users in **C2** and **C3**, but no significant difference was found, $F(2,18) = 1.444$, $p = .261$.

I. Discussion

As expected, the results from this study confirmed our prior assumptions, showing that the use of an HMD does improve overall task performance, perceived collaborative success and perceived quality of view in comparison to the fixed 2D camera monitor-based baseline condition.

Participants in both groups wearing the Oculus Rift (SHTV and MHTV conditions), perceived improved view quality. The average task completion time was about the same for all groups, but major differences could be found in the full task experience: in the driving part of the task, participants wearing the HMD were faster, and made less collisions, since they had a better view of the path. In the assembly part of the task, participants in **C1** were faster, but they also made more assembly mistakes. Because participants in **C1** had a lower viewing quality due to the fixed view, they had to stay further away from the table to visualize the entire working area, and thus did not have the same ability to look around and focus on different parts of the furniture as in **C2** and **C3**. These facts seem to have influenced the higher perceived collaborative success between users in **C2**, and **C3**.

The second user study also confirmed that there was no significant difference between participants in **C2** and **C3**. The average number of collisions and average number of assembly mistakes were about the same for both groups. The average ratings for perceived collaboration success and quality of view were also about the same for them. Contrary to our initial hypothesis, we discovered and confirmed that the addition of depth using stereo cameras played no major role for the task of collaboratively assembling a furniture, in comparison to the condition using a monocular head-tracked display.

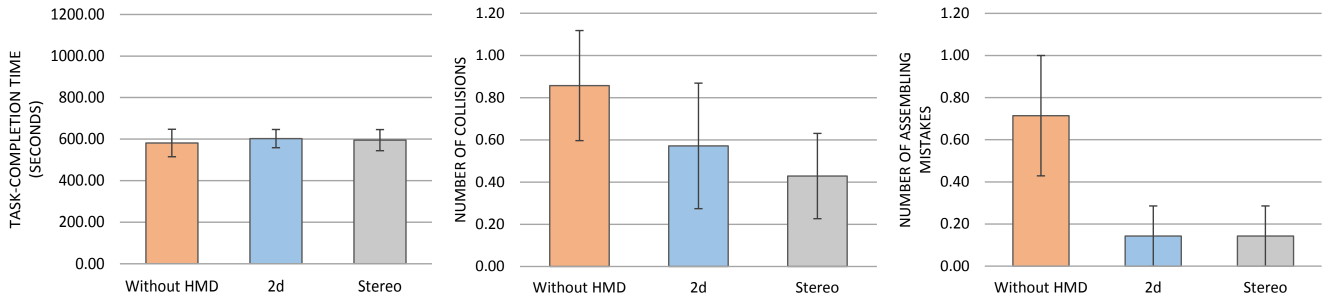

Fig. 6. User Study—Objective Measures

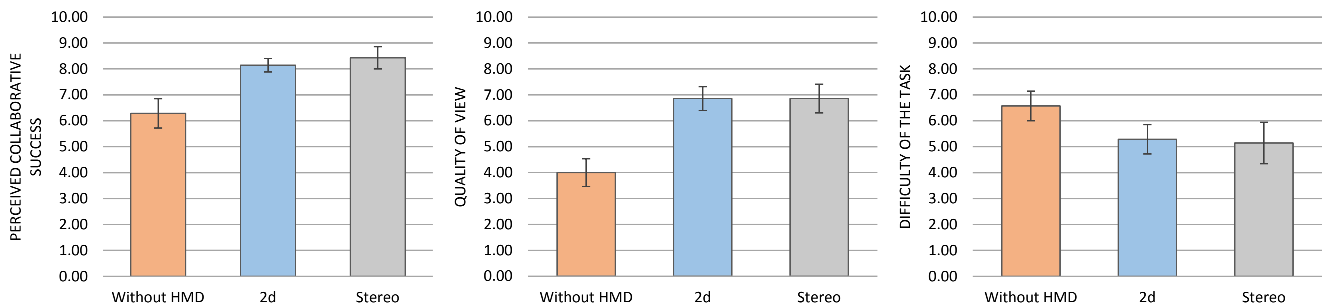

Fig. 7. User Study—Subjective Measures

IV. Conclusion

We presented the results of a user study with the goal of evaluating the use of head mounted devices for remote collaboration in robotic telepresence. More specifically, we evaluated the use of HMD with stereo and 2D camera feeds in comparison with each other and in comparison with a fixed 2D monitor-based baseline condition. We designed, proposed and implemented a mobile telepresence system that was used for assembling an IKEA furniture in collaboration with a remote confederate.

We conclude that the use of stereo and 2D cameras mounted on top of a pan tilt servo rig along with the use of an HMD does improve perceived collaborative success and perceived quality of view compared to the baseline condition. Task error rates (collisions and assembly mistakes) were also lowered—users wearing an HMD made fewer driving errors that resulted in collisions and, furthermore, made fewer assembly mistakes in the furniture assembly part of the task.

The ability to look at different points at the remote location by just moving the head, specifically looking down to see specific parts of the furniture, made it easier to complete the collaborative task with the HMD and the servo pan/tilt system setup. Participants in the baseline condition, however, were faster in the assembly part of the task, since they could easily read and interpret the paper manual without the HMD, but they also made more mistakes, because of the worse view of the remote location.

The perceived collaborative success was higher for users wearing an HMD, in comparison to users in the baseline condition. Results also showed a better quality of view, which definitely contributed to the perceived success of the task for HMD users.

In the user study, participants were provided with a switchable semi-transparent overlay of the assembly manual on top of the camera feed. The majority of the participants enjoyed the virtual manual. In particular, they liked the ability to control the page display via the keyboard and the ability to overlay it on the camera feed semi-transparently

A further interesting result was our finding that the use of stereoscopic cameras did not significantly affect the user experience for the collaborative task, in comparison to a monocular camera setup. There was no significant difference in all measured variables between **C2** and **C3**. In informal comments, participants reported no major usefulness of 3D for the specific given task. This may also be an indication that the task did not require stereoscopic depth cues to the extent we had imagined beforehand.

V. Future Work

The studies presented in this work have highlighted the advantages of using immersive, HMD-based viewing interfaces for remote collaboration task. Consequently, this also shows deficiencies of desktop-based approaches. In future work, it would be interesting to study similar HMD-based setups in different, richer environments (perhaps field deployments) with more complex tasks, in order to establish if the results of this paper hold up under such conditions. Secondly, we also aim to re-design approaches for desktop-based viewing interfaces, in order to achieve similar usability results in comparison to HMDs, which are not preferred by all users.

Thus, in a future publication, we plan to discuss and study new non-planar image mappings of panoramic video in order to increase view awareness desktop-based robot control user interfaces. We are also exploring new "through-the-screen" user interfaces for sending semi-autonomous movement commands to the robot, as an alternative to driving the robot manually. This may lower the overall view immersiveness requirements for successfully navigating a mobile robot through an environment.

References

[1] A. Kristoffersson, K. S. Eklundh, and A. Loutfi, "Measuring the quality of interaction in mobile robotic telepresence: a pilots perspective," *International Journal of Social Robotics*, vol. 5, no. 1, pp. 89–101, 2013.

[2] E. Paulos and J. Canny, "Delivering real reality to the world wide web via telerobotics," in *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, vol. 2. IEEE, 1996, pp. 1694–1699.

[3] G. Venolia, J. Tang, R. Cervantes, S. Bly, G. Robertson, B. Lee, and K. Inkpen, "Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 1049–1058.

[4] P. Boissy, H. Corriveau, F. Michaud, D. Labonté, and M.-P. Royer, "A qualitative study of in-home robotic telepresence for home care of community-living elderly subjects," *Journal of Telemedicine and Telecare*, vol. 13, no. 2, pp. 79–84, 2007.

[5] F. Michaud, P. Boissy, D. Labonte, H. Corriveau, A. Grant, M. Lauria, R. Cloutier, M.-A. Roux, D. Iannuzzi, and M.-P. Royer, "Telepresence robot for home care assistance." in *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*. California, USA, 2007, pp. 50–55.

[6] O.-H. Kwon, S.-Y. Koo, Y.-G. Kim, and D.-S. Kwon, "Telepresence robot system for english tutoring," in *Advanced Robotics and its Social Impacts (ARSO), 2010 IEEE Workshop on*. IEEE, 2010, pp. 152–155.

[7] R. Marín, P. J. Sanz, and A. P. Del Pobil, "The uji online robot: An education and training experience," *Autonomous Robots*, vol. 15, no. 3, pp. 283–297, 2003.

[8] G. H. Ballantyne, "Robotic surgery, telerobotic surgery, telepresence, and telementoring," *Surgical Endoscopy and Other Interventional Techniques*, vol. 16, no. 10, pp. 1389–1402, 2002.

[9] P. M. Vespa, C. Miller, X. Hu, V. Nenov, F. Buxey, and N. A. Martin, "Intensive care unit robotic telepresence facilitates rapid physician response to unstable patients and decreased cost in neurointensive care," *Surgical neurology*, vol. 67, no. 4, pp. 331–337, 2007.

[10] Suitable Technologies, "Beam Telepresence Robot," Retrieved on 9/17/15 from https://www.suitabletech.com/beampro/, 2015.

[11] VGO Communications Inc, "VGO Telepresence Robot," Retrieved on 9/17/15 from http://www.vgocom.com/, 2015.

[12] I. Rae, B. Mutlu, and L. Takayama, "Bodies in motion: Mobility, presence, and task awareness in telepresence," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 2153–2162.

[13] M. K. Lee and L. Takayama, "Now, i have a body: Uses and social norms for mobile remote presence in the workplace," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 33–42.

[14] S. Johnson, I. Rae, B. Mutlu, and L. Takayama, "Can you see me now? how field of view affects collaboration in robotic telepresence," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2397–2406.

[15] J. V. Draper, D. B. Kaber, and J. M. Usher, "Telepresence," *Human Factors: The journal of the human factors and ergonomics society*, vol. 40, no. 3, pp. 354–375, 1998.

[16] Oculus VR, "Oculus Rift," Retrieved on 9/17/15 from http://www.oculusvr.com/, 2015.

[17] I. Rae, G. Venolia, J. C. Tang, and D. Molnar, "A framework for understanding and designing telepresence," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2015, pp. 1552–1566.

[18] T. Fong and C. Thorpe, "Vehicle teleoperation interfaces," *Autonomous robots*, vol. 11, no. 1, pp. 9–18, 2001.

[19] T. W. Fong, F. Conti, S. Grange, and C. Baur, "Novel interfaces for remote driving: gesture, haptic, and pda," in *Intelligent Systems and Smart Manufacturing*. International Society for Optics and Photonics, 2001, pp. 300–311.

[20] B. House, J. Malkin, and J. Bilmes, "The voicebot: a voice controlled robot arm," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 183–192.

[21] S. Zhao, K. Nakamura, K. Ishii, and T. Igarashi, "Magic cards: a paper tag interface for implicit robot control," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 173–182.

[22] S. Thrun, W. Burgard, D. Fox, *et al.*, "Probabilistic robotics," vol. 1. MIT press Cambridge, 2005, pp. 237–278.

[23] L. Takayama, E. Marder-Eppstein, H. Harris, and J. M. Beer, "Assisted driving of a mobile remote presence system: System design and controlled user evaluation," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1883–1889.

[24] J. T. Biehl, D. Avrahami, and A. Dunnigan, "Not really there: Understanding embodied communication affordances in team perception and participation," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2015, pp. 1567–1575.

[25] S. Kratz, D. Kimber, W. Su, G. Gordon, and D. Severns, "Polly: "being there" through the parrot and a guide," in *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2014), Industrial Case Studies*. ACM, september 2014.

[26] M. Fiala, "Pano-presence for teleoperation," in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE, 2005, pp. 3798–3802.

[27] L. Zalud, "Orpheus–universal reconnaissance teleoperated robot," in *RoboCup 2004: Robot Soccer World Cup VIII*. Springer, 2005, pp. 491–498.

[28] H. Martins and R. Ventura, "Immersive 3-d teleoperation of a search and rescue robot using a head-mounted display," in *Emerging Technologies & Factory Automation, 2009. ETFA 2009. IEEE Conference on*. IEEE, 2009, pp. 1–8.

[29] S. W. Lawson, J. R. Pretlove, A. C. Wheeler, and G. A. Parker, "Augmented reality as a tool to aid the telerobotic exploration and characterization of remote environments," *presence: teleoperators and virtual environments*, vol. 11, no. 4, pp. 352–367, 2002.

[30] C. Pittman and J. J. LaViola Jr, "Exploring head tracked head mounted displays for first person robot teleoperation," in *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 2014, pp. 323–328.

[31] doraplatform.com, "DORA Platform," Retrieved on 05/26/16 from http://doraplatform.com/, 2016.

[32] S. Kratz, J. Vaughan, and D. Kimber, "Evaluating Stereoscopic Video with Head Tracking for Immersive Teleoperation of Mobile Telepresence Robots," in *10th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 2015.

[33] IKEA, "IKEA Vildapel Plant Stand (602.372.26) Assembly Manual," Retrieved on 9/17/15 from http://www.ikea.com/us/en/assembly_instructions/vildapel-plant-stand__AA-905292-2_pub.pdf, 2015.

[34] Mobile Robots, "Pioneer P3-DX," Retrieved on 9/17/15 from http://www.mobilerobots.com/ResearchRobots/PioneerP3DX.aspx, 2015.

[35] Oculus VR, "Oculus Rift SDK," Retrieved on 1/21/16 from https://developer.oculus.com/, 2016.

[36] Hauppauge, "Hauppauge USB-Live2," Retrieved on 2/17/16 from http://www.hauppauge.com/site/products/data_usblive2.html, 2016.

[37] Webcam XP, "Webcam XP—Webcam and Network Camera Surveillance Software," Retrieved on 1/24/16 from http://www.webcamxp.com/home.aspx, 2016.

[38] Arduino, "Arduino/Genuino UNO," Retrieved on 1/21/16 from https://www.arduino.cc/en/Main/ArduinoBoardUno, 2016.

[39] ROS Wiki, "RosSerial Package Summary," Retrieved on 1/21/16 from http://wiki.ros.org/rosserial, 2016.

[40] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an Open-Source Robot Operating System," *ICRA Workshop on Open Source Software*, 2009.

[41] Google, "Nexus 9 Tablet," Retrieved on 1/22/16 from http://www.google.com/nexus/9/, 2016.

[42] Google, "Google Hangouts," Retrieved on 1/22/16 from https://hangouts.google.com/, 2016.