Postprint

View Position Impact on QoE in an Immersive Telepresence System for Remote Operation

Elijs Dima[1], Kjell Brunnström[2,1], Mårten Sjöström[1], Mattias Andersson[1], Joakim Edlund[1], Mathias Johanson[3] and Tahir Qureshi[4]

[1]*Dept. of Information Systems and Technologies, Mid Sweden University*, Sundsvall, Sweden.
Email: {elijs.dima, marten.sjostrom, mattias.andersson, joakim.edlund}@miun.se
[2]*RISE Research Institutes of Sweden, Division ICT - Acreo*, Kista, Sweden. Email: kjell.brunnstrom@ri.se
[3]*Alkit Communications AB*, Mölndal, Sweden. Email: mathias@alkit.se
[4]*HIAB AB*, Hudiksvall, Sweden. Email: tahir.qureshi@hiab.com

Abstract—In this paper, we investigate how different viewing positions affect a user's Quality of Experience (QoE) and performance in an immersive telepresence system. A QoE experiment has been conducted with 27 participants to assess the general subjective experience and the performance of remotely operating a toy excavator. Two view positions have been tested, an overhead and a ground-level view, respectively, which encourage reliance on stereoscopic depth cues to different extents for accurate operation. Results demonstrate a significant difference between ground and overhead views: the ground view increased the perceived difficulty of the task, whereas the overhead view increased the perceived accomplishment as well as the objective performance of the task. The perceived helpfulness of the overhead view was also significant according to the participants.

Index Terms—quality of experience, augmented telepresence, head mounted display, viewpoint, remote operation, camera view

I. INTRODUCTION

Immersive technologies like Virtual Reality (VR) and Augmented Reality (AR) are gaining attraction for industrial applications. Together with immersive telepresence, these technologies help solve safety and accessibility in hazardous environments, and enable remote operation of industrial machinery.

In this study, we investigate the effect of camera view position for immersive telepresence systems, focusing on QoE [1], [2] aspects as well as a task completion metric. We distinguish *immersive telepresence* from *augmented telepresence* [3]–[5] only by the lack of augmentations. The target applications are immersive video-based telepresence systems for industrial uses such as forestry cranes [5], that let an operator control machines from a safe location using a Head-Mounted Display (HMD). The safety aspects and operator accuracy of such a system may be dependent on the camera position, which defines the remote operator's view of the environment. A telepresence system that reduces operator accuracy or their QoE may lead to higher on-site risk and loss of productivity.

A QoE test has been conducted to study the task completion rate in an immersive telepresence system with two different view positions, and to capture the subjective experiences from using the system. The different view positions were used to alter the operators' reliance on depth perception, which may be compromised in systems with fixed camera placement.

II. BACKGROUND AND RELATED WORK

The research for QoE on video presented in HMDs is currently quite intensive, where 360 video is very popular, see e.g. [6]–[9]. Previous research has investigated the role of stereo presentation in e.g. HMD [10], 3DTV [11] and as investigations into video quality of VR [12]. Less work has been done into the QoE for industrial application with real-time visual feedback and interactive control of machines as in this work. However, the influence on gaming QoE when simulating the operation of a forklift in VR was compared to that of a 2D-simulation by evaluating the perceived presence, usability and user emotion [13]. Other works related to this topic are [5], [14]. For the impact of camera position on tele-operation of a robot, earlier work has been done where different camera position and display presentations where used to improve the robot tele-operation in rough terrain conditions in terms of navigating and time for task completion [15].

III. METHOD

We present a formal subjective study with an immersive telepresence system for remote navigation. We have varied the view position, as shown in Fig. 1, to study its impact on operator performance and QoE in remote navigation tasks.

Fig. 1. The experiment setup, seen from the overhead (left) and ground-level (right) view positions. FoV is 110 degrees. RC vehicle is at the starting position of each test session.

A. Test Procedure

Test subjects were invited to perform a remote-navigation task in the telepresence system. After the participants' consent, the task of driving a RC excavator to accurately reach selected targets was explained and followed by a 5-minute training session on controlling the RC vehicle and using the telepresence system's HMD. After training, a 2-minute test run for each view position was conducted with random order of targets and positions. After each run, participants were asked to rate *A) their ability to precisely reach the targets* (Task Accomplishment), *B) the view position helpfulness in precisely reaching the targets* (Viewpoint Helpfulness), and *C) the difficulty of precisely reaching the targets* (Task Difficulty). Figure 2 shows the 5-point Likert scales for questions A, B (top) and C (bottom). The excavator and targets were tracked, giving the fourth, objective scale - number of targets reached.

Fig. 2. Scales for recording participant opinions on task accomplishment, view position helpfulness (top) and task difficulty (bottom).

B. Experiment Setup

The experiment setup, shown in Fig. 3, was based on a multi-camera real-time streaming system similar to [16], connected to an HTC Vive Pro HMD. This HMD was chosen due to its higher resolution and more accurate tracking, advantage of which was shown in [17]. The camera views were rectified and projected onto a virtual enclosing sphere surface to match the horizontal and vertical FoV of the cameras. Inter-camera distance in each stereo pair was 8 cm. The low view position was 10 cm above ground-level, and the overhead view position was 120 cm high, observing a 220 by 300 cm test area, with targets spaced 70 cm apart. The cameras recorded at 24 Frames per Second (FPS), and the virtual environment was rendered at 90 FPS, avoiding any delay on HMD rotation and translation. The RC vehicle was a Hulna 1507 1:14th size excavator.
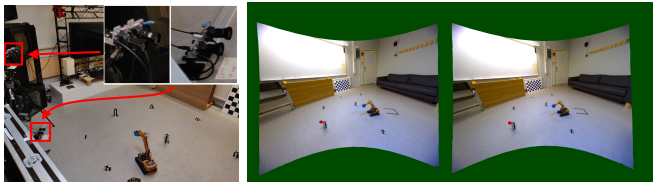

Fig. 3. Left: the experiment setup, with highlighted stereo camera pairs. Right: stereoscopic enclosing-sphere projection of the overhead view position, with a selected target. Left and Right eye views shown side to side, without HMD lens compensation.

C. Analysis Tools and Metrics

The Likert scale responses were converted to numerical 1 to 5 interval scales [18] prior to to calculating the Mean Opinion Score (MOS) of the participants. The scales were shown and described with equal distances between the categories. The analysis was conducted by applying a Bonferroni-corrected paired-sample t-test between view position responses for each individual scale. Significance is considered at 95% confidence level per scale. All scales were checked for distribution normality using Pearson Chi-square, Kolmogorov-Smirnov and Jarque-Bera tests, to verify the normality as suggested in [19]. A cross-scale Repeated-Measures Analysis of Variance (RMANOVA) was applied to examine the interaction between scales and view positions and to quantify the main effect of view position on scale responses.

IV. RESULTS AND ANALYSIS

The experiment was conducted in a controlled lab on university premises, with consistent lighting and isolation of outside noises. 27 test subjects from the local student and staff population took part, 5 females and 22 males, with a mean age of 31 (min: 20, max: 54). Some had previous experience in using HMDs for VR, none had previous experience in using the experiment system. One test subject did not finish the test, due to a battery failure in the RC vehicle. The ratings given up to the point of stopping have been included in the analysis. None of the participants indicated serious symptoms of simulator sickness. All participants were encouraged to adjust the HMD inter-lens distance and fit to their preference.
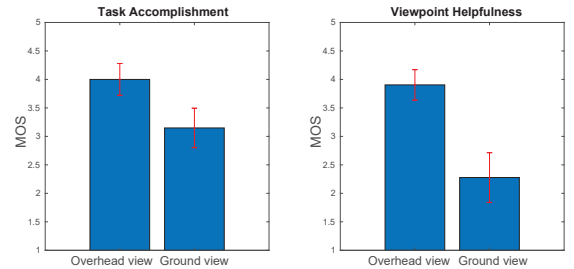

Fig. 4. The MOS on Task Accomplishment (left) and Viewpoint Helpfulness (right) for different view positions. Error bars show 95% confidence intervals.

The result of the test participants' MOS on Task Accomplishment and View Helpfulness is shown in Figure 4, and Task Difficulty in Figure 5 (left). The bar height shows the mean, and error bars indicate 95% confidence interval. MOS on Task Accomplishment is 4.0 for overhead view and 3.1 for ground view. View Helpfulness in overhead view has a MOS of 3.9, and the ground view has a MOS of 2.3. MOS of Task Difficulty also varies, from 2.4 for overhead view, to 3.5 for ground view. Chi-square, Kolmogorov-Smirnov and Jarque-Bera tests confirm that responses in all three scales fit a normal distribution. Paired sample t-tests ($\alpha = 0.05$) show a significant difference between overhead and ground view position MOS for Task Accomplishment scale with $p = 0.0015$, Viewpoint Helpfulness scale ($p = 3.7 * 10^{-6}$), and Task Difficulty scale ($p = 35 * 10^{-6}$). RMANOVA shows a significant main effect of view positions for the three opinion scales, with $F_{1,75} = 12.6$,

$p = 0.00066$ and a significant interaction between scales and view positions with $F_{2,75} = 32.2$, $p = 8 * 10^{-11}$.
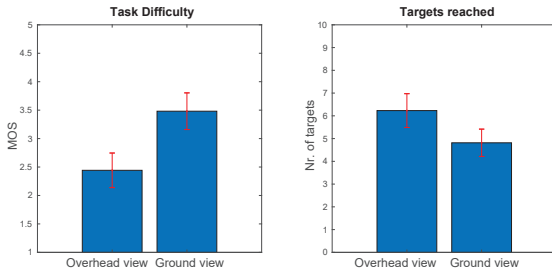

Fig. 5. The perceived task difficulty (left), and mean number of targets reached (right) for different view positions. Error bars show 95% confidence intervals.

Figure 5 (right) shows the mean number of targets reached by test subjects as the bar height for the overhead and ground view positions, and the 95% confidence intervals as the error bars. With overhead view, the mean number of targets reached was 6.2, compared to a mean of 4.8 for ground view. Normality tests show that these measurements are normally distributed. Paired sample t-test ($\alpha = 0.05$) shows a significant difference between the mean targets reached via overhead and ground view position cases, with $p = 5.9 * 10^{-6}$. RMANOVA across all four scales shows a significant main effect of view positions with $F_{1,100} = 35.8$, $p = 3.4 * 10^{-8}$, and a significant interaction between scales and view positions with $F_{3,100} = 24.7$, $p = 4.6 * 10^{-12}$.

V. DISCUSSION AND CONCLUSIONS

The results indicate that the view position has a significant effect on QoE aspects of immersive telepresence systems, and a significant effect on navigation and positioning task completion. By forcing a ground-level view position (equivalent to standing to the side of an excavator) where users must rely on stereoscopic depth cues through HMDs, the perceived difficulty of the navigation task increased. The view position's impact on the task was clearly noticed by the users, as the mean helpfulness of the ground view position was rated as slightly above Poor (MOS 2.3), compared to the Good rating (3.9) of the overhead position. This impact was confirmed by a significant difference in the measured task performance.

These findings confirm previous studies [15] for a new scenario by showing that immersive telepresence systems must consider the placement of cameras when used for non-entertainment purposes, and that stereoscopic depth perception through HMDs is not sufficient for remote operation if camera placement is poorly chosen.

The results and measurements of this study will serve as a baseline for further investigations of simulator sickness, stereo perception, remote navigation and remote operation tasks in augmented telepresence systems.

ACKNOWLEDGMENT

REFERENCES

[1] P. Le Callet, S. Möller, and A. Perkis, *Qualinet White Paper on Definitions of Quality of Experience. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, Lausanne, Switzerland, 2012.

[2] S. Möller and A. Raake, *Quality of Experience - Advanced Concepts, Applications and Methods*. Springer International Publishing, 2014.

[3] F. Okura, M. Kanbara, and N. Yokoya, "Augmented telepresence using autopilot airship and omni-directional camera," in *International Symposium on Mixed and Augmented Reality*. IEEE, 2010, pp. 259–260.

[4] V. V. Saxena, T. Feldt, and M. Goel, "Augmented telepresence as a tool for immersive simulated dancing in experience and learning," in *Proceedings of the India HCI 2014 Conference on Human Computer Interaction*. ACM, 2014, p. 86.

[5] K. Brunnström, E. Dima, M. Andersson, M. Sjöström, T. Qureshi, and M. Johanson, "Quality of experience of hand controller latency in a virtual reality simulator," in *Human Vision and Electronic Imaging*. Society for Imaging Science and Technology, 2019, pp. HVEI–218.

[6] R. G. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli, and P. Frossard, *Visual Distortions in 360-degree Videos*, 2018.

[7] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, "Towards subjective quality of experience assessment for omnidirectional video streaming," in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1–6.

[8] A. Singla, S. Fremerey, W. Robitza, and A. Raake, "Measuring and comparing qoe and simulator sickness of omnidirectional videos in different head mounted displays," in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1–6.

[9] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on qoe of 360 video for vr communication," in *19th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2017, pp. 1–6.

[10] J. Häkkinen, M. Pölönen, J. Takatalo, and G. Nyman, "Simulator sickness in virtual display gaming: a comparison of stereoscopic and non-stereoscopic situations," in *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*. ACM, 2006, pp. 227–230.

[11] K. Wang, M. Barkowsky, K. Brunnström, M. Sjöström, R. Cousseau, and P. Le Callet, "Perceived 3d tv transmission quality assessment: Multi-laboratory results using absolute category rating on quality of experience scale," *IEEE Transactions on Broadcasting*, vol. 58, no. 4, pp. 544–557 (10.1109/TBC.2012.2 191 031), 2012.

[12] J. Simsch, "Video quality in vr," M.Sc. Thesis, University of Twente and TNO, Enschede, The Netherlands, 2016.

[13] I. Hupont, J. Gracia, L. Sanagustín, and M. A. Gracia, "How do new visual immersive systems influence gaming qoe? a use case of serious gaming with oculus rift," in *Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2015, pp. 1–6.

[14] K. Brunnström, M. Sjöström, M. Imran, M. Pettersson, and M. Johanson, "Quality of experience for a virtual reality simulator," in *Human Vision and Electronic Imaging*, vol. 2018. Society for Imaging Science and Technology, 2018.

[15] J. Wang, M. Lewis, and S. Hughes, "Gravity-referenced attitude display for teleoperation of mobile robots," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48. SAGE Publications, 2004, pp. 2662–2666.

[16] E. Dima, M. Sjöström, R. Olsson, M. Kjellqvist, L. Litwic, Z. Zhang, L. Rasmusson, and L. Flodén, "Life: A flexible testbed for light field evaluation," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*. IEEE, 2018, pp. 1–4.

[17] M. Suznjevic, M. Mandurov, and M. Matijasevic, "Performance and qoe assessment of htc vive and oculus rift for pick-and-place tasks in vr," in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017, pp. 1–3.

[18] H. Wu and S.-O. Leung, "Can likert scales be treated as interval scales? a simulation study," *Journal of Social Service Research*, vol. 43, no. 4, pp. 527–532, 2017.

[19] K. Brunnström and M. Barkowsky, "Statistical quality of experience analysis: on planning the sample size and statistical significance testing," *Journal of Electronic Imaging*, vol. 27, no. 5, p. 053013, 2018.