https://doi.org/10.2352/EI.2022.34.11.HVEI-166

Augmented Remote Operating System for Scaling in smart mining applications: Quality of Experience aspects

Shirin Rafiei[a,b], Elijs Dima[a], Mårten Sjöström[a], Kjell Brunnström[a,b]
a: Dept. of Information Systems and Technology, Mid Sweden University, Sundsvall, Sweden ***b: RISE Research Institutes of Sweden, Kista, Sweden***

Abstract

Remote operation and Augmented Telepresence are fields of interest for novel industrial applications in e.g., construction and mining. In this study, we report on an ongoing investigation of the Quality of Experience aspects of an Augmented Telepresence system for remote operation. The system can achieve view augmentation with selective content removal and Novel Perspective view generation. Two formal subjective studies have been performed with test participants scoring their experience while using the system with different levels of view augmentation. The participants also gave free-form feedback on the system and their experiences. The first experiment focused on the effects of in-view augmentations and interface distributions on wall patterns perception. The second one focused on the effects of augmentations on the depth and 3D environment understanding.

The participants' feedback from experiment 1 showed that the majority of participants preferred to use the original camera views and the Disocclusion Augmentation view instead of the Novel Perspective views. Moreover, the Disocclusion Augmentation, that was shown in combination with other views seemed beneficial. When the views were isolated in experiment 2, the impact of the Disocclusion Augmentation view was found to be lower than the Novel Perspective views.

Introduction

Immersive telepresence systems are becoming viable tools for industrial applications and remote vehicle operation scenarios [1, 2], in part thanks to the emerging high-bandwidth mobile network technologies such as 5G. For instance, the mining industry is poised for a transition to telepresence and remote operation [3, 4]. Within telepresence systems, augmentation of captured and extrapolated views of the remote scene has been shown to be beneficial [5, 6].

Quality of Experience (QoE) is "the degree of delight or annoyance of the user of an application or service" as defined in ITU-T Rec. P.10 [7, 8]. The research on QoE has been moving from just video quality to more advanced, immersive applications, and has had to embrace methods involving user interaction and user experience to capture the full QoE [9]. This is an ongoing research transition, and this study is a part of that process. When for instance studying real systems in operation or systems that are normally used by professionals it may be hard or not economical to run many test persons making the quantitative analysis less reliable. Mixing qualitative and quantitative methods a.k.a mixed methods have therefore become important [10].

In this study, we explored how in-view augmentations and interface distributions influence the remote operator's ability to perceive wall patterns, and estimate the 3D environment through an Augmented Telepresence (AT) interface. The study was performed with the usage of our Augmented Remote Operating System for Scaling in Mining (AROSS) [11]. The AROSS was initially built to demonstrate the technical feasibility of introducing real-time in-view augmentations (object removal, i.e., "Disocclusion Augmentation") and Novel Perspective view generation for the mining context. This paper presents ongoing research on AROSS focusing on QoE. Specifically, we present two QoE experiments and discuss the subsequent implications towards AROSS. We present primarily the quantitative results, but the discussion is also based on the qualitative results.

Background and Related work

Augmented Telepresence

Augmented Telepresence denotes immersive video-mediated communication wherein additional data can be superimposed on or merged with the video, similar to Augmented Reality (AR) [12]. AT is like AR in that the environment shown to the user is augmented or mediated in some way. It differs from AR in that the user is present in a remote location and is observing the augmented view through an audio-visual communication channel [13].

QoE for AT in remote control

Jahromi et al. [14] investigated Telepresence Robot Systems. They performed a subjective study to assess remote navigation using such system live over the Internet with the aspect of QoE. The influence of network impairments (delay, bandwidth, and packet loss rate) was evaluated on the QoE. The results showed that users could separate quite well between the control and visual aspects of using a Telepresence Robot Systems..

Effects of perceivable delay were investigated in Brunnstr öm et al.[13] via a Virtual Reality (VR) simulator of a remotely operated forestry crane. It was found that user QoE starts to degrade at approximately 500 ms of hand control delay, and at as little as 30 ms delay in the rendering update of the VR presentation due to simulator sickness.

In Dima et al.[15] investigated how different viewing positions affect users' QoE and performance in an immersive telepresence system. The results indicated that the view position has significant effect on QoE aspects of immersive telepresence systems, and a considerable effect on navigation and positioning task completion. Non-headset based remote control systems with view augmentation have been studied in [5, 16] and [17], where the augmentations were shown to improve operator task performance.

In AT and AR, most augmentation types investigated are additive, superimposing artificial content on top of the real view (as seen in [5, 6, 13, 15, 16, 17]). A less explored area in AT

is view augmentation through selective content removal from the presented view. The AROSS [11] is an AT system with augmentation through selective content removal and Novel Perspective view generation. The interplay and the relative benefits of these factors towards operator QoE, especially for the mining context, have not been investigated previously.

Method

This study investigates a remote-control system with AR elements, that approximates a test system for future remote-control interfaces in underground mining machines. Two formal subjective studies have been performed. The aim of Experiment 1 (Exp 1) was to investigate the perception of the operator of small patterns. This could be cracks that found on the mine tunnel walls. Experiment 2 (Exp 2) investigated the operator's sense of depth and 3D when using the AROSS system. An additional goal of the two experiments was to investigate the impact of the specific augmentations provided by this test system and to determine the subsequent research and development direction.

Common procedures for the formal test

Test participants were invited to the laboratory to participate in an experiment with the AROSS system. On arrival, they were introduced to the system and were asked to read the instructions, which explained the task to perform. Then, they practiced remote operation with the controller and a default interface setup. Description about the remote-control systems with additional views was given verbally and the test participants were instructed to test all operations in the training session to get an understanding of how to operate the AROSS system.

Then participants followed the main test tasks: to locate defined targets through various graphical interface configurations and to use the robotic arm to point at identified targets for validation. In each experiment we have some scenarios, which we here call 'test run'. There were eight and five test runs in Exp 1 and Exp 2 respectively. Over the course of the experiment, the participants answered questionnaires divided into three parts that were related to:

1) Background information of participants (prior to test);

2) Participants' feedback about each test run (during test);

3) Their suggestions and overall feedback in a free-form response (after test).

All tests were performed on one occasion per participant. The specific questions varied between the two experiments.

Apparatus

AROSS was designed to support a better perception of the spatial configuration of the mine wall, and was tested in a laboratory approximation of an underground mine in the Mine Lab at Mid Sweden University, shown in Figure 1. A robotic arm was used as a remote manipulator to interact with the remote environment. It was used to point to a specific location on the mine wall. The graphical interface implemented in this experiment was a remote observation interface with views of the remote scene. In Exp 1 different combination of views were used, as e.g., shown in Figure 2 where all views were enabled in this case. In Exp 2 a single view at the time was presented, see Figure 3.

Figure 1: AROSS test set-up, with cameras, lights, remotely controlled robot arm, and the mine-like rock wall.

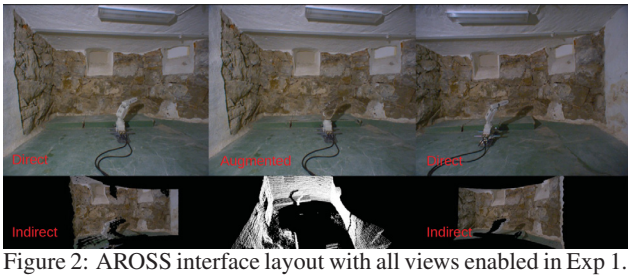
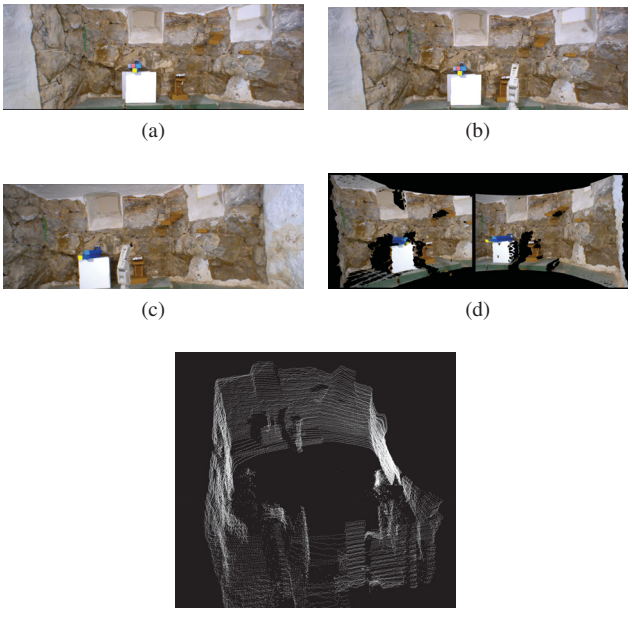


Figure 2: AROSS interface layout with all views enabled in Exp 1. In the left and right top row, two Direct camera views are shown, the top center view is a Disocclusion Augmentation view, in the left and right bottom row two Novel Perspective views are shown, and the bottom center view is a pure Lidar geometry view.



(a) (b) (c) (d) (e)

Figure 3: AROSS interface layouts in Exp 2. Layouts a) Disocclusion Augmentation view, b) Left Original view, c) Right Original view, d) Novel perspective views, e) Lidar view.

Safety and ethical considerations

Due to COVID-19, precautions for experiments with test persons in indoor labs were followed, based on the protocol in [18]. The precautions taken during the Exp 1 included that both test

leaders and participants wore face masks. In Exp 2, the precaution was relaxed a bit thanks to vaccination. Face masks were not used, but distancing were still maintained. Test participants answered the questionnaires digitally in Google forms instead of on paper. All equipment was disinfected by using surface disinfection before and after each test session. The participants were asked to use hand disinfection before and after the test. In this experiment the rules in the General Data Protection Regulation (GDPR) were followed. Participants were informed that the study is entirely voluntary and they had the right to leave the test at any time, without any explanation. The recorded responses were anonymized.

All participants in both experiments were selected from the Mid Sweden University staff. Since participants were selected from a predetermined group, instead of asking, in the background questionnaire, respondents to state their exact age, we created different age categories within the range of 10 years, starting from 21 until 60 years old.

Experiment 1

Procedures

In this experiment participants followed the main test task: to locate wall patterns ('targets') through various graphical interface configurations (e.g., additional views and view distributions), and to use the robotic arm to point at the identified target for validation. The targets were arranged in a randomly ordered sequence for each participant and test run, and were shown one at a time. Examples are shown in Figure 4. A total of 90 target images were used, which were wall patterns photographed at three different scales for obtaining variation, giving approximately 30 distinct wall patterns. The participants completed the task eight times, each time ('test run') took approximately 2.5 minutes. At the end of each run, test participants reported their immediate opinion on the interface configuration. The total time was estimated at approximately 75 minutes per participant.

(a) (b) (c) (d)

Figure 4: Target images used for wall structure detection and location task (randomly selected subset shown).

Apparatus

The interface consists of two Direct camera views (Dir) (left/right top row), a Disocclusion Augmentation view (DA) (top center), two Novel Perspective views (NP) (left/right bottom row), and a pure Lidar geometry view (Lidar) (bottom center), see Figure 2. "DA" denotes a view of the remote scene where an occluding object (such as a tool boom) has been partly or fully made transparent. "NP" denote views showing the scene from positions outside of the camera array at the remote location. Further technical details of AROSS are given in [11]. During Exp 1, an autostereoscopic 4K display was used for presenting the interface of the remote operation system. The interface was shown in monoscopic or 2D mode. This display was picked as the only large-panel, high-resolution display at hand and it was judged before the experiment that impact of the lenticular lenses would have a minor influence on the presentation of the interface.

Augmented elements and interface distributions

The experiment used eight test runs to cover the viable combinations of test conditions, listed in Table 1. During testing, the interface had different combinations of enabled views. The test runs were: *"No AROSS"*, *"SP(1)"*, *"DA(1)+SP(2)"*, *"DA(2)+SP(2)"*, *"DA(3)+SP(2)"*, *"DA(1)+SP(3)"*, *"DA(2)+SP(3)"*, *"DA(3)+SP(3)"*. SP means scene perspective view which is a combination of DA, Dir and NP views. Each test run was limited to 2.5 minutes. The *"No AROSS"* run was used first as a training session, wherein participants looked directly at the scene without the remote operation interface. The order of the remaining runs was randomized. The target sequences were randomized for each test run and participant, with a maximum of 90 possible targets in each sequence. The targets were presented to participants one at a time. Participants reported their opinion after each test run. The questions asked after each test run are as follows:

1- How would you rate your experience of identifying wall patterns visually via the interface?

2- How would you rate your experience of moving the robotic arm to a specific feature?

3- How was the support of the available views for your activity?

4- To what extent did you feel the impact of the different Novel Perspective views?

Task difficulty in Exp 1 was experienced as composed of both a detection problem to identify crack patterns on the wall, and a robot arm's control problem. Moreover, the amount of identified targets by participants under different interface configurations were investigated. Responses were recorded on a 5-point Likert scale. A graphical representation of the scale is shown in Figure 5. The scale was shown each time it was rated by the participants to give them a mental picture of the distances between scale levels.

Figure 5: Example of scales used for requesting participants to rate their experience with AROSS.

Experiment 2

Procedures

The main task for Exp 2 was to locate convex and concave parts of the mine wall ('targets'), i.e., structures sticking out of wall and going into the wall, respectively.

In each test run, participants worked through one enabled graphical interface configuration and used the robotic arm to point at the identified target for validation. The targets used are shown in Figure 6. They are located in the left and right parts of the mine wall shown in Figure 7. Target A, in Figure 7-(a) had five different levels of depth. Targets B, illustrated in Figure 7-(b) with two different levels of depth. To prevent memorization problems, the position of target A were changed with another target that had a

Table 1: Test conditions and levels in Exp 1. Disocclusion Augmentation view (DA), Direct camera views (Dir), Novel Perspective view (NP), Scene Perspective view (SP).

Independent variable	Condition level
Disocclusion Augmentation (DA)	1) 100% transparent 2) 50% transparent 3) 0% transparent
Scene Perspective (SP)	1) Direct camera views (Dir) 2) Dir + DA 3) Dir + DA + NP
Use of System	1) AROSS views (DA + SP) 2) No AROSS interface

different depth level between each test run. Also, the position of target B was changed randomly between each test run. The participants completed the task five times and each test run took approximately 3 minutes. After each test run, we asked participants to fill in their opinions in the questionnaire. After finishing an experiment we asked participants' suggestions and overall feedback in a free-form response.

The total time was estimated to about 75 minutes per participant.

Figure 6: AROSS mine-like rock wall in Exp 2.

(a) (b)

Figure 7: Targets in Exp 2. a) Target with five different levels of depth, b) Targets with two different levels of depth.

Apparatus

In this experiment five separate interfaces were presented to participants in different test runs: 1) Left Original view, 2) Right Original view , 3) Disocclusion Augmentation view (100% transparent), 4) Novel Perspective views, and 5) a pure Lidar geometry view, c.f. Figure 3. In this test, a regular non-autostereoscopic 4K display was used for presenting the interface of the remote operation system.This was a part of the improvements of the apparatus between the experiments. In Exp 2, in order to reduce the influence of monocular 3D cues from the ceiling and the floor in the depth perception tasks, all the views were zoomed in close to the mine walls to remove those monocular cues.

Augmented elements and interface

The experiment used five test runs. The test runs were "Disocclusion Augmentation view", "Left Original view", "Right Original view", "Novel Perspective view" and "Lidar view", shown in Figure 3. Left Original view was used as a training session. In order to prevent target memorization, the position of targets changed after each test run. The questions asked after each test run were as follows:

1- What is the effects of different interface views on depth estimation and 3D environment?

2- What is the effects of different interface views on controlling the robot arm?

3- How is user's experience using different interface views on depth estimation and 3D environment?

Responses were recorded on a 5-point Likert scale, see Figure 5, which was shown each the scale should be rated.

Results

Participant statistics

In both experiments the number of test persons were lower than would be ideal for quantitative statistical analysis [19], due to the pandemic.

In Exp 1 a total of 10 test participants participated, 8 males and 2 females. Youngest participant's age group was 21-30 and the oldest between 51-60 years old. Four of the test participants had experience in piloting remote-controlled vehicles/toys and two of them had some previous experience in driving trucks or heavy machinery. Three participants had participated in one of our previous studies. The visual status of the test participants were self-reported. No problems with performing the task were reported due to poor vision.

In Exp 2 a total of 11 test participants participated, 9 males and 2 females. Youngest participant's age group was 21-30 and the oldest between 41-50 years old. Three of the test participants had experience in piloting remote-controlled vehicles/toys. The visual status of the test participants were self-reported. Two participants reported their visual status:*"I have corrective glasses and slight red-green deficient color vision"* and *"I have Dyschromatopsia (deficiency in the perception of colours)"*.

Experiment 1

During the test procedure, one participant stopped the test and did not complete further test conditions, due to technical problems in the hardware of the robot arm. Therefore, only 9 users participated in the performance task to identify the targets.

The mean rate of identified targets by the participants with 95% confidence intervals in each test run are shown in Figure 8. The bar chart illustrates that the participants achieved the highest rate of identified targets when they looked directly at the scene without the remote operation interface, with a mean rate of 0.78 (test run 1). When using the AR views (test run 2-8) the results

were lower. The lowest performance were found for the combination of the Disocclusion with 50% transparent view and Direct camera view (test run 4). However, one way repeated-measures ANOVA showed that there was no statistically significant effect of different test runs on the number of correct targets identified by participants, $F(7, 56) = 1.482$, $p = 0.193$.

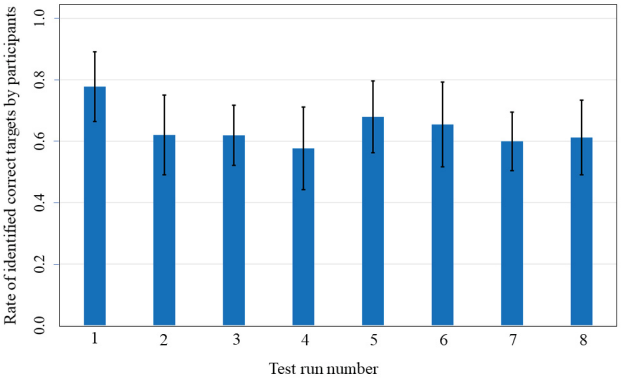


Figure 8: Rate of targets identified by participants in each test run.

Experiment 2

The participant Mean Opinion Score (MOS) with 95% confidence intervals are shown in Figures 10 – 12 on the left side of the participants' responses to the scales: "Controlling the robot arm", "Depth estimation" and "User's Impression". The figures illustrate the impact of the enabled views on the participants' ratings where 1 = Bad and 5 = Excellent. One way repeated-measures ANOVA showed that the different AR design factors had significant effects. The results for controlling the robot arm was $F(4,40) = 7.32$, $p = 0.00016$. There were also a significant effects on the depth estimation ($F(4,40) = 4.54$, $p = 0.004$) and on the user experience of using the difference interface views ($F(4,40) = 10.03$, $p = 0.000001$).

The post-hoc tests were performed using Tukey's Honest Significant Difference (Tukey's HSD) test with 95% confidence interval to determine the significant differences between the measurement. The Figures 10 – 12 shows on the right side graphically the pair-wise comparisons from Tukey's HSD of the Control of robot arm, Depth estimation and User's impression, respectively. Any confidence intervals that do not contain 0 provide an indication of a statistical significant difference in the groups. There are significant differences in controlling the robot arm between the Left Original and the Lidar view and the Right Original and the Lidar view. In depth estimation, there was a significant difference between the two Novel perspective and the Left Original pair views. Moreover, in User's impression response there were significant differences between Lidar and Left Original view as well as Lidar and Right Original view.

Experiment 1 and Experiment 2

Figure 9 shows a comparison results of the participants' MOS with 95% confidence intervals for Exp 1 and Exp 2. This illustrates the impact of Disocclusion Augmentation and Novel perspective views on participant opinions. The difference between the impacts of these two views are not statistically significant in both experiments. This was also confirmed with a T-test ($\alpha = 0.05$). The barchart shows that the users' MOS for Disocclusion Augmentation view is higher in Exp 1 than Exp 2. The reason for the decrease in the MOS level in Exp 2 is likely due to that in Exp 2 only one view at the time is presented in contrast to Exp 1 where combination of views were presented. Moreover, in Exp 2, just 100% Disocclusion Augmentation view was used, but in Exp 1 both 50% and 100% were used. In Disocclusion Augmentation view, there is a significant difference between participants' MOS in Exp 1 and Exp 2. In the Novel Perspective views, MOS of the users are almost the same in both experiments.

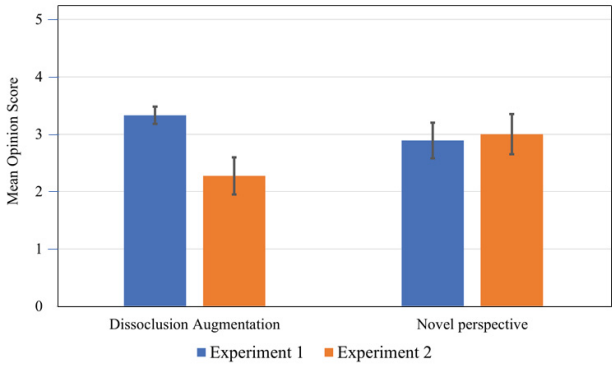


Figure 9: The MOS and 95% confidence intervals, for Disocclusion Augmentation view compares to Novel Perspective view in Experiment 1 and Experiment 2

Discussion

Experiment 1

The overall results of Exp 1 did not show a significant effect of either Disocclusion Augmentation level (100% or 50%) or the presence of Novel Perspective view on the participants' performances or opinions. The Disocclusion Augmentation were by most participants perceived more positively than the Novel Perspective views. However, this outcome may be due to a friendship bias — all test participants were colleagues at Mid Sweden University due to the ongoing COVID-19 pandemic. Moreover, the results and the free-form feedback indicate several limitations of the test setup and methodology, which were partly addressed in Experiment 2 and will be further revised in subsequent experiments.

The resolution of the cameras (1600 by 1200 pixels) were generally deemed as too low by the test participants. The lenticular cover of the autostereoscopic 4K display used to show the user interface degraded the apparent camera resolution and may have caused light nausea in one test participant. The display also caused visual artefacts in the Lidar view, turning solid white dots into an apparent cluster of red, green and blue dots that shifted colours whenever the viewers moved their head.

In both experiments the available remote manipulator (robot arm) was relatively small with limited reach, which severely restricted the options for user interaction with the remote environment, forcing the experiment task into a "point at ..." design. The Lidar used for the basis of the rendering geometry had a minimum measurement distance of approximately 0.5 m, which meant that the robot arm took up a small portion of the camera view, thus minimizing the impact of one of the independent variables (Disocclusion Augmentation) in the experiment and further worsening

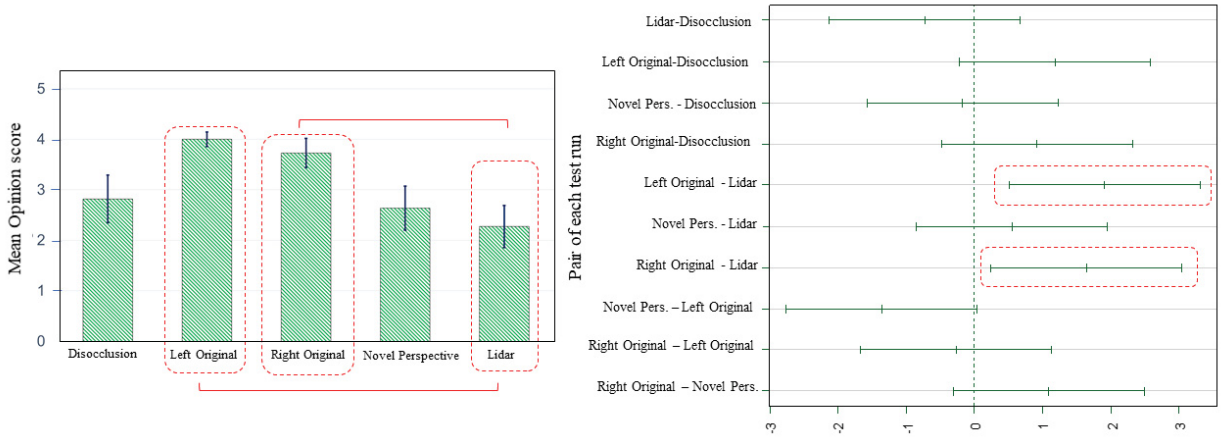


Figure 10: The left bar chart shows Mean Opinion Scores (MOS) for controlling the robot arm in Exp 2. From the left along the x-axis enabled views in each test run are shown, error bars indicate 95% confidence intervals. The right figure illustrates pair-wise comparisons from Tukey's HSD posthoc test with 95% confidence interval. If the pairwise comparison confidence interval does not contain zero, there is a significant difference between the pairs (marked also with dashed lines). The corresponding pairs are also marked in the bar chart to the left.

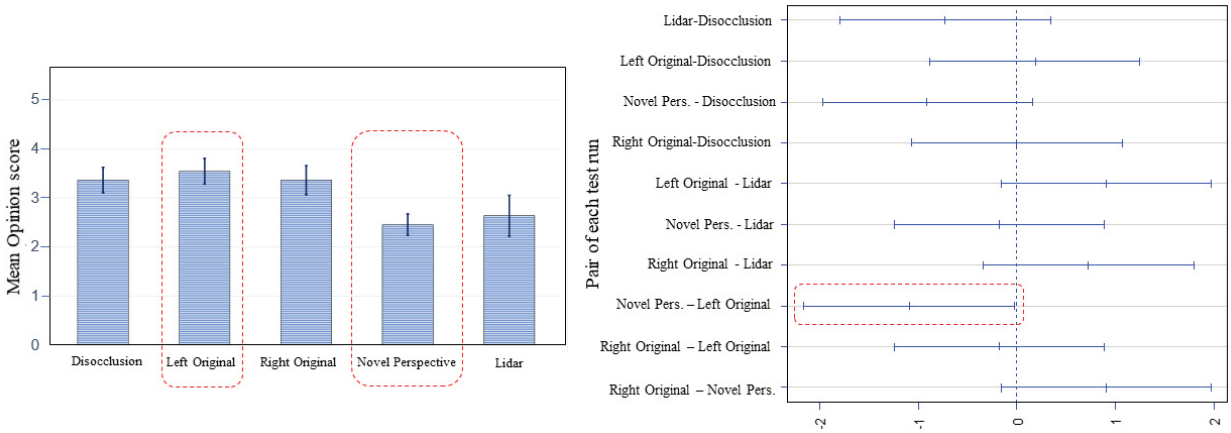


Figure 11: The left bar chart shows MOS for depth estimation in Exp 2, error bars indicate 95% confidence intervals. The right figure illustrates pair-wise comparisons from Tukey's HSD posthoc test with 95% confidence interval in depth perception. There is a significant difference between the pairs (marked with dashed lines) since confidence interval does not contain zero, the corresponding pairs marked in the left bar chart.

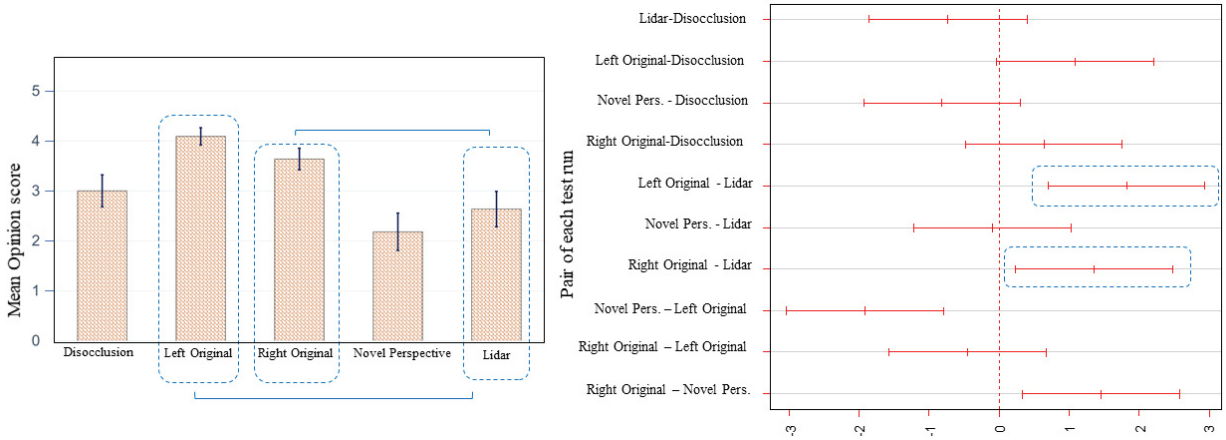


Figure 12: The left bar chart shows participants MOS for user's impression in Exp 2 with 95% confidence intervals. The right figure illustrates pair-wise comparisons from Tukey's HSD with 95% confidence interval user's impression using different enabled views.There are two significant differences between the pairs (marked with dashed lines) since confidence intervals do not contain zero, the corresponding pairs marked in the left bar chart.

the low-resolution issue.

The available targets (pictures of small regions of the rock wall) were mostly not obscured by the robot arm. Furthermore, they were co-planar and taken with a different camera than the cameras used in the AROSS system. The difference in camera settings made recognition of the structure more difficult. The lack of obstruction reduced the effects of in-view disocclusion. The co-planarity of the rock wall removed the need to understand its 3D structure, thereby reducing the impact of the Novel Perspective views.

With both of the independent variables in Exp 1 compromised, the experiment task became largely a memorization and pattern recognition problem. As one participant wrote, *"after some time I felt I memorized the targets; if target pictures remain fix then this can help to operate the system."* This shows that there were an unwanted emphasis in the experimental task of target-position recall, at the cost of on-the-fly experience and understanding of the remote environment.

There were 90 target images of approximately 30 distinct wall patterns in this experiment and although randomization of target order was used, the target variety was not sufficient to prevent pattern memorization across test runs.

Experiment 2

Results of Exp 2 show significant effects of the different interfaces on the opinions of the participants. The results of Lidar interface had significantly lower opinion scores by the users in controlling the robot arm compared to the Original (right/left) views. There was also a significant difference between the Left Original view and the Novel Perspective views, in perceiving depth and 3D environment. The Left Original view was perceived more positively than Novel Perspective view.

We asked participants to "point into the target behind of the robot arm". Most of the participants mentioned that the question was not clear, they assumed the position of themselves to be behind the robot and target located in front of the robot arm.

The majority of participants believed that the Disocclusion Augmentation interface had a minor effect on their experience and ability when working with the system. The robot arm used in both experiments were too small as an occluder, and did not obscure available targets, therefore the lack of obstruction reduced the effects of in-view disocclusion.

The participants were not satisfied with the quality of Novel Perspective view, black shadow of the robot arm, mixed colors and low quality of the view were reported in participants' feedback. In the Novel Perspective view, users could not follow the remote manipulator in the view and it reduced the usability of the system. In contrast, four users believed the interface was useful in finding geometry of targets and depth cues.

There is an asymmetry between the MOS of participants when they used the Left and the Right Original views in controlling the robot arm, depth estimation, and user's impression using the AROSS system. The reason might come from the position of the targets in the Left Original view, where all the depth levels were clearly visible. However, this was not the case in the Right Original view, where the angle of the camera made the depth cues less visible.

In the Lidar view, controlling the remote arm was hard since the users could not follow the robot arm movements in it. One of the participant gave the feedback that a combination of this view with one of the original views would be useful in using the system.

The scene showed in each interface did not include the ceiling and floor of the mine lab room. This was a design choice to remove these strong monocular 3D cues, for better testing the system's ability to convey the depth and 3D structure of the mine wall. This was done by zooming in close to the mine walls to reduce that the guidance came from the environment.

The quality of display was improved by replacing the autosterescopic display used in Exp 1 with a regular 43 inch 4K display. There were no feedback received from participants regarding the low quality of the display.

In order to prevent memorization in Exp 1, the position of targets were changed in each test run. The participants did not give any feedback regarding memorizing the pattern during the test runs. In this test design, the Right Original view was not shown in the first test run, since the view provided monocular depth cues that could help and therefore bias the participants' opinion. The other views were randomly showed in each test run.

Conclusion

In this paper, we reported on two evaluations of the influence of view augmentations and Novel Perspective views on the perception of depth and 3D space of a mine wall for remote control operators in a laboratory approximation of an underground mine. Our results show that based on task difficulty users can control the robot arm easily but finding the targets were hard. The participants' feedback show that the majority of participants preferred to use the original camera views and the Disocclusion Augmentation view instead of Novel Perspective views. In Exp 1, Disocclusion Augmentation view was percieved as beneficial. In Exp 2, Disocclusion Augmentation view was rated lower than the Novel Perspective views by the participants. In Exp 2, users were not satisfied with the quality of the Novel Perspective views, but finding depth cues were easier with these interfaces.

Future work

We foresee restructuring the experiment setup, design and methodology. The low-resolution issue discovered in Exp 1 and partly addressed in Exp 2 by upgrading the display, can be addressed via post-capture upscaling of the recorded camera views. Removal of compression artifacts (and handling frames dropped during transmission) should also be implemented. From the results of Exp 1 and Exp 2, there is a need to add a camera view attached to the robot arm. For the Disocclusion Augmentation view larger obscuring objects or robot in front of the targets are needed to show the ability of the interface. Keeping the geometry of the remote manipulator is important in the Lidar view, the Disocclusion Augmentation view and Novel Perspective views. The quality of the the Novel Perspective views should be improved in the future experiments. Using augmented depth guidance on top of each enabled feature would help participants in perceiving the depth and the 3D environment.

From a methodology standpoint, the test environment must be redesigned with more emphasis on the environment 3D structure and variation in depth. Furthermore, that variation has to be within interactive range of the remote manipulator, to enable interaction beyond merely tasks based on pointing. The experiment task should then be changed to emphasize seeing behind the oc-

cluders, and interacting with the remote environment in a manner that relies on understanding the 3D structure of said environment. If using predetermined targets for the test tasks, care should be taken to avoid repeating targets between different test run configurations within one participant's session.

Acknowledgement

The economic support from the Swedish Foundation for Strategic Research (grant nr FID18-0030) and Vinnova (Sweden's innovation agency; grant nr 2021-02107) are hereby gratefully acknowledged.

References

[1] Paolo Tripicchio, Emanuele Ruffaldi, Paolo Gasparello, Shingo Eguchi, Junya Kusuno, Keita Kitano, Masaki Yamada, Alfredo Argiolas, Marta Niccolini, Matteo Ragaglia, and Carlo Alberto Avizzano, A Stereo-Panoramic Telepresence System for Construction Machines, Procedia Manufacturing, 11, 1552–1559 (2017).

[2] Kim Gyun-Hyung, Kim Ki-Duck, Lee Hyeon-Seung, Choi Yun-Sung, Mun Ho-Seong, Oh Jae-Heun, and Shin Beom-Soo, Development of Wi-Fi-Based Teleoperation System for Forest Harvester, Journal of Biosystems Engineering, 1–11 (2021).

[3] Guglielmo Carra, Alfredo Argiolas, Alessandro Bellissima, Marta Niccolini, and Matteo Ragaglia, Robotics in the Construction Industry: state of the art and future opportunities, ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 35, pg. 1–8. (2018).

[4] Felipe Sánchez and Philipp Hartlieb, Innovation in the Mining Industry: Technological Trends and a Case Study of the Challenges of Disruptive Innovation, Mining, Metallurgy, and Exploration 37, 1385–1399 (2020).

[5] Fabio Bruno, Antonio Lagudi, Loris Barbieri, Domenico Rizzo, Maurizio Muzzupappa, and Luigi De Napoli, Augmented reality visualization of scene depth for aiding ROV pilots in underwater manipulation, Ocean Engineering, 168, 140–154 (2018).

[6] Bence Bejczy, Rohat Bozyil, Evaldas Vaičekauskas, Sune Baagø Krogh Petersen, and Simon Bøgh, Sebastian Schleisner Hjorth, and Emil Blixt Hansen, Mixed Reality Interface for Improving Mobile Manipulator Teleoperation in Contamination Critical Applications, Procedia Manufacturing, 51, 620–626 (2020).

[7] ITU-T. Vocabulary for performance, quality of service and quality of experience (ITU-T Rec. P.10/G.100). International Telecommunication Union (ITU), Place des Nations, CH-1211 Geneva 20, (2017).

[8] Patrick Le Callet, Sebastian Möller, and Adrew Perkis. Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003) (Version 1.2 (http://www.qualinet.eu/images/stories/QoE-whitepaper-v1.2.pdf)), Lausanne, Switzerland, (2012).

[9] Tobias Hoßfeld, Poul E Heegaard, Martín Varela, and Sebastianr Möller, QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS, Quality and User Experience, 1, 2 (2016).

[10] Judith Schoonenboom and R. Burke Johnson, (2017). How to Construct a Mixed Methods Research Design. Kolner Zeitschrift fur Soziologie und Sozialpsychologie. 69(Suppl 2): p. 107-131, DOI: 10.1007/s11577-017-0454-1.

[11] Elijs Dima, Mårten Sjöström, Camera and Lidar-Based View Generation for Augmented Remote Operation in Mining Applications, IEEE Access, 9, 82199–82212 (2021).

[12] Fumio Okura, Masayuki Kanbara, and Naokazu Yokoya, Augmented Telepresence Using Autopilot Airship and Omni-directional Camera, IEEE International Symposium on Mixed and Augmented Reality, pg. 259–260. (2010).

[13] Kjell Brunnström, Elijs Dima, Tahir Qureshi, Mathias Johanson, Mattias Andersson, and Mårten Sjöström, Latency impact on Quality of Experience in a virtual reality simulator for remote control of machines, Signal Processing: Image Communication, 89, 116005, DOI: 10.1016/j.image.2020.116005, (2020).

[14] Hamed Z. Jahromi, Ivan Bartolec, Edwin Gamboa, Andrew Hines, and Raimund Schatz, You Drive Me Crazy! Interactive QoE Assessment for Telepresence Robot Control, 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), pg. 1–6. (2020).

[15] Elijs Dima, Kjell Brunnström, Mårten Sjöström, Mattias Andersson, Joakim Edlund, Mathias Johanson, and Tahir Qureshi, Joint effects of depth-aiding augmentations and viewing positions on the quality of experience in augmented telepresence, Quality and User Experience, 5, 1–17, DOI: 10.1109/QoMEX.2019.8743147, (2020).

[16] Balazs P. Vagvolgyi, Will Pryor, Ryan Reedy, Wenlong Niu, Anton Deguet, Louis L. Whitcomb, Simon Leonard, and Peter Kazanzides, Scene Modeling and Augmented Virtuality Interface for Telerobotic Satellite Servicing, IEEE Robotics and Automation Letters, 3, 4241–4248 (2018).

[17] Michael E. Walker, Hooman Hedayati, and Daniel Szafir, Robot Teleoperation with Augmented Reality Virtual Surrogates, 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pg. 202–210. (2019).

[18] Kjell Brunnström, Börje Andrén, Bo Schenkman, Anders Djupsjöbacka, and Omars Hamsis, Recommended precautions because of Covid-19 for perceptual, behavioural, quality and user experience experiments with test persons in indoor labs, Digital systems networks, RISE report 84, DOI: 10.23699/j865-cz77, (2020).

[19] Kjell Brunnström and Marcus Barkowsky, Statistical quality of experience analysis: on planning the sample size and statistical significance testing. Journal of Electronic Imaging. 27(5): p. 11, DOI: 10.1117/1.JEI.27.5.053013., (2018).

Author Biography

Shirin Rafiei received her B.Sc. and M.Sc. degrees in electrical engineering from Babol Noshirvani University of Technology and Azad University, Iran, in 2009 and 2014, respectively. Since 2019, she has been a researcher and Ph.D. student at Research Institute of Sweden (RISE), and Mid Sweden University. Her research interests include Quality of Experience, User Experience, Augmented Telepresence, ive video streaming and Augmented Reality.

Elijs Dima received his B.Sc. and M.Sc. degrees in Computer Engineering from Mid Sweden University, Sweden, in 2013 and 2015. He received his Lic. degree in 2018 and his Ph.D. degree in 2021 from Mid Sweden University, focusing on Augmented Telepresence based on Multi-Camera Systems. At the time of writing this publication, he was a researcher, lab supervisor, teaching assistant and post-doc in the Realistic 3D research group at Mid Sweden University. His research interests include 360-degree video and light field capture, rendering and streaming, parallel data processing, and the synchronization, calibration, modeling, and development of Virtual Reality, Augmented Reality and multi-camera systems.

Mårten Sjöström received the M.Sc. degree in electrical engineer-

ing and applied physics from Linköping University, Sweden, in 1992, the Licentiate of Technology degree in signal processing from the KTH Royal Institute of Technology, Stockholm, Sweden, in 1998, and the Ph.D. degree in modeling of nonlinear systems from EPFL, Lausanne, Switzerland, in 2001. He was an Electrical Engineer with ABB, Sweden, from 1993 to 1994, and a Fellow with the CERN, from 1994 to 1996. In 2001, he joined Mid Sweden University, where he was appointed as an Associate Professor and a Full Professor in signal processing, in 2008 and 2013, respectively. He founded the Realistic 3D Research Group in 2007. He is the head of the research subject Computer and System Science, and Computer Engineering, since 2013 and 2020, respectively. He is part of the Faculty Board since 2021. His current research interests include multidimensional signal processing and imaging, system modeling, and identification.

Kjell Brunnström Ph.D., is a Senior Scientist at RISE Research Institutes of Sweden AB and Adjunct Professor at Mid Sweden University. He is an expert in image processing, computer vision, image and video quality assessment having worked in the area for more than 30 years. He is leading standardization activities for video quality measurements as Co-chair of the Video Quality Experts Group (VQEG). His current research interests are in Quality of Experience for visual media in particular video quality assessment both for 2D and 3D, including AR, VR and remote operation of machines, as well as display quality related to the TCO Certified.