Quality and User Experience (2020) 5:2
https://doi.org/10.1007/s41233-020-0031-7

RESEARCH ARTICLE

Joint effects of depth-aiding augmentations and viewing positions on the quality of experience in augmented telepresence

Elijs Dima[1] · Kjell Brunnström[1,2] · Mårten Sjöström[1] · Mattias Andersson[3] · Joakim Edlund[1] · Mathias Johanson[4] · Tahir Qureshi[5]

Received: 17 August 2019 / Published online: 10 February 2020

Abstract

Virtual and augmented reality is increasingly prevalent in industrial applications, such as remote control of industrial machinery, due to recent advances in head-mounted display technologies and low-latency communications via 5G. However, the influence of augmentations and camera placement-based viewing positions on operator performance in telepresence systems remains unknown. In this paper, we investigate the joint effects of depth-aiding augmentations and viewing positions on the quality of experience for operators in augmented telepresence systems. A study was conducted with 27 non-expert participants using a real-time augmented telepresence system to perform a remote-controlled navigation and positioning task, with varied depth-aiding augmentations and viewing positions. The resulting quality of experience was analyzed via Likert opinion scales, task performance measurements, and simulator sickness evaluation. Results suggest that reducing the reliance on stereoscopic depth perception via camera placement has a significant benefit to operator performance and quality of experience. Conversely, the depth-aiding augmentations can partly mitigate the negative effects of inferior viewing positions. However the viewing-position based monoscopic and stereoscopic depth cues tend to dominate over cues based on augmentations. There is also a discrepancy between the participants' subjective opinions on augmentation helpfulness, and its observed effects on positioning task performance.

Keywords Quality of experience · Augmented reality · Telepresence · Head mounted displays · Depth perception

Introduction

Applications of Virtual Reality (VR) and particularly Augmented Reality (AR) are becoming increasingly important for non-entertainment applications. Driver assistance systems based on volumetric AR are developing [1, 2] and likely to appear in consumer cars in the near future. Beyond

This work has been funded by the Knowledge Foundation (Grant No. 20160194) and the EU Regional Development Fund (Grant No. 20201888), which are gratefully acknowledged.

✉ Elijs Dima
Elijs.Dima@miun.se

Kjell Brunnström
Kjell.Brunnstrom@miun.se; kjell.brunnstrom@ri.se

Mårten Sjöström
Marten.Sjostrom@miun.se

Mattias Andersson
mattias.andersson@miun.se

Joakim Edlund
Joakim.Edlund@miun.se

Mathias Johanson
mathias@alkit.se

Tahir Qureshi
tahir.qureshi@hiab.com

[1] Department of Information Systems and Technology, Mid Sweden University, Sundsvall, Sweden

[2] Division ICT-Acreo, RISE Research Institutes of Sweden, Kista, Sweden

[3] Department of Design, Mid Sweden University, Sundsvall, Sweden

[4] Alkit Communications AB, Mölndal, Sweden

[5] HIAB AB, Hudiksvall, Sweden

teleconferencing, which already benefits from VR and stereoscopic displays [3], AR is expected to gain significant traction in various industry structures in the coming years [4]. By providing a form of telepresence and enhanced views of the world, AR and VR can improve worker safety in construction [5], improve satellite repairs via teleoperation [6], and expand marine and air traffic control in airports [7] and shipyards [8].

In particular, immersive telepresence systems for vehicle operation are fast becoming a reality in industrial applications [9, 10], thanks to the push towards 5G low-latency communications. AR is likewise gaining traction as a method for assisting operators in performing specialized tasks [5, 6], leading to Augmented Telepresence applications for industry, in use cases like the remote operation of truck mounted forestry cranes shown in Fig. 1. We use the term Augmented Telepresence (AT) [11] to denote applications where high-quality video-mediated communication is the enabling technology, but where additional data can be superimposed on or merged with the video as in AR. AT is similar to AR in that it tries to present additional information on top of the view seen by the user. It primarily differs from AR in that the user is present in a remote location seeing the augmented view of the location, with optional two-way audio or audio-visual communication. As such, AT is a form of immersive telepresence.

Based on the aforementioned cases and ongoing research, Head-Mounted Display (HMD) based AT has potential for improving user Quality of Experience (QoE) in industrial applications. However, if AR assistance is employed in industrial immersive telepresence systems as demonstrated in [9], then the operator QoE and task performance may be affected by the overlap of AR and stereoscopy in unknown ways to an unknown extent. Therefore, studies are required to address the impacts of AR and stereoscopy on operator QoE.

Fig. 1 Photo of VR-headset based crane operation from inside a truck cabin—a case of augmented telepresence in an industrial application [10]

This study focused on how depth-aiding AR affects remote positioning in immersive telepresence systems, where the task dependency on stereoscopic depth perception varies and different viewing positions are used. The individual and joint effects of these two factors—AR and stereoscopy dependence—were examined from a QoE perspective. Subjective participant opinions and task performance metrics are considered in a 27-participant study where a telepresence system is used for remote positioning tasks. The immersive telepresence system comprises real-time stereoscopic video from several stereo cameras with AR in a VR HMD. The positioning task is modelled from the problem of remote control of industrial machines such as cranes, excavators and loaders and is based on actual use of similar systems within respective industries. The industrial problem and context is replicated in a lab setting, to isolate the AR and stereoscopy factors.

An initial analysis of only the stereoscopy factor from this study was presented in [12]. Beyond that work, this paper covers the full experiment, includes analysis of the AR factor and the joint effect of AR and stereoscopy, discusses the study and related work in greater detail, and provides new results with new analysis of performance-based metrics and participant QoE, both independently and in connection to simulator sickness. Thus, the novelty and contribution of this paper is an assessment of the individual and joint influence of perception aiding factors (AR, stereoscopy) on QoE during a positioning task, in the context of remotely controlling industrial machinery through immersive AR–VR telepresence systems. The results of this study may be of general interest to all designers, engineers and researchers focusing on QoE, AR and mixed reality experiences, and telepresence systems, in particular using HMDs.

Related work

Quality of experience for virtual, augmented environments

In general, QoE is the degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility and enjoyment of the application or service in the light of the user's personality and current state [13, 14]. The above

definition of QoE, which is also pointed out by Möller and Raake in [15], goes beyond the traditional QoE and Quality of Service (QoS) research and makes an overlap with the User Experience (UX) research tradition, by incorporating the experience and behaviour of users when being confronted with systems or services [16–18].

Traditionally, in QoE research, the methods to gain insight into the delivered quality of a service and the users' experience of it have been done through controlled laboratory experiments, where the opinions of panels of users have been collected. The results are typically reported using Mean Opinion Score (MOS). These methods are very often referred to as subjective quality assessment methods and there are standardized ways of conducting them e.g. for visual quality, ITU-R Rec. BT.500-13 [19] or ITU-T Rec. P.910 [20]. These methods have been criticized for not providing enough ecological validity, i.e. that they insufficiently represented the real world situation [21]. Improvements in response to these claims have been done for example in ITU-T Rec. P.913 [22]. The investigations into 3-Dimensional (3D) video quality a few years ago, when the 3D-TV hype was the most intense, resulted in three ITU Recommendations [23–25]. It was discovered that if care was not taken, several user experience issues such as discomfort and visual fatigue may occur.

An attempt to build an experimental framework for QoE of AR was made by Puig et al. [26], who advocated a combination of subjective assessment (e.g. questionnaires, subjective ratings) and objective measurements (e.g. task completion time, error rates). Reliance solely on subjective assessment, also known as explicit metrics, has been criticised by Kroupi et al. in [27] and Hoßfeld et al. in [28], highlighting the importance of including objective (i.e. implicit) metrics in QoE assessment. Implicit metrics include measurements of participant-task interaction as suggested in [26], as well as measurements of participant psychophysical state during the experiment, as discussed in [27]. The connection between psychophysical measurements and perceived QoE in VR has been brought up by Keighrey et al. in [29]. Recently, Concannon et al. [30] used psychophysical measurements together with opinion scores to investigate the effects of network delay on remote operation in virtual reality, and Barreda-Ángeles et al. [31] used implicit metrics to investigate the sense of immersion and realism in 360-degree video based VR. Surveys of the state of psychophysical measurements in QoE research are given by Engelke et al. in [32], Barreda-Ángeles et al. in [33], and Bosse et al. in [34]. A further review of current status of QoE research for immersive media involving AR, VR and HMDs is given in the VQEG eLetter (vol 3 issue 1), via summary articles [35–38].

QoE studies for VR environments often explicitly measure cyber-sickness or simulator sickness in addition to visual quality and task performance. Brunnström et al. [10, 39] investigated simulator sickness in VR systems replicating varied delays in a tele-operation context while assessing the effects of presentation delay from user input delay during a remote operation task, in a simulation of the use case shown in Fig. 1. Schatz et al. [40] conducted a lab-based study to compare various rendering techniques for a VR training application. Tran et al. [41] assessed the QoE of 360-degree videos under different displays, video contents, and encoding qualities. Both studies highlight simulator sickness as part of the QoE factors, in addition to context-specific subjective metrics. A particular aspect in 360-degree video quality assessment is the QoE of streaming video. As Singla et al. point out in [42], there is no complete standard for assessing streaming 360-degree video, and few studies that focus on streaming have included simulator sickness as part of the assessment. Standardization of streaming 360-degree video is in progress. Recently, Curcio et al. [43] have adapted ITU-T 2D and 3D video testing recommendations to 360-degree video in VR HMD, Tran et al. [44] have assessed and categorized QoE aspects that affect 360-degree video streaming, and Pérez and Escobar [45] have proposed an assessment-aiding tracking tool for an upcoming ITU-T Recommendation for 360-degree VR video, ITU-T Rec. P.360-VR [46]. A broad survey of the state of the art in 360-degree video aspects, including streaming and QoE, is given by Fan et al. in [47].

Besides data compression and delivery, the display technology and scene presentation affects the user QoE. The effect of using VR and AR HMDs instead of regular displays was analysed in [48]. It was found that HMDs increase the cognitive load of a participant compared to conventional displays, particularly for headsets with narrow Field of View (FoV). In HMD-supported telepresence, as we demonstrated in [12], the choice of viewing position in a telepresence system has a significant impact on operator tasks, when those tasks depend on the operators' position-estimation abilities. This occurs because different viewing positions place different demands on the operators to estimate positions and depths via stereoscopy or other depth cues. Effects of depth estimation tasks may be aggravated in HMD, since inaccurate stereoscopic depth estimation in HMD is another known, display technology dependent issue [49–51].

Depth perception in virtual, augmented environments

Headsets and other stereo-enabled displays affect depth perception [49, 52], as people tend to underestimate depths in VR and AR. A survey on positioning in mixed reality [53] listed the types of tasks used to analyze depth perception in VR and AR: depth estimate verbal reporting, eye-body coordination tests such as walking or remote object

movement via joysticks, and object-depth interaction such as picking, placing, and throwing. The survey also noted the need for research on whether depth perception allows for better localization in real-world scenes shown through AR and VR headsets.

In a headset environment, Pointon et al. [49] showed that VR and AR headsets induce similar errors on estimated depths. Depth estimation in virtual environments was surveyed by Lin et al. [50], finding that egocentric distance estimation is about 80% accurate in virtual environments, compared to 94% in the real world. This decrease of accuracy may adversely impact depth estimation and positioning tasks. Cutolo et al. [51] suggested that HMD optics may be the cause for spatial perception errors, and that camera convergence improves near-distance 3D object positioning [54]. However, these publications did not address the QoE of keystone artefacts from converged cameras, or the influence of camera FoV. Such factors would need to be analyzed before the converged-camera recommendation can be accepted as an improvement over the conventional parallel stereo camera setup, especially for full-system QoE with see-through rendering between the cameras and the HMD devices.

In a non-headset environment with volumetric and 2-Dimensional (2D) AR see-through displays, Lisle et al. [2] found that AR shown on volumetric 3D displays outperformed AR shown on 2D panels for estimating depths greater than 5 m, which also implies a potential benefit of 3D AR in headset-based environments for specific task performance. In mixed depth cue environments, Berning et al. [55] showed that monoscopic cues dominate over stereoscopic cues, however stereoscopy assists in challenging tasks. These findings extended the earlier work of Nagata [56], which showed the impact of distance on real world depth cue dominance. Similarly, Diaz et al. [57] showed that in AR, monoscopic cues such as shadows significantly aid depth estimation. Rizek et al. [58] found a split result in a QoE study on stereoscopic 3D teleconferencing systems. By comparing a 2D and 3D system, it was shown that subjectively, participants preferred the QoE of the 2D system, whereas the objective performance in depth-based tasks was significantly better in the 3D system.

Improving perception or task performance with AR

Besides depth estimation, a number of works have focused on improving perception and task performance using AR cues, i.e. AR designs intended to convey supplementary information to the user. Albarelli et al. [59] have categorized AR cues as collaborative (see-through) and competitive (overlay), and assessed an AR guidance system. However, the study had only 10 participants, and the designed AR cues overlaid a large fraction of the display FoV. A different use of AR cues for guidance was shown by Bork et al. [60] to guide user attention to out-of-view objects. The AR design influenced the guidance effectiveness, and several types of flat, map-like, Heads-Up Display (HUD)-like and sphere-like designs were tested. However, guidance towards objects already within the camera and display FoV was not explored.

The camera FoV and viewpoint effect on AR environment interactions was examined by Taira et al. [61] and Sun et al. [62]. The term *viewpoint* is hereby meant as the position from which the observer is viewing the scene. Both studies found that camera placement had a significant effect on the resulting user interaction, however both studies focused only on desktop-adjacent environments for person-to-person video communication via digital interfaces. A similar study was performed by Lager et al. [63], comparing 2D and 3D views in desktop interfaces for remote piloting of unmanned boats. 3D views were advantageous for the remote control task, but it should be noted that this study did not include any telepresence or AR aspects besides the simulated interface viewpoint on a flat display.

Immersive AR for remote vehicle control was investigated by Vagvolgyi et al. [6], by describing an AR system for satellite cutting and welding in space. AR was used to show a replica model of the satellite, augmented by real camera view projections, and AR guides for optimal cutting and welding paths. Experiments showed a significant improvement in task completion time and accuracy based on the AR system. However, this system did not include a stereo camera system, and did not assess operator QoE. Walker et al. [64] used a similar system to aid flying drone operators, with AR guides for drone flight path and rendering virtual replicas of drones along the flight path. They found a slight objective and subjective benefit from using AR guides, but no significant difference between various types of AR. It should also be pointed out that the study focused on the virtual replicas of drones instead of augmenting the flying drone's real-time positioning. A real-time AR system was presented in [65], used to notify car drivers when to brake. Various AR real-world integrations were tested, showing the benefit of integrating AR into the world and leveraging monocular depth cues such as expected object size. However, the tests were not carried out in a headset-based immersive environment.

AR design in see-through immersive environments was analyzed by Volmer et al. in [66]. A special button-laced dome was built and shown through a VR game, with blinking, colored, and pointer-extending AR cues designed to direct the user to pressing a specific button. The study found AR to be beneficial, and in particular that the cues should clearly indicate the next action without interference between current and next action markers. However, the AR was used only to guide the user to a target, not to help with target 3D positioning or remote object control. Kytö et al. [67] used AR in a similar system to help users to make depth

judgements, by relying on the binocular disparity and relative size of the AR cues. This approach improved the depth judgements, even though the AR cues were fixed in space and did not track the user-manipulated objects. AR has been used to set a user-selected 3D position interactively in the real world by rendering guiding lines in AR, in the study by Lages et al. [68]. Nevertheless, that system was not connected to any remotely operated devices to manipulate the observed space or act upon the specified position.

Manipulation of the observed space via AR was analyzed in robot teleoperation studies by Brizzi and Peppoloni et al. in [69, 70]. A headset-based AR environment was used to give a first-person perspective of the robot arm actuation, with AR cues added for robot arm's motion vectors. In a pick-and-place task, the AR was found to be beneficial for task accuracy and QoE. However, these studies focused on humanoid robots and egocentric robot embodiment, mimicking the operator arm motions with the robot arm and providing a nearly top–down view perspective and minimal object depth variation. From a non-egocentric perspective, Uddin et al. [71] used AR for presenting additional position data for a Remote Controlled (RC) robot arm pick-and-place task. The AR was found to be useful, however the AR was used to render an additional view perspective instead of augmenting the primary real world view, and there was no stereoscopic headset to provide an immersive AR telepresence environment.

User study

Aiming to test the influence of two potentially competing factors—stereoscopic depth perception and AR cues—on positioning in a remote operation context, we designed a user study that tests these factors via a remote telepresence system with AR presented through a VR HMD. After training, participants performed a navigation-and-control task for all factor permutations, and reported their subjective experiences of using the system. The system was tracking the time and position of the remote-controlled object throughout the experiment.

System implementation

We used a VR headset that shows a live stereo-camera passthrough with AR overlays. This implementation choice was made because there is little difference between AR and VR headsets for 3D object manipulation in terms of task completion [72], and VR headsets have a higher FoV, which reduces cognitive load compared to narrow FoV headsets [48] such as most contemporary AR HMDs. A point of difference between VR and AR headsets is the evident delay visible to the user. In a see-through AR HMD, there is bound to be a delay between the visible parts of the real world and the rendered AR, leading to perceivable registration errors and system latency, as discussed by Kruijff et al. in [73]. In contrast, VR headsets display an environment that is equally delayed from the real world in all sections of the display, thereby presenting a more coherent environment. However, the isolation from seeing the real world leads to tighter constraints on the display rendering latency in order to prevent nausea and simulator sickness. The experiment system was built on our multi-camera real-time streaming system [74], and connected to a 2nd generation HTC Vive Pro HMD. This VR display was chosen because previous studies [75–77] showed a slight benefit to the HTC Vive's tracking and resolution compared to other headsets, and as a consumer product suited for immersive videogames, the HTC Vive Pro has acceptable levels of rendering delay.

Two pairs of synchronized Basler da-A1600-60uc cameras were used, with an 8 cm baseline in each pair. The cameras had 110-degree wide angle lenses, recording 1600-by-1200 pixels at 24 Frames per Second (FPS), with a 9.2 ms exposure. The camera views were rectified and projected on an enclosing sphere in the HMD virtual environment. AR overlays were projected in front of the sphere surface, anchored to lines between target points in camera views and the HMD eye positions, as illustrated in Fig. 2. This setup provided stereoscopic AR that reacts to omnidirectional 6-Degrees of Freedom (DoF) headset movement, as per DoF definitions in [78]. Since the camera positions were

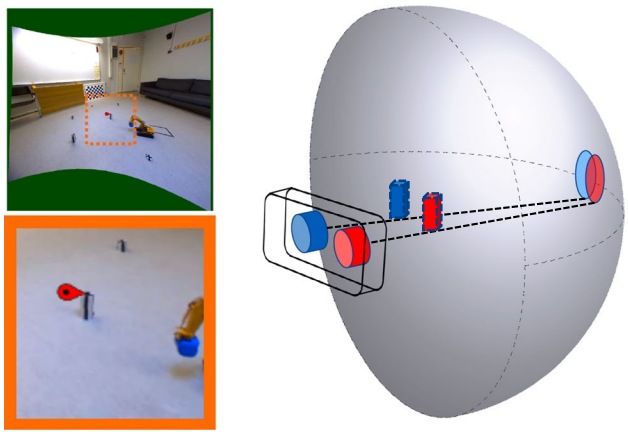

Fig. 2 Left: Single-eye view of an augmentation (red pin) in front of the sphere-projected camera view (distorted image). Right: Principle of stereoscopic augmentation rendering on enclosing sphere projection of camera views. Each eye of the headset observes a different texture projected on the enclosing sphere. The augmentation for each eye is anchored in the 3D virtual space on a line connecting the virtual eye position to a texture point on the sphere surface. This texture point is selected based on the visual content and objects detected in the image texture, as presented to the viewer. The augmentation therefore has correct perceived stereoscopy, regardless of headset position changes or content disparity variations

fixed, the scene rendering had 3 DoF. The importance of stereo-correct AR was showed by Sun et al. [79], demonstrating that stereo cursors are subjectively preferable in augmented environments, and stereo-correct AR markers may thus assist with depth perception. The in-headset VR was rendered at $\geqslant$ 90 FPS to avoid inducing delay-based nausea [39] in participants. The target points in camera views (such as the remote control crane tip) were detected via color separation and matching in HSV space. There exist more advanced color-based tracking methods such as [80]; we chose the basic approach to reflect an AR context based on standard computer vision tools. Since target points for the AR markers were identified in camera frames prior to rendering the frames, the perceived position of the AR matched the rendered scene at all frames. Thus, the system enforced coherence between the rendered AR and the rendered camera views.

Figure 3 shows the lab environment, with a RC Hulna 1507 1:14th-scale excavator, 6 targets of different heights placed approximately 70 cm apart from each other, and two observing camera pairs. The excavator was controlled via two joysticks and four buttons on the wireless remote shown next to the excavator in Fig. 3. The joysticks controlled the excavator movement, and four buttons controlled the excavator crane with two buttons (open, close) per crane joint. The low camera pair was placed 10 cm above the floor, to give a side view of the scene. The high camera pair was placed 120 cm above the floor, to give a three-quarters overview of the scene. Both camera pairs were pointed at the center of the 220-by-300 cm test area.

Participants

A total of 27 participants were recruited from the local university student and staff population, more than the recommended minimum of 24 participants [81]. All participants were volunteers and gave their informed consent to the experiment protocol and data analysis. Prior to the test, each participant was asked to fill out a pre-questionnaire to report on their age, gender, vision, eyewear, and sensitivity to sea sickness and motion (car) sickness. The sensitivity query was stated as two separate questions, "Do you get seasick easily?" and "Do you get car-sick easily?", each with either "Yes" or "No" answer. None of the participants had previous experience in remote operation of industrial machinery. Ten participants mentioned past experience in having used or tried a HMD previously. The degree of past experience or time spent with HMDs was not reported. Four participants had participated in a different subjective study involving a HMD. None of the participants had been otherwise involved in the field of VR development or research.

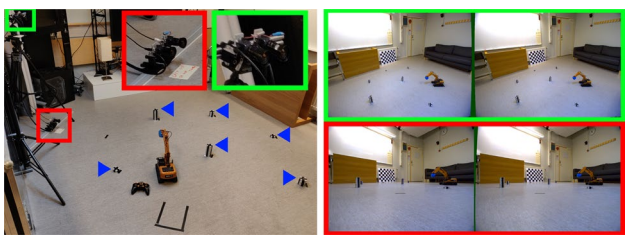

Fig. 3 Left: Experiment setup with cameras, targets, and RC excavator. Overhead and Ground camera pairs outlined with green and red frames, respectively. Target objects highlighted by blue triangles. Right: The rectified views from Overhead and Ground camera pairs

The participants, 5 women and 22 men, were aged from 20 to 54, with a mean age of 30.63 years (std. deviation: 8.98). 18 participants wore glasses or contact-lenses. All participants had normal or corrected-to-normal vision according to their self-reporting, and one participant had red-green color blindness. Due to time constraints, the participants did not undertake a vision test prior to the experiment. Two participants reported difficulties seeing 3D—one had corrective glasses, and the other was blind in the right eye. Both participants showed good ability to perform the task, and therefore their test data was included in the results. Five participants reported being easily car-sick and sea-sick, one participant reported being easily car-sick but not sea-sick, and three participants reported being easily sea-sick but not car-sick. All participants were allowed to cancel the test at any time for any reason; none chose to do so. One participant had to cancel the experiment after three attempts due to power outage, and one participant did not report their opinion on augmentation helpfulness. Their responses were excluded from analysis for the missing attempts and categories.

Experiment design

We wanted to see how stereoscopic depth perception and AR affect remote navigation and positioning. The task dependence on participant stereoscopic depth perception, hereafter called *stereoscopy dependence factor*, was varied by setting two viewing positions of the test environment—a high *Overhead* view, where participants are less reliant on depth perception, and a low *Ground* view, where depth perception via stereopsis is essential for performing the task. Stereo baseline of the cameras was the same in both view positions, thus the stereoscopy dependence factor described how much the users had to rely on stereopsis. This setup is similar to the 2D and 3D view comparison in [63]. We used third-person views i.e. the operated machinery is seen from outside, since first-person views i.e. views of the operated machinery from inside or ego-centric perspective have been covered in remote control teleoperation by [69, 70]. The AR factor was varied by three *Augmentation* designs—a target-specific marker (A1), a 2D grid-map (A2) and a 3D scaling cube

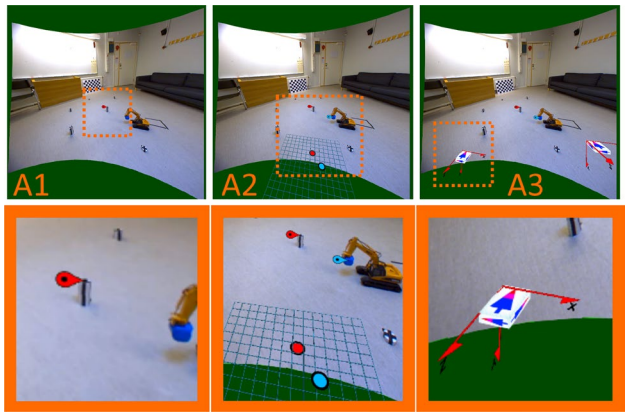

Fig. 4 AR designs used in test. Top: Left eye view prior to HMD distortion, cropped for visibility. Bottom: zoom-ins of indicated parts. A1: Highlight of the selected target only. A2: 2D grid map of selected target and excavator tip. A3: 3D 'cube' indicator, showing distance from excavator tip to selected target via cube's x, y, z scale

(A3), shown in Fig. 4. Therefore, 6 test runs were specified, for all permutations of view and augmentation. As recommended by Puig et al. in [26], we captured subjective and objective measurements for each test run by letting the participants fill out a subjective assessment form on paper. This form is further detailed in Sect. 3.4. The total test length was kept to approximately 30 min : 5 min for briefing, 5 min for HMD fitting and remote-control training, and six 2-min test runs with 1–2 min for subjective reporting after each run. As [82] showed, taking short breaks from VR headsets reduces visual fatigue, compared to continuous VR HMD use. The overall duration of the experiment was kept short to avoid fatiguing the participants, as suggested by Curcio in [36].

In each test run, the participants were asked to navigate and accurately touch highlighted targets in the test area, using the remote control excavator's arm. This is a combination of the locomotion and pick-and-place tasks commonly used for depth estimation and positioning assessment [53, 69–71, 76], adapted to the context of remotely controlling industrial machinery. The targets were highlighted in random order, and participants were scored on the number of targets reached. The order of the test runs was randomized. Only one target was highlighted at a time to avoid confusion, as suggested by Volmer et al. [66].

Measures and analysis

For subjective assessment of participant QoE, participants were asked to rate their "ability to precisely reach the targets" (Task Accomplishment), "the difficulty of reaching the targets" (Task Difficulty), "the viewing position helpfulness in precisely reaching the targets" (Viewpoint Helpfulness), and "the augmentation helpfulness in precisely reaching the targets" (Augmentation Helpfulness).

Fig. 5 Scales for Task Difficulty (top) and Task Accomplishment, Augmentation Helpfulness, Viewpoint Helpfulness (bottom)

The ratings were attained via 5-point Likert scales, shown in Fig. 5. The scales were designed specifically for this investigation, focusing on answering the research questions of our study.

The first scale in Fig. 5 was not strictly labelled as a quality scale, but we judged the chosen labels to be more natural in relation to the question that the subjects were asked to rate. During analysis, the scales were treated as numerical interval scales [83] ranging from 1 to 5, and used to calculate the MOS for each of the four scales. Paired-sample T-tests with Bonferroni-correction [84] were used to check the significance of MOS differences. For each scale, two-way Repeated-measures ANOVA was used to assess the impact of the stereoscopy factor (viewpoint position) and AR design.

For task performance measurements, the excavator crane tip position ($[x_c, y_c, z_c]$), highlighted target position ($[x_h, y_h, z_h]$), and number of targets reached N_{reached} were logged every 50 ms by the experiment system. The coordinates were resolved by the target point detection system described in Sect. 3.1. The *Number of Targets Reached* N_{reached} per run was the first of four task performance metrics, the other three being *Error to Target*, *Time to Target*, and *Time near Target*. The Error to Target metric $\epsilon_{\text{toTarget}}$, shown in Eq. (1), was the distance between the top center of the target object and the bottom of the excavator crane tip, at the moment when participants indicated they had touched the target object.

$$\epsilon_{\text{toTarget}} = \sqrt{(x_c - x_h)^2 + (y_c - y_h)^2 + (z_c - z_h)^2} \tag{1}$$

Time to Target $t_{h,p}$ measured the time participants spent between targets, from the moment t_p of touching a previous target p and the moment t_h of touching the highlighted target h, normalized by the distance between h and p. Therefore, this metric described how efficient the participants' driven route was from target to target.

$$t_{h,p} = \frac{t_h - t_p}{\sqrt{(x_h - x_p)^2 + (y_h - y_p)^2 + (z_h - z_p)^2}} \tag{2}$$

Similarly, Time near Target $t_{\text{near}\,h}$ measured the time participants spent with the crane tip in close proximity (within 20 cm) of the highlighted target. Whereas $t_{h,p}$ described

the overall route efficiency, $t_{\text{near } h}$ described how long participants spent on fine-tuning the excavator crane's position leading up to the touch event. In Eq. (3), t_{prox} is the moment when $\sqrt{(x_c - x_h)^2 + (y_c - y_h)^2 + (z_c - z_h)^2} = 20$ cm.

$$t_{\text{near } h} = t_h - t_{\text{prox}} \tag{3}$$

As recommended by Brunnström et al. in [81], the measurement distributions were checked for normality using Pearson Chi-square, Kolmogorov–Smirnov and Jarque–Bera tests. Bonferroni-corrected paired-sample *T*-tests were used to determine the significance of differences of each measurement. For multi-factor effects, Repeated-measures ANOVA test was used to examine the factor interactions.

Once before starting the experiment and once after completing the training session and all test runs, participants also filled out the Simulator Sickness Questionnaire (SSQ), in order to complete the QoE assessment and check that experiment results aren't influenced by significant simulator sickness. In our experiment, the participants were shown a 16-symptom, 5-point SSQ that had been used in [12], with 16 symptoms as defined by Kennedy et al. [85]. Of all participants, only 3 participants indicated a "Strong" response in a single symptom each, and no participants indicated any "Severe" response. During analysis, we merged the "Strong" and "Severe" responses to a single response "severe", thereby mapping the responses presented to participants ("None, Slight, Moderate, Strong, Severe") to the responses defined in the Kennedy et al. SSQ ("None, Slight, Moderate, Severe"). In analysis, each symptom's response was converted to a number by "None"= 0, "Slight"= 1, "Moderate"= 2,"Severe"= 3. The symptom values were aggregated into Nausea (N), Oculomotor (O), Disorientation (D) and Total Score (TS) groups, using the formula and weights defined in [85]. Each group was analyzed for differences between pre- and post-experiment scores via paired sample *T*-tests.

Results

SSQ results

The SSQ was analyzed according to procedure described in Sect. 3.4. Figure 6 shows the pre-experiment and post-experiment symptom scores in Nausea (N), Oculomotor (O), Disorientation (D) and Total Score (TS) groups. Paired sample *T*-tests showed that each of the 4 groups has a significant difference before and after the test (N: $p = 0.01$, O: $p < 0.01$, D: $p = 0.03$, TS: $p < 0.01$) at the

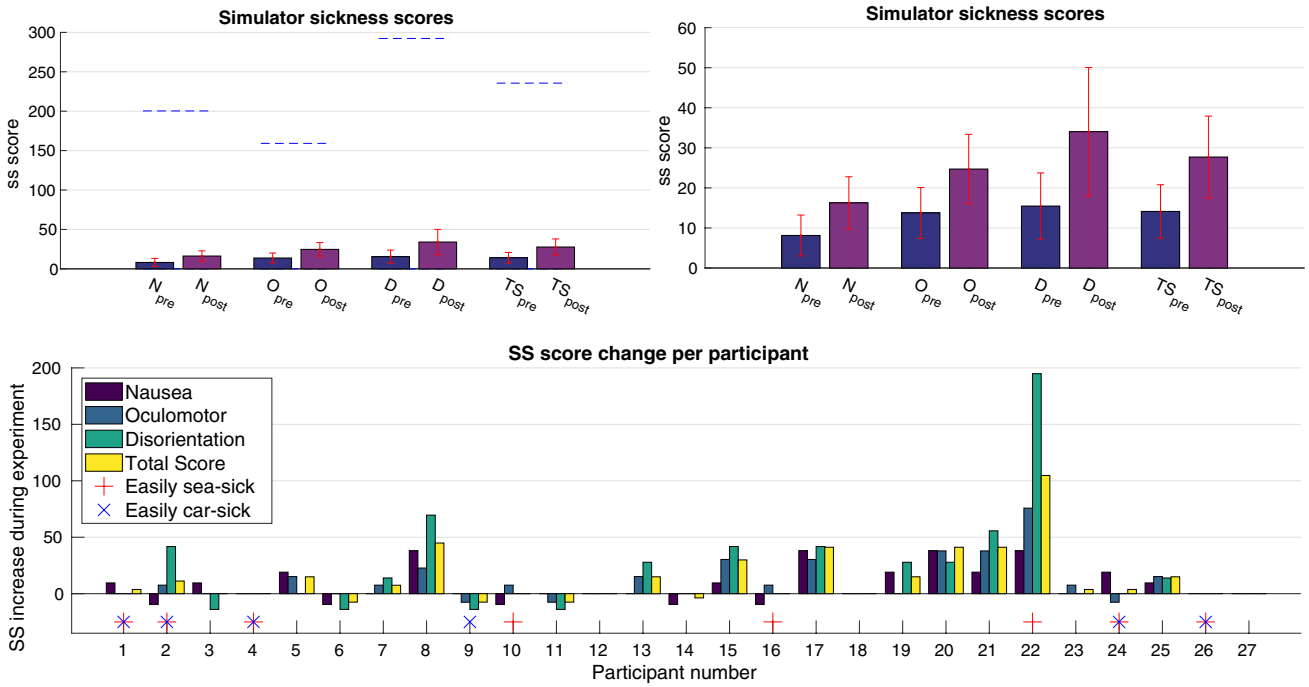

Fig. 6 Top: The mean and 95% confidence interval for pre-experiment and post-experiment simulator sickness scores. Symptom groups are (N)ausea, (O)culomotor, (D)isorientation and Total Score (TS). Top Left: dashed lines show maximum attainable scores for each symptom group, as per formula in [85]. Top Right: scores from the left side, without maximum lines. Bottom: the change from pre-experiment to post-experiment simulator sickness scores (SS increase = SS $\text{score}_{\text{post}}$ − SS $\text{score}_{\text{pre}}$), per participant, per symptom group. Missing bars indicate no change

$\alpha = 0.05$ level. A Repeated-measures ANOVA on each of the groups showed no significant interactions between time (pre or post) and participant age, gender, or use of eyeglasses. As shown in Fig. 6, one participant experienced relatively large increase in simulator sickness symptoms during the test (+38.16 Nausea, +75.8 Oculomotor, +194.88 Disorientation, +104.72 Total Score), some participants experienced minor increase in simulator sickness, and some participants experienced no change or minor decrease in simulator sickness. Of 27 participants, five showed no change at all in any symptom group, 13 showed an increase in all or some symptom groups, four showed decrease in all or some symptom groups, and five showed a mixed response (minor increase in some, minor decrease in other symptom groups). There were no statistically significant correlations between the participant responses and test sequence orders.

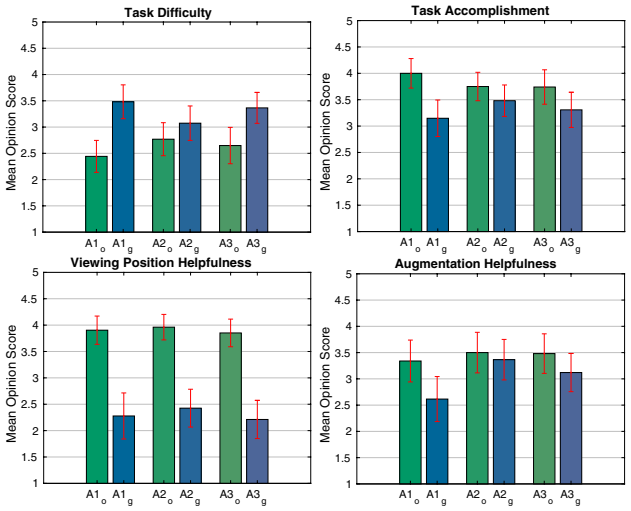

Fig. 8 The MOS and 95% confidence intervals, for three AR designs (A1, A2, A3) and two viewpoint positions ([o]verhead, [g]round)

Subjective participant opinion

The histograms of participant responses in Fig. 7 indicates that, on the whole, participants used the full range of the response scales. Pearson Chi-square and Jarque–Bera tests confirmed the response normality. The participant MOS with 95% confidence intervals are shown in Fig. 8. Repeated-measures ANOVA showed that the stereoscopy dependence factor (overhead and ground viewpoint position) was significant for the Task Difficulty scale ($F_{1,25} = 30.6$, $p = 9.3 * 10^{-6}$), Task Accomplishment ($F_{1,25} = 20.3$, $p = 1.3 * 10^{-4}$), Viewpoint Helpfulness ($F_{1,25} = 56.4$, $p = 7.3 * 10^{-8}$), and Augmentation Helpfulness scale ($F_{1,24} = 7.1$, $p = 0.013$). The AR design factor (A1, A2, A3) was significant only in the Augmentation Helpfulness response scale ($F_{2,48} = 4.0$, $p = 0.023$). The interaction of AR design and stereoscopy dependence factor was significant only in the Task Difficulty scale ($F_{2,50} = 4.7$, $p = 0.013$).

Based on T-tests for each scale and AR design, we found significant ($\alpha = 0.05$) differences between overhead and ground viewpoint MOS in Task Difficulty (A1: $p = 3.5 * 10^{-5}$, A3: $p = 2.5 * 10^{-4}$), Task Accomplishment (A1: $p = 0.001$, A3: $p = 0.017$), Augmentation Helpfulness (A1: $p = 0.006$) and Viewpoint Helpfulness (A1: $p = 3.6 * 10^{-6}$, A2: $p = 2.9 * 10^{-7}$, A3: $p = 11.6 * 10^{-7}$) scales. In T-tests for each scale and viewpoint, there were only two significant differences between AR designs. In Task Difficulty, overhead view, there was a significant difference between A1 and A2 ($p = 0.012$, at Bonferroni-corrected $\alpha = 0.017$). In Augmentation Helpfulness, ground view, again A1 and A2 were significantly different ($p = 0.010$).

Task performance measurements

The four objective metrics, described in Sect. 3.4, were analysed with Repeated-measures ANOVA and Bonferroni-corrected T-tests to determine the effect of stereoscopy and AR design factors. The position of the excavator tip was continuously tracked in camera image coordinates, from which disparity and real-world 3D position were estimated. Figure 9 shows that the disparity ranged from ≈ 20 to ≈ 45 pixels for overhead view, and from ≈ 30 to ≈ 70 pixels for

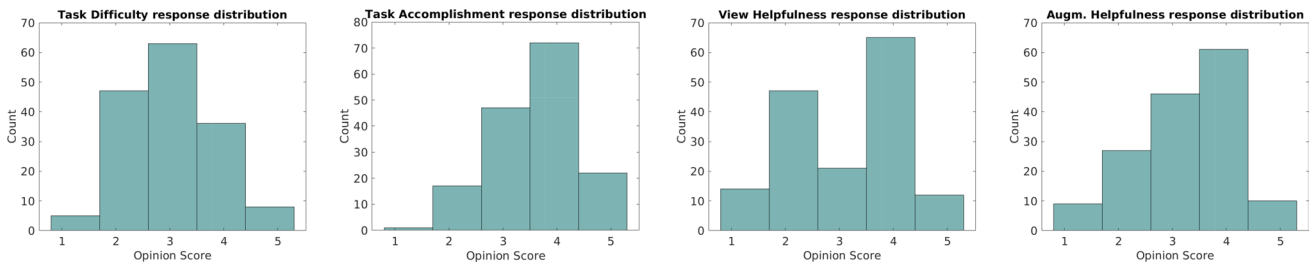

Fig. 7 Participant response histograms for Task Difficulty, Task Accomplishment, Viewpoint Helpfulness and Augmentation Helpfulness, after conversion from Likert scales to numerical scales with 1–5 range

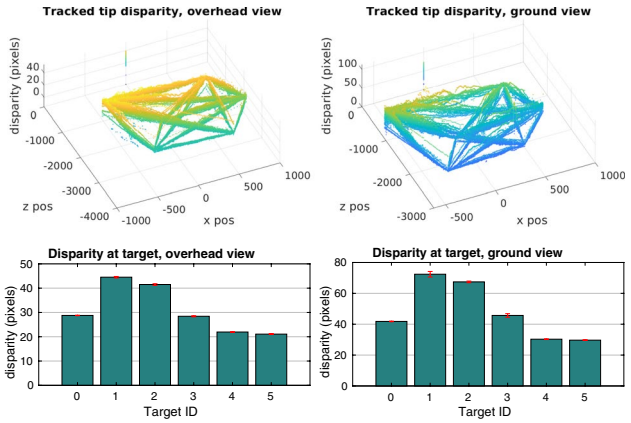

Fig. 9 Stereo disparity of the tracked excavator tip for overhead and ground views. Top: disparity mapped on the optimal path between targets. Point color based on disparity value. Bottom: disparity at target positions, when users 'touch' targets with excavator tip

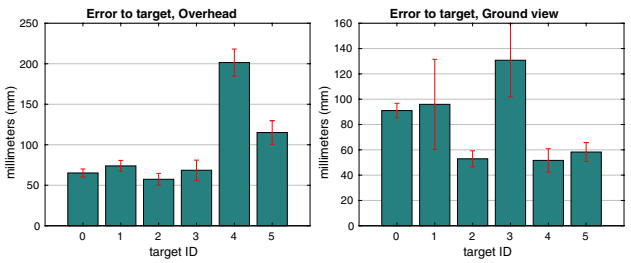

Fig. 10 Mean and 95% confidence interval of the measured positioning error, per target, for overhead view (right) and ground view (left)

ground view, where nearer targets corresponded to higher disparity. However, there was some disparity fluctuation, which together with camera calibration inaccuracies led to target-specific offsets in the estimated 3D positions, compared to the expected target positions. Each camera position had a separate stereo camera calibration, coordinate system, and fixed target positions. Figure 10 shows the measured error from excavator tip to target for each of the targets, for all participants. To reduce the impact of target-specific depth estimation errors on the analysis, only the three targets with smallest error of each view were considered for the Error to Target $\epsilon_{\text{toTarget}}$ analysis throughout this section. The other three metrics $N_{\text{reached}}, t_{h,p}, t_{\text{near }h}$ were not dependent on depth estimation accuracy, and were analysed for all targets.

2-Way repeated-measures ANOVA showed that the stereoscopy dependence factor (i.e. viewpoint factor) was a significant predictor for all four metrics: number of targets reached ($F_{1,25} = 65.7$, $p = 1.8 * 10^{-8}$), error to target ($F_{1,18} = 8.6$, $p = 0.009$), time to reach target ($F_{1,23} = 70.4$, $p = 1.9 * 10^{-8}$), and time spent near target ($F_{1,25} = 83.6$, $p = 1.9 * 10^{-9}$). Figures 11 and 12 show the corresponding measurements. Two of the ANOVA tests had reduced

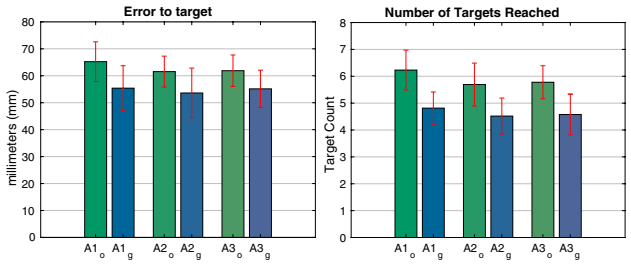

Fig. 11 Left: mean and 95% confidence interval for error to target. Right: mean and 95% confidence interval for number of targets reached per run. A1, A2, A3 are AR designs, 'o' and 'g' denote overhead and ground viewing position

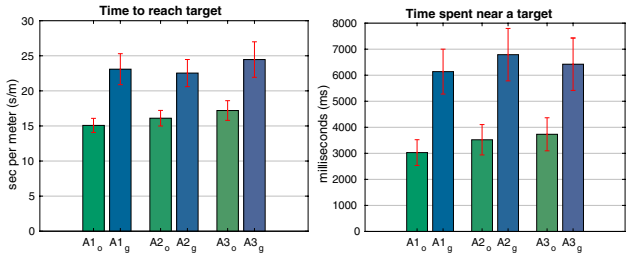

Fig. 12 Left: mean and 95% confidence interval for time to reach a new target, normalized by distance between targets. Right: mean and 95% confidence interval for time spent near a target (within 20 cm). A1, A2, A3 are AR designs, 'o' and 'g' denote overhead and ground viewing position

degrees of freedom because some participants' targets could not be tracked due to the excavator body occluding the crane tip; participants with the missing runs were excluded from the ANOVA. The AR design factor was not significant on any of the metrics. The *T*-tests confirmed that there were significant differences between overhead and ground views in number of targets reached (A1: $p = 5.9 * 10^{-6}$, A2: $p = 6.0 * 10^{-5}$, A3: $p = 2.1 * 10^{-4}$), time to target (A1: $p = 1.4 * 10^{-3}$, A2: $p = 9.1 * 10^{-6}$, A3: $p = 3.3 * 10^{-5}$), and time near target (A1: $p = 8.8 * 10^{-5}$, A2: $p = 1.7 * 10^{-4}$, A3: $p = 1.4 * 10^{-4}$) measurements. However, the *T*-tests did not find a significant difference between viewpoints in the error to target measurements. There were also no significant differences found between AR designs.

During the tests, we noticed that some participants had greater ability to learn the remote-controlled excavator controls than other participants, in a shorter time. Since participant training speed was not recorded, we instead decided to categorize participants into three groups based on their individual performance during the experiment, measured as each participant's mean number of targets per run. This performance was used as an approximate indicator of participant skill in both operating the remote-controlled excavator

and accomplishing the experiment task. A k-means clustering algorithm [86, 87] was used to identify three participant groups with no overlap and consistent participant grouping across repeated initializations of the clustering algorithm. Here we denote these groups as the Low skill (<4 targets), Medium skill (4–6 targets), and High skill (>6 targets) participant groups, shown in Fig. 13. One out of 6 participants in the low-skill group, 5 out of 14 participants in the medium-skill group, and 4 out of 7 participants in the high-skill group had used a VR headset before the experiment. The participant with red–green colour blindness was in the medium-skill group. The participant with a blind eye was in the low-skill group. The participant that reported difficulties seeing 3D without corrective glasses was in the medium-skill group. A repeated-measures ANOVA on participant skill showed significance in the viewpoint factor ($F_{1,25} = 24.1$, $p = 4.8 * 10^{-5}$) and, to a lesser extent, in the AR design factor ($F_{2,50} = 3.5$, $p = 0.0375$).

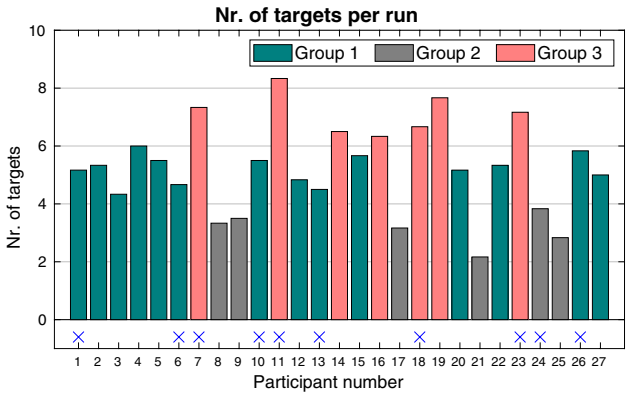

Fig. 13 Participants grouped in three clusters approximating participant skill, based on each participant's average number of targets per run. Groups 1, 2, 3 correspond to the so-called Medium-, Low-, and High-skill categories, respectively. Crosses below the bars indicate participants who had previously used a VR headset

From the above ANOVA tests, the AR design factor had sporadic effect on participant skill, Task Difficulty, and Augmentation Helpfulness. We used these three measures for participant grouping, to investigate group responses to AR designs. The participant skill groups are shown in Fig. 13. The Task Difficulty and Augmentation Helpfulness groups were categorized into Positive (MOS $\geqslant 4$), Ambivalent (MOS between 2 and 4), and Negative (MOS $\leqslant 2$) opinion groups. Bonferroni-corrected *T*-tests between AR designs within each group showed a significant difference at the 95% confidence level between AR designs A1 and A3 in the time taken to reach target, when grouping by participant skill. For low-skill participants, there was a significant difference ($p = 0.013$, $\alpha = 0.0167$) between A1 and A3 from overhead view. For medium-skilled participants, there was a significant difference ($p = 0.010$, $\alpha = 0.0167$) between A1 and A3 from ground view. For high-skilled participants, again there was a significant difference ($p = 0.010$, $\alpha = 0.0167$) between A1 and A3 from overhead view. In all these cases, A3 increased the mean time taken to reach the next target, compared to A1. Figure 14 shows the time to reach targets for each of the three participant skill groups. The other group categories did not show any significant differences between AR designs.

Discussion

Although this study did not test a commercial augmented telepresence system in a truck, the influence of stereoscopic depth perception dependence and different assisting AR designs on remote positioning was assessed from a QoE perspective. Therefore, the experiment results covered the participant opinions, task performance measurements, and simulator sickness factors.

We did find a significant increase in SSQ symptoms in all four simulator sickness dimensions (nausea, oculomotor,

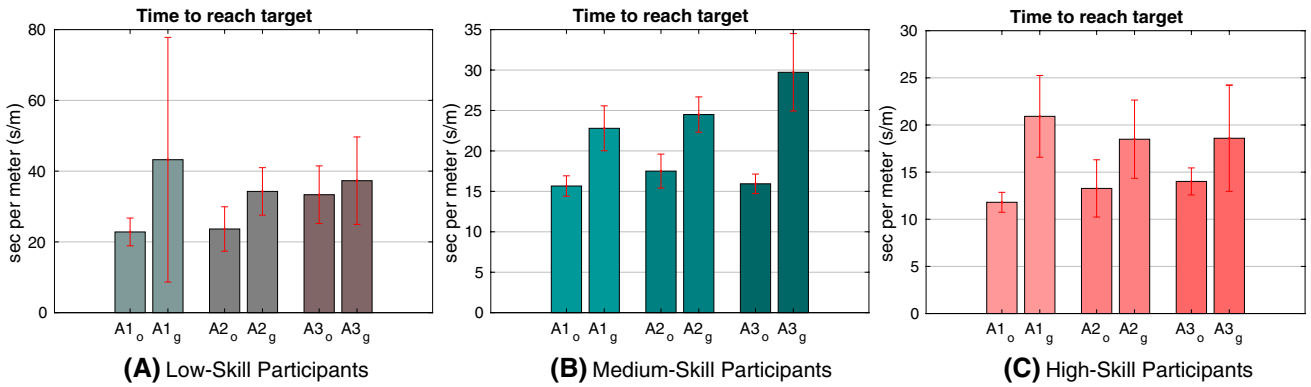

Fig. 14 Time to reach target, for low-skilled (A), medium-skilled (B) and high-skilled (C) participants. Time is normalized by the distance between targets. A1, A2, A3 are AR designs, 'o' and 'g' denote overhead and ground viewing position

disorientation, total score) from the test. On average, the largest impact was on participant disorientation when considering the group weighting proposed by Kennedy et al. [85]. However, the observed values were a small fraction of the actual ranges (at most, a change from 15.5 to 34.0 in the range of 0–292.3 for mean disorientation, an increase of 6.32% of the disorientation range), predominantly corresponding to None or Slight symptoms. Only one participant reported "Moderate" or higher in several symptoms in the post-test SSQ, and even then the effect was not strong enough to cause the participant to cancel the experiment despite the rise in their simulator sickness score. Curiously, several participants verbally remarked on a decrease in symptom severity after the test. However, in reported measurements the decreases corresponded to a change from Slight to None for a minority of symptoms, and did not indicate any particular trends with regard to specific symptoms or experiment sequences. Based on this, it is possible that these symptom decreases were psychosomatic, possibly connected to the difference in context and environment within the experiment area (isolated, calm, with clear tasks and goals) and outside (typical everyday work or learning facilities), or influenced by other unknown factors specific to the participants. Overall, the low magnitude of the simulator sickness factors for most participants indicates that the subjective and objective measurements in the study are based on the tested viewing positions and AR designs, not overwhelmed by effects of simulator sickness.

According to the subjective participant opinions, both the viewing position and the AR design had a noticeable effect on the experiment task. All four opinion scales showed that the QoE dropped by approximately 1–2 units when participants had to rely on stereoscopic depth perception (due to ground view) for performing the positioning task. However, AR assistance via the A3 (cube) and especially A2 (grid) designs reduced the difference between ground and overhead viewing positions, according to task difficulty and task accomplishment MOS. This suggests that depth-assisting AR can mitigate the negative influence of the stereoscopy dependence, and compensate for viewing positions that emphasize depth perception in telepresence systems.

By itself, the AR factor variance only affected the Task Difficulty, and only via a joint effect with the stereoscopy dependence factor. Nevertheless, participants generally rated the two active-assistance AR designs' helpfulness as "Fair" to "Good", implying at least some perceived benefit towards the overall QoE. Three participants mentioned that they used the shadows cast by the excavator crane, instead of relying on the AR guides. In related studies, Berning et al. [55] determined that monoscopic cues dominate over stereoscopic depth cues in depth perception, and Diaz et al. [57] specifically showed the importance of shadow cues for position estimation. This aspect may also explain why the influence of changing viewing position, which changes the visibility of monocular positioning cues, was much greater than the influence of using different AR designs.

The task performance measurements also showed that the viewing position had a significant, substantial impact on the participants' effectiveness in completing the task. When participants had to rely on stereoscopic depth perception, even with AR assistance, the participants reached fewer targets on average, spent significantly more time getting to the targets and took longer to do fine adjustments in target proximity. This verifies the implications of previous studies [49–51] on depth perception inaccuracies in HMD VR. Although adaptation and learning over time is possible, results suggest that operators would perform better in telepresence systems that do not place stereoscopic depth perception as a necessity for task completion. Contrary to the results from the subjective evaluation, the depth-assisting AR did not reduce the performance-based measurement gap between ground and overhead views as effectively, compared to the influence on the subjective measurements.

The AR design variance showed no significant effects in most of the metrics directly related to fine motor control. However, based on participant feedback and the tracking disparity oscillation shown in Fig. 9, it was evident that the remote vehicle controls and the optical tracking and depth estimation would have to be more precise, in order to increase participant confidence in the AR and prevent them from using monoscopic cues instead. When grouped by skill, participants showed a distinction between the A1 and A3 AR designs for time taken to reach the targets. A3 increased the time taken to reach a target, compared to 'no-ar' A1, which implies a less efficient pathing or routing from target to target. This might have been because A3 presented the directions to target as X, Y, Z component vectors instead of a single direction or spatial positions. One of the participants had remarked that it was hard to relate the X, Y, Z distance vectors to joystick controls on the vehicle's remote control. Even in this case, the impact of the stereoscopy dependence factor (i.e. viewpoint) was still dominant; the AR design comparison results motivate the need to use AR that provides directly useful information (such as the A2 grid design, which gives relative positions). With the potentially increased cognitive load from using VR HMD [48], complicated AR guides such as A3 can be detrimental to task performance in some capacity.

Overall, our results imply that a significant loss in subjective rating and positioning task performance can be seen when forcing users to perform stereoscopic depth estimation by themselves in HMD-based telepresence or teleoperation systems. While certain types of AR can be used to mitigate this loss, and users at least experience that they benefit from AR, the choice of camera placement and the resulting task dependence on stereoscopic depth perception

is more impactful for the overall QoE. Considering the contrast between neutral-to-positive subjective ratings of AR and the neutral-to-detrimental effects of certain AR designs on task metrics, an interesting problem arises in context of QoE—should more importance be placed on the subjective user opinion, or the task performance aspects? This may be an open issue especially for the QoE of long-term use of immersive telepresence, when effects such as cognitive load, adaptation, simulator sickness and task learning become prevalent.

Study limitations

Concerning limitations of the study, we did not include a baseline test mode without any AR or HMD. A no-headset, no-AR baseline would have introduced unaccounted variables into the experiment design, such as resolution mismatch between the display and the world, as well as mismatch between viewing positions and participant poses. Without a minimal AR such as the marker in A1 design, there would have been no consistent way to indicate the next target object in the baseline run, compared to the other runs. For similar reasons, related studies such as [60] had also chosen to omit a baseline mode.

Because of the selected components, the system inherently had a mixed-framerate baseline: the cameras recorded the scene at 24 FPS, whereas the virtual environment and all aspects of rendering and augmentation were performed at $\geq$ 90 FPS. In principle, such a discrepancy might affect participant QoE and simulator sickness. However, in this experiment, the effects of such framerate discrepancy would not be as evident to a test participant due to several factors - guaranteed alignment between AR and scene state, the static camera viewpoint of a mostly static scene, relatively slow movement of the remote controlled excavator, and no other physical feedback from the scene to the participant beyond the camera views. Due to these factors, the scene content did not change rapidly enough for the camera framerate to become problematic (no evident motion blur or multiple-pixel "jumping" of the moving excavator). It is possible that some participants could have been perceptive enough to notice the framerate difference regardless, despite the mitigating factors and absence of any such reports from the participants, however a dedicated investigation of this aspect was outside the scope of our study and technical setup.

The participants were given time to practice operating the excavator and using the system without any AR, but they were not given instructions on how to use the AR overlays, except for the next-target highlight marker, which was constant in all test runs. We withheld this information to avoid inducing our own bias, and to allow participants to decide whether and to what extent each AR design should be used. This, however, might have made participants more willing to decide to ignore the AR information, compared to e.g. a forced attention requirement.

Within the post-test questionnaires, we specified only one subjective rating per each measure of the QoE, for a total of four ratings. In general, having multiple subjective ratings per each measure with a summary rating would provide more robustness and enable self-consistency checking during analysis. We decided to reduce the ratings to one per each QoE measure to avoid causing boredom or frustration in participants, and taking too much of their time. We felt this was a legitimate risk due to already asking the participants to fill out participation forms, pre- and post-experiment SSQ, and six repeated subjective assessment forms, one per each test run. Other ways of administering these forms, e.g. by virtual questionnaire within the HMD as suggested by Regal et al. in [88], might be viable for maintaining participant engagement during the experiment.

Lastly, there were no elevated occluders placed in front of the overhead views, because we did not want to introduce environmental occlusion as a variable linked to viewing position, and we did not seek to replicate a particular environment with specific occluder distributions. For a completely application-related experiment, such as replicating a forested area with branches and leaves, occluders would have to be included in the experiment design.

Future studies within the real environments, using industrial machinery and pole- or drone-mounted cameras, are necessary to examine how environment factors such as occlusions, camera sway, increased background complexity, and communication delays affect the operator QoE. Likewise, a study on the effects of extended use of such systems may be necessary, to examine the interplay of AR, stereoscopy, cognitive load, and simulator sickness in long-duration sessions. Under extended use or more rapid in-scene movement, a discrepancy between camera recording and virtual rendering framerates may also impact operator QoE and contribute to simulator sickness. A dedicated investigation of the impact of such a discrepancy would be beneficial to inform the design of applied AT systems.

Conclusion

In this work, we investigated how different AR and viewing positions affect the QoE in immersive telepresence systems, by means of a 27-participant experimental study. A remote controlled excavator was used for a positioning and navigation task, while viewing and controlling the excavator through an immersive telepresence system. We studied the effects of two different kinds of position-assisting AR (2D grid and 3D scaling object) and two different viewing positions, one of which forced participants to rely heavily

on stereoscopic depth perception for positioning, the other transforming depth to a 2D projection in a perspective stereoscopic view.

In our experiments, the participant QoE and positioning task performance was significantly reduced by the viewing position that forced participants to rely on stereoscopic depth perception. Position-assisting AR was able to mitigate this negative effect for task performance metrics in the case of the 2D grid AR. However, AR that was difficult for participants to correlate to excavator controls (3D scaling cube) proved ineffective and reduced the participant navigation (routing) efficiency. Based on subjective ratings, both of the depth-assisting AR designs were rated neutral-to-positive, suggesting a beneficial perceived effect on participants' QoE.

Between viewing position and AR, the viewing position had a more pronounced effect on subjective ratings and all task performance metrics. From this, we conclude that viewing positions in telepresence systems significantly affect user QoE, and best positions should reduce user reliance on stereoscopic depth perception for task completion. AR can be used to partly compensate the impact of bad viewing positions, however the extent of such compensation is subject to AR design, complexity, and implementation specifics. These findings are directly applicable to the design and implementation of augmented telepresence systems intended for industrial machinery control systems. Moreover, the interaction between the AR and stereoscopy factors may be of general interest to all designers of headset-based mixed reality experiences.

Acknowledgements Open access funding provided by Mid Sweden University.

Funding This work has been funded by the Knowledge Foundation (Grant No. 20160194) and the EU Regional Development Fund (Grant No. 20201888).

Compliance with ethical standards

Informed consent Informed consent was obtained from all individual participants involved in the study.

Conflicts of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Merenda C, Kim H, Tanous K, Gabbard JL, Feichtl B, Misu T, Suga C (2018) Augmented reality interface design approaches for goal-directed and stimulus-driven driving tasks. IEEE Trans Vis Comput Gr 24(11):2875–2885
2. Lisle L, Tanous K, Kim H, Gabbard JL, Bowman DA (2018) Effect of volumetric displays on depth perception in augmented reality. In: Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications, ACM, New York, AutomotiveUI '18, pp 155–163
3. Anton D, Kurillo G, Bajcsy R (2018) User experience and interaction performance in 2d/3d telecollaboration. Future Gener Comput Syst 82:77–88
4. Kohn V, Hardborth D (2018) Augmented reality—a game changing technology for manufacturing processes? In: Twenty-sixth european conference on information systems (ECIS2018)
5. Li X, Yi W, Chi HL, Wang X, Chan AP (2018) A critical review of virtual and augmented reality (vr/ar) applications in construction safety. Autom Constr 86:150–162
6. Vagvolgyi BP, Pryor W, Reedy R, Niu W, Deguet A, Whitcomb LL, Leonard S, Kazanzides P (2018) Scene modeling and augmented virtuality interface for telerobotic satellite servicing. IEEE Robot Autom Lett 3(4):4241–4248
7. Gürlük H, Gluchshenko O, Finke M, Christoffels L, Tyburzy L (2018) Assessment of risks and benefits of context-adaptive augmented reality for aerodrome control towers. In: 2018 IEEE/AIAA 37th digital avionics systems conference (DASC), pp 1–10
8. Fraga-Lamas P, Fernández-Caramés TM, Blanco-Novoa Ó, Vilar-Montesinos MA (2018) A review on industrial augmented reality systems for the industry 4.0 shipyard. IEEE Access 6:13358–13375
9. Tripicchio P, Ruffaldi E, Gasparello P, Eguchi S, Kusuno J, Kitano K, Yamada M, Argiolas A, Niccolini M, Ragaglia M, Avizzano CA (2017) A stereo-panoramic telepresence system for construction machines. Proced Manuf 11:1552–1559 27th international conference on flexible automation and intelligent manufacturing, FAIM2017, 27–30 June 2017, Modena
10. Brunnström K, Dima E, Andersson M, Sjöström M, Qureshi T, Johanson M (2019) Quality of experience of hand controller latency in a virtual reality simulator. In: Chandler D, McCourt M, Mulligan J (eds) Human vision and electronic imaging 2019, Society for Imaging Science and Technology, pp HVEI–218
11. Okura F, Kanbara M, Yokoya N (2010) Augmented telepresence using autopilot airship and omni-directional camera. In: IEEE international symposium on mixed and augmented reality 2010, IEEE Xplore, vol IEEE international symposium on mixed and augmented reality 2010 science and technolgy proceedings, pp 259–260
12. Dima E, Brunnström K, Andersson M, Sjöström M, Edlund J, Johanson M, Qureshi T (2019) View position impact on QoE in an immersive telepresence system for remote operation. In: 2019 eleventh international conference on quality of multimedia experience (QoMEX) (QoMEX 2019), Berlin
13. ITU-T (2017) Vocabulary for performance, quality of service and quality of experience. Report ITU-T Rec. P.10/G.100, International Telecommunication Union (ITU), ITU Telecommunication Standardization Sector
14. Le Callet P, Möller S, Perkis A (2012) Qualinet white paper on definitions of quality of experience. European network on quality

of experience in multimedia systems and services (COST Action IC 1003). Qualinet, Lausanne
15. Möller S, Raake A (2014) Quality of experience–advanced concepts, applications and methods. T-labs series in telecommunication services. Springer, Switzerland
16. Bevan N (2008) Classifying and selecting UX and usability measures. In: COST294-MAUSE workshop: meaningful measures: valid useful user experience measurement, pp 13–18
17. Hassenzahl M, Tractinsky N (2006) User experience—a research agenda. Behav Inf Technol 25(2):91–97
18. Hassenzahl M (2008) User experience (UX): Towards an experiential perspective on product quality. In: Association francophone d'interaction Homme-machine, , vol 339. https://doi.org/10.1145/1512714.1512717
19. ITU-R (2012) Methodology for the subjective assessment of the quality of television pictures. Report ITU-R Rec. BT.500-13, International Telecommunication Union, Radiocommunication Sector
20. ITU-T (1999) Subjective video quality assessment methods for multimedia applications. Report ITU-T Rec. P.910, International Telecommunication Union, Telecommunication Standardization Sector
21. De Moor K, Fiedler M, Reichl P, Varela M (2015) Quality of experience: From assessment to application (dagstuhl seminar 15022). Report, DROPS (Dagstuhl Online Publication Service). https://doi.org/10.4230/DagRep.5.1.57
22. ITU-T (2014) Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment. Report ITU-T Rec. P.913, International Telecommunication Union, Telecommunication Standardization Sector
23. ITU-T (2016a) Display requirements for 3d video quality assessment. Report ITU-T Rec. P.914, International Telecommunication Union
24. ITU-T (2016b) Information and guidelines for assessing and minimizing visual discomfort and visual fatigue from 3d video. Report ITU-T Rec. P.916, International Telecommunication Union
25. ITU-T (2016c) Subjective assessment methods for 3D video quality. Report ITU-T Rec. P.915, International Telecommunication Union
26. Puig J, Perkis A, Lindseth F, Ebrahimi T (2012) Towards an efficient methodology for evaluation of quality of experience in augmented reality. In: Fourth international workshop on quality of multimedia experience (QoMEX 2012), IEEE Xplore, vol Proc fourth international workshop on quality of multimedia experience (QoMEX 2012), pp 188–193
27. Kroupi E, Hanhart P, Lee JS, Rerabek M, Ebrahimi T (2016) Modeling immersive media experiences by sensing impact on subjects. Multimed Tools Appl 75(20):12409–12429
28. Hoßfeld T, Heegaard PE, Varela M, Möller S (2016) QoE beyond the mos: an in-depth look at QoE via better metrics and their relation to MOS. Qual User Exp 1(1):2
29. Keighrey C, Flynn R, Murray S, Brennan S, Murray N (2017) Comparing user QoE via physiological and interaction measurements of immersive AR and VR speech and language therapy applications. In: Proceedings of the on thematic workshops of ACM multimedia 2017, ACM, pp 485–492
30. Concannon D, Flynn R, Murray N (2019) A quality of experience evaluation system and research challenges for networked virtual reality-based teleoperation applications. In: Proceedings of the 11th ACM workshop on immersive mixed and virtual environment systems, ACM, pp 10–12
31. Barreda-Ángeles M, Aleix-Guillaume S, Pereda-Baños A (2019) Users' psychophysiological, vocal, and self-reported responses to the apparent attitude of a virtual audience in stereoscopic 360-video. Virtual Real. https://doi.org/10.1007/s10055-019-00400-1
32. Engelke U, Darcy DP, Mulliken GH, Bosse S, Martini MG, Arndt S, Antons J, Chan KY, Ramzan N, Brunnström K (2017) Psychophysiology-based QoE assessment: a survey. IEEE J Sel Top Signal Process 11(1):6–21
33. Barreda-Ángeles M, Redondo-Tejedor R, Pereda-Baños A (2018) Psychophysiological methods for quality of experience research in virtual reality systems and applications. IEEE COMSOC MMTC Commun Front 4(1):14–20
34. Bosse S, Brunnström K, Arndt S, Martini MG, Ramzan N, Engelke U (2019) A common framework for the evaluation of psychophysiological visual quality assessment. Qual User Exp 4(1):3
35. Alnizami H, Scovell J, Ong J, Corriveau P (2017) Measuring virtual reality experiences is more than just video quality. Video Qual Experts Group 3(1):9–17 www.vqeg.org
36. Curcio I (2017) On streaming services for omnidirectional video and its subjective assessment. Video Qual Experts Group 3(1):26–32 www.vqeg.org
37. De Simone F, Frossard P, Brown C, Birkbeck N, Adsumilli B (2017) Omnidirectional video communications: new challenges for the quality assessment community. Video Qual Experts Group 3(1):18–25 www.vqeg.org
38. Milovanovic D, Kukolj D (2017) An overview of developments and standardization activities in immersive media. Video Qual Experts Group 3(1):5–8 www.vqeg.org
39. Brunnström K, Sjöström M, Imran M, Pettersson M, Johanson M (2018) Quality of experience for a virtual reality simulator. Electron Imaging 14:1–9
40. Schatz R, Regal G, Schwarz S, Suettc S, Kempf M (2018) Assessing the QoE impact of 3D rendering style in the context of VR-based training. In: 2018 Tenth international conference on quality of multimedia experience (QoMEX), pp 1–6
41. Tran HTT, Ngoc NP, Pham CT, Jung YJ, Thang TC (2017) A subjective study on QoE of 360 video for VR communication. In: 2017 IEEE 19th international workshop on multimedia signal processing (MMSP), pp 1–6
42. Singla A, Göring S, Raake A, Meixner B, Koenen R, Buchholz T (2019) Subjective quality evaluation of tile-based streaming for omnidirectional videos. In: Proceedings of the 10th ACM multimedia systems conference, ACM, pp 232–242
43. Curcio ID, Toukomaa H, Naik D (2017) 360-degree video streaming and its subjective quality. In: SMPTE 2017 annual technical conference and exhibition, SMPTE, pp 1–23
44. TT Tran H, Ngoc NP, Pham CT, Jung YJ, Thang TC (2019) A subjective study on user perception aspects in virtual reality. Appl Sci 9(16):3384
45. Pérez P, Escobar J (2019) Miro360: A tool for subjective assessment of 360 degree video for ITU-T P. 360-VR. In: 2019 Eleventh international conference on quality of multimedia experience (QoMEX), IEEE, pp 1–3
46. ITU-T ((Under Study)) Subjective test methodologies for 360 degree video on HMD (p.360-vr). Report ITU-T Rec. P.360-VR, International Telecommunication Union, Telecommunication standardization sector
47. Fan CL, Lo WC, Pai YT, Hsu CH (2019) A survey on 360 video streaming: acquisition, transmission, and display. ACM Comput Surv 52(4):71
48. Baumeister J, Ssin SY, ElSayed NAM, Dorrian J, Webb DP, Walsh JA, Simon TM, Irlitti A, Smith RT, Kohler M, Thomas BH (2017) Cognitive cost of using augmented reality displays. IEEE Trans Vis Comput Gr 23(11):2378–2388
49. Pointon G, Thompson C, Creem-Regehr S, Stefanucci J, Bodenheimer B (2018) Affordances as a measure of perceptual fidelity in augmented reality. In: 2018 IEEE VR 2018 workshop on perceptual and cognitive issues in AR (PERCAR) pp 1–6

50. Lin CJ, Woldegiorgis BH (2015) Interaction and visual performance in stereoscopic displays: a review. J Soc Inf Disp 23(7):319–332
51. Cutolo F, Fontana U, Ferrari V (2018) Perspective preserving solution for quasi-orthoscopic video see-through HMDs. Technologies 6(1):9
52. Swan JE, Singh G, Ellis SR (2015) Matching and reaching depth judgments with real and augmented reality targets. IEEE Trans Vis Comput Gr 21(11):1289–1298
53. Jamiy FE, Marsh R (2019) Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. IET Image Process 13(5):707–712
54. Cutolo F, Ferrari V (2017) The role of camera convergence in stereoscopic video see through augmented reality displays. In: Future technologies conference (FTC), pp 295–300
55. Berning M, Kleinert D, Riedel T, Beigl M (2014) A study of depth perception in hand-held augmented reality using autostereoscopic displays. In: 2014 IEEE international symposium on mixed and augmented reality (ISMAR), IEEE, pp 93–98
56. Nagata S (1991) How to reinforce perception of depth in single two-dimensional pictures. In: Pictorial communication in virtual and real environments, Taylor & Francis, Inc., pp 527–545
57. Diaz C, Walker M, Szafir DA, Szafir D (2017) Designing for depth perceptions in augmented reality. In: 2017 IEEE international symposium on mixed and augmented reality (ISMAR), pp 111–122
58. Rizek H, Brunnström K, Wang K, Andrén B, Johanson M (2014) Subjective evaluation of a 3D videoconferencing system. In: Proceedings Vol 9011, stereoscopic displays and applications XXV
59. Albarelli A, Celentano A, Cosmo L, Marchi R (2015) On the interplay between data overlay and real-world context using see-through displays. In: Proceedings of the 11th biannual conference on Italian SIGCHI chapter, ACM, pp 58–65
60. Bork F, Schnelzer C, Eck U, Navab N (2018) Towards efficient visual guidance in limited field-of-view head-mounted displays. IEEE Trans Vis Comput Gr 24(11):2983–2992
61. Taira GMN, Sementille AC, Sanches SRR (2018) Influence of the camera viewpoint on augmented reality interaction. IEEE Lat Am Trans 16(1):260–264
62. Sun H, Liu Y, Zhang Z, Liu X, Wang Y (2018) Employing different viewpoints for remote guidance in a collaborative augmented environment. In: Proceedings of the sixth international symposium of Chinese CHI, ACM, New York, ChineseCHI '18, pp 64–70
63. Lager M, Topp EA, Malec J (2019) Remote supervision of an unmanned surface vessel-a comparison of interfaces. In: 2019 14th ACM/IEEE international conference on human-robot interaction (HRI), pp 546–547
64. Walker ME, Hedayati H, Szafir D (2019) Robot teleoperation with augmented reality virtual surrogates. In: 2019 14th ACM/IEEE international conference on human-robot interaction (HRI), pp 202–210
65. Kim H, Gabbard JL, Anon AM, Misu T (2018) Driver behavior and performance with augmented reality pedestrian collision warning: an outdoor user study. IEEE Trans Vis Comput Gr 24(4):1515–1524
66. Volmer B, Baumeister J, Von Itzstein S, Bornkessel-Schlesewsky I, Schlesewsky M, Billinghurst M, Thomas BH (2018) A comparison of predictive spatial augmented reality cues for procedural tasks. IEEE Trans Vis Comput Gr 24(11):2846–2856
67. Kytö M, Mäkinen A, Tossavainen T, Oittinen PT (2014) Stereoscopic depth perception in video see-through augmented reality within action space. J Electron Imaging 23(1):011006
68. Lages WS, Li Y, Bowman DA (2018) Evaluation of environment-independent techniques for 3D position marking in augmented reality. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR), pp 615–616
69. Brizzi F, Peppoloni L, Graziano A, Stefano ED, Avizzano CA, Ruffaldi E (2018) Effects of augmented reality on the performance of teleoperated industrial assembly tasks in a robotic embodiment. IEEE Trans Hum-Mach Syst 48(2):197–206
70. Peppoloni L, Brizzi F, Ruffaldi E, Avizzano CA (2015) Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing. In: Proceedings of the 21st ACM symposium on virtual reality software and technology, ACM, New York, VRST '15, pp 237–240
71. Uddin W, Sakr M, Quintero CP, der Loos HMV (2018) Orthographic vision-based interface for robot arm teleoperation. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)
72. Krichenbauer M, Yamamoto G, Taketom T, Sandor C, Kato H (2018) Augmented reality versus virtual reality for 3d object manipulation. IEEE Trans Vis Comput Gr 24(2):1038–1048
73. Kruijff E, Swan JE, Feiner S (2010) Perceptual issues in augmented reality revisited. In: 2010 IEEE international symposium on mixed and augmented reality, IEEE, pp 3–12
74. Dima E, Sjöström M, Olsson R, Kjellqvist M, Litwic L, Zhang Z, Rasmusson L, Flodén L (2018) Life: a flexible testbed for light field evaluation. In: 2018-3DTV-conference: the true vision-capture, transmission and display of 3D video (3DTV-CON), IEEE, pp 1–4
75. Geršak G, Lu H, Guna J (2018) Effect of VR technology matureness on VR sickness. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-6969-2
76. Suznjevic M, Mandurov M, Matijasevic M (2017) Performance and QoE assessment of HTC vive and oculus rift for pick-and-place tasks in VR. In: 2017 Ninth international conference on quality of multimedia experience (QoMEX), pp 1–3
77. Singla A, Fremerey S, Robitza W, Raake A (2017) Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays. In: 2017 Ninth international conference on quality of multimedia experience (QoMEX), pp 1–6
78. Wien M, Boyce JM, Stockhammer T, Peng WH (2019) Standardization status of immersive video coding. IEEE J Emerg Sel Top Circuits Syst 9(1):5–17
79. Sun J, Stuerzlinger W, Riecke BE (2018) Comparing input methods and cursors for 3D positioning with head-mounted displays. In: Proceedings of the 15th ACM symposium on applied perception, ACM, New York, SAP '18, pp 8:1–8:8
80. Villegas-Hernandez YS, Guedea-Elizalde F (2017) Marker's position estimation under uncontrolled environment for augmented reality. Int J Interact Des Manuf 11(3):727–735
81. Brunnström K, Barkowsky M (2018) Statistical quality of experience analysis: on planning the sample size and statistical significance testing. J Electron Imaging 27(5):053013
82. Guo J, Weng D, Zhang Z, Liu Y, Duh HBL, Wang Y (2019) Subjective and objective evaluation of visual fatigue caused by continuous and discontinuous use of HMDs. J Soc Inf Disp 27(2):108–119
83. Wu H, Leung SO (2017) Can Likert scales be treated as interval scales? -a simulation study. J Soc Serv Res 43(4):527–532
84. Bland JM, Altman DG (1995) Multiple significance tests: the bonferroni method. Bmj 310(6973):170
85. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3(3):203–220
86. Lloyd S (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28(2):129–137
87. Arthur D, Vassilvitskii S (2007) k-means++: the advantages of careful seeding. In: SODA '07: Proceedings of the eighteenth

annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, pp 1027–1035
88. Regal G, Voigt-Antons JN, Schmidt S, Schrammel J, Kojić T, Tscheligi M, Möller S (2019) Questionnaires embedded in virtual environments: reliability and positioning of rating scales in virtual environments. Qual User Exp 4(1):5