Multi-User Collaboration on Complex Data in Virtual and Augmented Reality

Adrian H. Hoppe[1], Kai Westerkamp[2], Sebastian Maier[2], Florian van de Camp[2], and Rainer Stiefelhagen[1]

[1] Karlsruhe Institute of Technology (KIT)
{adrian.hoppe,rainer.stiefelhagen}@kit.edu
[2] Fraunhofer IOSB
{kai.westerkamp,sebastian.maier,florian.vandecamp}@iosb.fraunhofer.de

Abstract. With increasing task and system complexity, it becomes necessary to support workers, e.g. performing repair tasks, from a remote location. Current approaches utilize images or a video stream combined with annotations and speech to allow collaboration with remote users. We propose a technique that gives the remote supporter the ability to see a high fidelity point cloud of a real world object in Virtual Reality (VR). The VR user can indicate points of interest via a laser pointer. The local worker sees these indications on top of the real object with an Augmented Reality (AR) headset. A preliminary user study shows that the proposed method is faster and less error-prone regarding the comprehension of the object and the communication between the users. In addition to that, the system has a higher usability. This work shows that even non-virtual, collaborative tasks can be supported by new forms of user interaction using different technologies like VR and AR.

1 Introduction

There are several tasks where collaboration can be assisted by Virtual Reality. Complex machines, software and data make it hard to be understood by a single user. Many tasks, like setup, service or repair of real machines need to be performed by a specialist or expert. Urgent or distant tasks can be performed by a remote expert with the help of annotated images or videos and speech communication. However, it is quite cumbersome for the expert to explain how the local worker should move and what he needs to do. This collaboration task can be improved by Virtual and also Augmented Reality. The collaborative Virtual Environment (VE) [4] allows multiple users to analyze and discuss information as well as interact with the VE and each other [12, 3, 13, 1, 7]. VR allows the expert to see the object of interest from a view point, independent from the worker. AR makes it possible to show indications and annotations directly inside the real world instead of an image. Furthermore, VR and AR technologies are very mobile and cheap.

2 Related Work

Remotely supported collaboration can be achieved using different forms of technology. Kuzuoka [9] used a video stream to convey the intentions of the expert. The video is captured by the local user and then send to the remote expert who can annotate the viewed content. The annotated video is then displayed to the local worker. Bauer et al. [2] extended this approach and showed a mouse cursor that is controlled by the expert in an AR Head Mounted Display (HMD) which is worn by the local worker. However, the mouse location is only 2D and it's position is volatile if the HMD moves. Chastine et al. [6] used a 3D cursor to show the expert's intention. Still, the 3D cursor movement is difficult and slow. The system by Botteccia et al. [5] allows to place 3D animations in the field of view of the local worker. The goal is to demonstrate to a user, how a task should be solved. However, the predefined animations are not very flexible. Tachia et al. [14] used static depth sensors to capture the dynamic environment of the users. The 3D scene of the local user and the hands of the remote expert were combined and presented to both users. This system allows the expert to utilize hand gestures for his assistance. Kurata et al. [8] placed a camera and a laser pointer on the shoulder of the local worker. The remote expert saw the video stream and could control the laser to highlight a point of interest in the real world. Lanir et al. [10] expanded this idea and let a movable robotic arm carry a camera and a projector. The robotic arm could be controlled by the expert and the expert's annotations in the 2D video were projected on top of the real world. Oda et al. [11] tracked predefined local objects and represented these as virtual proxies to the remote expert. The expert could create copies of these objects and move them to the correct positions. The local worker saw the virtual copies in an AR environment.

3 Virtual and Augmented Reality Collaboration

VR allows users to interact with complex virtual data collaboratively. In addition to that, it is possible to extend the collaboration to the real world using AR. We propose a VR/AR collaboration system to aid complex tasks through remote collaboration. In order to supply the remote expert with the problem area, a virtual representation is needed. As a first step, a local worker captures a point cloud of the object/region of interest and sends it to a remote expert. The point cloud consists of several filtered Kinect v2 point clouds with color information. The extrinsic camera transform is calculated using the Lighthouse Tracking System of the HTC Vive. The recorded point cloud is displayed in VR for the expert using a HTC Vive (see Fig. 1). The expert can freely inspect the object from any angle and indicate locations using a laser pointer on a tracked controller. The local worker sees the laser pointing on the real object in AR with the Microsoft HoloLens (see Fig. 2). Furthermore, the remote expert and the local worker can engage through a speech communication system. The VR and AR world are calibrated using an anchor point, a HTC Vive Tracker, that is in

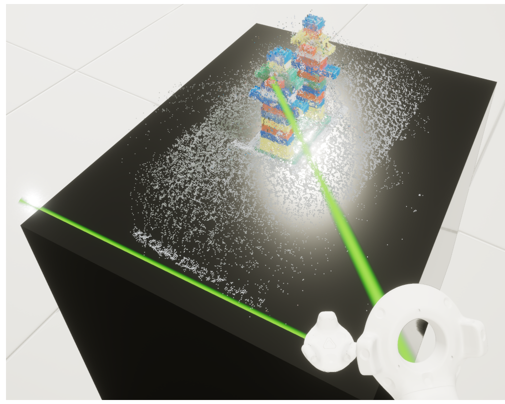

Fig. 1. The VR view from the remote expert with a HTC Vive. The expert highlights a red block using a laser pointer attached to a tracked controller.

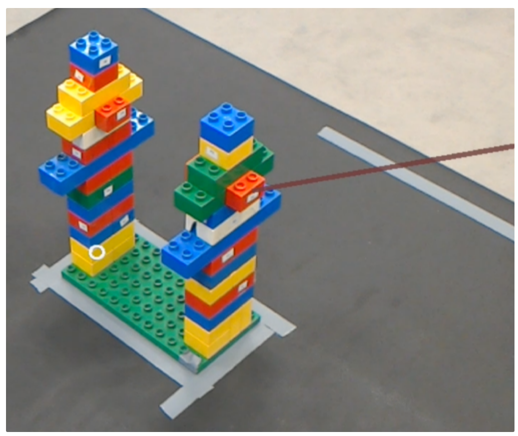

Fig. 2. The AR view from the local worker with a Microsoft HoloLens. The laser pointer highlights a red block.

a fixed location relative to the object. In AR, a coordinate system is placed on top of the anchor point to calibrate the different coordinate systems (see Fig. 3).

4 Evaluation

To evaluate the proposed concept a preliminary user study was performed. In this user study the system was compared to a system that contained pre-recorded images and a live video stream, as well as speech communication. The pre-recorded images serve for the preparation of the remote expert and the additional

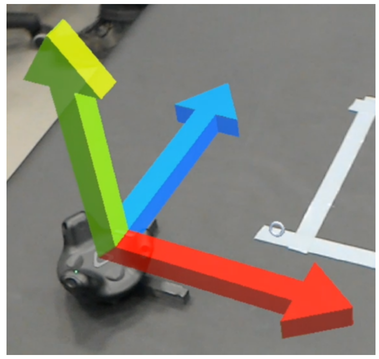

Fig. 3. The calibration of the AR coordinate system relative to the VR world.

live video aids during the support of the local worker. The task for the expert in both setups was to locate a specific block in a set of two towers (see Fig. 2). To convey knowledge to the expert, he or she was given a list of four colors with a place holder for the target at the start or end of that list (e.g. red, green, blue, red, XXX). The varying lists represent sequential tower blocks from top to bottom. All sequences are unique. In a first step, the expert should locate the desired block. Secondly, the expert activates the communication with the local worker and indicates the block. When activated, the live stream is shown to the expert or the laser pointer is shown to the local worker depending on the current setup. Speech communication content is not limited, except for the unique color sequence. To finish one round the local worker confirms the block by reading a text label printed on it. For each setup, a pair of participants performed five training and ten timed rounds. The participants performed both setups. To minimize learning and fatigue effects in the results, the order of the two setups was mixed, the user switched roles on setup change and two different sets of towers were used. 26 people in pairs of two participated in the user study. Two teams were excluded from the evaluation because of tracking issues with the VR/AR setup. The participants had a medium experience with VR and a low experience with AR. Two subjects declared they suffer from a red green color deficiency. However, both reported that the block colors were strong enough to distinguish between them.

When locating the block, users were about 1.12s faster with VR (Ø 9.90 $\pm$ 5.58 s) than with the images (Ø 11.02 $\pm$ 7.67 s). The difference in time for the second part of the task is only 0.73s with Ø 12.24 $\pm$ 5.36 s for the VR/AR setup and Ø 12.97 $\pm$ 5.23 s for image/video. Both differences are not significant. When asked how easy it was to locate the block the VR expert rated the task significantly harder than the expert with images (see Fig. 4). On a scale from -3 (very hard) to 3 (very easy) users rated the location task with a median of 1 with VR and 3 with the images. It was easier for the participants to locate

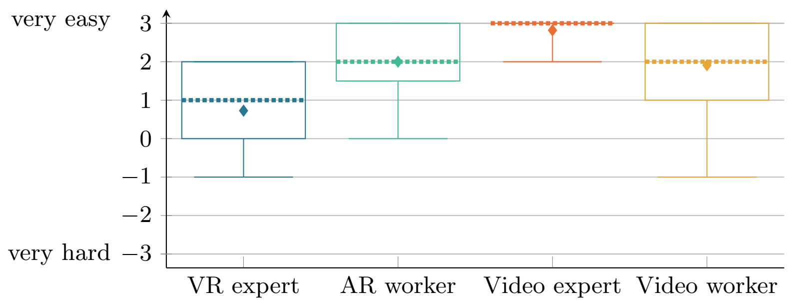

Fig. 4. Question: How easy/hard was it to locate the block?

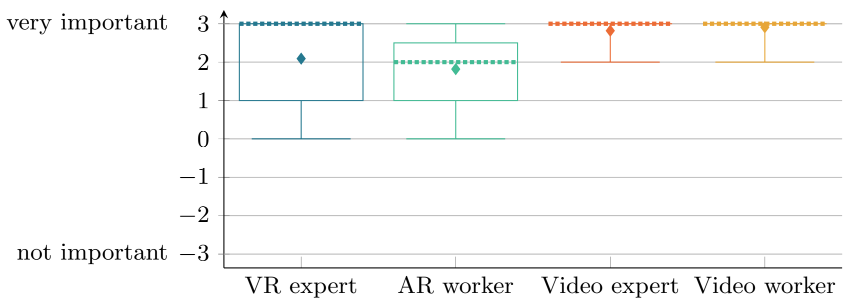

Fig. 5. Question: How important/unimportant was the use of speech communication?

blocks in the images. The VR experts reported it was difficult to see the point cloud, because it was pixelated and imprecise.

The second phase of the task contained the collaboration between the two roles. Experts perceived the video stream mostly as confusing, since they could not control the perspective of the camera and the video was shaking, when the worker moved. This lead to the users often ignoring the video and focusing on the images. They coordinated themselves using unique color features of the two towers or their left and right location. Subjects that made use of the video used it to confirm locations by pointing with the finger on the blocks. When asked how important the speech communication for the task execution was, a significant difference between the local AR and video user occured (see Fig. 5). The participants made almost no errors with both setups. 8 out of 11 teams were error-free with VR/AR and 6 teams did not make a mistake with image/video. The other teams made up to 1 error with VR/AR and up to 3 errors with image/video.

The questionnaires NASA Raw-TLX and UEQ (see Fig. 6) show that there are the following significant differences between the two setups. The physical demand is lower for the image/video expert compared to the VR/AR expert ($p = 0.025$). The performance of the AR worker is higher than the VR expert ($p = 0.039$) and his or her frustration is lower ($p = 0.020$). The UEQ ratings

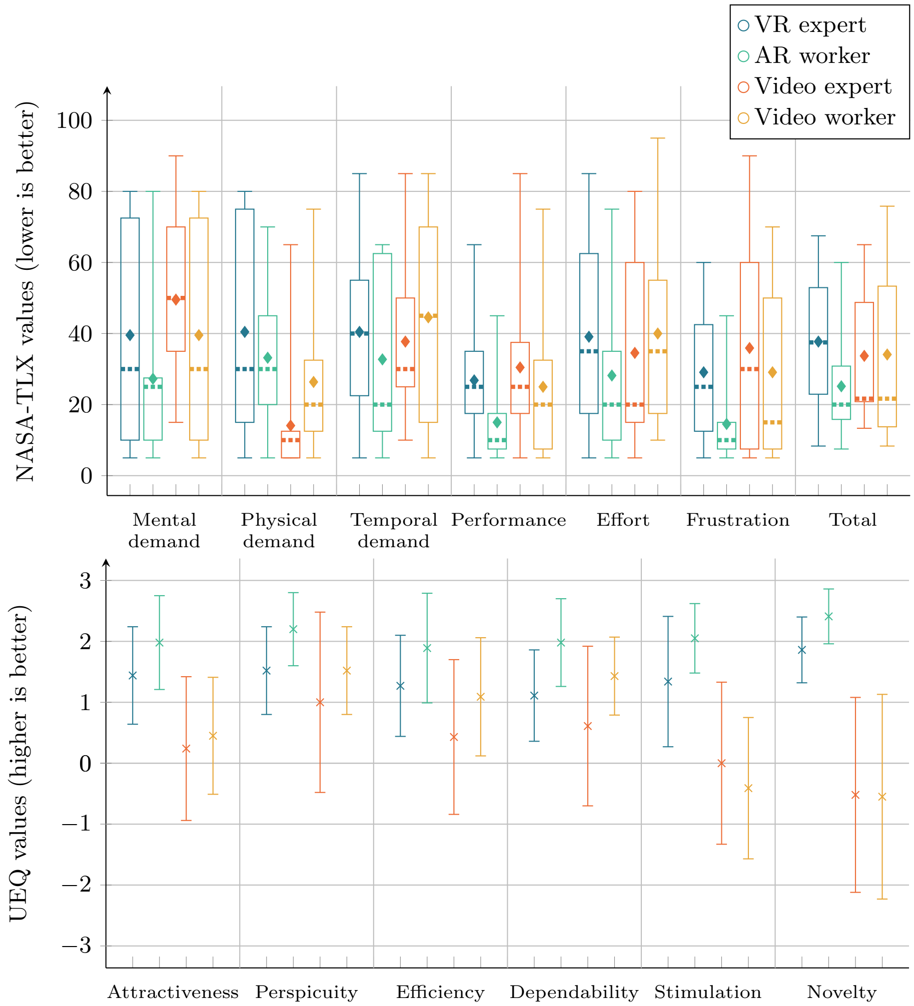

Fig. 6. NASA Raw-TLX ratings with box-and-whisker plots (diamond indicates average) and UEQ ratings with average and standard deviation.

show significant differences for both setups when comparing the two roles. Attractiveness ($p \leq 0.012$), stimulation ($p \leq 0.017$) and novelty ($p \leq 0.001$) are ranked higher for the VR/AR setup compared to the image/video setup.

In addition to that, participants were asked if the independent perspective had any (dis-)advantages. On a 7 point Likert scale from -3 (disadvantageous) to 3 (advantageous), users rated the system with a median of 2 (1. quartile = 1 and 3. quartile = 3).

5 Discussion

The evaluation of the proposed system shows that it is beneficial to use VR and AR technologies for the support of a local worker with a remote expert. Locating a block is reportedly harder with VR, but faster. With further hardware improvements and adjusted data visualization the issues with the visibility of the point cloud should be solved. A problem that impairs the performance of the VR/AR setup is the calibration between the two systems. The manual calibration is error-prone and both tracking systems do seem to have slightly different distance measurements. This leads to a changing offset in the location of the laser pointer beam and therefore a shift of the indicated block. Furthermore there was no collision test from the laser with the object. Because of that, the indicated end location is ambiguous. If these issues are fixed the benefit of the VR/AR system might be not only a tendency, but a significant difference.

6 Conclusion

Our work for connecting two users shows that collaboration can be enhanced using VR/AR technology. Although there were some issues that resulted in an inaccurate laser beam, the system showed improved performance and user experience. For future work, we want to engage more than two users with full-body avatars in VR and AR. In addition to that, it would be interesting to determine how a collaboration of more than two users can be enhanced with new interaction techniques in VR.

References

[1] Kevin Arthur et al. "Designing and building the pit: a head-tracked stereo workspace for two users". In: *2nd International Immersive Projection Technology Workshop*. 1998, pp. 11–12.

[2] M. Bauer, G. Kortuem, and Z. Segall. ""Where are you pointing at?" A study of remote collaboration in a wearable videoconference system". In: *Digest of Papers. Third International Symposium on Wearable Computers*. Oct. 1999, pp. 151–158. DOI: 10.1109/ISWC.1999.806696.

[3] Stephan Beck et al. "Immersive group-to-group telepresence". In: *IEEE Transactions on Visualization and Computer Graphics* 19.4 (2013), pp. 616–625.

[4] Steve Benford et al. "Collaborative virtual environments". In: *Communications of the ACM* 44.7 (2001), pp. 79–85.

[5] Sébastien Bottecchia, Jean-Marc Cieutat, and Jean-Pierre Jessel. "T.A.C: Augmented Reality System for Collaborative Tele-assistance in the Field of Maintenance Through Internet". In: *Proceedings of the 1st Augmented Human International Conference*. AH '10. Megève, France: ACM, 2010, 14:1–14:7. ISBN: 978-1-60558-825-4. DOI: 10.1145/1785455.1785469. URL: http://doi.acm.org/10.1145/1785455.1785469.

[6] J. Chastine et al. “Studies on the Effectiveness of Virtual Pointers in Collaborative Augmented Reality”. In: *Proceedings of the 2008 IEEE Symposium on 3D User Interfaces*. 3DUI ’08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 117–124. ISBN: 978-1-4244-2047-6. DOI: 10.1109/3DUI.2008.4476601. URL: http://dx.doi.org/10.1109/3DUI.2008.4476601.

[7] Alfred Kranstedt et al. “Measuring and reconstructing pointing in visual contexts”. In: *Proceedings of the brandial* (2006), pp. 82–89.

[8] T. Kurata et al. “Remote collaboration using a shoulder-worn active camera/laser”. In: *Eighth International Symposium on Wearable Computers*. Vol. 1. Oct. 2004, pp. 62–69. DOI: 10.1109/ISWC.2004.37.

[9] Hideaki Kuzuoka. “Spatial Workspace Collaboration: A SharedView Video Support System for Remote Collaboration Capability”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’92. Monterey, California, USA: ACM, 1992, pp. 533–540. ISBN: 0-89791-513-5. DOI: 10.1145/142750.142980. URL: http://doi.acm.org/10.1145/142750.142980.

[10] Joel Lanir et al. “Ownership and Control of Point of View in Remote Assistance”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’13. Paris, France: ACM, 2013, pp. 2243–2252. ISBN: 978-1-4503-1899-0. DOI: 10.1145/2470654.2481309. URL: http://doi.acm.org/10.1145/2470654.2481309.

[11] Ohan Oda et al. “Virtual Replicas for Remote Assistance in Virtual and Augmented Reality”. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST ’15. Charlotte, NC, USA: ACM, 2015, pp. 405–415. ISBN: 978-1-4503-3779-3. DOI: 10.1145/2807442.2807497. URL: http://doi.acm.org/10.1145/2807442.2807497.

[12] Holger Salzmann, Jan Jacobs, and Bernd Froehlich. “Collaborative Interaction in Co-Located Two-User Scenarios”. In: *Joint Virtual Reality Conference of EGVE - ICAT - EuroVR*. Ed. by Michitaka Hirose et al. The Eurographics Association, 2009. ISBN: 978-3-905674-20-0. DOI: 10.2312/EGVE/JVRC09/085-092.

[13] Zsolt Szalavári et al. “Studierstube: An environment for collaboration in augmented reality”. In: *Virtual Reality* 3.1 (1998), pp. 37–48.

[14] Franco Tecchia, Leila Alem, and Weidong Huang. “3D Helping Hands: A Gesture Based MR System for Remote Collaboration”. In: *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI ’12. Singapore, Singapore: ACM, 2012, pp. 323–328. ISBN: 978-1-4503-1825-9. DOI: 10.1145/2407516.2407590. URL: http://doi.acm.org/10.1145/2407516.2407590.