Perspektive Vermitteln

Alexander Baur
alexander.baur@†

Pedram Berendjy Jorshery
pedram.berendjy.jorshery@†

Lennart Diedrichsen
lennart.diedrichsen@†

Oliver Manka
oliver.manka@†

†*informatik.uni-hamburg.de*
Department of Informatics, Human Computer Interaction
University of Hamburg
Hamburg, Germany

ABSTRACT

In the context of a HCI project work for our studies, we examined three distinct methods to convey perspective in a multiplayer virtual reality environment and compared them with each other. The core idea is to enhance the field of view of a player, at a particular position, with visual cues from a partner at a different position. The additional information enables the player to perform an otherwise hard to complete task. Derived from this our suggested methods are Object-Highlighting, a Visual-Cone and Picture-In-Picture (live broadcast view of partner to player). We conducted an experiment to assess differences, regarding precision, performance and ease of adaption after failure as well as VR sickness, in an interactive interdependent multiplayer task. Following our results, Object-Highlighting was the inferior of all techniques, as adjustments after mistakes within the task were significantly harder in comparison.

Author Keywords

Virtual reality, 3D interaction, Multiplayer, Conveying perspective, Comparison, Collaboration

INTRODUCTION

Virtual reality systems have been around since the early days of Sutherland and his students [8]. Since then, as time passed by, those system have evolved and become increasingly interesting in several different fields. In the framework of our studies we developed and tested an application for conveying perspective in a multiplayer virtual reality (MVR) environment. Due to the fact, that there are many possible ways in conveying the perspective inside a MVR setup, we wanted to compare three methods of conveyance that are very common in 3D interaction [10, 9, 1] and determine which of them is most suited for future MVR applications.

- Object-Highlighting (henceforth: *OH method)*:
  The object of interest is being highlighted by a player (e.g. with an salient color) and is clearly visible even through other objects that might be in front of it. Now all players which have the object of interest in their field of view (henceforth: FoV) can see that it is highlighted. (Figure 1)
- Picture-In-Picture (henceforth: *PIP method)*:
  The entire FoV of a particular player is being broadcast live on a surface or an area (often down-scaled, monitor size). Now all other players which have this surface or area in their FoV can see exactly what that player sees. (Figure 1)
- Visual-Cone (henceforth: *VC method)*:
  Other players can see the head of a certain player represented as a visual cone (not necessarily a cone, every object which is suitable to visualize the line of sight is conceivable). The cone moves synchronously on all axes with the head movement of that player and clarifies his line of sight by the direction in which its open end points. (Figure 1)

As people are using verbal communication when collaborating, which would have an effect on the methods presented, we have decided that verbal communication was prohibited within the study. Participants were therefore not allowed to talk. This allowed us to ensure that the interaction between the Observer and Crane Operator relied only on the particular method which was tested. However, an exception was made to the introduction and tutorial, so that the participants had no ambiguities before starting the study. Additionally, verbal communication is not always helpful and can lead to many mistakes and misunderstandings especially inside MVR environments with a high degree of collaboration and without representing an avatar for each player (which is the case in our study) as Marks et al. [5] is explaining in context of virtual-environment-based surgeries.

As a forecast of the contents:

- Designing and constructing a suitable MVR environment in which we could test the three perspective conveying methods and compare them afterwards.
- Performing a user study and describing detailed the *WHO* (who was participating on our study), *HOW* (how we set up our study) and *WHAT* (what have we actually measured and what were the results).
- How we classify our results, where we had limitations and where there still is scientific work to do.

RELATED WORK

Just like in the ordinary VR sector, it didn't take long for the head-mounted display VR sector until in applications like games and simulators (applications in which 3D interaction is possible) multiplayer setups appeared. Books, like the one of Smed et al. [6] were already reviewing a summary of the key factors of networking in physical, logical and application levels in 2002 and that is even before "the announcement of the Oculus Rift in 2013 ha[d] brought back the mainstream interest of companies as well as the general public in VR."

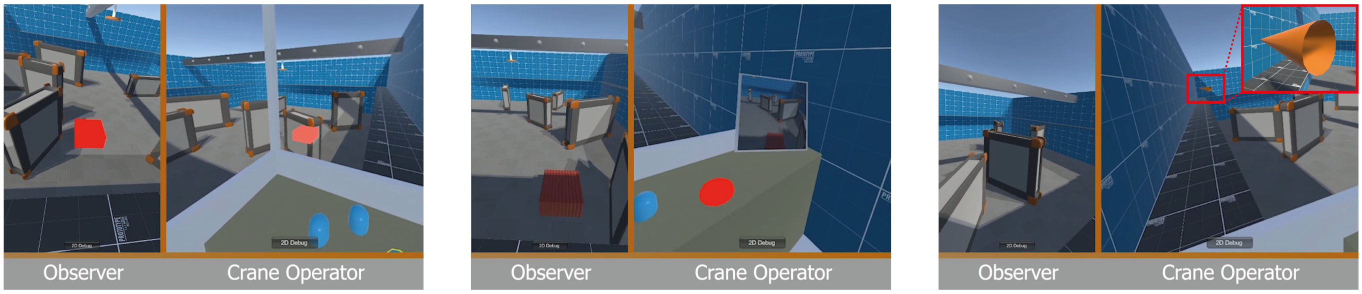

Figure 1:
OH method: Observer is highlighting the Pick-Up Object by looking straight at it and Crane Operator is clearly seeing it through the obstacle (left).
PIP method: Crane Operator is seeing broadcast live FoV of Observer on a monitor inside the control room (center).
VC method: Crane Operator is seeing the orange cone of the Observer, while the Observer moves on the catwalks (right).

[7]. Each of our three perspective conveying methods are a direct or indirect interaction form between user-user or user-environment. There are many possible perceivable interaction forms as Manninen [4] defines them: "Interaction forms are perceivable actions that act as manifestations of the user-user and user-environment interaction. These forms are used to convey the actions of the user to oneself and to others. They enable awareness of actions by offering mutually perceivable visualisations and auralisations within the environment. In addition, the feeling of presence and the level of psychological immersion are increased due to the communication, co-ordination and collaboration aspects these forms bring forward." Some of the top level interaction form categories in Collaborative Virtual Environments (CVE) he is describing are: Avatar Appearance, Kinesics, Occulesics, Facial Expressions, Environmental Details and Language-based Communication. Avatar Appearance contain all forms of indirect interaction that is emanating from e.g. the clothes, hair, physique or adornments of the avatar while Kinesics describes postures, body movement, head movement or gestures. Because eye movement, eye contact and line of sight are such important aspect of interaction they have their own category: Occulesics. Environmental Details contain all sorts of manipulation or interaction with objects or the environment itself e.g. our *OH method*. Moreover, there is a lot of ongoing research in the field of collaborative learning in VR. For example the article "Technology and applications for collaborative learning in virtual reality" [2] is introducing six presentations on this topic, which are all showing the benefits and drawbacks of collaborative learning in VR, as well as findings and aspects to consider when dealing with different tasks to perform in a collaborative way inside VR.

PERSPEKTIVE VERMITTELN

In this section we describe the design and development of the project. Our goal was to find the superior method of conveying the perspective of a player B to another player A in a MVR environment. In order to test and compare our three methods we considered a scene with a theme that suited our methods. We decided to pick a scene theme which is close to reality and thus transferable to the real world. For this purpose we constructed an environment of a virtual warehouse and a crane, in which player A is the Crane Operator (henceforth: *CO*) who is standing in a Control Room (henceforth: *CR*) and has to pick up an object (Pick-Up Object; henceforth: *PUO*; red box, Figure 2). The *CO* can do that by using the crane to steer a gripper over a target area (*TA*, Figure 2). Steering the crane is done by pressing buttons (blue) assigned to directions left, right, forth and back. Another button (red) was implemented to lift and put down the object. *Obstacles* in form of walls inside the warehouse restrict the direct sight of the *CO* on the *PUO*. Player B (henceforth: *Observer*) then tries to compensate the handicap of the *CO* by enhancing the *FoV* of *CO* via one of the perspective conveying methods. The *Observer* can move freely along the catwalks, which are attached along the walls of the first floor of the warehouse (see Figure 2, OP), and is not restricted to the *CR* like the *CO*.

Designing the Warehouse

In order to provide a comprehensive test environment in VR, we designed three warehouses of different sizes (small, medium and large) with three different levels for each (different positions of the *PUO*, *TA* and *Obstacles*). The sizes of the warehouse vary between a size unit of $100m^2$ to $400m^2$. The small one has the size of $10x10m$, the medium one of $15x15m$ and the large warehouse of $20x20m$. The scaling is a trade-off between warehouse sizes in reality, the limitation within VR (e.g. resolution) and a solvable task setup.
We tried to keep the design simple so that there are as few distractions as possible. The crane, the warehouse and the *CR* were kept minimalist as illustrated in Figure 2. As an aid to assessing the depth, we have decided to provide the floor and walls with visual guidelines (see Figure 1, white lines). Furthermore, we provided the *Observer* with teleportation points along the catwalks, to allow a optimal vantage point for all our level setups.

EQUIPMENT AND SOFTWARE

For the setup we used two VR-capable PCs, each with one HTC Vive Pro. Unity3D was used to render the warehouse scene, as well as the SteamVR asset for HTC Vive feature implementation.

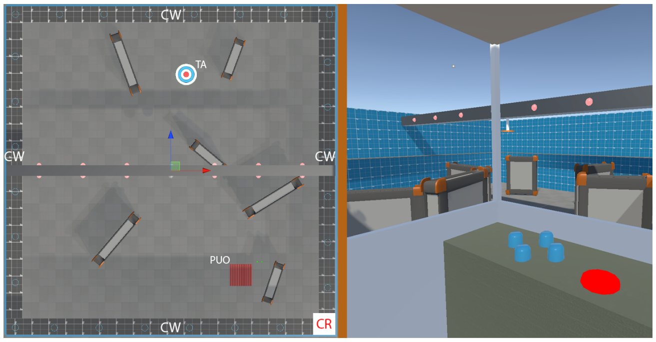

Figure 2:
Left:Warehouse design with catwalk (CW), control room (CR), target area (TA) and pick-up object (PUO).
Right: control room view and steering unit.

Multiplayer and Network-Script

A crucial part for the multiplayer setup is a real-time data synchronisation; it enables two players to interact with each other. As working with Unity we tried the inbuilt "UNet" networking solution and the third party networking solution "Photon". We discarded both options as they quickly became too extensive and left us with no control over the data that was exchanged. Therefore, we wrote a networking script in C# that allowed to transmit the positional and rotational data of the objects it was attached to. The script communicated between both PCs via LAN and was used to synchronize the cranes' and observers' movement in both scenes, allowing us to implement the methods. Writing our own networking script enabled us to find a lightweight solution for the networking problem.

STUDY

This section deals with the actual user study. Our hypothesis was the *PIP method* being the superior method as we thought it would nullify most of the challenge posed by the *Obstacles*. The *OH method* was predicted to perform worse due to lack of depth perception. *VC method* was expected to result in the worst performance as the aid seemed to provide the least direct information about the *PUO* or *TA*, thus resulting in a higher dependence on the *Observer*. Verifying our hypothesis required a test setup was that allowed us to compare the three different perspective conveying methods. Hence, we designed a task in which the participants were asked to steer the crane while being restricted to standing in a control room *CR* in the VR-environment. The participants were instructed to use the crane to pick up the *PUO* and put it down on the *TA* as precisely as possible. The task had to be completed with help of each method, resulting in a 3x3-setup: The three methods were tested with three different scalings of the warehouse. By designing three levels with varied obstacle placement for each size, no setup was passed twice and the *CO* was therefore fully depended on the *observer*.

Participants

16 participants ($M = 22.88$, $SD = 2.31$, 5 female) took part in the experiment. Mean time for the experiment per participant was approx. 45 minutes. Participants varied widely in 3D video game experience, time and frequency of play per week.

Procedure

Starting with a demographic data and simulator sickness questionnaire (SSQ) [3]. The experiment continued with a briefing about the nature of the task and an explanation of the different perspective conveying methods. The introduction to the task provided information about how to steer the crane, what the pick-up object looks like (see Figure 2), as well as facts and functions about the different perspective conveying methods, followed by a tutorial where subjects could get used to the task specific controls and the roll of the *CO*. Steering of the crane could be executed through pressing the buttons left, right, forth and back. In order to pick up and put down the object a red button next to the steering buttons, on the console in the *CR* (see Figure 2), had to be pressed. The action of pressing a button was carried out by hovering over a certain button combined with a vertical down motion of the Vive-controller. This implementation was chosen with the objective to restrict the participant to the *CR*, so he would be fully dependent on the *Observers'* cues and therefore forced to use the methods for task completion.
Once the participant understood relevant functionalities, the experiment proceeded with the first of three data obtaining conditions. All method-trials contained three different warehouse sizes: small, medium and big, with a randomized level setup. The order of the methods were randomized for each subject, as was the order of warehouse sizes within each method. Level order was also randomized for all sizes. Thus, learning effect should be minimized in the final data. In each trial, the participant spawned at the center of the *CR*. The *Observer* was positioned next to the *CR* and was able to move around the warehouse catwalks (see Figure 2, *CW*) via teleportation, as this enabled fast movement, which is key for conveying perspective in our design. The role of the observer was held by the study conductors. After each trial, participants were asked to answer the SSQ again as well as to subjectively rate their VR-experience with the preceding method and state ratings regarding degree of loneliness and to which extent the perspective conveying method helped completing the task. In our between-subject design, every participant was tested in all conditions (*VC,OH,PIP*). Task completion was accomplished, if the subject *CO* managed to pick up the *PUO* (red-box), within 5 minutes time and put it down anywhere in the warehouse. Exceeding the time limit was recorded as a out-of-time failure and the trail was cut off. Picking up the object required the gripper of the crane to be within an one meter radius around the center of the box, otherwise the crane lights (red) were activated and a fail attempt was recorded.

Measures

Three distinct measures were taken for each task. The time from start of the task until its completion (*Performance*), distance from the center of the pick-up object or target area when pressing the gripping button (*Precision*) and the number of times the gripping button was pressed when not hovering over the pick-up object (*Failure rate*). A lower failure rate speaks for better adjustments after mistakes. In addition, we

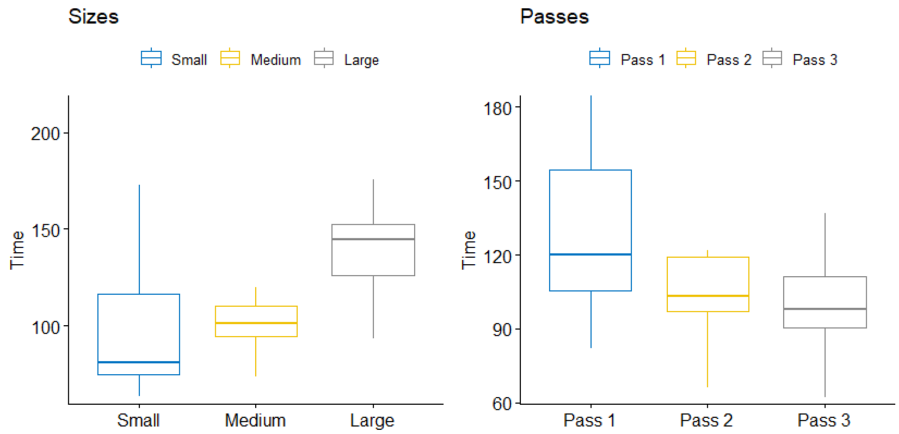

Figure 3: Time in seconds across sizes of warehouse and passes. Lower is better.

started the experiment by obtaining demographic data and a self-reported VR sickness measure with the SSQ [3], which consists of 16 items divided into 3 subcategories (nausea, disorientation and oculomotor), with each item on a 5-point Likert scale from 0 to 4. Further, participants were asked to state their degree of loneliness as well as the degree of aid the specific perspective conveying method provided. Both items can be rated on a 5-point Likert scale ranging from 1-5. After completing all trails, subjects could assess which of the three methods they liked the most (single-choice) (Figure 5).

RESULTS

The gathered data was aggregated into the three perspective conveying methods, the three sizes of the warehouse and the three trials. The measures where then tested for normal distribution using Shapiro-Wilk test. If the test could not confirm normal distribution, we proceeded with a Friedman rank sum test and, depending on the result, checked for differences pairwise with a posthoc Friedman-Nemenyi test. For normally distributed measures, a repeated measure ANOVA was conducted, followed by a paired t-test for significant results.

For performance, significantly more time was needed to complete the task in the large setup ($M = 141.11$ seconds, $SD = 25.12$) compared to the small setup ($M = 98$ seconds, $SD = 31.98$, $p = 0.0003$). Furthermore, participants performed significantly faster on their last pass ($M = 101.07$ seconds, $SD = 22.84$) compared to their first pass ($M = 131.27$ seconds, $SD = 37.83$,$p = 0.0218$) (see Figure 3).

Precision was best for the small setup ($M = 0.97$ meters, $SD = 0.62$), where participants put the pick-up object significantly closer to the target area in comparison to the large setup ($M = 2.6$ meters, $SD = 2.13$, $p = 0.0081$).

Participants failure rate was highest for *OH method* ($M = 4.33$ failures, $SD = 5.55$), whereas significantly fewer mistakes were made using *VC method* ($M = 0.98$ failures, $SD = 1.18$, $p = 0.028$) and *PIP method* ($M = 1.17$ failures, $SD = 1.48$, $p = 0.028$) (see Figure 4).

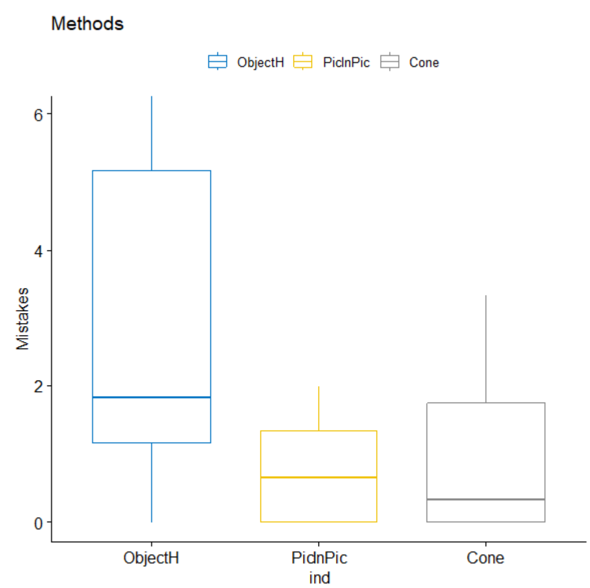

Figure 4: Failures across methods. Lower is better.

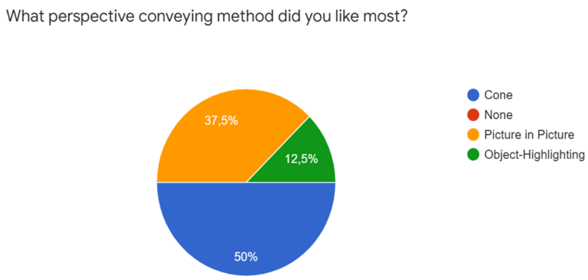

Figure 5: Participants rating for preferred conveying method. *Extracted from GoogleForms*

Questionnaires

In terms of SSQ, there were no significant differences across the conditions. Differences from the SSQ gathered before and after each condition were compared with a Wilcoxon signed-rank test. Results for methods: *VC method* ($M = -3.04, SD = 15.56$), *OH method* ($M = -1.87, SD = 18.11$) and *PIP method* ($M = -0.70, SD = 21.21$). Regarding "degree of loneliness" we found, using the Mann-Whitney U Test, a significant difference between *VC method* and *OH method* ($z = -2.50631, p = 0.01208 at p < .05$). After the experiment was finished, participants where asked to state their preferred "perspective conveying" method. Here the *VC method* scored highest ($.5$), followed by *PIP method* ($.375$) (Figure 5).

DISCUSSION

The results of the study point toward the importance of a wealth of variables when it comes to efficiently and conveniently conveying perspective in VR. We showed, that highlighting objects through walls *OH method* was the least favorable method. Not only was it hard to adjust after errors (Participants often reported the lack of depth was making task difficult and in the large setup nearly impossible) but also did participants feel more alone in the virtual environment. This could be explained by the lack of any dynamical and humanoid movement that the cone of the *VC method* and the monitor content of the *PIP method* provide. This confirms part of the hypothesis as the *OH method* was expected to perform worse than the *PIP method*.

Unexpected, however, was the strong performance of the *VC method*. Contrary to our hypothesis did the Visual Cone perform much better than the *OH method* and even as well as the *PIP method*. We presume an important role of social interaction at play as the Visual Cone was reminiscent of a physical human head. Not all participants made that connection immediately and felt helpless at first. This in turn would suggest social soft skills being a confounding variable.

As the time, scope and resources of the project within which this study was conducted were confined, two economical decisions have to be considered: We were limited to out team being the *observers* and hence limiting variety of social interplay. We therefore tried to remove that variable all together by using consistent behavior patterns when observing across all methods. Additionally, we had to enforce a five minute cut-off for each trial. Failure to complete the task in that time forced us to ignore these trials in our time records. Due to the upper bound in time data, distributions are skewed favorably for worse performing methods, thus making differences in performance more conservative.

Significant differences in task precision and performance due to size of warehouse and number of trials were found. By randomizing order of methods and sizes within each method, we hope to have minimized any confounding effects.

CONCLUSION

We presented three methods for Perspective Taking in VR and compared them in performance, precision and the ease of adjusting to the others perspective over time (by recording failure rate). We showed that *PIP method* and *VC metod* are superior to the *OH method* in regards to that last point. Additionally, participants felt most isolated using *OH method*, suggesting it to be the inferior method for most MVR applications.

Even though participants were not satisfied nether performed well with the *OH method*, modifying the method with depth cues such as color fading in the distance or using a more complex texture to be highlighted might offset its current weaknesses.

Future Work

Following our findings in the study, we would like to improve on our aforementioned shortcoming by conducting a larger scale study with more participants and more time per participant. Furthermore, we see high potential in determining how the methods would interact with each other. Could the combination of the methods let us get significant and better results or would it just be too distracting from the main task for the *CO*?
Moreover, we formed the hypothesis that misunderstanding of the *VC method* could correlate with the soft skills of the player? We would like to investigate these hypotheses in further works.

ACKNOWLEDGMENTS

We want to thank our tutor of this project Jann Philipp Freiwald who supported us where he could, our project leader Prof. Dr. Frank Steinicke and the University Hamburg which gave us the resources for this project.

REFERENCES

1. Trevor J Dodds and Roy A Ruddle. 2008. Using teleporting, awareness and multiple views to improve teamwork in collaborative virtual environments. In *Virtual Environments 2008*. Eurographics Association, 81–88.
2. S Greenwald, Alexander Kulik, André Kunert, Stephan Beck, B Frohlich, Sue Cobb, Sarah Parsons, and Nigel Newbutt. 2017. Technology and applications for collaborative learning in virtual reality. (2017).
3. Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying

simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

4. Tony Manninen. 2002. Interaction forms in multiplayer desktop virtual reality games. In *VRIC2002 Conference*, Vol. 223. 232.
5. Stefan Marks, John Windsor, and Burkhard Wünsche. 2009. Enhancing virtual environment-based surgical teamwork training with non-verbal communication. (2009).
6. Jouni Smed, Timo Kaukoranta, and Harri Hakonen. 2002. *A review on networking and multiplayer computer games*. Citeseer.
7. Frank Steinicke. 2016. *Being really virtual*. Springer.
8. Ivan E Sutherland. 1968. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 757–764.
9. Matthias Trapp, Christian Beesk, Sebastian Pasewaldt, and Jürgen Döllner. 2011. Interactive rendering techniques for highlighting in 3d geovirtual environments. In *Advances in 3D Geo-Information Sciences*. Springer, 197–210.
10. Jason Wuertz, Sultan A Alharthi, William A Hamilton, Scott Bateman, Carl Gutwin, Anthony Tang, Zachary Toups, and Jessica Hammer. 2018. A design framework for awareness cues in distributed multiplayer games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.