Perceiving Depth: Optical versus Video See-through

Daniel Medeiros*[,1,2], Maurício Sousa[1], Daniel Mendes[1], Alberto Raposo[3], Joaquim Jorge[1]
[1]INESC-ID Lisboa, Técnico Lisboa, Universidade de Lisboa; [2]CAPES Foundation, Brasil; [3]Tecgraf, PUC-Rio Brasil

Abstract

Head-Mounted Displays (HMDs) and similar 3D visualization devices are becoming ubiquitous. Going a step forward, HMD see-through systems bring virtual objects to real world settings, allowing augmented reality to be used in complex engineering scenarios. Of these, optical and video see-through systems differ on how the real world is captured by the device. To provide a seamless integration of real and virtual imagery, the absolute depth and size of both virtual and real objects should match appropriately. However, these technologies are still in their early stages, each featuring different strengths and weaknesses which affect the user experience. In this work we compare optical to video see-through systems, focusing on depth perception via exocentric and egocentric methods. Our study pairs Meta Glasses, an off-the-shelf optical see-through, to a modified Oculus Rift setup with attached video-cameras, for video see-through. Results show that, with the current hardware available, the video see-through configuration provides better overall results.These experiments and our results can help interaction designers for both virtual and augmented reality conditions.

Keywords: Depth Perception, See-through system, Augmented Reality, User Evaluation

Concepts: •**Human-centered computing** → *User studies;*

1 Introduction

The recent popularization of virtual reality (VR) devices such as the Oculus Rift and the GearVR head-mounted displays (HMDs), brings us challenges of reducing the gap between virtual and real objects. This is possible with the use of see-through systems, which are able to capture the real world and combine it with virtual objects within the same shared space.But, to provide consistent perception within the 3D space, the absolute depths and sizes of objects in the two images must correspond appropriately.

This see-through capability can be accomplished using either an optical or a video see-through HMD [Rolland and Fuchs 2000]. When using optical-see-through (OSTs) HMDs the real world is seen through semi-transparent mirrors placed in front of the user's eyes. These mirrors are also used to reflect he computer generated images into the user's eyes, thereby combining the real- and virtual-world views. A recent example of this kind of device is the Google Glass, that allows the visualization of interactive content on a portion of the user's left-eye field of view. Another example is the Meta Glass Headset, a binocular optical see-through device that uses a depth camera to detect gesture movements in front of the user's field of view.

On video see-through (VST) systems the real-world is captured by one or multiple cameras, normally located in front of the HMD [Edwards et al. 1993]. The computer generated images are electronically combined with the representation of the real world with devices such as a modified Oculus Rift [Steptoe et al. 2014] and the use of the monoscopic see-through embedded on the GearVR. One important characteristic on both types of see-through HMDs is that both virtual and real objects are aligned in a way that the relation between real and virtual is minimal.

Depth perception is the visual ability to spatially perceive the world in three dimensions. This sensation is perceived by the information from both monocular cues, such as distance, occlusion and size, and binocular cues, such as convergence and motion parallax [Cutting 2003]. According to Cutting [Cutting 2003], the relative importance of different depth cues is determined by the distance of the objects to the user. There are three different areas: Personal space (0 to 2 meters), action space (2 to 20 meters) and vista space (more than 20 meters). In the personal space, binocular disparity provides the most accurate depth judgments. It is the most important depth cue provided by stereoscopic vision and particularly useful to resolve ambiguities created by other perceptual cues.

On see-through systems, a way of evaluating depth perception is through open loop tasks, with procedures such as blind-walking [Swan II et al. 2007]. These studies are based on cognitive aspects of how the human-eye perceives distance, inspired on the concepts of egocentric and exocentric distance measurements, which are the distance measured by the distance from the user point of view to an object, and the perceived distance from two objects, respectively [Kelly et al. 2004b; Loomis and Knapp 2003].

Even though the evaluation of depth perception issues are widely studied, the majority of the works focus on such factors on a particular type of see-through device, either video or optical see-throughs. Some works try to categorize and compare aspects of both of them but do not compare precision tasks based on depth cues, using similar conditions on both devices [Lizandra and Calatrava 2011; Rolland and Fuchs 2000].

This paper evaluates users' depth perception by combining egocentric and exocentric methods. The proposed task evaluates the perception of depth by enabling a user to connect two objects using a line, by interacting with a wand. In this paper we evaluate and compare two different types of see through, the video see-through and optical see-through considering both hardware and software differences. For the evaluation we use low-cost devices such as a modified Oculus Rift with attached digital video cameras and the Meta Glass SDK1 Headset[1], an off-the-shelf optical see-through HMD (Figure 1b).

2 Task Design

Normally, as seen on the literature [Iwamoto and Ishikawa 2013; Swan II et al. 2007], the distance is evaluated via egocentric methods, i.e. evaluating how users perceive distance between their point of view and an object. On the other hand, Kelly et al. [Kelly et al.

*email: daniel.medeiros@tecnico.ulisboa.pt

VRST '16, November 02-04, 2016, Garching bei München, Germany
ISBN: 978-1-4503-4491-3/16/11...$15
DOI: http://dx.doi.org/10.1145/2993369.2996348

[1]https://www.metavision.com/

2004b] use an exocentric method to evaluate how users perceive distance between different objects (two or more), and conclude that their perception depends on the position, size and orientation of the users' bodies towards the objects in a real scenario. Further work from the same authors [Kelly et al. 2004a] indicates that user's depth perception is similar in both virtual and real environments. Our proposed evaluation combines aspects from both exocentric and egocentric approaches. We also take into account both the advantages and limitations of the devices used on the design process. First, we present both the setup used and the task design details.

2.1 Setup

We used two different types of see-through devices, an Optical see-through and a Video see-through. Also, Optitrack [2] motion capture reflective markers were attached to each device to provide accurate positioning input. The optical motion capture system used includes twelve Flex 3 cameras operating at 100 FPS. The Optitrack system is responsible for computing six degrees-of-freedom (6DoF: 3 for position and 3 for orientation) of the tracked see-through devices.

(a) (b)

Figure 1: *See-Through approaches evaluated: (a) Video See-Though with Oculus Rift; (b) Optical See-Though with Meta Glasses.*

The video see-through, shown in Figure 1a, is a reproduction of the device presented by Steptoe et al. [Steptoe et al. 2014], an immersive head-mounted video see-through AR display comprised of commercially available low-cost components. The HMD used is an Oculus Rift DK1 with two modified webcameras attached with FoV of 110 degrees on the vertical and 90 degrees horizontal with a perceived distance of approximately 3.66 meters. The Rift has a 7" panel with a image formed for each eye (each of them rendered on half of the Oculus panel) with a resolution of 640x800 pixels per-eye. We also implemented a shader to correct the known effects of radial distortion caused by the webcam lenses.

The optical see-through used in the experiment was the Meta Glasses(Figure 1b), an off-the-shelf Optical see-through device. This OST presents a stereoscopic glass screen with an aspect of 16:9 and HD resolution (1280x720) that is shown in the eyes of the user, in the middle of its field of view. This device can be used in both monoscopic and stereoscopic approaches, with an image formed for each eye, improving the depth sensation for the use in virtual environments. Additional specifications of this system are a FoV of 23 degrees and 35.9 degrees, with normal and FoV expander lenses, respectively. In our test we used the Meta Glasses with the FOV expander lenses. The distortion caused by the lenses is corrected using the provided SDK. Beyond the visualization system, the Meta glasses also include a depth-based sensor and a gyroscope to detect user's head orientation. For replicability purposes we chose not to use the sensors embedded on the Meta Glasses.

2.2 Performed Task

On the proposed task, participants need to draw a line with a wand between two colored spheres split by a distance, situated at different depths inside the user's personal space [Cutting 2003]. The proposed task is divided in two phases: in the first users need to reach the initial sphere (egocentric) and on the second, the exocentric phase, participants have to move their hands to draw a line between spheres. To evaluate and compare the two different see-through systems we have to consider their differences both on hardware and software. Despite that, both setups shared the same variables and evaluation conditions.To avoid depth misinterpretation, the physical environment of the experiment is composed by an open space, without additional depth cues.

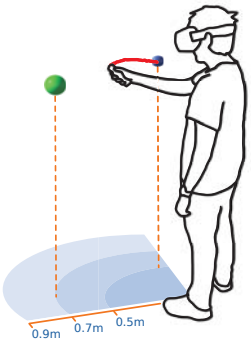


Figure 2: *Distances used and the relation with the frustum's deformation.*

Due to the low FOV on the Meta in comparison to the VST device (35.9 versus 110 degrees), we chose to keep all object within the Meta's FoV with objects in the user's personal space. This space is divided in three stages up to 0.9 meters, separated by 20 centimeters on the user's view-axis. The choice of these distances are based on how the head movements of an user affect the notion of distance [Kelly et al. 2004b].We also present the objects with different sizes, for a better understanding of distance, thus one can underestimate it by the size of the object [Gogel and Da Silva 1987]. The sphere diameter varies between 3, 6 and 9 centimeters. The objects are placed using perspective distortion(Figure 2). This condition makes some objects appear the same size in some conditions. Also, the spheres are rendered in two different colors, blue for the initial target and green for the final. So, briefly the variables used in each turn are: a) *object separation* according to perspective deformation, b) *object depth* (z coordinate) , c) *size* of the objects and d) *stroke direction* (right to left or left to right).

To compare the precision between the two types of see-through systems three types of data were collected: 1) The error between the initial objects center and the initial stroke position; 2) The error between the final object's center and the final stroke position; 3) Time elapsed of each stroke;This data was used to evaluate the precision of the task and thus the depth perception, as used by Iwamoto et al. [Iwamoto and Ishikawa 2013] and Swan et al. [Swan II et al. 2007]. The data used indicates the error between the real distance of the object to the perceived distance of the user.

An additional rigid-body was used to track a Wiimote. Since none of the devices have occlusion between real and virtual objects, and this is considered one of the most significant depth cues, we chose to use a virtual cursor of 1 centimeter of diameter to provide a better relation between virtual and physical objects.

3 Evaluation

We evaluated the system in a controlled environment, using an intragroup approach where each subject performed tasks using both OST and VST devices. The group was composed by 31 participants, all of them with a major in Computer Science, five of which were female. Participants were on average twenty-four years old with a standard deviation of five. Most subjects reported previous experience with 3D systems but 16 (51.67%) participants didn't have any experience with VR/HMD systems.

At first, users were greeted with a description of the test objectives, then, they were asked to fill up a profile questionnaire to assess pre-

[2] NaturalPoint Optitrack: http://www.optitrack.com/

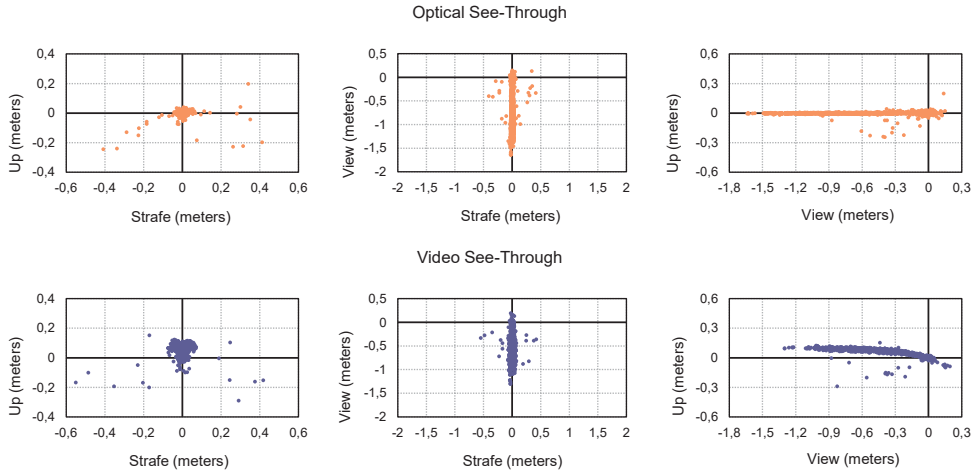


Figure 3: *Error distribution around target separated by axis. OST in orange and VST in purple.*

vious experience with either of the devices used. Subsequently, we gave subjects a brief description of the application and the devices. Next, we placed the users in the initial position of the test, facing the virtual test objects. Then, in order to familiarize users with the procedures, they performed a task in a training scenario, composed by some of the combinations of the main test, where they could explore the virtual environment and learn the necessary skills to execute the test task. The combinations on the training were randomly varied for each participant to avoid bias.

After performing the training task, we gave additional task-specific information and asked users to execute the tests. For each of the conditions mentioned before, users had the opportunity to do it again if they found that they had made a mistake. We also randomly varied the starting order, with 16 subjects starting the test using the OST and 15 with the VST.

Finally we asked subjects to fill a user-experience related questionnaire. To gather user profiles, preferences and factors such as comfort and satisfaction. The questionnaire contained a list of statements followed by a 6–point Likert Scale, where 1 meant that the user didn't agree at all with the statement and 6 means she/he fully agreed with it, as summarized in Table 1. In addition to the questionnaire, we conducted a semi-structured interview in order to capture participants' perceptions about performed tasks, clarify their answers to questionnaire and elicit suggestions for improvement.

3.1 Results and Discussion

We present the main observations made during the evaluation sessions.Additionally we discuss the analysis and the results obtained.

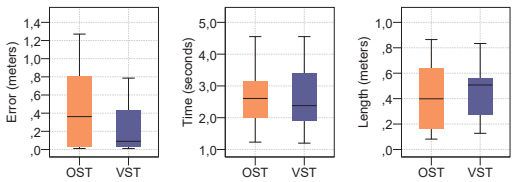
3.1.1 Task Performance

To better visualize the distribution around the spheres we opted on grouping the initial and end spheres on a single graphic. This is explained because the obtained data does not follow a normal distribution, according to the Shapiro-Wilk test and there are non-statistical differences between the error on any axis of both targets by conducting the Mann-Whitney non-parametric Test. Figure 3 illustrates the error distribution separated by axis around the target.

To better analyse and understand the results we chose to use the mean unit of each of the 84 combinations (42 by device) of the users. These combination correspond to variations of the variables described on the task design section.Furthermore, we used the Wilcoxon Signed-Rank Test to compare each of the device data regarding time, error and stroke length, grouped by device.

About the magnitude of the error (Figure 4a), we can say that the error is smaller on the VST (233 mm) in comparison to the OST (mean = 427 mm), we can also emphasize a larger distribution on distances near the object target (10 cm) on the VST. Consequently, we can also say that the error magnitude is statistically favorable to the VST compared to the OST ($Z = -3.557, p < 0.01$). Decomposing the error on strafe, Up and View axis, we can emphasize the positive results of the VST (mean = 226 mm) comparing to the OST (mean = 424 millimeters) on the View axis ($Z = -3.687$ $p < 0.01$). On the Strafe and Up axis we find that the smaller FoV on the OST (mean = 15 and 6 mm, respectively) provoked an expressive smaller error than the VST (16 and 37 mm), but statistically significant only on the Up axis ($Z = -3.989$ $p < 0.01$). Taking into account that the average diameter of the targets were 6mm, we can even disregard the error on the Up axis. Another observation taken from the tests is that in some cases the users walked on the opposite direction of the spheres, justifying some outliers on the view axis.

Regarding stroke length (Figure 4c), we did not found statistical significant differences between both devices. Despite that, the stroke on the OST (mean = 412 mm) has a smaller length in comparison with the VST (mean = 445 mm). This reaffirms the lack of accuracy on the View axis, making the spheres to appear to be nearer from each other. We also found that the mean time needed to accomplish the tasks were higher on the OST (mean = 3000 ms) in comparison with the VST (mean = 2729 ms) (Figure 4b), even though without statistical significance.



(a) *Error Plot* **(b)** *Time Plot* **(c)** *Stroke Length Plot*

Figure 4: *Task performance - median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers) for error, time and stroke length. OST in orange and VST in urple.*

Question	OST	VST
It was easy to locate myself on the environment*	4 (2)	5 (1)
It was easy to coordinate my movements*	3 (1)	5 (0)
It was easy to draw*	4 (2)	5 (1)
I felt fatigue*	4 (2)	2 (2)

Table 1: *User preferences: Median (Interquartile Range). * indicates statistical significance.*

3.1.2 User Preferences

The questionnaire had eight questions, four for each specific device and each of them containing a relevant part of the user experience.To compare the results obtained, we used a Willcoxon Signed Ranks Test to compare each issue (Table 1).

In general, users preferred the VST over the OST. Regarding comfort issues, the users experienced less fatigue when using the VST in comparison to the OST ($Z = -3.396$ $p<0.01$), when asked about this on the post-test interview the users related the extra fatigue to the limited field of view of the Meta glasses. About the test task experience, the users found it easy to locate themselves ($Z = -2.321$ $p = 0.02$), draw ($Z = -4.091$ $p<0.01$) and coordinate movements while performing the drawing task ($Z = -4.304$ $p<0.01$) on the augmented environment using the VST.

Further, when asked on the post-interview about these issues, participants said that it was related to full-immersion sensed on the VST, while they found the superposition of the virtual objects somewhat artificial on the OST. Because of this, some of the participants related of not correctly perceiving the depth of virtual objects and getting lost on the augmented environment, as previously discussed.But, for them, the ability of walking around virtual objects helped them to better understand the relation between virtual and real objects, specially on the OST device. About the virtual cursor, users reported that it helped them to relate their movements and establish better scale understanding with the virtual objects.

4 Conclusions

In this work we presented our approach to evaluate depth perception in two different configurations of head-mounted see-through displays. The main contribution lies in implementing and evaluating a task procedure that combines both egocentric and exocentric approaches to estimating the distance applied to the comparison of both OST and VST regarding their limitations and advantages.

We also emphasize the differences between devices used on the experiment and what makes each of them unique. When developing applications for VR/AR, one of the chief concerns is building on the technique that best suits a given device. Regarding the Meta Glasses, because of their limited FoV, the recommended usage lies in applications with hand gestures or augmented scenes where 3D objects are either inside or close to the user's field of view. From the results found on this paper, we conclude that when objects lie far from the user's FoV, people wearing Meta Glasses tend to get lost, which may compromise the user experience. Another issue worth point out is that optical see-through configurations still remains a prototype, which will most likely be improved over time. On the other hand, the Rift see-through configuration is more suited to precise augmented virtual-reality applications, using optical tracking.

Results show that depth perception was statistically better on the modified VST as compared to the OST. Another issue highlighted by user preferences is that users felt both more immersed and required significantly less time to perform tasks using the VST configuration. Because of this, VST users are able to accomplish the same task on less time and with less effort.

In more complex applications typical of engineering and architectural modelling, accomplishing tasks may be very laborious and troublesome to less experienced users. However, see-through technologies prove beneficial in allowing most people to establish close relationships between virtual objects and the real physical world. Indeed, by being able to keep track of their position in the physical environment, designers can more effectively explore a virtual environment, by feeling less constrained even in cluttered offices.

Acknowledgements

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with ref. UID/CEC/50021/2013, through projects TECTON-3D (PTDC/EEI-SII/3154/2012) and IT-MEDEX (PTDC/EEISII/6038/2014), and doctoral grant SFRH/BD/91372/2012. Daniel Medeiros would like to thank CAPES Foundation, Ministry of Education of Brazil for the scholarship grant (reference 9040/13-7).

References

CUTTING, J. E. 2003. Reconceiving perceptual space. In *Looking into Pictures*, H. Hecht, R. Schwartz, and M. Atherton, Eds. The MIT Press, Cambridge, MA.

EDWARDS, E., ROLLAND, J., AND KELLER, K. 1993. Video see-through design for merging of real and virtual environments. In *IEEE Virtual Reality Annual International Symposium*.

GOGEL, W. C., AND DA SILVA, J. A. 1987. Familiar size and the theory of off-sized perceptions. *Perception & Psychophysics 41*.

IWAMOTO, K., AND ISHIKAWA, J. 2013. High resolution binocular video see-through display for interactive work support - development of system and evaluation of depth perception and peg-in-hole tasks. In *IEEE International Conference on Systems, Man, and Cybernetics*.

KELLY, J. W., BEALL, A. C., AND LOOMIS, J. M. 2004. Perception of shared visual space: Establishing common ground in real and virtual environments. *Presence: Teleoperators and Virtual Environments 13*, 4.

KELLY, J. W., LOOMIS, J. M., AND BEALL, A. C. 2004. Judgments of exocentric direction in large-scale space. *Perception 33*.

LIZANDRA, M. C. J., AND CALATRAVA, J. 2011. An Augmented Reality System for the Treatment of Phobia to Small Animals Viewed Via an Optical See-Through HMD: Comparison With a Similar System Viewed Via a Video See-Through HMD. *International Journal of Human-computer Interaction 27*.

LOOMIS, J. M., AND KNAPP, J. M. 2003. Visual perception of egocentric distance in real and virtual environments. *Virtual and adaptive environments*, 11.

ROLLAND, J. P., AND FUCHS, H. 2000. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments 9*, 3.

STEPTOE, W., JULIER, S., AND STEED, A. 2014. Presence and discernability in conventional and non-photorealistic immersive augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality*.

SWAN II, J. E., JONES, A., KOLSTAD, E., LIVINGSTON, M. A., AND SMALLMAN, H. S. 2007. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics 13*, 3 (May).