Object-Alignment Performance in a Head-Mounted Display Versus a Monitor

Pavlo Bazilinskyy, Natália Kovácsová, Amir Al Jawahiri, Pieter Kapel, Joppe Mulckhuyse, Sjors Wagenaar,
and Joost de Winter
Department of BioMechanical Engineering
Delft University of Technology
Delft, the Netherlands
p.bazilinskyy@tudelft.nl

Abstract—**Head-mounted displays (HMDs) offer immersion and binocular disparity. This study investigated whether an HMD yields better object-alignment performance than a conventional monitor in virtual environments that are rich in pictorial depth cues. To determine the effects of immersion and disparity separately, three hardware setups were compared: 1) a conventional computer monitor, yielding low immersion, 2) an HMD with binocular-vision settings (HMD stereo), and 3) an HMD with the same image presented to both eyes (HMD mono). Two virtual environments were used: a street environment in which two cars had to be aligned (target distance of about 15 m) and an office environment in which two books had to be aligned (target distance of about 0.7 m, at which binocular depth cues were expected to be important). Twenty males (mean age = 21.2, SD age = 1.6) each completed 10 object-alignment trials for each of the six conditions. The results revealed no statistically significant differences in object-alignment performance between the three hardware setups. A self-report questionnaire showed that participants felt more involved in the virtual environment and experienced more oculomotor discomfort with the HMD than with the monitor.**

Keywords—*human-machine interaction; Oculus Rift; depth perception; user-centered design; human factors; ergonomics*

I. INTRODUCTION

Recently, a large number of head-mounted displays (HMDs) have become available on the market. HMDs may hold promise for a variety of applications in entertainment, training, and research.

It has been found that participants who wear an HMD experience the environment to be immersive and emotionally intensive [1]. The immersed feeling is caused by the large field of view (i.e., the viewer's visual field covered by the display), the large field of regard (i.e., the coverage of the virtual environment provided by the display when head motion is considered), and the fact that an HMD occludes visual access to the physical world [2]. Moreover, HMDs can present a unique image to each eye (also called binocular disparity), giving rise to stereoscopic vision.

To elicit realistic task performance, the spatial relations in the virtual environment have to be perceived by the user in the same manner as they would be perceived in the real world. However, there are vital differences in human perception between real and virtual environments. For example, a large number of studies have found that people underestimate distance in virtual environments [3]–[5]. Moreover, HMDs may bring cybersickness manifested by oculomotor discomfort, nausea, or disorientation [6],[7],

which may have a negative effect on task performance as well.

A. Human depth perception

There are three types of visual cues that influence the perception of depth: (1) pictorial depth cues, (2) oculomotor depth cues, and (3) binocular depth cues [8]. Pictorial cues are two-dimensional sources of information that produce an impression of depth. The best known pictorial cues are relative size, occlusion, height in the visual field, relative density, aerial perspective, and motion perspective [9]. Experiments in virtual environments have shown that user's positioning actions are more accurate when shadows and perspective cues are added [10],[11]. Although pictorial cues are generally important, they may contaminate depth perception at certain tasks. For example, it has been found that when objects of the same shape had different sizes, the placement errors were larger compared to when the objects were of equal size [12].

Oculomotor cues include accommodation and convergence, in which the muscles and orientation of the eyes provide depth cues. Under normal vision, accommodation and convergence act in synchrony. However, when perceiving the environment through an HMD, the eyes accommodate to a fixed viewing distance while eye convergence varies according to the distance to the virtual object, due to binocular disparity [13].

The third type of visual cues are binocular cues, which are dependent on information from both eyes. The depth sensation is provided by a process called stereopsis. Stereopsis is the mechanism whereby the brain creates a three-dimensional mental model from the combination of images received from both eyes.

Cutting and Vishton [9] categorized depth cues as follows: cues that are invariant with distance (e.g., occlusion, relative size, relative density), cues that diminish with distance (e.g., binocular disparity, accommodation, convergence), and a cue that increases with distance (i.e., aerial perspective). Accordingly, Cutting and Vishton divided the space around the observer into three functional regions: the personal space (0–2 m), the action space (2–30 m), and the vista space (beyond 30 m). Binocular and oculomotor cues play an important role in depth perception particularly for distances up to 2 m, whereas pictorial cues can be effective at farther distances as well [9].

B. Effects of stereoscopic displays on task performance

At near distances, stereoscopic viewing has been shown to have a positive effect on participants' performance in placing an object on a static surface [10] and in various types of other alignment and placement tasks [12],[14],[15]. However, these experiments were usually performed with elementary objects such as cubes, blocks, or balls in virtual environments that do not include many pictorial cues. As HMDs are gaining popularity, for example, in driving and flight simulators or for simulating manipulation tasks in industrial settings, it becomes important to understand depth perception when multiple pictorial cues are available, as is the case in the real world.

Creem-Regehr et al. [16] investigated the effect of binocular viewing on judgments of absolute egocentric distance in a full pictorial-cue real environment. Participants were asked to walk while blindfolded to targets at distances ranging from 2 to 12 m in monocular (the non-dominant eye was covered with an eye patch) and binocular viewing conditions. No significant differences in performance were found between the two conditions. In a similar study [17], distance judgments in the action space in virtual environments were underestimated by approximately 45%. The results did not show significant differences in distance judgments between HMD conditions with stereoscopic viewing, monoscopic (bi-ocular) viewing, and monocular (i.e., one-eye) viewing, suggesting that stereopsopic depth cues are of lesser importance if the environment features ample pictorial depth cues.

C. Aim of this research

The aim of this study was to investigate how a stereoscopic HMD affects object-alignment performance as compared to the same HMD with monoscopic settings and a conventional computer monitor. In addition, self-reported immersion and cybersickness were compared between a conventional monitor and an HMD.

Three setups were compared: a computer monitor, an Oculus Rift with monoscopic viewing (same image presented to both eyes), and an Oculus Rift with stereoscopic viewing (disparity between both eyes). This design allowed us to examine whether binocular disparity (HMD stereo vs. HMD mono) and the high immersion (HMD mono vs. monitor) contribute to object-alignment performance. As mentioned above, the utility of depth cues depends on viewing distance. Participants in the present experiment aligned objects at two viewing distances: an office environment with a target distance of 0.7 m (i.e., personal space) and a street environment with a target distance of 15 m (i.e., action space). Both environments were rich in pictorial depth cues, such as occlusion, height in the visual field, and relative density. However, we used slightly different objects that have to be aligned. This way, the participants could not use the size/symmetry of the objects as a cue to align them.

II. METHOD

A. Apparatus

Two virtual environments were created in Unity v5.0.1: an office environment (Fig. 1) and a street environment (Fig. 2).

Fig. 1. Office environment. In each trial, the green book (which is 40% larger than the dark red book) was at the left side of the table, and the dark red book was at the right side of the table. The distance between the camera and the target book (right book in this screenshot) is 0.7 m.

Fig. 2. Street environment. In each trial, the dark red car was in the left lane and the black car was in the right lane. The distance between the camera and the target car (left car in this screenshot) is 15 m.

Three hardware setups were used: 1) a computer monitor, 2) an Oculus Rift with monoscopic settings, meaning that each eye received the same image (HMD mono), and 3) an Oculus Rift with binocular settings with an inter-pupillary distance of 64 mm (HMD stereo). The camera always looked straight ahead and remained in the same orientation. Camera motion was disabled for all setups and no head position tracking was used.

The computer monitor was the 15.6-inch screen of an HP EliteBook laptop with a resolution of 1920x1080. The distance between the monitor and the eyes of the participant was approximately 50 cm. The HMD was an Oculus Rift Development Kit 2 with a resolution of 960x1080 per eye. Fig. 3 and 4 illustrate the difference of the image presented to each eye. The two environments combined with the three hardware setups yielded six conditions.

In the office environment, participants had to align the nearest edge of a controllable book with the nearest edge of a fixed target book. The camera position was the middle of the table at a height of 0.15 m above the table. The distance in *x* direction (i.e., along the direction of the table) from the camera to the nearest edge of the target book was 0.665, 0.687, 0.700, 0.713, or 0.735 m. This slight variation was used to prevent the participants from memorizing points of reference (e.g., specific textures on the table) that could be used to align the objects. The controllable book was either the left book or right book. The starting distance between the camera and the controllable book was 0.4 m.

Fig. 3. Overlay of screens for left and right eyes in the office environment. This figure illustrates the degree of binocular disparity.
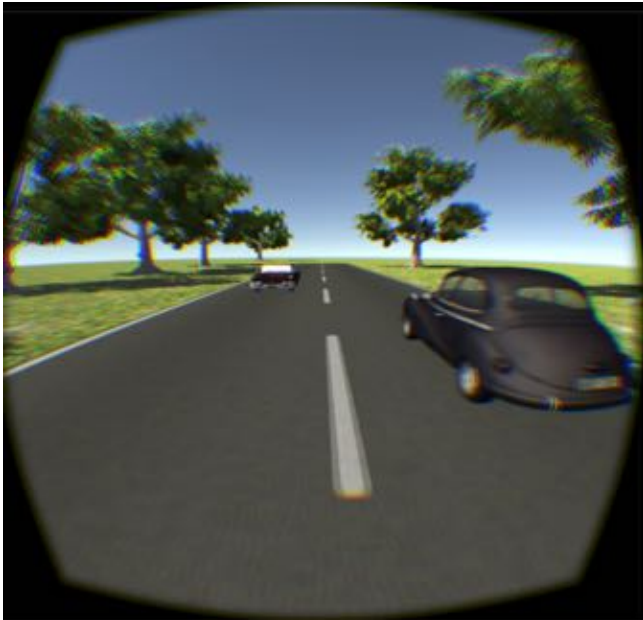

Fig. 4. Overlay of screens for left and right eyes in the street environment. This figure illustrates the degree of binocular disparity.

In the street environment, a controllable car had to be aligned with a fixed target car. The camera position was the middle of the two-lane road at a height of 1.80 m. The distance in x direction from the camera position to the bumper of the target car was 14.25, 14.50, 15.00, 15.50 or 15.75 m. The target car was either the left car or the right car. The starting distance in x direction between the camera and the rear bumper of the controllable car was about 2.5 m. Fig. 5 illustrates a situation when the two cars are perfectly aligned with each other.

Fig. 5. Illustration of perfect alignment of the target car and the controllable car. For this figure, the camera was positioned at a height of 10 m and faced downward. Such top view was not shown to participants.

The participant could move the controllable car/book along the x direction by pressing the upper and lower arrow keys of the keyboard. Because the starting position of the controllable object was always closer to the camera than the target object (see Fig. 1 and 2), the participant first had to press the upper arrow key, in order to move the controllable object away from them. Pressing the spacebar recorded the position of the controllable car and loaded the next trial. Participants could take as much time as they needed to complete the trial. After all, our goal was to measure how precisely the participants could align the objects, not the extent to which they were able to perform the task within a given time.

B. Participants

Twenty males (mean age = 21.2 years, *SD* age = 1.6 year, no self-reported visual handicap) participated in the experiment.

C. Procedure

At the start of the experiment, the participants read and signed a consent form and demographics questionnaire. The form introduced the experiment to the participants and explained the goal of the research. Afterwards, the six measurement sessions started.

Each session consisted of approximately 10 trials. For each trial, a random condition was selected out of the 10 possible conditions (i.e., 5 possible distances x 2 sides [i.e., left vs. right book/car being the controllable object]). After each of the six sessions, a questionnaire was offered, in which participants were asked to estimate the distance to the target car or the target book in meters ("What was the distance between you and the target (the car or book?)" and to describe the strategy they used to fulfil the alignment task ("Did you use a specific strategy to perform the task? Please comment."). The questionnaire also contained questions about:

- Performance ("How successful were you in accomplishing what you were asked to do?", from *Perfect* to *Failure*),

- Effort ("How hard did you have to work to accomplish your level of performance?", from *Very low* to *Very high*),

- Involvement ("How much did the visual aspects of the environment involve you?", from *Not at all* to *Very much*),

- Correctness ("The objects appeared to be correct in size and distance", from *Not at all* to *Very much*),

- Three-dimensional impression ("I had a three-dimensional impression of the displayed environments and objects", from *Not at all* to *Very much*),

- General discomfort ("I felt uncomfortable", from *Not at all* to *Very much*), and

- Oculomotor discomfort ("I experienced oculomotor discomfort (eyestrain, difficulty focusing, blurred vision or headache)", from *Not at all* to *Very much*).

Each of these items was rated on a 21-point horizontal scale.

For each trial, the percentage distance error between the positioned object and the target object was determined. For each session, the mean error across the 10 trials (a measure of bias) and the standard deviation of the error across the 10 trials (a measure of precision) were calculated. The non-parametric Friedman test was used to assess differences between the three conditions.

III. RESULTS

Trials yielding an error score greater than 10% were removed. Such large errors may be due to the participant accidentally pressing the spacebar (e.g., in the beginning of the trial). Data remained for on average (*SD*) 11.0 (1.5), 10.4 (1.6), 9.8 (1.5), 10.8 (2.0), 10.3 (0.9), and 10.5 (1.2) trials for the Office monitor, Office mono, Office stereo, Street monitor, Street mono, and Street stereo conditions, respectively.

The results for the mean error (bias) are shown in Fig. 6. There were no statistically significant differences among the three setups for the office environment ($p = 0.951$, $\chi^2(2,38) = 0.10$) and street environment ($p = 0.449$, $\chi^2(2,38) = 1.60$). The results regarding the standard deviation of the alignment error (precision) are shown in Fig. 7. Again, there were no significant differences among the three setups for the office environment ($p = 0.951$, $\chi^2(2,38) = 0.30$) and for the street environment ($p = 0.861$, $\chi^2(2,38) = 0.10$). However, from Fig. 6 and 7 it can be seen that participants performed better in the office environment than in the street environment. Specifically, the median of the standard deviation of the error was 1.48% for the office environment and 2.54% for the street environment (The scores were first averaged per participant across the three setups). This difference was statistically significant according to a signed rank test ($p = 0.003$, $z = -2.95$, $n = 20$). The median of the mean error was 0.01% for the office environment and 0.72% for the street environment, a difference which was significant as well ($p < 0.001$, $z = -3.40$, $n = 20$).

Fig. 8 shows the results for the self-report questionnaire. The HMD yielded higher ratings of involvement and three-dimensional impression than the monitor. However, the HMD also yielded higher ratings of general discomfort and oculomotor discomfort than the monitor. In line with the measured actual performance, there were no statistically

significant differences between the three hardware setups in self-rated task performance.

After each session, participants provided a verbal estimate of the distance to the target. Participants overestimated the distance to the target book by a median of 114% (IQR = 161%), 114% (IQR = 143%), and 114% (IQR = 143%) for the monitor, mono and stereo conditions, respectively. However, participants underestimated the distance to the target car by a median of 33% (IQR = 47%), 20% (IQR = 33%), and 10% (IQR = 60%), respectively (Fig. 9). There were no statistically significant differences between the three setups for the office environment ($p = 0.118$, $\chi^2 = 4.27$) and the street environment ($p = 0.758$, $\chi^2 = 0.55$).

Out of 120 completed sessions (6 trials x 20 participants), 83 times a comment was provided regarding the strategy to perform the tasks. In the majority of comments (58 out of 83), participants made use of specific visual features (e.g., stripes on the road, edge of the book, wheels of the cars) or an imaginary horizontal line which they tried to use as a reference. In 16 of 83 comments, participants used more holistic cues, such as 'just guessing', 'feeling', or 'looking at the whole picture'. In 7 of the 83 comments, participants indicated they made use of pixels. This always occurred for the HMD conditions and never for the conventional monitor.
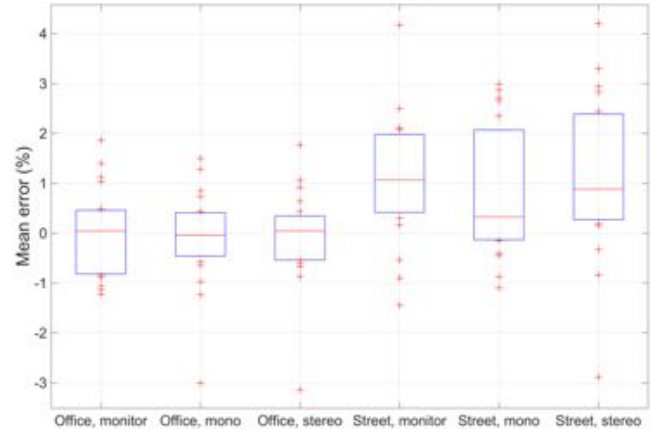

Fig. 6. Boxplot (showing the median, 25th percentile, 75th percentile, and data points outside the inter-quartile range) for the mean error for each of the six conditions ($n = 20$). A positive value means that the participant positioned the controllable car closer to the camera than the target.
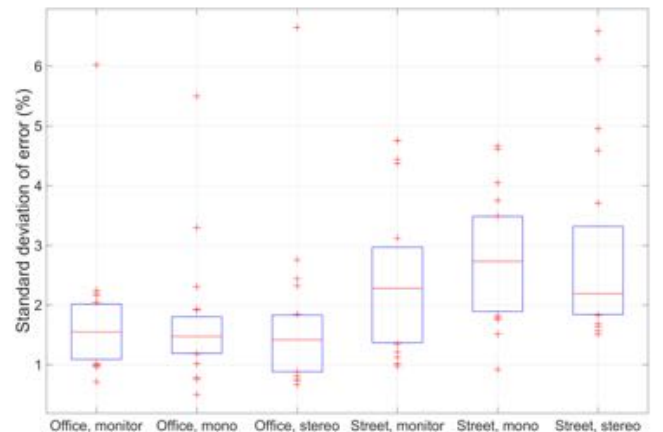

Fig. 7. Boxplot (showing the median, 25th percentile, 75th percentile, and data points outside the inter-quartile range) for the standard deviation of the error for each of the six conditions ($n = 20$).
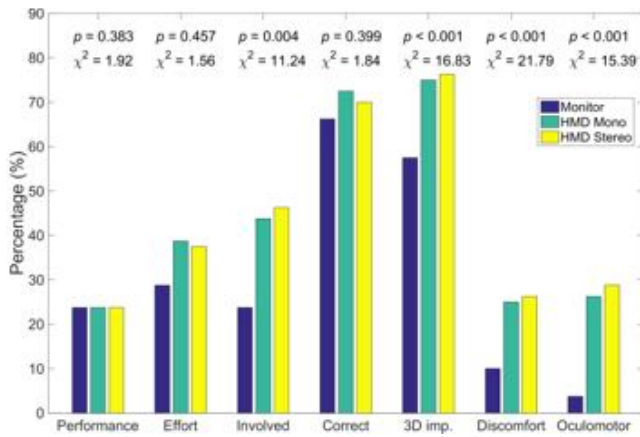
Fig. 8. Median self-reported ratings across the 20 participants. The *y*-axis is on a scale from 0% (*Perfect* for the performance item, *Very low* for the Effort item, *Not at all* for the other items) to 100% (Failure for the performance item, *Very high* for the Effort item, *Very much* for the other items). The scores were first averaged per participant across the office and street environments.
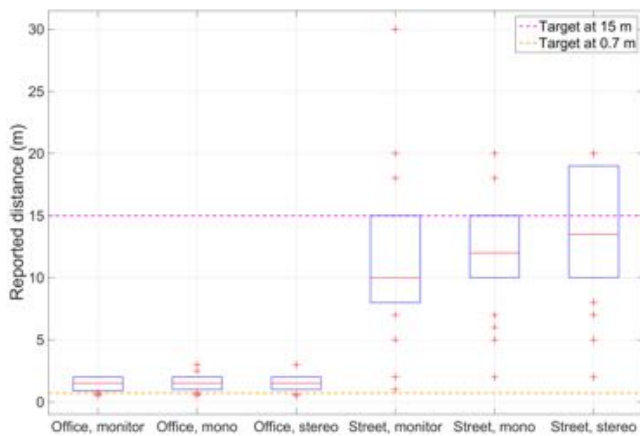

Fig. 9. Boxplot (*showing* the median, 25th percentile, 75th percentile, and data points outside the inter-quartile range) for the reported distance to the target object.

IV. DISCUSSION

The aim of this study was to compare participants' object-alignment performance between three hardware setups: a conventional monitor, an HMD with binocular disparity (HMD stereo), and the same HMD without binocular disparity (HMD mono). The results showed no statistically significant differences in object-alignment performance between these three conditions, neither for the office nor for the street environment. The former environment measured performance in the personal space for which binocular depth cues are known to be important for object-alignment performance [14], whereas the latter environment represents the action space where the degree of disparity is relatively minor (see Fig. 3 vs. Fig. 4). Our results suggest that neither binocular disparity (HMD stereo vs. HMD mono) nor conditions of immersion (HMD mono vs. monitor) improve object-alignment performance in virtual environments that are rich in pictorial depth cues.

The equivalent performance between the three conditions may have been caused by the fact that pictorial cues such as shadows and perspective (table, road) were strong in both environments. Moreover, participants could align the front edges of the object, which had the same rendered height, without having to use binocular cues. The self-reported strategies indicate that participants strongly relied on visual features (edges, pixels, lines) in order to achieve successful object-alignment performance.

Participants underestimated the distance to the target car but overestimated the distance to the target book (Fig. 9). We have no particular explanation for this phenomenon. It is possible that the relatively high declination angle of the camera with respect to the table surface was a cause of distance overestimation. It is also possible that pictorial cues such as familiar size (e.g., expected size of books) and texture or color of the environment, as well as perspective (e.g., the protruding table) are contributing factors to the overestimation of distance in the office environment.

Our results should be replicated with larger sample sizes, since statistical power with 20 participants and 10 trials per condition is probably limited. In addition, it cannot be ruled that specific hardware features, such as the resolution, aspect ratio, and luminance of the screens, or specific experimental features, such as the self-paced nature of the task, have been causes of the present null results. However, research suggests that the realism of graphics rendering is not associated with the accuracy of distance judgments [18].

The participants reported some discomfort (e.g., eye strain, sickness) when exposed to the Oculus. This can be explained by the fact that the fixed viewing distance imposed by an HMD (which is about 1.3 m [19]) may result in accommodation-convergence mismatch.

In our experiment, we did not make use of the full potential of the Oculus; no head movement was used, and hence no effect of motion parallax [20] could occur. The effect of head movement on distance perception and task performance may be addressed in future research. Moreover, for future studies, we recommend to employ a tighter experimental control of pictorial cues, to focus more strongly on measuring spatial relationships (e.g., verbal estimations of distance), to explore the effects of inter-camera distance (e.g., adjusted for each participant [17], hyperstereo, or hypostereo [14],[19]), and to investigate the effects of discomfort on task performance.

REFERENCES

[1] T. Reiners, H. Teräs, V. Chang, L. C. Wood, S. Gregory, D. Gibson, N. Petter, and M. Teräs, "Authentic, immersive, and emotional experience in virtual learning environments : The fear of dying as an important learning experience in a simulation," *Proc. 23rd Annu. Teaching Learning Forum*, Jan. 2014.

[2] W. R. Sherman and A. B. Craig, *Understanding Virtual Reality: Interface, Application, and Design*. San Francisco, CA: Elsevier Science, 2003.

[3] B. R. Kunz, L. Wouters, D. Smith, W. B. Thompson, and S. H. Creem-Regehr, "Revisiting the effect of quality of graphics on

distance judgments in virtual environments: A comparison of verbal reports and blind walking," *Atten. Percept. Psychophys.*, vol. 71, no. 6, pp. 1284-1293, Aug. 2009.

[4] R. Messing and F. H. Durgin, "Distance perception and the visual horizon in head-mounted displays," *ACM Trans. Appl. Percept.*, vol. 2, no. 3, pp. 234-250, July 2005.

[5] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson, "The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments," *ACM Trans. Appl. Percept.*, vol. 6, no. 2, Feb. 2009.

[6] J. Polcar and P. Horejsi, "Knowledge acquisition and cyber sickness: a comparison of VR devices in virtual tours," *MM Sci. J.*, pp. 665-669, Jun. 2015.

[7] F. Steinicke and G. Bruder, "A self-experimentation report about long-term use of fully-immersive technology," *Proc. 2nd ACM Symp. Spatial User Interaction*, Oct. 2014, pp. 66-69.

[8] J. D. Pfautz, "Depth perception in computer graphics," Univ. Cambridge, Cambridge, UK, Tech. Rep. UCAM-CL-TR-546, Sep. 2002.

[9] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances : the integration, relative potency, and contextual use of different information about depth," in *Handbook of Perception and Cognition, Vol. 5; Perception of Space and Motion*, W. Epstein and S. Rogers, Eds. San Diego, CA: Academic Press, 1995, pp. 69-117.

[10] H. H. Hu, A. A. Gooch, S. H. Creem-Regehr, and W. B. Thompson, "Visual cues for perceiving distances from objects to surfaces," *Presence*, vol. 11, no. 6, pp. 652-664, Dec. 2002.

[11] L. C. Wanger, J. A. Ferwerda, and D. P. Greenberg, "Perceiving spatial relationships in computer-generated images," *IEEE Comput. Graph. Appl.*, vol. 12, no. 3, pp. 44-51, 54-58, May/Jun. 1992.

[12] M. Poyade, A. Reyes-Lecuona, and R. Viciana-Abad, "Influence of binocular disparity in depth perception mechanisms in virtual environments," in *New Trends on Human–Computer Interaction*, J. A. Macías, T. Granollers, and P. L. Andrés, Eds. London: Springer-Verlag, 2009, pp. 13-22.

[13] J. P. Wann, S. Rushton, and M. Mon-Williams, "Natural problems for stereoscopic depth perception in virtual environments," *Vision Res.*, vol. 35, no. 19, pp. 2731-2736, Oct. 1995.

[14] L. B. Rosenberg, "The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks," *Proc. IEEE Virtual Reality Annu. Int. Symp.*, Sep. 1993, pp. 27-32.

[15] J. P. McIntire, "Investigating the relationship between binocular disparity, viewer discomfort, and depth task performance on stereoscopic 3D displays," Ph.D. dissertation, Wright State Univ., 2014.

[16] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson, "The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments," *Perception*, vol. 34, no. 2, pp. 191-204, Feb. 2005.

[17] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr, "Effects of stereo viewing conditions on distance perception in virtual environments," *Presence*, vol. 17, no. 1, pp. 91-101, Feb. 2008.

[18] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall, "Does the quality of the computer graphics matter when judging distances in visually immersive environments?," *Presence*, vol. 13, no. 5, pp. 560-571, Oct. 2004.

[19] R. Yao, T. Health, A. Davies, T. Forsyth, N. Mitchell, and P. Hoberman, "Oculus VR best practices guide (July 23, 2014 version)". Oculus VR, Inc., July 2014.

[20] S. H. Ferris, "Motion parallax and absolute distance," *J. Exp. Psychol.*, vol. 95, no. 2, pp. 258-263, Oct. 1972.