Perception of Depth in a Virtual Environment with Altered Perceptual Aspects

Katherine Kowalski
Stanford University
kasia4@stanford.edu

Sara Olsson
Stanford University
sarol@stanford.edu

Abstract

A study to test the effect of depth cues on a user's ability to navigate a virtual environment, measured by perceived distance and time to reach a target. Another study to test the effect of depth cues and view matrix selection to most accurately estimate distance from user to objects in scene and relative distances between objects in the scene. The goal of these studies is to evaluate how effective depth cues are in a Unity environment, and how accurate the distances perceived are.

1. Introduction

Accurately representing depth in a virtual environment is a continually researched problem. Humans rely on depth cues(monocular, binocular, accommodation, vergence, etc) to perceive the relative distance between objects.The biggest challenge currently, is creating this virtual environment with adequate depth cues, while accounting for vergence-accommodation, so as to decrease the vergence-accommodation conflict, motion sickness, and overall discomfort associated with immersing oneself in a virtual environment via a head mounted display.

2. Background

2.1. Wayfinding

The three types of knowledge that are essential to wayfinding are survey knowledge, procedural knowledge, and landmark knowledge [2]. Survey knowledge develops over a longer period of time of familiarizing oneself with the environment he or she is positioned in. Survey knowledge pertains to one's spatial orientation, perceived distances between objects and landmarks, and orientations of oneself and surrounding features. Survey knowledge lends itself to one's "map-like view" of the environment [2]. When navigating an environment, not necessarily virtual, a user utilizes procedural knowledge by memorizing their sequence of directions taken in order to achieve a specific outcome, whether that relate to finding a target, navigating to an exit, or finding an object in the scene. The steps taken to reach a target would be memorized as a series of turns relative to landmarks. Lastly, landmark knowledge pertains to one's memorization of distinct locations or objects having a specific spatial orientation relative to other objects in the scene. Landmark knowledge lends itself to procedural knowledge when navigating an environment.

2.2. Proprioceptive Feedback

The ability to physically walk through a virtual environment provides proprioceptive feedback, helping to inform a user of the position and orientation of one's head and limbs. Wayfinding studies have concluded that the ability to physically walk through a depicted environment results in more successful navigation in a virtual environment, due to the vestibular feedback that lends to having a sense of translation and rotation of oneself in space. The success was measured in terms of time to navigate to a target [1]. Another research group also found that allowing a user to physically rotate and translate in virtual environments was more beneficial than digitally maneuvering through an environment [5]. For these reasons, we will be using a lighthouse to allow a user to physically walk through a scene.

2.3. Perception

A person, upon observing a scene, will perceive the distance of an object based on many depth cues, such as eye convergence, stereopsis, disparity, etc. It has been found that on average, a person will underestimate the distance of an object farther away and overestimate the distance of an object nearer the observer [3]. VR displays attempt to implement a satisfactory number of depth cues to simulate a 3D environment, mostly using stereoscopy, but lack in accounting for many other visual cues, which lead to discrepancies between the visual cues and vestibular system, causing motion sickness and discomfort.

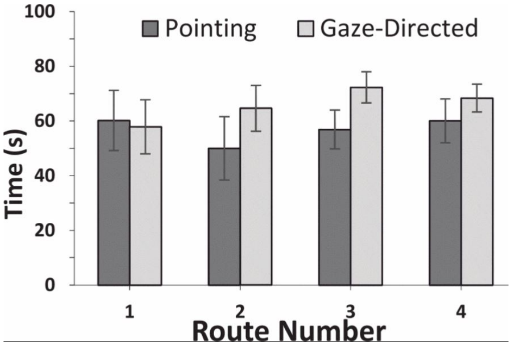
Furthermore, changes in stereoscopic depth, which are necessary to create depth in a scene, also cause a change in focus. In near-eye displays, a user must change their vergence angle to focus on the scene, but the asymmetry lends itself to the vergence-accommodation conflict. This conflict results in double vision, distorted visual clarity, visual discomfort, and fatigue [4].

To determine distance, humans rely on proprioceptive and visual cues. Proprioceptive cues(accommodation, vergence) are those that the visual system accounts for and are dependable for short distances(up to 2 meters). Visual cues are comprised of monocular and binocular cues. Although well researched, there is much improvement still to be made in current VR headsets, as it remains as one of the main issues, more specifically, distance estimation isn't as accurate as it should be [3].

3. Previous Work

3.1. Way-finding

A research group ran a study to compare a users' ability to navigate a virtual environment by steering according to their gaze direction, and with a method to decouple the gaze directions from the direction of travel... the pointing method allows the user to pick the translation direction by pointing (with a tracked hand or pointing device)" [2]. Motion control is usually gaze-directed, meaning the the user must rotate the camera in the desired direction. The group measured success of a method by measuring the time it took for a user to locate a target in the virtual environment, as well as the success rate of finding the targets. The study found that participants were more successful with the pointing method of control motion, which resulted in users arriving at the target location faster and with fewer errors [2].



3.2. Depth

There are many ways to evaluate a human's perception of depth in a virtual system. These include specifying distance verbally, using an "internal scale" to describe in units (meters, feet), using a scale of measurement, such as using a slider, a measurement between two objects (object A is 2x farther than object B, and through triangulation [3].

> In the absence of distance cues, the subject locates the object at a default distance, called dark vergence in visual perception, which is located 1.9 m from the subject [3].

A lab group crafted a user study to observe the effect of a standard virtual environment, a blurred, and a faded one on a user's altered perception of depth, sense of presence, and task completion time [7]. The study consisted of a user and a hand tracker to study the perceived distance necessary to grab an object in the virtual scene.

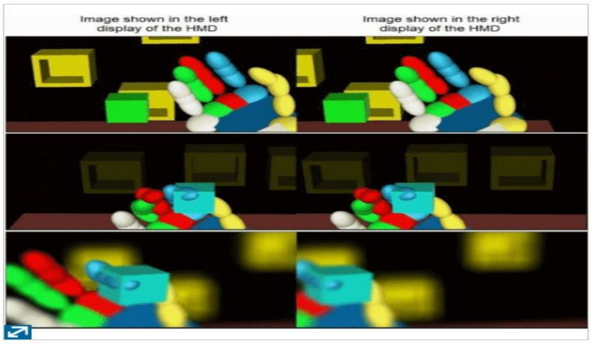
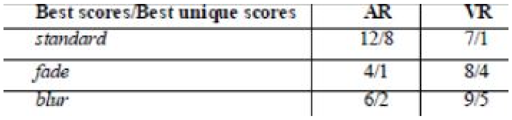


Image shown in the left display of the HMD
Image shown in the right display of the HMD

Figure 2:
Screen captures of stereo visualization for each sub-condition in AR: (Upper) Standard; (Middle) Fade and (Bottom) Blur.

The research group tested their environment on 16 participants and used guiding questions after each condition for each participant. The group used destination boxes (those that the participant was meant to reach for) to be 6x6x6 cm placed 5-10 cm in front of the target marker. The group used an alpha fade filter, and a depth-of-field blur effect from the Unity 3D standard assets attached to the cameras. The point of focus stood 70cm from the cameras (focus point positioned just behind the destination boxes for the blur effect). The study found that most people felt more "present" in the standard environment, but there was no significant difference in the exploration times (time that the user took exploring the virtual environment before reaching for the correct box). In analyzing which case users performed best in, researchers focused only on users that performed best statistically in one of the three environments.



Best scores/Best unique scores	AR	VR
standard	12/8	7/1
fade	4/1	8/4
blur	6/2	9/5

Of these users, the greatest number of unique best score were achieved under fade and blur environments. Because these results seem to contradict common perception of when a user should perform best in a clear and sharp environment, the group concludes that further research is necessary to study the effect or visual effects on depth perception in VR [7].

Another research group found that people partly rely on the size-distance invariance property to estimate relative depth. [8]

4. Methods

All virtual environments are created with Unity, and using a ViewMaster VR Starter Pack, a Topfoison 6” 1080p LCD, a InvenSense MPU-9250 IMU, an Arduino Teensy 3.2 microcontroller, 2 USB cables (for Arduino and LCD driver board power) and HDMI, and a light house. Participants volunteered to take part in the studies described below. 10 male and 5 female participants within the age range of 19-23 years old participated in the studies. Previous experience in VR varied, though the majority had little to no previous experience. Participants all were given study 1 followed by study 2, but the scene (A or B) they started with was randomized.

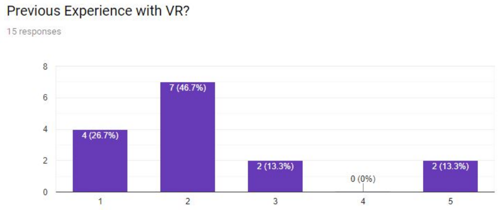


Figure 1. Participant prior VR experience

1. Study 1: Participants will volunteer to wear a head mounted display(HMD) to navigate a virtual environment in several situations. Light houses will be used to enable the user to physically walk, as they navigate the environment. One run will be tested with depth cues(environment A), and another without (environment B). The target is 2.97 meters away in both scenes. The user’s perception of distance and time to navigate to a target destination is measured.

2. Study 2: In a still environment, a scene with a perspective view matrix (A) and orthographic view matrix (B) will be compared to measure a user’s ability to discern the distance of an object relative to themselves and between objects. The user is intended to estimate distance from oneself to each of the blocks. The user is then asked to estimate the relative distances between blocks. The user is given a rod 1m away from them as a reference.

4.1. Guiding Research Questions

1. Study 1

   (a) Before the user begins the study: How far away do you think your destination is?

   (b) How long did it take a user to navigate from point A to point B with and without depth cues?

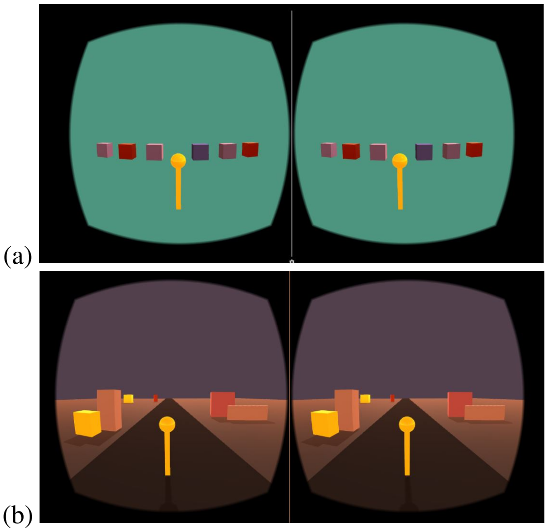


Figure 2. A virtual environment depicted with all Unity depth cues (a) and without depth cues (b). The user is intended to estimate distance to target and walk towards it. The time it takes to arrive at the destination will be measured and compared. Because we are only able to work with one lighthouse with the hardware we are working with, we are limited to walking forward.

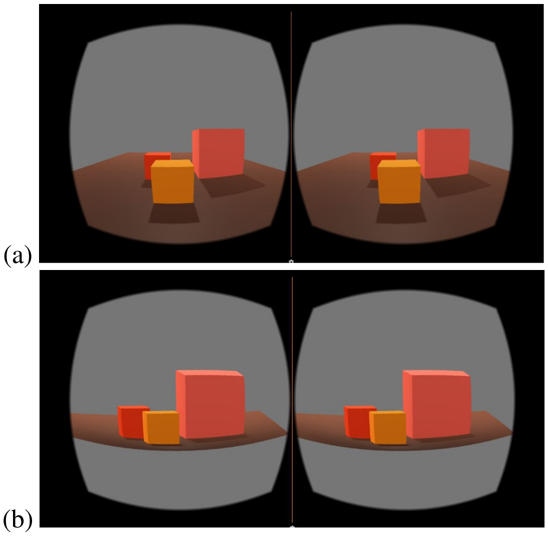


Figure 3. A virtual environment depicted with perspective view matrix (a) and orthographic view matrix (b).

   (c) On a scale of 1-5, how difficult was it to navigate environment A,B?

   (d) What made it easier/more difficult to navigate the environments?

2. Study 2

   (a) How far away do you think the object is in scenario A,B?

(b) How far are objects C and D from each other?

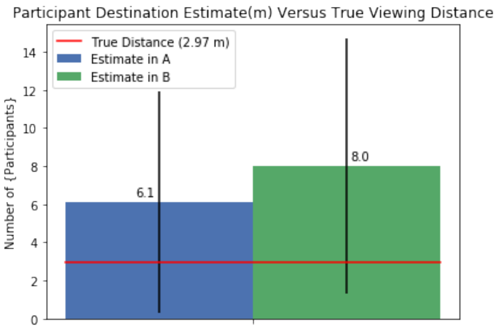
(c) Which scene (A or B) was easier to estimate the distances in?

(d) Which scene (A or B) are you more confident of your distance estimates in?

5. Results

5.1. Study 1

$\frac{11}{15}$ participants reported environment A (depth cues) was easier to navigate because of the road and greater number of objects to help with one's spatial orientation. $\frac{3}{15}$ participants reported environment B (no depth cues) was easier to navigate because there were fewer distractions, and the user was able to focus on just the target destination. $\frac{1}{15}$ participants reported that neither environment was easier/harder to navigate.



(a)

	Environment A	Environment B
Average Error(m)	3.1	5.0

(b)

Figure 4. (a) User estimates of target destination in both environments. The red line shows the ground truth. (b) Average error between user distance estimate and true distance (2.97 m).

The time it took for a participant to navigate the environment from the starting position to the target destination was measured in seconds. The user knew they had reached their destination when a "ping" sounded. The results of each environment are shown in figure 5. We did not derive any significance of these results, as is most likely due to the fact that the total walking distance was 2.97 meters in a straight line. The spikes we see in certain times may be attributed to the number of people who started with one environment versus another, and became accustomed to the walking environment, thus decreasing the time it took to navigate the next environment that was presented to them.

In estimating the target distance in environments A and B, a t-test resulted in a p-value of 0.4, which is not significant. The user is likely to estimate a similar distance in the Unity environment with and without depth cues. This seems contrary to what we would've expected to see. This may be due to the fact that our sample size may not have been large enough to effectively evaluate how accurate a user is on average in estimating a target distance in each of the environments.

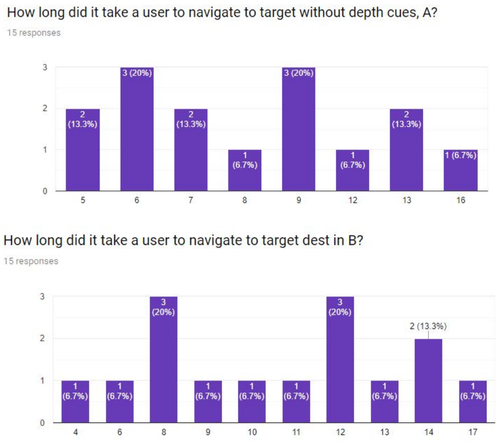


Figure 5. (a) Average time: 8.7 sec (b) Average time: 8.0 sec

5.2. Study 2

In environment A, users tend to underestimate distances of objects and distances between objects, whereas in environment B, users tend to overestimate the distances of objects and relative distances between objects. Figure 6 shows these results.

Figure 7 shows that more users were confident in their estimates in environment A. The corresponding estimates on average are closer to the true distance estimate. On the other hand, those more confident in their reported distance estimates in environment B tended to be further from the true distance. The implies that user depth perception is more accurate in an environment rendered with a perspective view matrix, which is exactly what we would expect.

Figure 8 seems contradictory because while 66.7% of users reported environment B being easier to estimate distances in, only 40% of users reported being more confident in their estimates in environment B. This leads us to believe that some users may not have been confident in either of their estimates.

6. Discussion

Through this study, our goal was to assess the accuracy of Unity depth rendering. We found in our study that users tend to increase their estimation error for objects farther from their POV, as well as in estimating relative distances

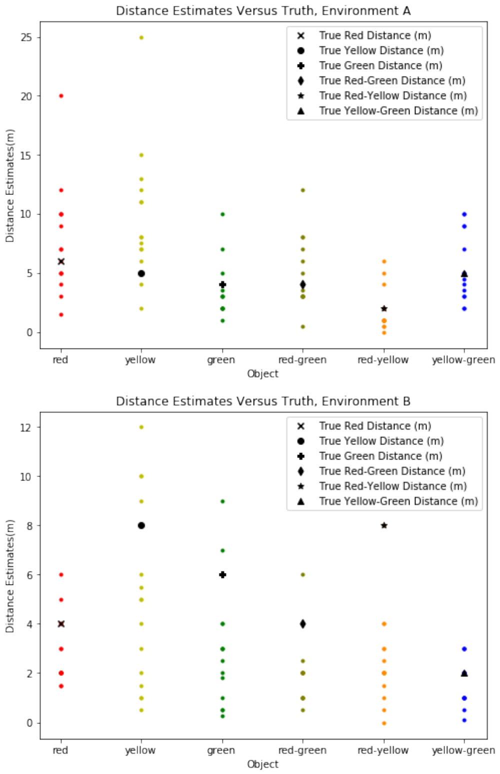


Figure 6. User distance estimates from self to given objects, and distances of objects relative to each other in the scene. All distance estimates are given in meters

between object. While the average estimate values for objects in a scene are not too far from the truth, we do see a wide range of distances that different users perceive objects at in an environment regardless of a perspective or orthographic view matrix.

6.1. Study 1

The navigation study did not show a significant disparity between distance estimates with and without depth cues, although the average user more accurately estimated the distance to the destination, referencing the road and other objects as helpful tools to become spatially oriented in the scene. We conclude that for a target about 3 m away, a user will tend to overestimate the distance in a virtual environment, regardless of depth cues. Time to navigate the environment showed no significance, as this would be due to the simplicity of the task (walking 3 m in a straight line). This contradicts the findings of [3]. The users in this study did not tend to overestimate closer distances and underestimate farther distances.

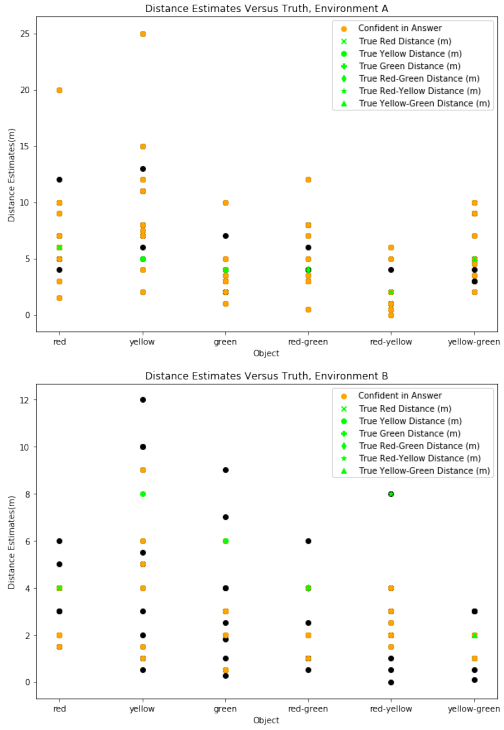


Figure 7. User distance estimates relative to which environment the user reported being more confident in his or her estimates in.

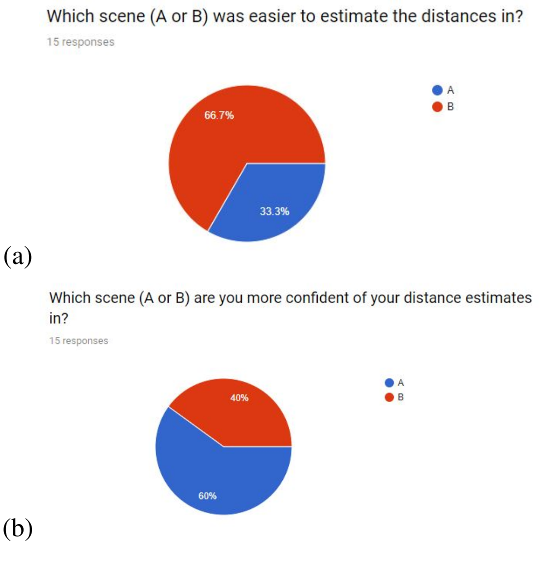
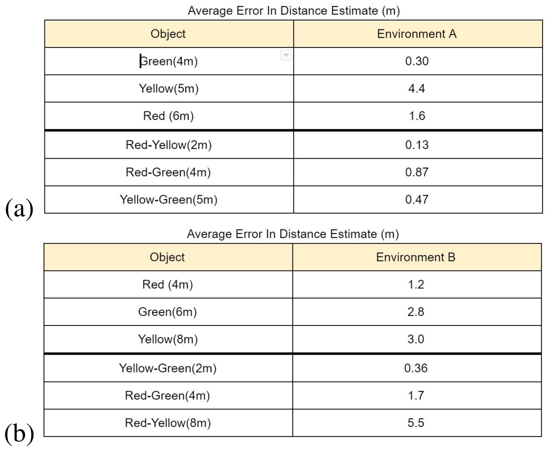


Figure 8. (a) Users report which environment they felt was easier to estimate distances in. (b) Users report which environment they are more confident of their distance estimates in.

6.2. Study 2

Figure 6 shows the wide range of distance estimates that users guessed for the various objects. In environment A



(a)

Average Error In Distance Estimate (m)

Object	Environment A
Green(4m)	0.30
Yellow(5m)	4.4
Red (6m)	1.6
Red-Yellow(2m)	0.13
Red-Green(4m)	0.87
Yellow-Green(5m)	0.47

(b)

Average Error In Distance Estimate (m)

Object	Environment B
Red (4m)	1.2
Green(6m)	2.8
Yellow(8m)	3.0
Yellow-Green(2m)	0.36
Red-Green(4m)	1.7
Red-Yellow(8m)	5.5

Figure 9. Average error of participant's distance estimate from true distance for each environment.

(perspective view matrix), as the distance of the object increased, the margin of error for the user's estimate increased as well. As the distance between objects increased, there was also an increase in average estimate error. On average, a user estimated an object 4m away from himself/herself 66% less of the time than estimating a relative distance of 4m between objects. In estimating an object 5m away from the user, the user's estimate on average was more than 8 times less accurate than in estimating that same distance between two objects in the scene. For object at distances further than 2 meters, user's were amble to more accurately estimate relative distances between objects than estimating the distance from oneself in the environment to the object in question.

In environment B, there is a similar trend in that distance estimates tend to err more as the distance of an object from the user increases. Similarly, as the distance between objects in the scene increases, the average error in estimation increases as well. For a distance of 4m, a user on average reported an estimate that erred 30% less than estimating the same distance between objects in the scene. For a distance of 8m, a user erred 45% less than when estimating that distance between objects in the scene.

In comparing the two environments, the environment rendered with the perspective view matrix produced results suggesting that estimating distance from oneself to the object in question was easier than in environment B. Similarly, in estimating distances between objects users performed better in environment A than in environment B.

7. Future Work

There are many things we would've liked to do had we had more time.

- We would run a user study with a greater variety of objects at various locations. It would be more effective to test distance estimates with distances ranging from 0m-15m, as well as testing distance estimates between objects 0m-15m apart. We would need to take into consideration the distance from the user that the objects were.
- Another parameter we would like to test would be the effect of different kinds of shading on depth perception. We would run similar tests comparing Phong shading, Gouraud shading, smooth shading, flat shading.
- Another study we would want to conduct, would be to create a much more involved environment, in which the user would need to reach a series of checkpoints, while needing to turn corners, in order to better study the effect of certain depth cues on the user's ability to navigate an environment. This would need to be done using the HTC Vive in order to track a greater range of motion.
- We would have also liked to replicate the study with a hand tracker, so as to see how users interact with the environment physically, more than just walking.
- Lastly, we would like to test the effect of a user's ability to estimate distances by removing one depth cue at a time to analyze the user's performance. This study would help to determine the weight of each variable in rendering a realistic scene that a user feels fully immersed in and spatially oriented well.

8. Conclusion

In conclusion, we found that depth estimates are more accurate in an environment with a perspective view matrix rather than a scene with an orthographic view matrix, as expected. The navigation study showed some interesting insight with a user's experience in traversing an environment with depth cues. While the majority reported an environment with depth cues, such as a road to travel down (vergence) and objects casting shadows for depth perception, some users did express that these depth cues made the scene more difficult to navigate because they were overwhelmed with the number of objects to focus on in the scene.

9. Acknowledgements

We would like to thank the course staff of EE267, especially Professor Gordon Wetzstein and our mentor, Nitish Padmanaban. In this class we learned about depth cues, stereo rendering, view matrices, transforming world coordinates into a 2D image to be streamed to the VR headset, among many other topics that helped influence this study.

References

[1] Chance, S., Gaunet, F., Beall, A. and Loomis, J. (1998). Locomotion Mode Affects the Updating of Objects Encountered During Travel: The Contribution of Vestibular and Proprioceptive Inputs to Path Integration. *Presence: Teleoperators and Virtual Environments, 7(2),* pp.168-178.

[2] Christou, C., Tzanavari, A., Herakleous, K. and Poullis, C. (2019). *Navigation in virtual reality: Comparison of gaze-directed and pointing motion control - IEEE Conference Publication.* [online] Ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/document/7495413 [Accessed 25 May 2019].

[3] F. El Jamiy and R. Marsh, ”Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality,” in *IET Image Processing*, vol. 13, no. 5, pp. 707-712, 18 4 2019. doi: 10.1049/iet-ipr.2018.5920

[4] Padmanaban, N., Konrad, R., Stramer, T., Cooper, E. and Wetzstein, G. (2017). Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences,* 114(9), pp.2183-2188.

[5] Ruddle, R. and Lessels, S. (2006). For Efficient Navigational Search, Humans Require Full Physical Movement, but Not a Rich Visual Scene. *Psychological Science, 17(6),* pp.460-465.

[6] Stanney, K., Mourant, R. and Kennedy, R. (2019). [online] Web.mit.edu. Available at: http://web.mit.edu/16.459/www/Stanney.pdf [Accessed 30 May 2019].

[7] Cidota, M., Clifford, R., Lukosch, S. and Billinghurst, M. (2019). *Using Visual Effects to Facilitate Depth Perception for Spatial Tasks in Virtual and Augmented Reality - IEEE Conference Publication.* [online] Ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/document/7836491 [Accessed 30 May 2019].

[8] Naceri, A. and Chellali, R. (2011). Depth Perception Within Peripersonal Space Using Head-Mounted Display. *Presence: Teleoperators and Virtual Environments,* 20(3), pp.254-272.