This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3382616

Impact of Background, Foreground, and Manipulated Object Rendering on Egocentric Depth Perception in Virtual and Augmented Indoor Environments

Matthew McQuaigue, Kalpathi Subramanian, Paula Goolkasian, Zachary Wartell

***Abstract*—This research investigated how the similarity of the rendering parameters of background and foreground objects affected egocentric depth perception in indoor virtual and augmented environments. We refer to the similarity of the rendering parameters as visual 'congruence'. Study participants manipulated the depth of a sphere to match the depth of a designated target peg. In the first experiment, the sphere and peg were both virtual, while in the second experiment, the sphere is virtual and the peg is real. In both experiments, depth perception accuracy was found to depend on the levels of realism and congruence between the sphere, pegs, and background. In Experiment 1, realistic backgrounds lead to overestimation of depth, but resulted in underestimation when the background was virtual, and when depth cues were applied to the sphere and target peg. In Experiment 2, background and target pegs were real but matched with the virtual sphere; in comparison to Experiment 1, realistically rendered targets prompted an underestimation and more accuracy with the manipulated object. These findings suggest that congruence can affect distance estimation and the underestimation effect in the AR environment resulted from increased graphical fidelity of the foreground target and background.**

***Index Terms*—Depth Perception, Augmented Reality, Virtual Reality, Rendering, Egocentric, Perceptual Matching**

I. Introduction

The ability to perceive depth is one of the most important behaviors in virtual environments, and it is currently a popular field of study. Consistently throughout past research, misestimations of distance in virtual environments have been reported [1]. To accurately represent the environment in future visual systems and to improve a range of human behaviors influenced by depth judgments, such as reaching for an object at a perceived distance, moving virtual objects to match a target (perceptual matching), or understanding the distance relationship between groups of objects, it is important to understand the source of the misestimation. These applications have become effective ways to train new employees on their day-to-day tasks. Blümel [2] States that these technologies have become important in the ecosystem of these companies.

One major area of popularity for augmented and virtual systems is that of medical imaging and medical research. Medical applications can benefit from VR/AR systems by providing more efficient and intuitive ways to understand information about their patients and the tasks they are performing in the form of 3D imaging data. Gsaxner et al. [3] developed a markerless Image-to-face registration for untethered AR in head and neck surgery. This allows the physicians wearing the AR to be untethered and see medical visualizations precisely on the patient. VR as a clinical tool to train users to perform tasks has been the subject of many studies. According to Sutherland, et al. [4] two groups of cardiologists performed the same training in learning carotid artery procedures. One group learned the traditional way and the other used VR. The group training with VR overall performed much better.

Since the popularity of research in medical imaging and systems has grown, the amount of tools and methods of viewing/learning information has grown as well. More importantly, medical systems must be extremely accurate to convey correct information for the clinical user to complete their task. Understanding depth perception in AR compared to VR environments and how different aspects of the visual field affect depth perception performance is important when designing a system to be used accurately and reliably.

Many depth judgments are egocentric, which is the ability to perceive depth between an observer and a reference object. In the virtual environment (VE), the provided depth cues influence distance misperception; the fewer the visual cues, the more egocentric distance misperception is observed. A number of recent studies have reported an underestimation of the distance between an observer and an object [1], [5], and their capability to transfer to behavior in real environments [6]. Misperceptions also seem to vary by viewer [7], [8]. Waller et al. [9] and Swan et al. [10] examined estimations with VR Head Mounted Displays (HMD) with tasks such as blind walking. They reported that distances in VEs were consistently underestimated.

A question that has sparked significant debate is whether advanced rendering and/or the use of realistic VEs play a role in lessening misperceptions. It was believed that realism in VEs was associated with distance estimation and influenced misperceptions. However, previous studies were conducted before consumer headsets became widely available. It's crucial to reassess these findings considering advancements in VR/AR technologies for a better understanding of depth perception in modern immersive environments. Murgia and Sharkey [11] found that depth estimation was underestimated in both poor and rich cue environments with greater underestimation in the poor cue environment. Thompson et al. [12] found that there was no correlation between visual quality and egocentric distance perception. Benjamin [13] also reported similar findings.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3382616

However, work by Díaz et al. [14] and Phillips et al. [15] investigated different rendering styles of volumes and environments respectively and contradicted earlier findings. Díaz et al. compared multiple shading models (No shading, Phong shading, Half-angle slicing, Direction occlusion shading) on volume data sets and found that Direction Occlusion shading performed significantly better than No Shading in user tasks, suggesting that advanced volume illumination improved depth perception. Phillips et al's work had similar results, and found that people tend to underestimate distance by a greater margin, when comparing renderings based on line drawings to its high-fidelity counterpart.

We conducted two experiments designed to explore how changes in depth cue information and graphical attributes of background, foreground, and manipulated objects in virtual and augmented reality environments affect depth perception performance. Independently manipulating these three attributes of the scene allows for testing how congruence of graphical fidelity may play a role in depth estimation performance. This research builds upon earlier work by offering a robust experimental framework that can test various experimental conditions and compare performance between virtual-to-virtual and virtual-to-real matching.

II. Related Work

A. Depth Perception in Virtual Environments

A number of recent studies have reported an underestimation of the distance between an observer and an object [1], [5], and their capability to transfer to behavior in real environments [6]. Misperceptions also seem to vary by viewer [7], [8]. Waller et al. [9] and Swan et al. [10] examined estimations with VR Head Mounted Displays (HMD) with tasks such as blind walking. They reported that distances in VEs were consistently underestimated.

Some studies looked at comparing the differences in distance estimates between the VEs and real environments. Feldstein et al., [16] performed an extensive review of earlier works (spanning 40 years) on this problem; while their review did not show a significant difference, many factors increased the discrepancy between the two environments. Plumert et al. [6] conducted experiments and found that people's time-to-walk estimates were similar across real and virtual environments up to 60 ft and underestimations were found after that distance. The lack of underestimation was due to the use of large-screen immersive displays instead of an HMD, which has a larger horizontal field of view (FOV). Visually directed tasks might also account for the similarity between real and virtual environments. In contrast, Feldstein et al's work reports an average ratio of virtual to real distance estimates of 77% using HMDs. More recently, Jamiy et al. [17] studied whether feeding real-world images or live video into an HMD would influence distance estimation. They found that distance compression with live video averaged 80.2% accuracy and the real-world images averaged 81.4%.

1) VR Rendering Quality: Virtual Reality environments in early depth perception studies were simple. This was because GPUs were not as powerful as they are now. Early studies 'high-fidelity rendering' conditions were not as photo-realistic as they would be with today's GPUs. Modern hardware can create much more complex meshed and shading models. Due to this increase in computational power, there has also been an increase in screen resolution and environment texture resolution.

a) Screen/Environment Resolution: Thompson et al [12] found no difference in distance estimations when comparing wireframe graphics versus panorama images. Iuliu et al. [18] investigated the influence of texture fidelity in VR and found that high-fidelity textures had a positive influence on precision but no significant influence on accuracy. Vienne et al. [19] found that bias in depth perception is proportional to the accommodation-vergence conflict size. The conflict had less influence on bias when multiple depth cues were available in a rich virtual CAVE environment. Buck et al. [20] results showed significant distance compression in the Vive Pro with higher resolution, similar to the original Vive. This suggests display resolution does not impact distance compression.

b) Environment Quality: Knapp et al. [21] believed that the increase in rendering environment complexity and quality of graphics would improve perceived depth in VEs using VR HMDs. However, research findings that tested Knapp's hypothesis were not conclusive. This conflicted with the work by Phillips et al. [15] who created multiple VEs, one with a photorealistic VE and one with a non-photorealistic VE. Testing participants in blind walking and verbal responses showed significantly better results in the photorealistic environment. Kunz et al. [13] conducted experiments, where two graphical quality versions of a lab room were constructed. The low-quality version had low-resolution polygonal models of virtual objects and low-spatial-frequency generic textures applied to the walls, ceiling, and floor. The high-quality rendering had high-spatial-frequency generic textures and realistic lighting. In the first experiment, a blind walking task was used. Higher-quality and low-quality rendering had similar results with higher-quality rendering performing slightly better. In the second experiment, the methodology was the same except participants used verbal reporting. This experiment resulted in significantly more accurate results with the high-quality rendering version. Gerig et al. [22] found that additional rendered depth cues (texture gradient, shadows, aerial perspective) perform slightly worse than participants without in a VR reaching task. They believe this is because the additional cues were not only unnecessary but distracting. Vaziri et al. [23] found a small but significant difference between real-world and video/NPR conditions in distance underestimation, but no significant difference between regular video and NPR video.

B. Depth perception in Augmented Reality

In AR systems, depth cues such as lighting models, shadows, and texture gradients are crucial for blending virtual items seamlessly into the real environment. However, transferring these properties to the virtual objects in real-time can be challenging, as they are generated by the environment itself and begin to reach the limitations of the hardware in terms of performance and capabilities. To measure distances in the action space, researchers often use OSTHMD and blind walking, throwing, perceptual matching, and other tasks to examine virtual objects and judge their depth in the surrounding environment. A large part of the work in depth perception has focused on egocentric depth judgments which focus on the distance estimation from an observer to a perceived distance [1], [10], [24]–[34].

Jones et al. [35] and Swan et al. [10] conducted studies and found that the egocentric depth of virtual objects is underestimated. Swan et al. found that the results were consistent with previous studies that implicated a restricted field of view and inability to scan the ground plane, but the misestimations were smaller than previously found. In a more recent study, Ping et al. [24] compared depth estimation between VR and AR and found that AR estimation is more accurate than VR but still underestimated with an error that grows with distance. Ishio and Miyao [26] revisited old research that found convergence and accommodation in the human eyes occur in response to a virtual image. They found that it is necessary to take into account both lens accommodation and convergence when using binocular see-through smart glasses. Adams et al. [36] compared depth perception and distance estimation in two AR headsets. Their results showed that video see-through AR displays may induce more underestimation than optical see-through. They also found that drop shadows and object height impact perceived depth. Krajancich et al. [37] developed a gaze-contingent stereo rendering algorithm that significantly improved the alignment of virtual objects with physical objects when compared to standard rendering using calibrated IPD.

1) AR Rendering Quality: AR environments are quite complex because the background is of the highest fidelity. Ping et al. [31] conducted multiple studies that tested combinations of rendering techniques to study how they contributed to depth perception performance. They concluded that realistic lighting models that use environment lighting and shading applied to realistic objects can help the participant relate the real-world depth cues with the virtual object's position. This was also found in a study by Diaz et al. [25] where drop shadows/cast shadows and shading from objects played a large factor in determining the depth of an AR object. In isolation and combination, these cues improved the participant's depth performance. Shading models can play an important role in a user's perception of depth in AR because the virtual object appears alongside real-world objects. Because the virtual object is on a 2D screen it may not appear in the correct position. Ping et al. [31] continued exploring the effect of opacity and three shading models on aligning virtual objects while also comparing the relationship of color and size in distance perception. Their conclusion showed that users perform better in depth matching tasks with green and yellow virtual spheres rather than blue ones. In the three shading models, users had significantly more accuracy in the depth matching with the Cook-Torrance shading model, in comparison to the Half-Lambert and Blinn-Phong models.

III. Experiments

The experiments presented in this work aim to better understand how depth perception performance is influenced by depth cues and graphical attributes applied to three aspects of a participant's visual field.

- foreground target (pegs): stationary objects whose distance the participant is trying to match.
- foreground manipulated object (sphere): an object whose position the participant controls and whose distance must be matched to the target
- background and peripheral objects: all other points in the participant's visual field excluding the target and manipulated object are termed the background.

In AR and OSTHMD displays, any of the three visual field aspects can be real or virtual because the light field of any real part comes directly from the participant's physical environment. There are 8 possible combinations of conditions, but the combination of real-real-real involves no display technology, therefore an AR perceptual matching experiment has 7 possible combinations.

To our knowledge, no prior perceptual matching experiment has systematically explored the 7 combinations we have identified using a common display system. To address this gap in knowledge, we have designed a set of experiments that will allow us to investigate how depth estimation errors change when switching between these different combinations while holding all other variables constant (FOV, display resolution, tracking latency, pixel illumination levels, pixel contrast ratio, etc.). At present, one can only attempt to answer these questions by comparing and reconciling results across different prior works, each of which used different display technology, and then trying to intuit in what way the differing display technologies are responsible for multiple prior works' contradictory results. This is not satisfactory.

By varying the 7 combinations using a common display, can one answer questions such as:

- How do depth estimate errors change between the "virtual manipulated object"-"virtual target" condition and a "virtual manipulated object"-"real target" condition?
- Do the above results change if the background is virtual or real?
- Do the above results change if the virtual background is a virtual representation of the real physical background versus a virtual background that is a uniformly illuminated white sphere (rendered as an environment map)?

We can also begin to understand how the visual congruence between the three aspects of the visual field affects depth perception accuracy. For example, how does depth accuracy compare between all aspects being virtual and some aspects are real and others are virtual? We also explore different levels of visual fidelity effects on depth perception. Visual fidelity refers to the accuracy and quality of visual representation. The virtual background/foreground targets have the lower visual fidelity and the real background/foreground targets have the highest visual fidelity.

A perceptual matching framework was designed for testing multiple depth cues and rendering methods to the background, foreground, and manipulated object across two experiments.

We chose an AR OSTHMD, the Microsoft HoloLens, for this exploration. In particular the two experiments provide a direct comparison between virtual-to-virtual and virtual-to-real perceptual matching performance.

- Virtual-to-virtual: refers to a virtual object matched to a virtual target (pegs).
- Virtual-to-real: refers to a virtual object matched to a real target.

Below is the general procedure we followed for each experiment. Specific variations for each study are described in their respective sections.

A. Experimental Hardware

The Hololens 2 was used to render the virtual stimulus (background, pegs, sphere). All participants used the same Hololens to view the 3D environment. The virtual renderings were created in Unity [38], a popular game engine. Once the environment was constructed on Unity, the program was compiled and deployed onto the Hololens 2. All participants conducted the experiments in the same room under the same lighting conditions. We have full control over the position of objects in the environment and the customization of the object's properties. The objects' properties can be manipulated in real-time for users to interact and familiarize themselves with the environment.

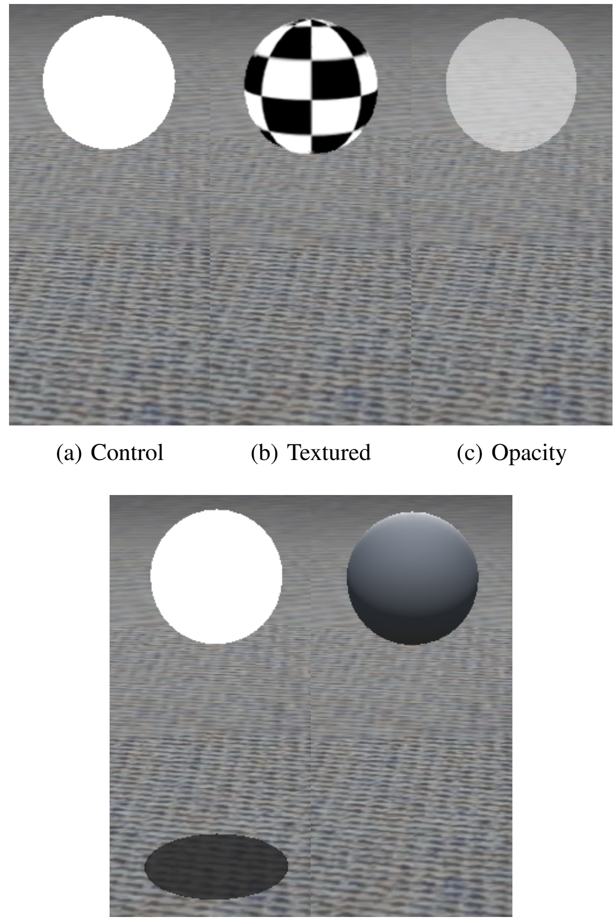
The participants used a Microsoft Xbox controller that was connected through Bluetooth to manipulate the sphere and submit a response. The left analog stick was used to move the sphere forward and backward (away from and toward the participant respectively). The A button was used to submit a trial response.

B. Perceptual Matching Task

The perceptual matching task was designed to investigate depth and rendering cues when applied to background, foreground, and objects. Participants, who were seated, were asked to use a joystick to move a virtual sphere to match the location of a pair of pegs that were positioned at one of three locations (close, middle, or far) from them. The perceptual matching task was adapted from Diaz et al's [25] work, where they used virtual spheres and varied the conditions as follows: drop shadows, texture gradient in a checkerboard pattern, shading, opacity, and a control case with no depth cues. The task allows for the isolation of head movement and other possible confounds that may affect results, such as motion sickness.

Figure 1 shows the five sphere conditions that we investigated. Three of the conditions added depth cues to the sphere while the other two (opacity and white) were used for comparison.

- *Texture gradient* is a monocular depth cue in which there is a gradual change in the appearance of objects so that the closer they are the more distinct the texture elements. The texture becomes less apparent the farther away it is.
- *Shadows* are a dark shape that appears on a surface when someone or something moves between the surface and a source of light. It is also an important monocular depth cue.
- *Object shading* is the process of altering the color of an object/surface/polygon in the 3D scene, based on things like (but not limited to) the surface's angle to lights, its distance from lights, its angle to the camera and its material properties.



(a) Control (b) Textured (c) Opacity

(d) Shadow (e) Shading

Fig. 1: Sphere Experimental Conditions.

Billboarding, in which the virtual object is rendered perpendicular and facing the camera, from [25] was not used because our participants were stationary. The sphere was 1m away from the participant in the starting position and 0.7m off the ground to be centered with the peg's height. The sphere had a diameter of 0.4m.

Each experiment had three pairs of pegs. The pegs are 0.2m in width, 0.4m in depth, and 1m in height. Each pair has one on the left and one on the right of the participant's front FOV. From the center of the FOV, the pegs are offset by 1m to the right and 1m to the left. The sets of pegs are spaced equidistant apart starting from 2.44m - 5.64m. The distances were chosen to replicate previous work from Diaz et al. [25], where these distances represent an equal interval throughout the action space.

There were two peg conditions (rich and impoverished). The rich pegs had added depth cues while the impoverished condition did not, as shown in Fig 2.

The pegs are modified versions from Swan et al. [10], located on both sides of the participant's FOV. Swan et al. pegs were adapted to avoid 2D solvable geometry, such as overlapping objects and same-size comparisons. Since the peg and sphere are completely different shapes, participants must rely on information from the isolated depth cues to judge the depth rather than the inherent attributes of the shapes.

To log participant performance data, the Universal Window Platform (UWP) was used to write data to a JSON file describing rendering and depth information of the environment along with metrics such as signed distance judgment error, time to complete, and distance from the user.

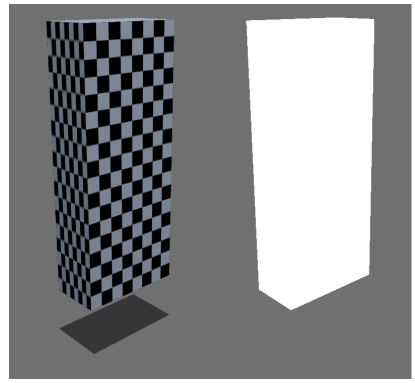
C. Experimental Procedure

Students were recruited from the University of North Carolina at Charlotte SONA system, which allows psychology undergraduate students to participate for research credit. Each participant was awarded one research credit for 45-60 minutes of their time. Participants were at least 18 years old and had stereoscopic vision. After filling out an informed consent sheet, they had to pass an anaglyph stereo test, where they had to identify a set of shapes from colored stereo images.

After receiving directions about the task and how to operate the hardware, the participants were given 10 practice trials to familiarize themselves with the task. Participants were run individually in 45-minute sessions held in a room that was 7.0104m $\times$ 9.7536m. Participants were seated in the center of the room against the front wall opposite to the screen shown in Fig 3.

On each trial, participants would see one of the randomly rendered scenes with different depth cues applied depending on the experiment. They were instructed to move the sphere to match its center with the center of a specific pair of pegs. Participants were also instructed to stay stationary with minimal head movement. A word *Near*, *Middle*, or *Far* was displayed in the center of the participant's field of vision on each trial to denote the target set of pegs. *Near* corresponds to the depth 2.44m, *middle* corresponds to 4.04m, and *far* corresponds to 5.64m. Peg sets were randomly chosen and appeared an equal number of times at each of the three distances. In the top left of the participant's FOV was a number that displayed how many trials were completed. The sphere would always start 1m from the participant but have different variable speeds to avoid the confound of memorizing movement timing. This velocity was randomized between v=3.0 m/s and v=5.0 m/s. The movement was then calculated using the following formula: movement = z * Joystick Axis * v. z is the current z position of the object and Joystick Axis is the magnitude of directional movement of the joystick in the positive z or negative z direction. Pulling backward on the joystick results in values from [-1, 0) and pushing forward results in values from (0, 1]. After 75 trials had been completed, participants were asked to take a 5-minute break to minimize eye fatigue and possible motion sickness.

This research was conducted in accordance with the ethical guidelines set forth by the Office of Research Protections and Integrity, and all participants provided informed consent. Approval for this study was granted under IRB protocol IRB-22-1025.



(a) Rich (b) Impoverished

Fig. 2: Peg Experimental Conditions.

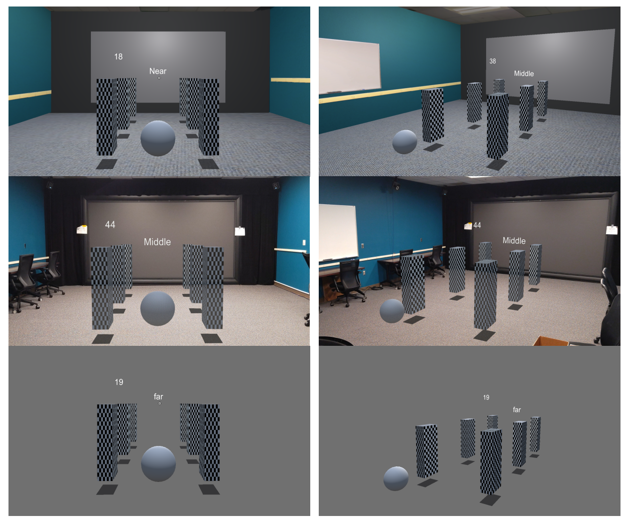


Fig. 3: Experimental setup of 3 trials showing different background conditions from the POV of the participant and from the side; *Top Row:* Experiment 1 with a shaded sphere, rich pegs, and VR background, *Middle Row:*. Experiment 1 with a shaded sphere, rich pegs, and an AR background, *Bottom Row:* Experiment 1 with a shaded sphere, rich pegs, and no (VI) background.

IV. Virtual-to-Virtual Study

A. Experiment 1

In this experiment, the virtual spheres were matched to one of the three pairs of virtual pegs presented in the foreground. The depth cues and graphical details described by Howard [39] and Diaz et al [25] were applied to the sphere and pegs (texture, shading, drop shadow, opacity) and viewed against VR, AR, and no background conditions. By splitting the scene into three different aspects for independent rendering: manipulated object, foreground frame of reference, and background, we could test multiple depth cues and rendering methods to determine if they had a singular effect or interacted across different aspects of the scene. In particular, having a virtual or real background was used to see if higher levels of information and fidelity can help with depth perception through interaction effects.

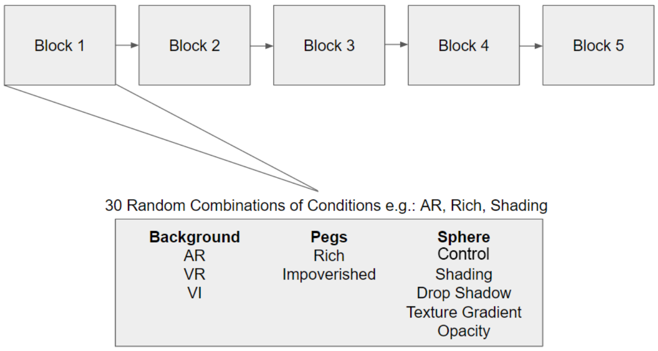


Fig. 4: Virtual-to-Virtual Experimental Design

Previous depth perception studies typically used virtual-to-real perceptual matching because, in augmented reality, users are mapping virtual objects to some position in the real world. In VR environments, there is only virtual space so all perceptual matching tasks are virtual to virtual. That being said, AR still requires making virtual-to-virtual distance judgments when not matching to a real-world object.

We focused on using a virtual sphere and virtual pegs with different depth cues and a background that can be virtual or real to test the following hypotheses.

H1: The AR background is expected to provide the most accurate perceptual matching compared to the VR and the No Background conditions because of the realistic background.

H2: Depth cues when added to the sphere are expected to improve accuracy in the perceptual task when compared to the control sphere.

H3: The rich pegs are expected to improve performance relative to the impoverished pegs.

H4: Interaction effects are expected among the background, peg, and sphere conditions.

B. Participants

Thirty-one participants (20 males, 11 females) (White: 18, Black: 6, Asian: 6, Hispanic: 1) took part in this experiment. Each participant completed 150 trials for a total of 4,650 trials in the whole experiment.

C. Experimental Conditions

During each trial, participants are presented with a randomized combination of rendering conditions for the background (3 conditions), the pegs (2 conditions), and the sphere (5 conditions). The 3 background conditions are called VR (“Virtual Reality”), VI (“Visually Impoverished”), and AR (“Augmented Reality”), They are illustrated in Fig. 3. In the VR condition, the background is a virtual rendering of the physical room with texture mapping, lighting, and shading. In the VI condition, the background is rendered with a uniform, blank gray environment map. In the AR condition, no background pixels are rendered, so the user sees the physical room through the OSTHMD display.

Each of these conditions represents a different level of graphical fidelity, having a different level of congruence with the rendering conditions of the pegs and sphere. The VR background condition emulates a VR headset on the HoloLens OSTHMD. Switching to an actual VR headset would add undesired covariates into the experiment because the VR headset would differ from the HoloLens in FOV, resolution, focal distance, weight, etc.

The physical room’s lighting level is fixed for all participants so that the VR and VI conditions’ virtual backgrounds generally occlude the details of the physical room and remain sharp and in focus.

Peg Conditions	
Rich	Impoverished
Lambertian Shading	Ambient Shading
Texture Gradient	None
Drop Shadows	None

TABLE I: Two different conditions of the virtual pegs and their depth cues.

The virtual pegs have 2 rendering conditions, “rich” and “impoverished”. Table I lists the rendering attributes of each condition. The two conditions are shown in Fig. 8.

The sphere has 5 rendering conditions labeled as follows:

- *Control*: Ambient shading only. This is the control condition with minimal depth cues.
- *Drop Shadow*: Ambient shading plus a drop shadow.
- *Opacity*: Ambient shading plus transparency using alpha blending (alpha = 0.5).
- *Shading*: Lambertian shading.
- *Texture*: Ambient shading plus a checkerboard texture image.

The 30 experimental conditions are blocked and presented 5 times for each participant. Each combination of background, pegs, and sphere condition is displayed in a randomized order as seen in Fig. 4. At the beginning of each block, the combination is re-randomized for a new order. In total, there are 150 trials for each participant.

D. Design and Data Analysis

The distance estimation errors from each participant were averaged across the 5 replications for each of the 30 experimental conditions. The data from any given trial were trimmed if the response speed exceeded 2.5 times the standard deviation from each participant’s mean and if the distance error was greater than 1 meter.

The mean trimmed distance estimation errors were analyzed with a $3 \times 2 \times 5$ repeated measures of analysis of variance (ANOVA), which tested for the main and interaction effects of the background (AR, VI, and VR), peg (Rich and Impoverished), and sphere conditions (Control, Drop Shadow, Opacity, Shading, Texture). A significance level of 0.05 was used for all statistical tests and the Greenhouse-Geisser correction was made to the p-value where appropriate to protect against possible violations of assumptions of sphericity. Bonferroni post-hoc pair-wise tests were used when the main effects were found to be significant.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3382616

E. Results

On average 1.3% of the trials were trimmed from a given participant. Two participants were removed; one for failing the stereo vision pretest, and another for not following instructions to move the sphere on every trial. For the remaining 31 participants, the mean trimmed distance estimate errors were calculated for each of the 30 experimental conditions across 150 total trials.

a) Distance Estimations: All main effects were found to be significant. The analysis of responses to different backgrounds showed considerable differences in distance estimation errors among the three AR, VR, and VI backgrounds ($F(2,60) = 96.8$, $p < 0.001$, $\eta_p^2 = 0.763$). The means were 0.079m, -0.002m, and 0.031m respectively. Follow-up Bonferroni tests showed statistically significant differences among all three background conditions p's < 0.001. Both the AR and VI conditions resulted in overestimated distance errors while the VR condition had a slight underestimation. The VR condition was the most accurate and the AR condition had the largest distance errors.

Sphere Condition	Mean	Standard Deviation
Control	0.027 m	0.020 m
Drop Shadow	0.031 m	0.017 m
Opacity	0.020 m	0.022 m
Shading	0.040 m	0.022 m
Texture	0.061 m	0.023 m

TABLE II: Breakdown of sphere conditions distance estimations errors

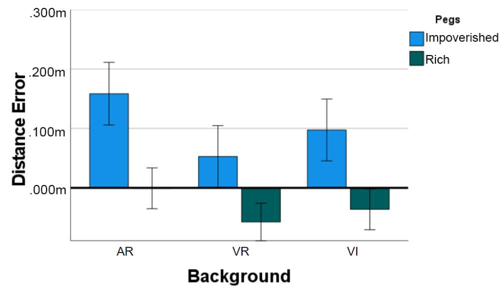
The main effect of sphere conditions was also significant showing an impact on distance estimations ($F(4, 120) = 4.9$, $p = 0.008$, $\eta_p^2 = 0.141$). Table II presents the average distance error rates for the sphere conditions. Follow-up Bonferroni tests showed that the texture gradient condition resulted in larger error rates in comparison to the control ($p = 0.001$), and opacity condition ($p < 0.001$). Surprisingly, all sphere conditions averaged an overestimation.

The two peg conditions had a strong and opposite effect on perceived depth ($F(1, 30) = 107.8$, $p < 0.001$, $\eta_p^2 = 0.782$). The rich pegs resulted in underestimated depth judgments (M = -0.31) while the impoverished pegs (M = 0.102) were associated with an overestimation of depth.

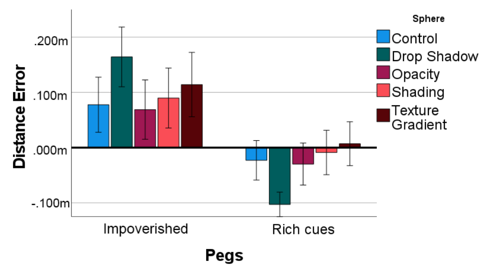
The peg conditions were also found to interact significantly with both background conditions ($F(2, 20) = 6.85$, $p = 0.002$, $\eta_p^2 = 0.186$) and sphere conditions ($F(4, 120) = 39.059$, $p < 0.001$, $\eta_p^2 = .566$). As shown in Fig 5a, the background conditions with impoverished pegs showed an overestimation while all background conditions with rich pegs showed an underestimation.

Similarly, for the impoverished pegs, all estimations for sphere conditions were overestimations; but for the rich pegs, the sphere depth judgments were much more accurate with some slight underestimations. Surprisingly, the interaction between pegs and drop shadows resulted in the largest misestimation.

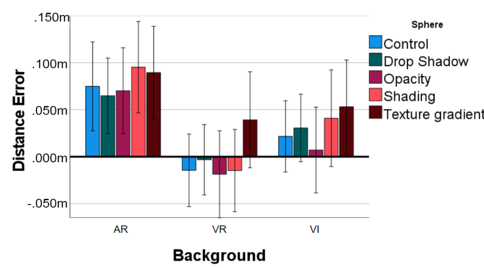
As shown in Fig 5, the third two-way interaction between background and sphere was also significant ($F(8,240) = 2.23$, $p = 0.026$, $\eta_p^2 = 0.069$). In the AR condition, all sphere conditions resulted in an overestimation of depth. The VI background condition also was associated with overestimation but to a much lesser extent than in AR. With the VR background, however, depth estimations were quite accurate across all sphere conditions except the sphere texture condition. The only interaction effect that was not significant was the three-way interaction between background, sphere, and peg ($F(8, 240) = 1.858$, $p = 0.086$, $\eta_p^2 = 0.058$)



(a) Peg and Background Distance Estimates



(b) Sphere and Pegs Distance Estimates.



(c) Background and Sphere Distance Estimates

Fig. 5: Distance estimates for two-way interactions

b) Discussion: In the AR condition, depth judgment is more accurate than VR for the rich pegs. For the impoverished pegs, AR and the VI conditions become less accurate than the VR condition. These effects can be seen in Fig 5a. This partially supports H1. A possible explanation is that accurate depth estimation requires congruence among the three independent rendering variables of the scene. That is the rendering conditions of the foreground pegs, sphere, and background must match. Past research shows that AR is usually more accurate than VR when performing virtual-to-real matching. That would mean that the pegs and background are congruent, but not the manipulated sphere (virtual-real-real). There was only congruence of the pegs and sphere (virtual-virtual-real).

Also, VR and VI conditions switched to an underestimation when more cues were applied in the rich pegs. This seems to be a pattern where when more depth cues are applied to a target, the more underestimation.

There was no support for H2 when looking at background and sphere interaction effects. As seen in Fig 5c Many of the sphere conditions performed worse or similar compared to the control condition for a given background condition. In VR, drop shadow had a positive influence and in the VI background, only opacity improved judgments. This is the opposite of what was reported by Diaz et al [25]. The main difference between our study and Diaz et al's study is that our study used virtual pegs to indicate the target distance while they used physical markers to indicate the target distance. Also, Diaz et al had the physical markers directly below the manipulated object; therefore drop shadows made distance matching very easy.

For all of the sphere conditions, the VR and VI conditions outperformed the AR condition. This again hints at some rendering congruence needed between the environment and sphere. In Fig 5b, interestingly, when the pegs switched from impoverished (having no cues) to rich (having depth cues), average estimations became an underestimation. This underestimation was also more accurate than the impoverished condition. This aligns with H3 and further supports the idea that objects rendered more realistically with more depth cues yield better distance estimations.

H4 is also supported by the significant two-way interaction effects. For the background and peg interaction, the AR background with rich pegs was the most accurate. This was a significant increase in accuracy compared to AR with no depth cues (AR Impoverished M = 0.158 m, AR Real M = -0.007 m). The VR condition switched from overestimation to underestimation when the pegs went from impoverished to rich. It seems more depth cues present in the pegs yield more underestimation. When the pegs had rich cues, almost all sphere conditions switched to an underestimation. This aligns more closely with past research. This experiment verified that more realistic pegs yield underestimation. It seems that the more realistic the virtual pegs look, the closer the results get to the results of experiments with a virtual manipulated object and physical targets (virtual-to-real). Drop shadow had the largest error in both peg conditions, which is odd because the drop shadow only needs to be aligned with the drop shadow of the peg. The VR background also prompted underestimation for sphere interactions while AR prompted overestimation. This and previous interactions suggest that more depth cues have a higher chance for underestimation when congruence occurs and overestimation when congruence does not occur. The VR condition may also have been more accurate because of the sphere and background congruence.

V. Virtual-to-Real Study

A. Experiment 2

Virtual-to-real depth perception tasks are very common in the AR research area. Augmented reality has the added difficulty of mapping virtual objects to their real-world positions. All objects are immersed in the real world through

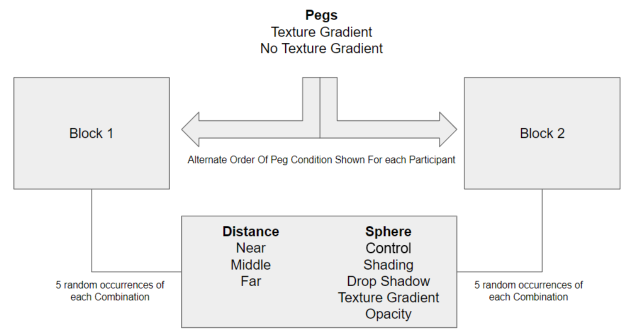


Fig. 6: Experiment Design for Virtual-to-Real Task.

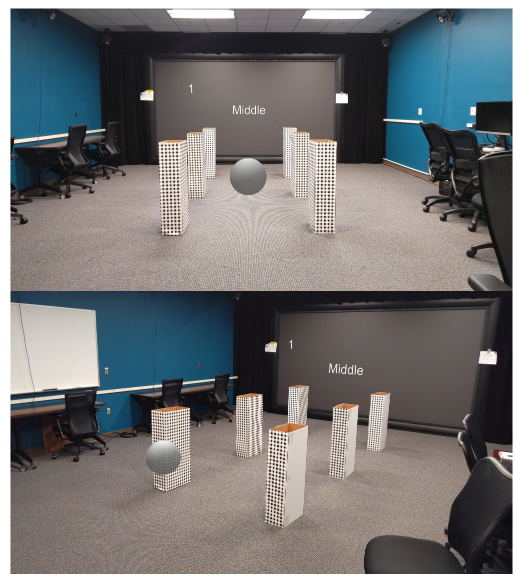


Fig. 7: Setup up for experiment 2 displaying trials from the POV of the participant and from the side; *(Top to Bottom)*

the HoloLens 2 spatial mapping to make sure the objects are in the correct position. This also means users must make comparisons of a virtual object to a real object to accurately determine its position in the real world. The virtual-to-real task design seen in Fig 7 is an expansion of the first experiment described in section IV-A. The background is always real (the AR condition in Experiment 1) along with real pegs rather than virtual pegs. All sphere conditions are the same.

From the results of our past virtual-to-virtual experiment, the following hypotheses were proposed:

H1: The real background and real pegs are expected to prompt an underestimation of sphere and distance conditions

H2: All sphere rendering conditions are expected to have a large distance error due to incongruence between peg and sphere realism

H3: Depth estimations are expected to be least accurate for the far pegs and most accurate with the near pegs

B. Participants

Twenty-nine participants (10 females and 19 males) (White: 15, Black: 3, Asian: 6, Hispanic: 5) were recruited from the

University of North Carolina at Charlotte SONA system. Each participant completed 150 total trials. That is a total of 4,350 trials across the whole experiment.

C. *Experimental Conditions*

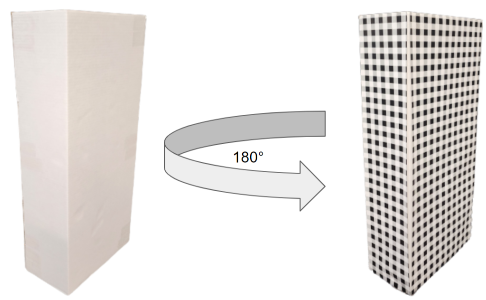
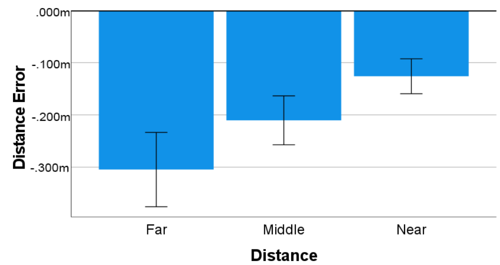


Fig. 8: Different textures are applied to opposite sides of the real pegs to easily switch conditions by rotating the pegs 180 degrees.

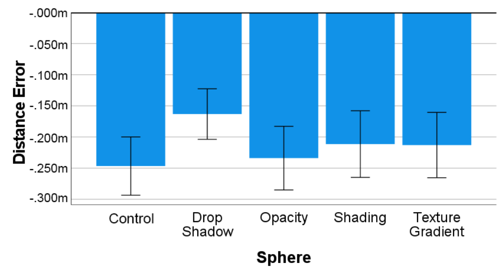
This experiment used real pegs that were the same dimensions as the virtual pegs in Experiment 1. Each of the pegs had two sides with checkerboard wrapping paper (representing a texture gradient) on one side and a blank white texture (no texture gradient) on the other (Fig. 8). This allowed us to rotate the pegs 180 degrees to switch between no texture and texture conditions. To make sure the pegs were placed in the correct positions (the same as in the virtual-to-virtual Experiment 1), virtual markers were rendered on the floor at the same locations as the virtual pegs in Experiment 1. At the halfway point in each participant's experiment, the pegs were rotated 180 degrees to show the opposite condition. The real replicas of the pegs were moved to the virtual marker positions and then the markers were hidden. The sphere had the same five rendering conditions and size as in Experiment 1 (Table II).

This experiment, unlike the first, did not use a virtual background. Experiment 2's background condition always matched Experiment 1's AR background condition, where the real room's background was visible. This was to test the performance of a virtual-to-real distance comparison when both the pegs and background were real. In Experiment 1, the virtual pegs were tested with a virtual background and a real background.

As seen in Fig 6, experiment 2 used a repeated measure design with the two peg conditions (with and without texture gradient) blocked and presented in counterbalanced order across participants. Within each block of 75 trials, the 5 sphere conditions were randomly presented at each of the three distances (near, middle, and far) 5 times. At the beginning of each block, the combination was re-randomized for a new order. In total, there were 150 trials for each participant. Experiment 2 differed from Experiment 1 by expanding the number of trials at each of the near, middle, and far distances from 3 to 5 and adding distance into the design as a factor.



(a) Mean estimations for distance main effect



(b) Mean estimations for sphere main effect

Fig. 9: Distance estimations for the main effects of the peg distance and sphere experimental conditions

After 75 trials had been completed, participants were asked to take a five-minute break to minimize eye fatigue and possible motion sickness. Then the peg condition was switched by rotating the pegs 180 degrees. After the break, the participant continued with the remaining 75 trials.

D. *Results*

On average 1.3% of the trials were trimmed from a given participant due to low response speed and distance error greater than 1 meter. For the 29 participants, the mean distance estimate errors were calculated for each of the 30 experimental conditions across 150 total trials. The mean trimmed distance estimation was analyzed with a repeated measures analysis of variance (ANOVA). A significance level of 0.05 was used for all statistical tests and the Greenhouse-Geisser correction was made to the p-value where appropriate to protect against possible violations of assumptions of sphericity. Bonferroni post hoc test were used when appropriate and when the main effects were found to be significant.

a) Depth Estimations: The distance of the pegs had a significant effect ($F(2, 56) = 36.254$, $p < 0.001$, $\eta_p^2 = 0.564$) on participant's depth judgments. As shown in Fig. 9a, there was a significant linear trend showing increased underestimation with distance ($F(1,28) = 39.69$, $p < 0.001$, $\eta_p^2 = .586$). The near condition was significantly more accurate compared to the middle and far conditions. The far condition was 0.179 m less accurate on average than the near condition. However, the peg conditions with or without the texture did not influence the degree of underestimation ($F(1, 28) = 2.081$, $p = 0.160$, $\eta_p^2 = 0.069$).

As in Experiment 1, the sphere conditions had an impact on participant performance ($F(4, 112) = 14.449$, $p < 0.001$ η_p^2

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3382616

= 0.34). Follow-up Bonferroni tests (at the $p < 0.05$) showed that the control (No Cue) sphere was less accurate than all conditions except opacity. Opacity's mean difference with drop shadows and texture gradient was significant, however. Drop shadow had the lowest mean underestimation while the control had the highest (Fig. 9b). Sphere shading and texture gradient had very similar performances.

Sphere conditions were also found to interact with peg distance ($F(8, 224) = 9.937$, $p < 0.001$, $\eta_p^2 = 0.262$). The two-way interaction appears to be due to the fact that adding cues to the sphere had a greater effect with distance. Figure 10 shows at the near distance, the effect of the sphere conditions is slight but the effect grows with distance.

None of the other interactions were significant. Peg distance was not found to interact with peg condition ($F(1, 34) = 2.569$, $p = 0.112$, $\eta_p^2 = 0.084$); or with sphere conditions ($F(4, 112) = 0.006$, $p = .547$, $\eta_p^2 = 0.019$). The three way interaction of distance × pegs × sphere ($F(8, 224) = 0.367$, $\eta_p^2 = 0.037$) was not significant.

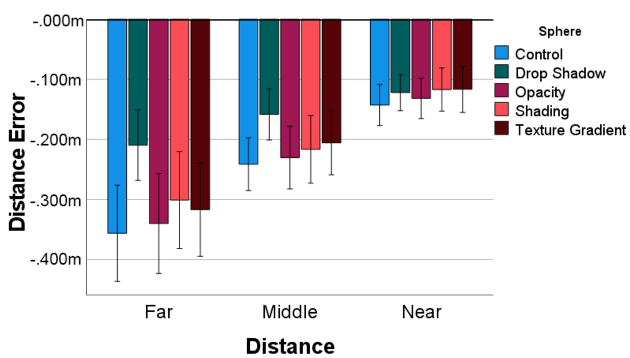


Fig. 10: Interaction effects of mean distance errors for distance and sphere

b) Discussion: All sphere conditions prompted an average underestimation error. This is also true for all three distance conditions. This supports H1 and previous research that used virtual-to-real perceptual matching tasks. Each sphere condition also had an estimation error ($M < -0.10$) for all conditions regardless of distance, supporting H2. Using real pegs rather than the virtual pegs as in Experiment 1, the foreground and the background conditions were congruent in Experiment 2, resulting in larger estimation errors than in Experiment 1. Perhaps congruence with sphere and peg is a more important factor in depth estimation than background and peg congruence. The 6 pegs provide a frame of reference for the sphere's movement and help to define the scene's foreground.

Supporting H3, the results showed that the peg at further distances has greater error. This is true for all sphere conditions and the textured peg. The increase in error based on distance was a linear trend for all conditions. This is consistent with past research that used verbal response tasks [40] where participants' distance underestimations increased linearly with distance.

In Experiment 2, the peg condition did not have an effect as it did in Experiment 1. It is possible that the surface properties of the plain white peg, such as real shading and very subtle texture gradients of the plain white paper, added some depth cues that made the plain white peg and textured peg more equivalent than intended. That being said, the checkerboard pattern did provide a very small amount of accuracy discrepancy in tasks. A new experiment is needed where the sphere is also realistically rendered to understand if the congruence of all three components will make improvements that are as accurate as the VR virtual-to-virtual condition in Experiment 1.

VI. Conclusions

In this article, we presented two experiments to gain insights into how depth perception performance is affected by changes in the depth cue information and graphical attributes of the background, foreground, and manipulated objects in virtual and augmented reality environments. We have built upon earlier work a robust experimental framework to test a multitude of experimental conditions, and directly compare performance between virtual-to-virtual and virtual-to-real matching.

Experiment 1 showed that depth estimations can switch from over- to underestimations depending on the background condition and its congruence with pegs or sphere. Specifically, participants showed greater overestimation in scenes with more realistic backgrounds with nonrealistic pegs. This result switched to underestimation when more depth cues were applied to the pegs, specifically when congruent with the VR background. Experiment 2 further supports the underestimation effect resulting from peg and background congruence in AR environments. Also, it reveals that congruence between objects in the scene is an important factor that affects distance estimation. Additionally, results supported the idea that more realistic pegs prompt underestimation and that access to depth cues differing from the sphere has a higher chance for underestimation when congruence occurs for the peg and background and overestimation when congruence does not occur.

With our use of both virtual-to-virtual and virtual-to-real perceptual matching in the same testing framework across experiments, we also found a difference in depth perception performance based on the peg condition in an AR environment. There was mostly overestimation for virtual pegs and underestimation for real pegs. The results indicate that participants' estimation errors increased with distance, and that texture gradient, drop shadow, and shading may have an effect on distance estimation accuracy.

These findings have implications for the design of VR and AR environments and the understanding of how humans perceive distances for different background/foreground contexts. Our contributions help to better understand the factors that affect depth perception in AR and VR background/peg congruence and highlights the importance of considering depth cue and graphical attribute combinations when designing virtual environments.

We are currently designing a third experiment that will include photo-realistic-to-real and photo-realistic-to-photo-realistic matching tasks. The sphere and pegs will be rendered using physically based lighting after we capture the physical room's illumination map. This creates greater congruence between the manipulated object with the real environment's foreground and background. Once the third experiment is completed, a comparison of data across all three experiments will be made, analyzing how low/high fidelity rendering and different depth cues applied to sphere, pegs, and backgrounds affect depth perception.

VII. Limitations

While this study contributes valuable insights to the field, it is essential to recognize its inherent limitations. A comprehensive understanding of the research findings necessitates a critical examination of the constraints and potential factors that may impact the generalizability and robustness of our conclusions. In this section, we elucidate the limitations of our study, providing transparency and context for the interpretation of the results.

One limitation was the use of an OSTHMD to simulate VR conditions. We chose to do this to avoid the confounds that using multiple types of headsets would introduce, as the HoloLens cannot give the same optical conditions as a VR HMD. One way to improve on this research is to use a newer headset that provides the ability to switch between a pass-through AR mode and an enclosed VR mode.

Also, all pegs were in static positions. This could possibly induce some learning effect on participants as they would place a sphere in the same position multiple times in one experiment. We tried to mitigate this by introducing variable speeds in sphere movement, and randomized order in target instructions. An easy solution for virtual targets is to use a continuous random variable to control the peg distances, but this solution is difficult for real targets, which requires changing the physical pegs distances continuously between trials.

We also did not render the peripheral geometry (desks and chairs) in the physical room in the VR condition. Though this was a slight oversight, the small FOV in the HoloLens would not have allowed the participants to see such geometry when fixated on the sphere. Also, participants remained seated and were instructed not to move their heads, which kept the pegs in their field of view during the experiment.

Each participant was tested for stereopsis using an anaglyph stereo test. While advantageous for simplicity and accessibility, anaglyph tests exhibit limitations, including color distortion and potential discomfort for some individuals.

References

[1] F. El Jamiy and R. Marsh, "A survey on depth perception in head mounted displays: Distance estimation in virtual reality, augmented reality and mixed reality," *IET Image Processing*, vol. 13, 01 2019.

[2] E. Blümel, "Global challenges and innovative technologies geared toward new markets: Prospects for virtual and augmented reality," *Procedia Computer Science*, vol. 25, pp. 4–13, 2013, 2013 International Conference on Virtual and Augmented Reality in Education. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050913012076

[3] C. Gsaxner, A. Pepe, J. Wallner, D. Schmalstieg, and J. Egger, *Markerless Image-to-Face Registration for Untethered Augmented Reality in Head and Neck Surgery*, 10 2019, pp. 236–244.

[4] J. Sutherland, J. Belec, A. Sheikh, L. Chepelev, W. Althobaity, B. Chow, D. Mitsouras, A. Christensen, F. Rybicki, and D. La Russa, "Applying modern virtual and augmented reality technologies to medical images and models," *Journal of Digital Imaging*, vol. 32, 09 2018.

[5] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert, "The perception of egocentric distances in virtual environments - a review," *ACM Comput. Surv.*, vol. 46, no. 2, Dec. 2013. [Online]. Available: https://doi.org/10.1145/2543581.2543590

[6] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker, "Distance perception in real and virtual environments," *ACM Trans. Appl. Percept.*, vol. 2, no. 3, p. 216–233, Jul. 2005. [Online]. Available: https://doi.org/10.1145/1077399.1077402

[7] A. Peer and K. Ponto, "Perceptual space warping: Preliminary exploration," in *2016 IEEE Virtual Reality (VR)*, 2016, pp. 261–262.

[8] K. Ponto, M. Gleicher, R. G. Radwin, and H. J. Shin, "Perceptual calibration for immersive display environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 691–700, 2013.

[9] D. Waller and A. Richardson, "Correcting distance estimates by interacting with immersive virtual environments: Effects of task and available sensory information," *Journal of experimental psychology. Applied*, vol. 14, pp. 61–72, 04 2008.

[10] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, "Egocentric depth judgments in optical, see-through augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 3, pp. 429–442, 2007.

[11] A. Murgia and P. Sharkey, "Estimation of distances in virtual environments using size constancy," vol. 8, pp. 67–74, 01 2009.

[12] W. Thompson, P. Willemsen, A. Gooch, S. H. Creem-Regehr, J. Loomis, and A. Beall, "Does the quality of the computer graphics matter when judging distances in visually immersive environments?" *Presence: Teleoperators and Virtual Environments*, vol. 13, pp. 560–571, 2004.

[13] B. R. Kunz, L. Wouters, D. Smith, W. 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," *Attention, Perception, and Psychophysics*, vol. 71, pp. 1284–1293, 2009.

[14] J. Díaz, T. Ropinski, I. Navazo, E. Gobbetti, and P.-P. Vázquez, "An experimental study on the effects of shading in 3d perception of volumetric models," *Vis. Comput.*, vol. 33, no. 1, p. 47–61, Jan. 2017. [Online]. Available: https://doi.org/10.1007/s00371-015-1151-6

[15] L. Phillips, B. Ries, V. Interrante, and M. Kaeding, "Distance perception in npr immersive virtual environments, revisited," 01 2009, pp. 11–14.

[16] I. T. Feldstein, F. M. Kölsch, and R. Konrad, "Egocentric distance perception: A comparative study investigating differences between real and virtual environments," *Perception*, vol. 49, no. 9, pp. 940–967, 2020.

[17] F. El Jamiy, A. Ramaseri Chandra, and R. Marsh, "Distance accuracy of real environments in virtual reality head-mounted displays," 08 2020.

[18] A. Lucaci, M. Bach, P. Jensen, and C. Madsen, "Influence of texture fidelity on spatial perception in virtual reality," 01 2022, pp. 244–251.

[19] C. Vienne, S. Masfrand, C. Bourdin, and J.-L. Vercher, "Depth perception in virtual reality systems: Effect of screen distance, environment richness and display factors," *IEEE Access*, vol. 8, pp. 29 099–29 110, 2020.

[20] L. Buck, R. Paris, and B. Bodenheimer, "Distance compression in the htc vive pro: A quick revisitation of resolution," *Frontiers in Virtual Reality*, vol. 2, 12 2021.

[21] J. Knapp and J. Loomis, *Visual Perception of Egocentric Distance in Real and Virtual Environments*, 06 2003, vol. 11, pp. 21–46.

[22] B. K. W. F. R. R. W. P. Gerig N, Mayo J, "Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements," 2018.

[23] K. Vaziri, P. Liu, S. Aseeri, and V. Interrante, "Impact of visual and experiential realism on distance perception in vr using a custom video see-through system," in *Proceedings of the ACM Symposium on Applied Perception*, ser. SAP '17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3119881.3119892

[24] J. Ping, Y. Liu, and D. Weng, "Comparison in depth perception between virtual reality and augmented reality systems," 03 2019, pp. 1124–1125.

[25] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir, "Designing for depth perceptions in augmented reality," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2017, pp. 111–122.

[26] H. Ishio and M. Miyao, "Importance of visual distance adjustment for ar application of binocular see-through smart glasses," in *2019 14th International Conference on Computer Science Education (ICCSE)*, 2019, pp. 1086–1088.

[27] C. Jerome and B. Witmer, "The perception and estimation of egocentric distance in real and augmented reality environments," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, 09 2005.

[28] J. II, L. Kuparinen, S. Rapson, and C. Sandor, "Visually perceived distance judgments: Tablet-based augmented reality versus the real world," *International Journal of Human-Computer Interaction*, vol. 33, 12 2016.

[29] C. S. Rosales, G. Pointon, H. Adams, J. Stefanucci, S. Creem-Regehr, W. B. Thompson, and B. Bodenheimer, "Distance judgments to on- and off-ground objects in augmented reality," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 237–243.

[30] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger, "Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr?" in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 585–592.

[31] J. Ping, B. H. Thomas, J. Baumeister, J. Guo, D. Weng, and Y. Liu, "Effects of shading model and opacity on depth perception in optical see-through augmented reality," *Journal of the Society for Information Display*, vol. 28, no. 11, pp. 892–904, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jsid.947

[32] J. E. Swan, G. Singh, and S. R. Ellis, "Matching and reaching depth judgments with real and augmented reality targets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1289–1298, 2015.

[33] C. J. Lin and B. H. Woldegiorgis, "Egocentric distance perception and performance of direct pointing in stereoscopic displays," *Applied Ergonomics*, vol. 64, pp. 66–74, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687017301187

[34] S. E. Kirkley and M. A. Siegel, "Augmented reality performance assessment battery (arpab): object recognition, distance estimation and size estimation using optical see-through head-worn displays," 2003.

[35] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, ser. APGV '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 9–14. [Online]. Available: https://doi.org/10.1145/1394281.1394283

[36] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer, "Depth perception in augmented reality: The effects of display, shadow, and position," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022, pp. 792–801.

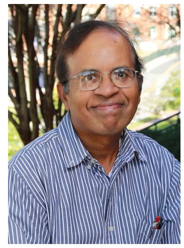
[37] B. Krajancich, P. Kellnhofer, and G. Wetzstein, "Optimizing depth perception in virtual and augmented reality through gaze-contingent stereo rendering," *ACM Trans. Graph.*, vol. 39, no. 6, nov 2020. [Online]. Available: https://doi.org/10.1145/3414685.3417820

[38] Unity Technologies, "Unity." [Online]. Available: https://unity.com/

[39] I. Howard, *Perceiving in depth. Volume 1 Basic Mechanisms*, 03 2012.

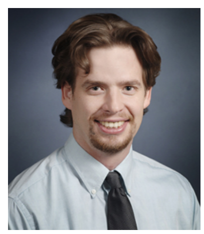
[40] A. Beall, J. Loomis, J. Philbeck, and T. Fikes, "Absolute motion parallax weakly determines visual scale in real and virtual environments," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 2411, 08 2002.

Matthew McQuaigue is a doctoral candidate at the University of North Carolina, Charlotte. His research interests include interactive AR/VR depth perception/rendering, computer graphics, interactive learning technologies, and scientific visualization. He is particularly interested in depth perception in virtual and augmented environments, exploring the impact of various rendering methods on accuracy.



Kalpathi Subramanian is an Associate Professor of Computer Science at the University of North Carolina at Charlotte. He received a BE (Honors) in Electronics and Communication engineering from the University of Madras in 1983, followed by an MS (1987) and PhD (1990) in Computer Science at the University of Texas at Austin. Between 1991 and 1993, he was a post-doctoral member of technical staff at AT&T Bell Laboratories. His research interests span the areas of interactive computer graphics, scientific and information visualization, spatial data structures, and computer science education.

Paula Goolkasian is a Professor Emeritus at UNC Charlotte. She retired from her position as Professor of Psychology and Director of Cognitive Science in 2016, where she taught graduate and undergraduate courses in Research Methods, Cognition, Sensation & Perception, and Research Design & Quantitative Analysis for 42 years. Her research interests are in the area of perception and attention with over 100 research articles and presentations at professional conferences. Currently, she is an assistant editor for *Frontiers in Psychology: Cognition* and in January 2021 retired as the Executive Editor for *The Journal of General Psychology*.



Zachary Wartell is an associate professor of the Computer Science department and co-Director of The Ribarsky Center for Visual Analytics at UNC Charlotte. His research interests are 3D user interfaces, stereoscopic and virtual reality displays, interactive 3D computer graphics and visualization.