Invisible Mesh: Effects of X-Ray Vision Metaphors on Depth Perception in Optical-See-Through Augmented Reality

Haoyu Tan* Tongyu Nie† Evan Suma Rosenberg‡

University of Minnesota

(a) Control (b) Visible Mesh (c) Invisible Mesh (d) Tramlines (e) Virtual Window

Figure 1: Representative images of the X-ray vision metaphors evaluated in the experiment. (a) Control: a standard overlay that always displays the virtual target at the correct depth, regardless of whether it would be occluded by the physical environment. (b) Visible Mesh: a grid mesh layer that partially occludes the virtual target if it is behind the wall. (c) Invisible Mesh: an invisible grid mesh structure that becomes visible only when the virtual target is behind the wall. (d) Tramlines: a pair of linear perspective lines that are attached to the ground. (e) Virtual Window: a virtual 3D model that simulates a real window. Note that the spherical target object is floating in the mid-air and placed at different distances in each image. In (a), (c), and (d), the target is positioned behind the wall, while in (b) and (e), the target is placed in front of the wall.

Abstract

This paper investigates the influence of X-ray vision metaphors on distance estimation in optical-see-through augmented reality (AR) in action space. A within-subjects study (N = 30) was conducted to evaluate depth judgments across five conditions, including a novel "invisible mesh" technique. Participants performed a series of blind walking tasks that required estimating the depth of AR objects displayed at multiple distance ranges in front or behind a physical occluding surface. Although quantitative results regarding the impact of different X-ray vision metaphors on distance perception were inconclusive, participant feedback revealed a diversity of strategies and preferences. Overall, the findings suggest that no single metaphor was considered universally superior, and multiple X-ray vision metaphors may be suitable for different users and situations. This research contributes to understanding of X-ray vision techniques and informs the design considerations for AR systems aiming to enhance depth perception and user experience.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Computing methodologies—Computer graphics—Graphics systems and interfaces—Mixed / augmented reality;

1 Introduction

Augmented reality (AR) technologies can enrich our perception of the real world by overlaying it with computer-generated content. This fusion of digital information with the user's environment happens in real-time, permitting simultaneous interaction with both the real world and the synthesized data. An AR head-mounted display can be classified as either Optical See-Through (OST) or Video See-Through (VST). VST displays utilize cameras to capture and superimpose computer-generated content onto images of the real world. Subsequently, the synthesized image is viewed by the user. Conversely, OST displays employ semi-transparent mirrors to project computer-generated content directly into the user's eyes, allowing for a direct view of the real environment [63]. VST displays have the advantage of handling the real world as images, which enables extensive manipulation through image processing techniques. However, OST displays allow users to directly observe the real world without the resolution downsampling and artificial latency from cameras [62, 63]. Although both technologies are potentially useful, this paper specifically focuses on OST displays.

One of the fascinating features provided by AR is "X-ray vision", which refers to the ability to see hidden information through physical barriers - a trait often associated with superheroes [70]. However, a naive approach of merely overlaying AR content onto the real world may create incorrect occlusion cues. Ideally, an object closer to the observer should hide the one behind it, but without proper occlusion handling, virtual content may seem to float above the real surface (see Figure 1(a)). This creates an illusion of the virtual content being closer than the actual surface, resulting in ambiguity in the perceived distance from the observer to the virtual content [67].

Various studies have explored visual solutions to address occlusion issues in AR [4, 6, 7, 23, 29, 46, 58, 72]. Nevertheless, only a few papers have conducted an extensive user study to compare different solutions in the same experimental setting, specifically focusing on depth perception with OST AR. In particular, Martin-Gomez et al. evaluated multiple X-ray vision techniques and compared users' performance on OST and VST displays within personal space in a medical setting [52], and Gruenefeld et al. performed an evaluation using blind walking with a Microsoft HoloLens v1 [29]. Usually, participants are required to walk towards the target without any feedback on their performance (e.g., [3, 14, 28, 43, 67, 68]). However, in the latter study, participants were provided with feedback that allowed them to make adjustments, and the distance between the wall and the participant was explicitly provided. While the HoloLens v1

*e-mail: tan00213@umn.edu
†e-mail:nie00035@umn.edu
‡e-mail:suma@umn.edu

used in their study was a state-of-the-art OST device at that time, it has a lower field-of-view (FOV) compared to current OST displays, potentially impacting depth perception [11, 41, 53, 71]. Therefore, there is a need for more systematic experiments that can compare AR occlusion techniques and improve scientific understanding of their effects on depth perception in action space.

In this paper, we introduce a novel AR occlusion visualization technique called "invisible mesh." Additionally, we describe a user study that evaluated the effects of four X-ray vision metaphors, along with a control condition, on distance judgments in OST AR using a HoloLens v2. The visual metaphors in this work incorporate or were inspired by depth cues that provide depth information; visualizations that explicitly label the depth of objects were not considered. Depth perception, a crucial element of interaction and decision-making in 3D environments, was evaluated using a blind walking task without feedback. The study was conducted in a controlled lab setting with a single layer of a physical barrier and one virtual object at a time. The evaluation examines the accuracy of egocentric distance judgment in action space, the range of medium-field distances in which users are most likely to interact with occluding objects.

2 Related Work

Depth perception, which is critical for humans in navigating and interacting with their environment, has traditionally been considered as an inference of the spatial layout from a set of depth cues. This inference is necessary because the 3D information of the environment is projected onto the 2D retinal images in our eyes, causing a loss of depth information [15, 27]. Many mixed reality experts have adopted this viewpoint [11, 33–35, 37, 67, 68]. Cutting and Vishton divided the environment into three regions, including personal space, which is within and slightly beyond arm's reach; action space, which is the space where individuals publicly interact; and vista space, which encompasses anything beyond the action space [15]. Gibson suggested that depth perception is closely tied to the perception of continuous ground texture [27]. Building on this, Adams et al. investigated the impact of shadows and ground contact on depth perception [1, 2]. Research over several decades has demonstrated that absolute egocentric distances tend to be underestimated in virtual environments [1, 47, 67]. Although various factors, including past experiences, field-of-view (FOV), and weight of the head-mounted display (HMD), are known to influence depth perception, the underlying cause of distance underestimation in virtual and augmented reality remains an open research question [13].

2.1 Egocentric Distance Estimation Protocols

Because egocentric distance perception cannot be directly observed, various methods have been developed to infer perceived distance [12, 67]. Action-based protocols require the observer to perform a physical action to indicate distance. They can be further subcategorized as open-loop methods (e.g., forced choice and visually directed actions), where there is no feedback on distance estimation accuracy during the task, and closed-loop methods (e.g., perceptual matching) that do provide feedback [44, 67].

Verbal Report Verbal report is an open-loop method in which observers directly communicate an estimated distance in units of measurement to the experimenter. This method is often employed due to its simplicity [1, 12, 29, 54, 67]. However, cognitive factors like prior knowledge and anchoring effects can introduce biases, making this method less reliant on pure perception [49, 57].

Forced Choice In the forced choice protocol, observers must choose from a set of options that make statements about depth judgments, such as whether one object is closer, equidistant, further away, or within specific distance ranges [17, 20, 24, 45, 64]. This open-loop method establishes an ordinal depth relationship among objects [44]. However, it is not suitable for our study, which involves only one layer of occlusion, as participants may too easily deduce the depth relationship between the physical obstruction and a virtual object after several trials.

Perceptual Matching In perceptual matching, observers adjust the position of an object until they perceive it to be at the same location as the reference object [16, 33, 44, 64, 67]. This method is considered closed-loop because the observer can receive feedback on the relative size of the object being adjusted [44].

Walking Protocols Blind walking is the action-based distance judgment protocol most widely employed in virtual and augmented reality research [3, 14, 28, 29, 37, 43, 54, 59, 67, 68]. In this method, the observer initially views a target object and subsequently walks without vision towards it, stopping when they believe they are at the target's location. Blind walking has been shown to be reliable for estimating perceived distances of up to 20 meters [48, 49], and we therefore employed this distance estimation protocol in our study.

Triangulation by walking is a similar technique that requires the observer to walk along a path at an oblique angle to the target. After stopping, the observer turns to face or walk towards the target, and the distance and angle are then utilized to calculate the perceived distance. Variations of visually directed walking exist that do not strictly adhere to these constraints. For example, in imagined walking tasks, the observer imagines walking to the target, and a timer stops when they believe that they would have reached the target. Intriguingly, this method can yield accurate distance judgments [44] and has been successfully utilized in perception research [60].

2.2 AR X-Ray Visualization Metaphors

In this subsection, we describe visualization metaphors that have been proposed to address occlusion issues for AR X-ray vision.

Alpha Blending Alpha blending involves adjusting the alpha channel of an occluding structure, thereby presenting occluded objects as they would appear behind a semi-transparent surface. This method is intuitive and has been widely used, especially with VST displays [8, 23, 40, 44, 45, 50, 51]. Purely image-based techniques may also employ visual saliency factors such as environment illumination [65], color [65, 72], material properties [72] and motion [56, 65] to determine the blending equations. Conversely, model-based techniques use geometric features to pinpoint essential regions of the occluder [39]. Compared to uniform blending, these variants retain more information from the occluding structure through context analysis. However, alpha blending has limitations; transparency diminishes the occlusion cue, and overlapping transparent surfaces can cause depth ambiguity [24, 44]. Moreover, as it necessitates direct manipulation of video images, it is impractical to implement on OST displays which cannot modify the transparency of the physical environment [44].

Virtual Window Also known as virtual hole or cutaway, the virtual window metaphor generates a synthesized opening in the occluding structure to reveal the occluded object. The human visual system often perceives a continuous surface as a singular entity. Directly overlaying an occluded object disrupts this continuity, making it appear to be in front even if it is actually behind the surface [19, 44]. The virtual window resolves this by breaking the real surface and conveying depth information through occlusion and motion parallax [24, 44]. This virtual window metaphor initially found widespread use in medical AR [5, 8, 22, 66] and has also been explored in other contexts. Feiner and Seligmann explored a set of algorithms for rendering cutaway effects [21]. Furmanski et al. found that rendering an occluded object in a cutaway box enhanced depth perception using forced choice [24]. Phillips et al. compared four different window viewing conditions using a triangulation by walking task [59]. Guo et al. built a dynamic virtual window that can track the user's position, movement, and eye gaze in real-time and compared it with a traditional minimap in a target tagging game.

They found both virtual window and minimap were effective, and the virtual window was more user-friendly for individuals without gaming experience [30]. Liao et al. found cutaway leads to poor performance when the user needs to see both the occluder and the occluded object [42]. In our study, we implemented a virtual window frame on the surface of the occluder (see Figure 1(e)) to further investigate its effectiveness relative to other techniques.

Tramlines Tramlines use linear perspective, an important depth cue, to assist users in inferring an object's distance [15, 27]. This metaphor renders two parallel lines on the ground leading to the object [44]. Bane and Höllerer introduced tramlines in their interactive toolkit for AR X-ray vision [6], and Livingston et al. found that tramlines are more effective in outdoor environments compared to indoor ones [47]. Unlike other X-ray vision metaphors which mainly offer ordinal depth cues, tramlines can provide metric depth cues due to their continuous use of linear perspective. We were therefore interested in investigating whether this continuous information can improve depth judgments compared to other metaphors. Therefore, we implemented a tramline condition in our study, drawing a pair of parallel lines on the ground originating from the same point and extending 11 meters into the distance (see Figure 1(d)).

Virtual Mask Previous X-ray vision study explored techniques by adding a layer of mask on the surface of the occluder. Otsuki et al. utilized a layer of random black dots to create a pseudo-transparency effect, allowing the front surface to be perceived as transparent [55]. This technique relies on using the random dot mask to help the observer reconcile the incongruity between binocular disparity and occlusion cues, convincing them that the virtual target is behind the surface. Ghasemi et al. further investigated the effect of relative dot size and density on the impression of surface transparency [26]. Martin-Gomez et al. implemented the virtual mask on both VST and OST displays along with several X-ray vision techniques [52]. The results showed that the virtual mask technique exhibited the lowest errors compared to other techniques on both displays. Since techniques compared in their study were initially designed for VST display, they did a second study and proposed solutions to adapt those technique to OST display. However, virtual mask did not benefit from their solutions.

Visible Mesh Edges are frequently employed as low-level features to represent an object's basic structure. Webster et al. developed a system that renders the occluded structure in wireframe style [69]. Other researchers have employed edge detection on the occluding structure to enhance ordinal depth perception between occluding and occluded object [4, 38]. This partial-occluding effect not only conveyed the correct ordinal depth relationship, but also preserved information about the occluder's structure. Livingston et al. proposed a simplified metaphor based on edge overlay, called virtual wall, and compared it with five other X-ray vision techniques [46]. The virtual wall added a layer of synthesized edges over the occluded object, and results showed this technique reduced depth estimation error. In our study, we adapted concepts from this technique in the "visible mesh" condition by applying a grid mesh layer with a fixed density over the surface of the occluder (see Figure 1(b)).

3 Invisible Mesh Metaphor

To address the challenges of visualizing multiple overlapping surfaces, Interrante et al. introduced the concept of using sparse opaque texture to enhance relative depth perception and convey shape of a transparent surfaces [35]. Heinrich et al., drawing inspiration from this work, implemented the texturing technique by rendering a wireframe layer on top of the entire occluding surface [32]. They compared this approach with other X-ray vision techniques, including the virtual mask technique mentioned earlier, on a patient's body in projective AR. Results suggested that the errors in distance deviation for the virtual mask and texturing techniques were slightly smaller than those for other techniques.

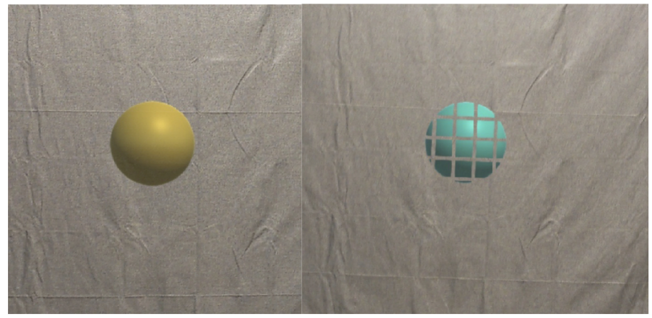

Figure 2: A comparison of the invisible mesh X-ray vision metaphor when the virtual target is displayed in front of the wall (left) and behind the wall (right). Note that the targets have different colors because it was selected randomly for each trial.

Figure 3: Photos of the physical environment setup used in the experiment. (left) A participant sitting at the starting point, wearing the HoloLens 2, and observing the virtual target object. (right) Virtual objects were displayed in front or behind a curtain, which was moved to the side by a motorized rail during the blind walking task.

As discussed earlier, displaying too much information at once can increase cognitive load and overwhelm observers. Edges preserve basic structure, but rendering the entire edge structure of an occluding object could be distracting. To address this issue, we propose the "invisible mesh" technique (see Figure 2), which was inspired by Interrante et al. [35]. We adapted the grid mesh model from the visible mesh metaphor described in section 2.2; however, the grid can only be perceived in the area where it overlaps with the target object and becomes completely invisible if there is no target object positioned behind the occluding structure.

The invisible mesh technique aims to create the perception that the occluding structure is a grid that obstructs visibility of the virtual object in AR. To achieve this effect, we first adjusted the stencil buffer to determine which area of the mesh does not need to be drawn. We set the virtual target to be rendered prior to the occluding structure and enabled the stencil test for both objects. In the occluding structure's shader, the stencil test only passed if it had the same reference value as that of the target and we kept the value if it passed. The logic of the stencil test may vary depending on the number of virtual objects and the number of occluding layers. In the next step, we colored the grid mesh black. Since black appears transparent and does not add any additional light in an OST display, the visible portion of the invisible mesh appears non-existent. Therefore, the mesh is not visible itself, but its presence is perceived as it occludes the object behind it.

4 EXPERIMENT

4.1 Study Design

In this experiment, participants performed a series of distance estimation tasks using blind walking. The study used a 5x3 within-subjects design with two independent variables:

- **X-ray vision metaphor:** Each participant experienced five conditions, as shown in Figure 1, including: tramlines (TL), virtual window (VW), visible mesh (VM), invisible mesh (IM), and a control condition (C) in which no additional information was displayed.
- **Distance range:** The target virtual object was positioned at three distance ranges: near ($3.0 \pm 0.5m$), medium ($5.5 \pm 0.5m$), and far ($7.2 \pm 0.5m$).

We carefully chose these distance ranges to ensure a balanced placement of the virtual object in relation to the room, wall, and participant. All of the ranges were utilized in each experiment, and the target virtual object appeared randomly within one of the target ranges for each trial. Each combination of factors was repeated twice, resulting in a total of 30 experimental trials. Tasks with the same X-ray vision metaphor were grouped into blocks to avoid context switching between metaphors. We chose not to use discrete distances because we were concerned that participants might figure out that target distances were repeated after completing a few tasks, given the large number of trials in the experiment. Several previous works also employed a continuous distance range with random target distances [36, 61]. We used a balanced Latin Square to counterbalance the order of the metaphor blocks [9]. This design not only controls the order of metaphor occurs, but also guarantees that each metaphor precedes and follows every other metaphor an equal number of times. In our study, the Latin Square resulted in 10 unique orders, each of which was experienced by three participants. The study protocol was approved by our university's Institutional Review Board (IRB).

4.2 Participants

We recruited 30 participants (17 men and 13 women) from university students and the general population. The participants were from various age groups, but were all under 44 years old (20 participants aged 18-24, 8 participants aged 25-34, and 2 participants aged 35-44). All of them reported normal or corrected-to-normal vision. 27 participants had prior experience with with video games, and among them, 13 had previously used an AR or VR headset.

4.3 Apparatus and Environment

We used a Microsoft HoloLens 2 for the study, which has a 52° diagonal field-of-view with 2048×1080 pixels of resolution per eye and weighs $566g$. We developed the AR application using Unity 2020.3 and the Microsoft Mixed Reality Toolkit 2.8. We connected the HoloLens 2 to a wireless keyboard and a laptop to control the HMD and collect some of the post-experiment questionnaire responses.

Figure 3 provides an overview of the experimental setup. The study took place in a controlled lab space with a walkable length of 8.7m. The room was illuminated by several ceiling lights that remained stable and consistent throughout the study. The brightness of the HoloLens 2 also remained consistent throughout the study. We used a gray blackout fabric curtain as the occluding structure controlled remotely using a motorized curtain rail. The curtain rail's noise while operating was less than 20 dB, which we believe neither interfered with the study nor provided depth cues to the participants. The curtain opened in a single direction and formed a flat surface when fully expanded. A previous study conducted by Gains and Kuhl also employed a movable obstacle in their direct blind walking task [25]. We marked the starting point with duct tape at one end of the room and placed a stool at that point. The curtain system was placed 4.2 meters away from the starting location.

To refine our study protocol, we performed informal testing with six volunteers; three provided suggestions on the properties of the target and experimental setting, and four provided feedback on the entire study, with one volunteer contributing to both aspects, to assist in finalizing our study design. Based on feedback from the initial testing, we opted for a sphere as the virtual target object. The volunteers suggested that identifying the center of a sphere was more intuitive compared to other shapes. We rendered all X-ray visualizations in light gray, except for the invisible mesh. This color was selected because it integrated unobtrusively with the fabric curtain. To avoid using relative size as a depth cue, which can be learned through experience, we randomized the size of the target object in each trial. Comments during pilot testing indicated that a singular size range made the target appear noticeably smaller at further distances. Consequently, we established distinct size ranges for each distance: for near distances, the radius was between 0.25 and $0.4m$; for middle distances, between 0.35 and $0.5m$; and for far distances, between 0.45 and $0.6m$. This ensured that targets appeared roughly consistent in size across all distances. Additionally, the color of the target was randomized for each trial.

4.4 Task and Procedure

At the beginning of the study, the experimenter instructed the participant to sit on the stool and explained the task and procedure. Participants were required to remain seated throughout the experiment, except during the blind walking phase. However, they were allowed to make exploratory movements while seated, such as leaning side-to-side to adjust their viewing angle to the target.

The experimenter initiated the AR application on a laptop using the HoloLens Device Portal. Once the application was running, the participant performed a measurement task to determine the ground height in the device's coordinate system. The participant initially scanned the surroundings to ensure the spatial awareness mesh encompassed the entire area. Following this, the participant gazed at the floor while keeping their head relatively upright. The AR application projected a ray from the participant's eyes to the ground. Once the participant confirmed that the ray did not intersect with any obstacles, the experimenter recorded the point where the ray met the ground, which indicated the ground height. We then positioned a cube 2 meters in front of the participant. If the cube appeared to rest on the floor without floating or sinking, the ground height was measured accurately. Otherwise, this procedure could be repeated.

Previous research has indicated that participants may be initially hesitant to walk without vision in unfamiliar surroundings [37]. Therefore, we included three practice trials before the experimental trials. These practice tasks followed the same blind walking procedure, except that no X-ray vision metaphors were presented.

In each experimental trial, participants observed a virtual object displayed at eye level along with an X-ray vision metaphor. They were instructed to estimate the distance to the center of the object without any explicit time constraints. When ready, the participant informed the experimenter and closed their eyes. After three seconds, the AR content vanished and the motorized rail began to move the curtain. The participant then remained seated for 22 seconds until the pathway was clear and a voice prompt instructed them to commence blind walking. As the participant began walking with their eyes closed, the experimenter closely monitored them prevent accidental collisions. If the participant opened their eyes before completing the walking task, a warning message would be displayed. When the participant believed they had reached the location of the target, they stopped walking and notified the experimenter. The position data was then recorded by the AR application.

After completing the blind walking task, the participant was instructed to keep their eyes closed, and the experimenter guided them back to the starting position, following a zig-zag path to avoid providing feedback about the distance traveled. During the return walk, the motorized curtain began moving the curtain back to its original configuration. Upon reaching the starting position, the experimenter tapped the stool to audibly signal its location to the participant. The participant was instructed to open their eyes once they were seated and the curtain was fully closed.

This procedure was consistent across all trials. The participant completed questionnaires after each 6-trial block for each X-ray vision metaphor and at the conclusion of the experiment, as detailed in section 4.5. The experiment lasted approximately 75 minutes, with the participant wearing the AR headset for about 65 minutes.

4.5 Measures

Depth Judgments For each trial, we subtracted the participant's final position from their starting point using the HMD's tracking data as the judged distance. First, we calculated the **signed error** of perceived depth using equation 1.

$$Signed\,Error = Judged\,Distance - Target\,Distance \tag{1}$$

We also examined the **relative error**, calculated by equation 2. This approach allowed us to gain insight into the depth judgment tendencies, regardless of changes in distances.

$$Signed\,Relative\,Error = \frac{(Judged\,Distance - Target\,Distance)}{Target\,Distance} \tag{2}$$

Observation Time The duration of observation was determined by the point at which the participant felt confident about their depth estimation. This observation time, defined as the interval between the onset of the trial and eye closure, was systematically recorded. The observation time might be considered an objective indicator of the system's usability and intuitiveness. If an X-ray vision metaphor is intuitive and easy to understand, then participants would likely need to spend less time observing the scene to make depth judgments.

Questionnaire Responses After each block, the participant completed the System Usability Scale questionnaire (**SUS**) [10], the NASA Task Load Index questionnaire **(TLX)** [31], and rated their **confidence** on a scale of 1=*not confident at all* to 7=*extremely confident*. After the experiment, the participant filled out a demographic questionnaire and a feedback questionnaire. The feedback questionnaire contained the four open-ended questions listed below, presented along with screenshots of each X-ray vision metaphor to provide visual aids for answering the first two questions.

- Were there any features you noticed that helped you complete the task?
- Were there any features you noticed that hindered you from completing the task?
- Were there any strategies that helped you complete the task?
- Do you have any suggestions for improvement or anything else you'd like to tell us?

4.6 Hypotheses

In this study, we investigated the following research questions regarding the effects of X-ray vision metaphors.

How do X-ray vision metaphors influence depth judgments compared to the control condition?
We hypothesized that all four X-ray vision metaphors would help participants make more accurate depth judgments compared to the C (control) condition, measured using either signed error or relative error. Based on prior research, we would also expect depth judgments to become less accurate as distance increases, and we were interested in exploring the potential interaction effects between X-ray vision metaphors and object distance ranges [18, 36].

Do the mesh-based techniques support better depth judgments than the other two X-ray vision metaphors?
Various sources of information about layout metrically reinforce and contrast with each other, forming a powerful network of constraints [15]. Although each metaphor provides multiple depth cues, they may vary significantly in presentation and magnitude. In the VM (visible mesh) and VW (virtual window) conditions, as well as the IM (invisible mesh) condition when the object is positioned behind the wall, participants can leverage occlusion, which is considered as the most powerful depth cues regardless of distance. Nevertheless, in the VW condition, participants usually need to make exploratory movements and adjust their perspective to observe whether the target is occluded by the virtual window. In the VM and IM conditions, the grid mesh directly overlays the target, which could produce greater motion parallax compared to VW and TL (tramlines). TL provides additional linear perspective cues, which involve the combination of multiple sources of information such as texture gradient. Given the substantial differences in depth cues, we were particularly interested in comparing distance judgments in the mesh-based conditions to the other two metaphors.

Are depth judgments underestimated in VM compared to IM?
The VM provides a virtual frame of reference, while the IM is more naturally integrated with the physical obstruction. Because distance judgments tend to be generally underestimated in AR/VR environments, the presence of an underestimated virtual mesh could further compress the perceived distance of the target. We therefore hypothesized that the perceived depth of targets may be shorter in the VM condition compared to the IM condition.

How do X-ray vision metaphors influence the observation time needed for depth judgments?
The VM, IM, and VW metaphors partially exploit motion parallax to different extents, which are generated by head movement, while TL provides additional linear perspective cues that provide continuous depth information without motion. We speculated that when using these metaphors, participants would need more time to observe before making depth judgments compared to the TL metaphor. Conversely, since the C condition itself does not offer any additional information beyond the real environment, we hypothesized that observation time would be the longest.

How do X-ray vision metaphors compare in terms of usability and subjective workload?
We expected that the C condition would receive the least favorable ratings among the five conditions in terms of usability and workload because it did not provide any visual augmentation to assist participants in resolving conflicting depth cues between the virtual target and the physical occluding surface. Additionally, we hypothesized that participants would prefer the seamlessly integrated invisible mesh over the more obtrusive visible grid mesh.

5 Results

Shapiro-Wilk tests of normality were conducted for all variables. When significant violations were found, non-parametric analyses were used, and descriptive statistics are reported using median and interquartile range. Otherwise, the data were analyzed using a repeated-measures analysis of variance (ANOVA). All statistical analyses used a significance value of $\alpha = .05$, except for post-hoc tests that were adjusted for multiple comparisons.

During preliminary analysis of signed distance errors, we identified one outlier which deviated more than two standard deviations ($-2.10m$) from the mean. Further examinations of this participant's data revealed that the walking distances for approximately half of the sessions were extremely short, regardless of the target's position. This was particularly prevalent during the second half of the session, suggesting that participant stopped following the instructions. We therefore omitted this outlier from the analyses and report results from the remaining 29 participants. After exclusion, we had $5 \times 3 \times 2 \times 29 = 870$ data points in total. We averaged the data points within the same distance range, resulting in only one data

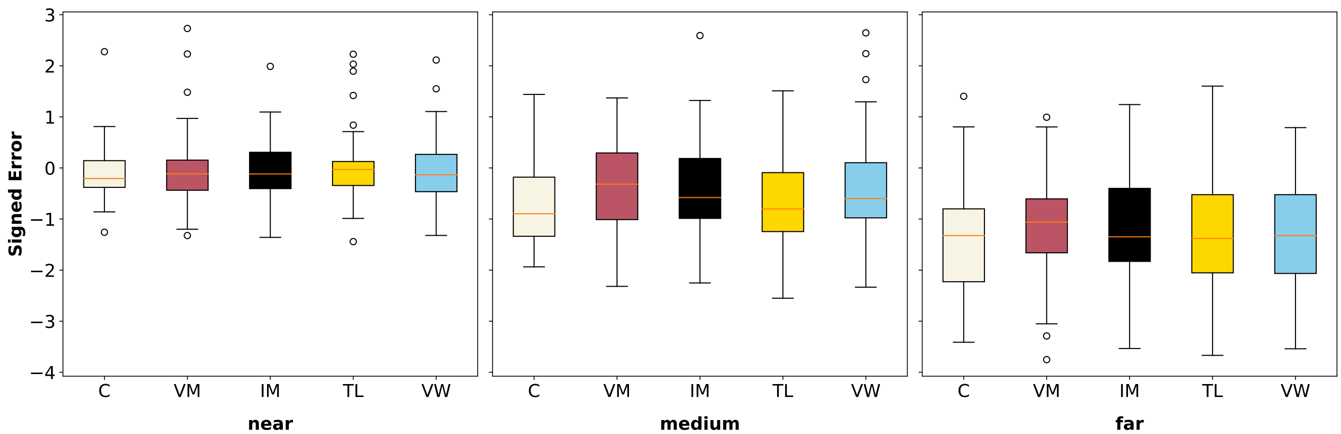

Figure 4: Box plots of signed depth judgment error for each five X-ray vision metaphors (control, visible mesh, invisible mesh, tramlines, and virtual window) under three distance ranges (near, medium, and far).

Table 1: Signed error results for all distance/metaphor combinations, showing the *M* (top) and *SD* (bottom) in each row.

	Condition				
Distance (m)	C	VM	IM	TL	VW
Near	-0.13 0.45	-0.06 0.68	-0.05 0.46	-0.03 0.56	-0.08 0.57
Medium	-0.72 0.74	-0.35 0.79	-0.44 0.74	-0.67 0.80	-0.43 0.88
Far	-1.41 0.91	-1.10 0.98	-1.24 1.03	-1.25 1.03	-1.29 1.02
Overall	-0.75 0.57	-0.50 0.71	-0.58 0.67	-0.65 0.69	-0.60 0.72

point per distance range per condition for each participant. This yielded a final count of 435 data points.

Signed Error of Depth Judgments Figure 4 shows the signed error results for the three distance ranges in each of the five x-ray vision metaphors. A 3x5 repeated-measures ANOVA revealed significant differences among distance ranges, $F(2,56) = 75.68$, $p < .001$. Post-hoc tests using a Holm-Bonferroni adjustment revealed differences between all comparisons between near ($M = -0.07, SE = 0.12$), medium ($M = -0.52, SE = 0.12$), and far ($M = -1.26, SE = 0.12$) distance ranges, all of which were significant at $p < .001$. The main effect for X-ray vision metaphor was not significant, $F(4,112) = 1.50$, $p = .208$, nor was there an interaction effect, $F(8,224) = 1.17$, $p = 0.32$.

Relative Error of Depth Judgments Figure 5 shows the relative error results for the three distance ranges in each of the five x-ray vision metaphors. A 3x5 repeated-measures ANOVA revealed significant differences among distance ranges, $F(2,56) = 34.20$, $p < .001$. Similar to the signed error results, post-hoc tests using a Holm-Bonferroni adjustment revealed differences between all comparisons between near ($M = -0.02, SE = 0.02$), medium ($M = -0.09, SE = 0.02$), and far ($M = -0.17, SE = 0.02$) distance ranges, all of which were significant at $p < .001$. The main effect for X-ray vision metaphor was not significant, $F(4,112) = 1.14$, $p = .34$, nor was there an interaction effect, $F(8,224) = 0.95$, $p = .48$.

Observation Time Shapiro-Wilk tests indicated possible violations of normality for nearly all observation times except those in the control condition. Therefore, we analyzed these data using a Friedman test. The effects of X-ray vision metaphor was not significant, ($\chi^2(4) = 6.10$, $p = .19$), and there was substantial variability in all five conditions: C ($Mdn = 11.87$, $IQR = 8.32$), VW ($Mdn = 13.71$, $IQR = 9.42$), TL ($Mdn = 13.21$, $IQR = 9.67$), VM ($Mdn = 9.96$, $IQR = 15.86$), and IM ($Mdn = 11.53$, $IQR = 9.35$).

TLX Scores Shapiro-Wilk tests indicated possible violations of normality for TLX overall workload. A Friedman test did not reveal any significant differences between X-ray vision metaphors, ($\chi^2(4) = 8.19$, $p = .085$). The overall workload scores in each condition were: C ($Mdn = 72.67$, $IQR = 13.38$, VW ($Mdn = 75.34$, $IQR = 11.37$), TL ($Mdn = 74.40$, $IQR = 13.77$), VM ($Mdn = 73.37$, $IQR = 12.67$), and IM ($Mdn = 71.04$, $IQR = 12.80$).

SUS Ratings Shapiro-Wilk tests indicated possible violations of normality for SUS ratings. A Friedman test did not reveal any significant differences between X-ray vision metaphors, $\chi^2(4) = 5.95$, $p = .20$. The usability ratings in each condition were: C ($Mdn = 80.0$, $IQR = 17.5$), VW ($Mdn = 80.0$, $IQR = 12.5$), TL ($Mdn = 80.0$, $IQR = 20.0$), VM ($Mdn = 80.0$, $IQR = 20.0$), and IM ($Mdn = 77.5$, $IQR = 15.0$).

Confidence Ratings Shapiro-Wilk tests indicated possible violations of normality for confidence ratings. A Friedman test did not reveal any significant differences between X-ray vision metaphors, $\chi^2(4) = 5.99$, $p = .20$. The confidence ratings in each condition were: C ($Mdn = 5.0$, $IQR = 1.0$), VW ($Mdn = 5.0$, $IQR = 1.0$), TL ($Mdn = 5.0$, $IQR = 2.0$), VM ($Mdn = 5.0$, $IQR = 1.0$), and IM ($Mdn = 5.0$, $IQR = 1.0$).

6 Discussion

6.1 Depth Judgments

Surprisingly, none of the X-ray vision metaphors significantly improved depth judgements compared to the *C* condition. All five conditions demonstrated underestimated depth, as evidenced by relative error, which aligns with previous AR depth perception research [1, 16, 33, 67]. However, the depth underestimation in the *C* was not as severe as we had anticipated. On average, participants in our study underestimated the distance by 12% in the *C* condition, with a 16.1% underestimation behind the wall. Although this was the highest magnitude of underestimation among the five conditions, it was still within the expected range of about 15% for viewing virtual objects on a HoloLens 2, as recent research suggests [1]. Phillips encountered similar findings in his previous X-ray vision

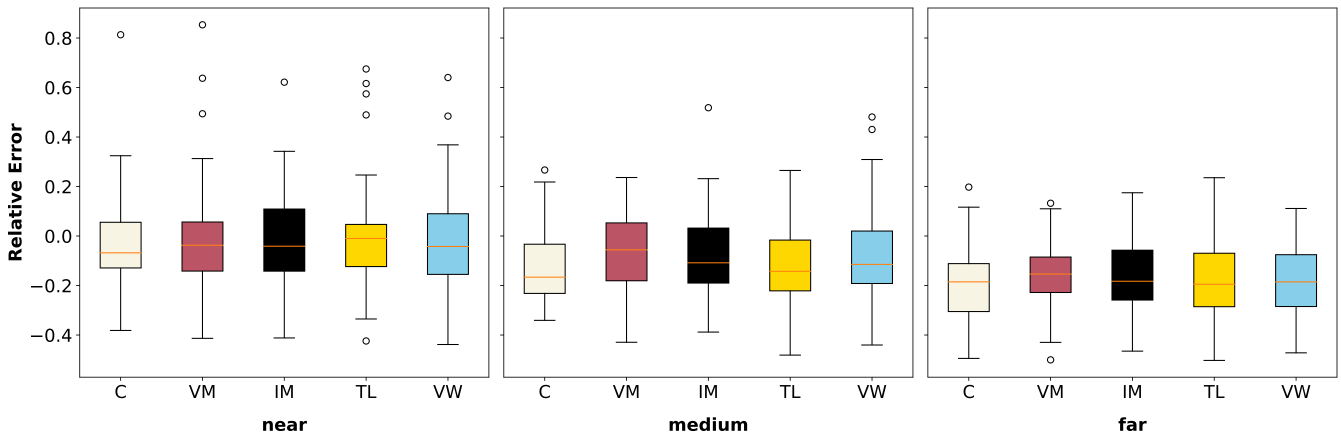

Figure 5: Box plots of relative depth judgment error for each five x-ray vision metaphors (control, visible mesh, invisible mesh, tramlines, and virtual window) under three distance ranges (near, medium, and far).

Table 2: Relative error results for all distance/metaphor combinations, showing the *M* (top) and *SD* (bottom) in each row.

	Condition				
Distance	C	VM	IM	TL	VW
Near	-0.04 0.15	-0.02 0.21	-0.02 0.15	-0.01 0.18	-0.03 0.19
Medium	-0.13 0.13	-0.07 0.14	-0.08 0.14	-0.12 0.14	-0.08 0.16
Far	-0.19 0.13	-0.15 0.13	-0.17 0.14	-0.17 0.14	-0.18 0.14
Overall	-0.12 0.11	-0.08 0.14	-0.09 0.13	-0.10 0.13	-0.09 0.14

research, wherein no significant differences were found when comparing virtual object viewing through an opaque wall and a real or virtual window [58]. This relatively low underestimation rate poses limitations to possible enhancements that can improve depth judgment accuracy.

Overall, participants' depth judgments in this experiment exhibited greater variability than we expected. As shown in Figure 5, the standard deviation of relative error indicated large variations for each condition and distance combination, even at short distances. Furthermore, overall depth judgment accuracy in both Tables 1 and 2 was consistently the worst in the *C* condition. Taken together, these characteristics imply that future studies of X-ray vision metaphors may need to more tightly control the distance estimation task, and similarly complex experimental designs will likely require larger sample sizes to achieve sufficient statistical power.

No statistical evidence was found to suggest that mesh-based techniques were superior to non-mesh-based X-ray vision techniques. We speculate that the directly overlaid mesh did not provide sufficient additional assistance to enhance motion parallax during the observation as we expected. Informal observation of Tables 1 and 2 show that the depth judgment error of *VM* was slightly, but consistently, smaller than that of *IM*, but these differences were too small to be statistically significant. In general, the variability in depth judgment error between four X-ray vision metaphors was relatively small. This suggests that selecting an appropriate X-ray vision metaphor from among the specific techniques evaluated in this study may be determined based on the fit for a specific application or influenced by subjective preferences.

As expected, the results showed that depth judgment error was significantly influenced by object distance. Figure 6 displays the relationship between actual target distance and blind walking distance for all data points in the three distance ranges. The green line represents the veridical walking performance, while the red dotted line indicates the position of the wall. Different colors are utilized to represent various metaphors. By looking at the graph, it is obvious that in the far distance range, participants tend to underestimate the target distance in most of the trials. This graph further demonstrates that as the target distance increases, there is a noticeable increase in distance underestimation. This result might imply that participants had difficulty estimating the target distance when it was too far away from the obstruction, particularly in the far distance range. However, most current X-ray vision metaphors focus on an ordinal judgment of whether the target is positioned in front of or behind the obstruction. Future X-ray vision metaphors could provide more cues or assist in improving depth judgment especially for distances behind the wall. Unfortunately, no significant interactions between X-ray vision metaphors and object distance ranges were found.

In summary, the absence of statistically significant differences in depth judgment error should not be interpreted as evidence that X-ray vision metaphors cannot provide meaningful benefits, and these data suggest the need for further investigation. The lack of significance among conditions aligns with previous X-ray vision studies conducted in personal space and medical settings, though the techniques compared in studies were different [32, 52].

6.2 Observation Time

We had originally hypothesized that observation times would vary significantly between X-ray vision metaphors. However, the results did not support our expectations. Therefore, we are unable to draw any conclusion about whether and to what extent TL helps expedite observation. An analysis of the observation time data, supplemented by the experimenter's direct observations, suggests that the strategies used for for distance judgments varied greatly between individuals. Some participants were capable of making swift judgments, typically within a few seconds, particularly after completing several trial blocks. However, other participants adopted a meticulous approach and consistently spent more time to make an accurate depth judgment, regardless of the X-ray vision metaphor being used.

6.3 Qualitative Feedback

Although the subjective ratings from **Confidence**, **SUS** and **TLX** questionnaires did not reveal significant effects, the qualitative feed-

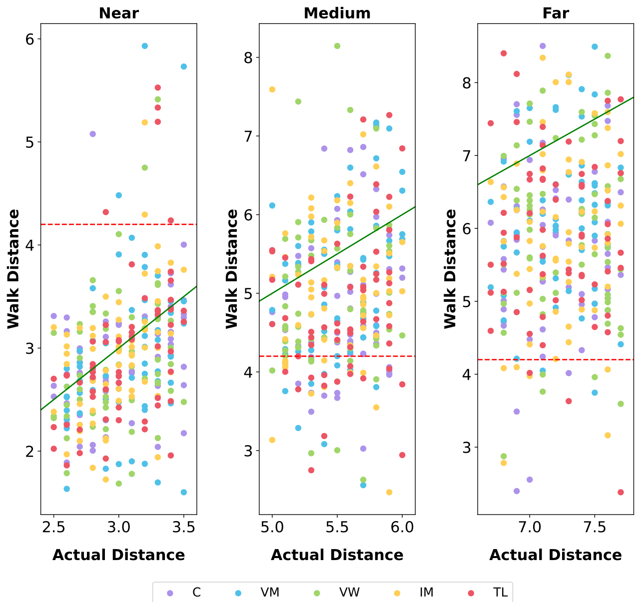

Figure 6: Scatter plots of Actual Distance vs. Walking Distance for all data points in Near, Medium, and Far distance ranges. The green line represents the veridical walking performance, while the red dotted line indicates the position of the wall. Note that as the target distance increases, data points become more scattered, with many of them falling below the green line, particularly in the third graph when compared to the first two graphs. This figure shows all 870 data points

back from participants provided valuable insights. As described in section 4.5, participant feedback was collected concerning both the metaphors and the experiment itself. In response to questions concerning the extent to which metaphor features assisted or impeded task completion, participants offered a variety of insightful comments.

When discussing the *C* condition, thirteen participants found it lacking cues or benchmarks to assist in distance judgment, with four considering it the least effective. Interestingly, however, two participants reported that the *C* condition was the most effective, with one noting, "no other things hindered or confused me."

Nineteen participants conveyed that the *VM* condition was a useful reference for determining whether the target was situated in front or behind the wall because it occluded the target. One participant mentioned feeling that the mesh was "very stable" because "it touched the ground." However, several participants articulated concerns about the target being partially occluded by the mesh, which they found unhelpful and distracting.

Feedback for the *IM* condition echoed that of the *VM* but was often more critical. Participants' comments on the mesh included "masked information," "a bit thick," "complicated the distance judgement," and made the observation task "trickier." It appears that our design of the invisible mesh confused some participants, as they did not mention the thickness of the mesh in the *VM* condition. We speculate that this may have occurred because the mesh was only visible in the overlapped area. The limited visible region accentuated the thickness of the mesh, creating an impression that the target was segmented by a set of "bars" rather than being occluded by a layer of mesh when participants did not move their bodies during observation. Future designs could address this issue by incorporating visual effects that extend the visibility of the mesh. Additionally, future designs should critically evaluate the thickness of the mesh rather than strictly replicating the appearance of the *VM*. In this study, we chose the latter approach to avoid introducing extra factors to our study design. Despite these criticisms, eight participants described the *IM* condition as beneficial, with two of them identifying it as one of the most useful metaphors.

Twenty-one participants described the *TL* condition as advantageous for making distance judgments, with seven participants classifying it as one of the most helpful metaphors. They shared comments such as the line told them how far the target was away from them, guided them to the target and helped them walk in a straight line. Some participants expressed a negative view of *TL*, noting that "the tracks were not level with the ground in the distance." They made this comment because the ground was blocked behind the wall, and we did not provide any reference, causing the feeling that the line was floating in the air. Although some participants expressed the favor of this unnatural placement gave them "a better idea of how far the spheres past the curtain," future designs could incorporate the visualization of ground behind the wall.

The *VW* condition was viewed as helpful by fifteen participants, with eight participants labeling it as one of their most preferred metaphors. These participants made remarks such as "like a real window," "gave me a strong perception and spatial guidance of the spatial relationship of me, the window square, and the ball," and "when blind walking, I could fell that I was walking through the window." Negative comments mainly focused on when target was position behind the wall. For instance, participants commented the window "hindered my ability to contrast the outline of the sphere with the wall." This feedback may suggest that some participants did not consider the window as a reference on the surface of the wall and thought it limited the observation area, somewhat contradicting the positive comments.

The diversity of subjective feedback echoes the results of depth judgment to some degree and reflects the wide range of strategies that participants employed , each having different interpretations for each metaphor. This suggests that there may not be a "one-size-fits-all" X-ray vision metaphor that would be considered generally superior for all potential users and situations. However, based on the feedback from participants, we identify several important considerations for the design of these techniques. For example, some participants underscored the importance of the ground as a reference point during observation, while others highlighted the stability of the visualization when objects appeared to touch the ground. Furthermore, when asked about their strategies for task completion, three participants cited the use of an imaginary vertical line extending from the objects to the ground as a basis for distance estimation. These insights suggest that a stable reference point to the ground plane should be a significant consideration when designing X-ray vision techniques. Additionally, participants found a direct comparison of the occlusion cues between the X-ray visualization and the virtual target object useful for depth judgment. However, their responses also suggest the need to consider the proportion of the virtual object that is being occluded.

6.4 Limitations and Future Work

Although the experiment was carefully designed using best practices from the literature, we can identify several limitations and methodological improvements for future work. In this study, we relied on the HoloLens 2's integrated Simultaneous Localization and Mapping (SLAM) system. While overall, we did not detect any noticeable instability of the system or inconsistency in the participants' performance and data, it is possible that the HoloLens 2's tracking system could have negatively impacted the precision of the recorded distance judgments in ways that are difficult to observe.

In the traditional blind walking procedures, participants remain standing while observing the target object. However, given the relatively long duration of our study, we opted to have them sit on the chair to conserve their energy for the walking tasks. Additionally, the target was suspended in the air. The variations between our approach and traditional blind walking could make it somewhat difficult to directly compare our results with those of other studies.

We chose not to include an additional control condition where the occluder is absent in our study because we believe it would pose challenges in setting up hypotheses and comparing the results of this condition with those of other X-ray vision metaphor conditions. However, future studies could incorporate another experiment that compares a "do-nothing' condition (*C* condition) with a condition where the occluder is not present before evaluating the X-ray vision metaphors. This approach would enable the assessment of the extent to which the introduction of the occluding structure might impact the results.

For practical reasons, the interval between the closing of participants' eyes and the onset of the blind walking task was longer than ideal, and several participants suggested increasing the curtain's speed. When constructing the movable wall for the study, we selected a motorized system that would be sufficiently quiet to avoid intrusive auditory cues, but no technical details regarding the curtain's slide time were available at the time of purchase. The participant feedback suggests a need to more carefully consider the trade-off between the sliding time and the ensuing noise. These delays extended the overall duration of task, and some participants noted that they felt fatigued in the latter stages of the experiment.

Additionally, due to space limitations, the lab space was not completely empty. Despite clearing the area around the tripod and ensuring a safe path for participants to traverse, there were objects presents along the sides of the room that couldn't be fully obscured by the wall. One participant disclosed utilizing the room's physical objects as references during observation, although the overall impact of these external reference points is unknown.

The primary motivation behind the design of the invisible mesh was to address potential distractions arising from the visible grid mesh structure. Although the quantitative results were inconclusive, participant feedback suggests that the visible mesh was generally preferred over its invisible counterpart, at least for the specific implementations tested in this experiment. Comments suggest that the density of the invisible mesh grid or the dynamic nature of the invisibility effect may warrant further refinement. In the future, studies that more deeply investigate the design parameters of the visible and invisible mesh techniques are needed to answer these open questions.

7 Conclusion

This paper presented a novel X-ray vision metaphor, "invisible mesh," and a within-subjects user study that investigated the effects of multiple techniques on distance estimation in action space using optical-see-through augmented reality. Although the quantitative depth judgment results were largely inconclusive, qualitative feedback from participants provides valuable insights that can enrich current understanding of AR X-ray vision techniques. Furthermore, the results indicate that users employ a variety of strategies and often have conflicting opinions about which visualization techniques are most helpful. These results suggest the need for multiple types of X-ray vision metaphors that can be selected based on user preferences, and grid-based techniques such as visible and invisible mesh are both viable candidates for future study and refinement.

Acknowledgments

The authors would like to thank Victoria Interrante for providing valuable suggestions on the study design and J. Adam Jones for constructive feedback on the paper. GPT-4 from OpenAI was used for proofreading and grammatical editing of the completed manuscript.

References

[1] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Depth perception in augmented reality: The effects of display, shadow, and position. pp. 792–801. IEEE, 3 2022.

[2] H. Adams, J. Stefanucci, S. Creem-Regehr, G. Pointon, W. Thompson, and B. Bodenheimer. Shedding light on cast shadows: An investigation of perceived ground contact in AR and VR. *IEEE Transactions on Visualization and Computer Graphics*, 28:4624–4639, 12 2022.

[3] J. Andre and S. Rogers. Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics*, 68:353–361, 4 2006.

[4] B. Avery, C. Sandor, and B. H. Thomas. Improving spatial perception for augmented reality X-ray vision. In *2009 IEEE Virtual Reality Conference*, pp. 79–82, 2009.

[5] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *SIGGRAPH Comput. Graph.*, 26(2):203–210, jul 1992.

[6] R. Bane and T. Hollerer. Interactive tools for virtual X-ray vision in mobile augmented reality. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 231–239, 2004.

[7] Y. Bao, J. Wang, Z. Wang, and F. Lu. Exploring 3D interaction with gaze guidance in augmented reality. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 22–32, 2023.

[8] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab. Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 129–138, 2007. doi: 10.1109/ISMAR.2007.4538837

[9] J. V. Bradley. Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association*, 53(282):525–528, 1958. doi: 10.1080/01621459.1958.10501456

[10] J. Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.

[11] L. E. Buck, M. K. Young, and B. Bodenheimer. A comparison of distance estimation in HMD-based virtual environments with different HMD-based conditions. *ACM Transactions on Applied Perception*, 15:1–15, 7 2018.

[12] Z. Choudhary, M. Gottsacker, K. Kim, R. Schubert, J. Stefanucci, G. Bruder, and G. F. Welch. Revisiting distance perception with scaled embodied cues in social virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 788–797, 2021.

[13] S. H. Creem-Regehr, J. K. Stefanucci, and B. Bodenheimer. Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378, 1 2023.

[14] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception*, 34:191–204, 2 2005.

[15] J. E. Cutting and P. M. Vishton. Chapter 3 - perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein and S. Rogers, eds., *Perception of Space and Motion*, Handbook of Perception and Cognition, pp. 69–117. Academic Press, San Diego, 1995.

[16] 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)*, pp. 111–122, 2017.

[17] T. D. Do, J. J. LaViola, and R. P. McMahan. The effects of object shape, fidelity, color, and luminance on depth perception in handheld mobile augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 64–72, 2020.

[18] F. El Jamiy, A. N. Ramaseri Chandra, and R. Marsh. Distance accuracy of real environments in virtual reality head-mounted displays. In *2020 IEEE International Conference on Electro Information Technology (EIT)*, pp. 281–287, 2020. doi: 10.1109/EIT48999.2020.9208300

[19] S. R. Ellis and B. M. Menges. Localization of virtual objects in the near visual field. *Human factors*, 40(3):415, 09 1998. Copyright - Copyright Human Factors and Ergonomics Society Sep 1998; Last

updated - 2017-11-09; CODEN - HUFAA6.

[20] M. T. Eren and S. Balcisoy. Evaluation of X-ray visualization techniques for vertical depth judgments in underground exploration. *Vis. Comput.*, 34(3):405–416, mar 2018.

[21] S. K. Feiner and D. D. Seligmann. Cutaways and ghosting: satisfying visibility constraints in dynamic 3D illustrations. *The Visual Computer*, 8:292–302, 2005.

[22] H. Fuchs, M. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. Crawford, P. Rademacher, S. Drake, and A. Meyer. Augmented reality visualization for laparoscopic surgery. pp. 934–943, 10 1998.

[23] T. Fukiage, T. Oishi, and K. Ikeuchi. Reduction of contradictory partial occlusion in mixed reality by using characteristics of transparency perception. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 129–139, 2012.

[24] C. Furmanski, R. Azuma, and M. Daily. Augmented-reality visualizations guided by cognition: perceptual heuristics for combining visible and obscured information. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 215–320, 2002.

[25] D. Gaines and S. Kuhl. Methods for evaluating depth perception in a large-screen immersive display. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*, SUI '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3385959.3418447

[26] S. Ghasemi, M. Otsuki, P. Milgram, and R. Chellali. Use of random dot patterns in achieving X-Ray vision for near-field applications of stereoscopic video-based augmented reality displays. *Presence*, 26(1):42–65, 2017. doi: 10.1162/PRES_a_00286

[27] J. J. Gibson. *The Ecological Approach to Visual Perception*. Psychology Press, 11 1979.

[28] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception*, 7:1–18, 7 2010.

[29] U. Gruenefeld, Y. Brück, and S. Boll. Behind the scenes: Comparing X-ray visualization techniques in head-mounted optical see-through augmented reality. In *19th International Conference on Mobile and Ubiquitous Multimedia*, pp. 179–185. ACM, 11 2020.

[30] H.-J. Guo, J. Z. Bakdash, L. R. Marusich, O. E. Ashtiani, and B. Prabhakaran. User evaluation of dynamic X-Ray vision in mixed reality. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 851–852, 2023.

[31] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988.

[32] F. Heinrich, L. Schwenderling, M. Streuber, K. Bornemann, K. Lawonn, and C. Hansen. Effects of surface visualizations on depth perception in projective augmented reality. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*, pp. 1–6, 2021. doi: 10.1109/ICHMS53169.2021.9582452

[33] J. Hertel and F. Steinicke. Augmented reality for maritime navigation assistance - egocentric depth perception in large distance outdoor environments. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 122–130, 2021.

[34] H. Hu, A. Gooch, W. Thompson, B. Smits, J. Rieser, and P. Shirley. Visual cues for imminent object contact in realistic virtual environments. pp. 179–185. IEEE, 2000.

[35] V. Interrante, H. Fuchs, and S. Pizer. Conveying the 3D shape of smoothly curving transparent surfaces via texture. *IEEE Transactions on Visualization and Computer Graphics*, 3:98–117, 1997.

[36] V. Interrante, J. Lindquist, and L. Anderson. Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. In *2007 IEEE Virtual Reality Conference*, pp. 11–18, 2007. doi: 10.1109/VR.2007.352458

[37] 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*, APGV '08, p. 9–14. Association for Computing Machinery, New York, NY, USA, 2008.

[38] D. Kalkofen, E. Mendez, and D. Schmalstieg. Interactive focus and context visualization for augmented reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 191–201, 2007.

[39] D. Kalkofen, E. Veas, S. Zollmann, M. Steinberger, and D. Schmalstieg. Adaptive ghosted views for augmented reality. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–9, 2013.

[40] Y. Kameda, T. Takemasa, and Y. Ohta. Outdoor see-through vision utilizing surveillance cameras. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '04, p. 151–160. IEEE Computer Society, USA, 2004.

[41] J. W. Kelly. Distance perception in virtual reality: A meta-analysis of the effect of head-mounted display characteristics. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–13, 2022. doi: 10.1109/TVCG.2022.3196606

[42] S. Liao, Y. Zhou, and V. Popescu. AR interfaces for disocclusion—a comparative study. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 530–540, 2023.

[43] Q. Lin, X. Xie, A. Erdemir, G. Narasimham, T. P. McNamara, J. Rieser, and B. Bodenheimer. Egocentric distance perception in real and HMD-based virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, pp. 75–82. ACM, 8 2011.

[44] M. Livingston, A. Dey, C. Sandor, and B. Thomas. *Pursuit of "X-Ray Vision" for Augmented Reality*. 09 2012.

[45] M. Livingston, J. Swan, J. Gabbard, T. Hollerer, D. Hix, S. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 56–65, 2003.

[46] M. A. Livingston, Z. Ai, K. Karsch, and G. O. Gibson. User interface design for military ar applications. *Virtual Real.*, 15(2–3):175–184, jun 2011.

[47] M. A. Livingston, Z. Ai, J. E. Swan, and H. S. Smallman. Indoor vs. outdoor depth perception for mobile augmented reality. In *2009 IEEE Virtual Reality Conference*, pp. 55–62, 2009.

[48] J. Loomis, J. A. Da Silva, N. Fujita, and S. Fukusima. Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):906–921, 1992.

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

[50] M. C. d. F. Macedo and A. L. Apolinario. Occlusion handling in augmented reality: Past, present and future. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021.

[51] M. C. d. F. Macedo and A. Lopes Apolinario. Improving on-patient medical data visualization in a markerless augmented reality environment by volume clipping. In *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 149–156, 2014.

[52] A. Martin-Gomez, J. Weiss, A. Keller, U. Eck, D. Roth, and N. Navab. The impact of focus and context visualization techniques on depth perception in optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4156–4171, 2022. doi: 10.1109/TVCG.2021.3079849

[53] S. Masnadi, K. Pfeil, J.-V. T. Sera-Josef, and J. LaViola. Effects of field of view on egocentric distance perception in virtual reality. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–10. ACM, 4 2022.

[54] R. Messing and F. H. Durgin. Distance perception and the visual horizon in head-mounted displays. *ACM Trans. Appl. Percept.*, 2(3):234–250, jul 2005.

[55] M. Otsuki, H. Kuzuoka, and P. Milgram. Analysis of depth perception with virtual mask in stereoscopic AR. In *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments*, ICAT - EGVE '15, p. 45–52. Eurographics Association, Goslar, DEU, 2015.

[56] A. Padilha and V. Teichrieb. Motion-aware ghosted views for single layer occlusions in augmented reality. In *2015 IEEE International Symposium on Mixed and Augmented Reality Workshops*, pp. 60–67, 2015.

[57] C. C. Pagano and G. P. Bingham. Comparing measures of monocular distance perception: Verbal and reaching errors are not correlated. *Jour-*

nal of Experimental Psychology: Human Perception and Performance, 24:1037–1051, 1998.

[58] N. Phillips. *X-ray vision at action space distances: depth perception in context*. Mississippi State University, 2022.

[59] N. Phillips, F. A. Khan, M. S. Arefin, C. L. Bethel, and J. E. S. IP. Measuring virtual object location with X-Ray vision at action space distances. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 820–821, 2022.

[60] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker. Distance perception in real and virtual environments. *ACM Trans. Appl. Percept.*, 2(3):216–233, jul 2005.

[61] B. Ries, V. Interrante, M. Kaeding, and L. Anderson. The effect of self-embodiment on distance perception in immersive virtual environments. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, VRST '08, p. 167–170. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1450579.1450614

[62] J. P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators & Virtual Environments*, 9:287–309, 2000.

[63] J. P. Rolland, R. L. Holloway, and H. Fuchs. Comparison of optical and video see-through, head-mounted displays. In H. Das, ed., *Telemanipulator and Telepresence Technologies*, vol. 2351, pp. 293 – 307. International Society for Optics and Photonics, SPIE, 1995.

[64] J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence*, 11(6):610–625, 2002.

[65] C. Sandor, A. Cunningham, A. Dey, and V.-V. Mattila. An augmented reality X-ray system based on visual saliency. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 27–36, 2010.

[66] A. State, D. Chen, C. Tector, A. Brandt, H. Chen, R. Ohbuchi, M. Bajura, and H. Fuchs. Case study: Observing a volume rendered fetus within a pregnant patient. 09 1995.

[67] 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*, 13(3):429–442, 2007.

[68] K. Vaziri, M. Bondy, A. Bui, and V. Interrante. Egocentric distance judgments in full-cue video-see-through VR conditions are no better than distance judgments to targets in a void. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 1–9. IEEE, 3 2021.

[69] A. Webster, S. K. Feiner, B. MacIntyre, W. Massie, and T. Krueger. Augmented reality in architectural construction, inspection, and renovation. *Computing in Civil Engineering*, pp. 913–919, 1996.

[70] W. Willett, B. A. Aseniero, S. Carpendale, P. Dragicevic, Y. Jansen, L. Oehlberg, and P. Isenberg. Perception! immersion! empowerment! superpowers as inspiration for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):22–32, 2022.

[71] B. Wu, T. L. Ooi, and Z. J. He. Perceiving distance accurately by a directional process of integrating ground information. *Nature*, 428:73–77, 3 2004.

[72] S. Zollmann, D. Kalkofen, E. Mendez, and G. Reitmayr. Image-based ghostings for single layer occlusions in augmented reality. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 19–26, 2010.