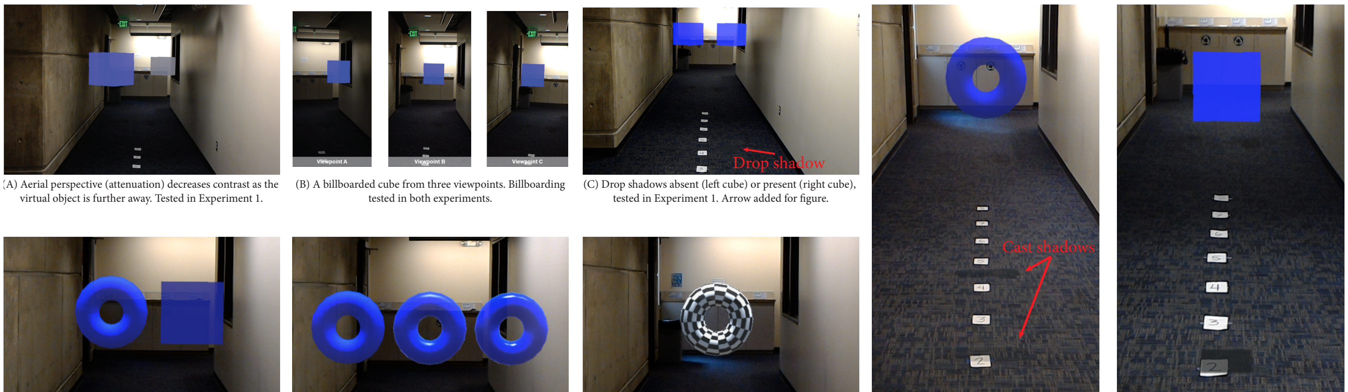
Designing for Depth Perceptions in Augmented Reality

Catherine Diaz* Michael Walker† Danielle Albers Szafir‡ Daniel Szafir§
University of Colorado Boulder, 1111 Engineering Drive, Boulder, CO 80309 USA



(A) Aerial perspective (attenuation) decreases contrast as the virtual object is further away. Tested in Experiment 1.

(B) A billboarded cube from three viewpoints. Billboarding tested in both experiments.

(C) Drop shadows absent (left cube) or present (right cube), tested in Experiment 1. Arrow added for figure.

(D) Dimensionality compared a 3D torus (left) against a 2D plane (right). Tested in Experiment 2.

(E) Lambertian (left), Blinn (middle), and Phong (right) shading. Shading models tested in both experiments.

(F) A checkerboard pattern was used to evaluate texture as a depth cue. Tested in Experiment 2.

(G) Cast shadows via raytracing from each light source, tested in Experiment 2. Arrows added for figure.

(H) An example of combined techniques; the cube is rendered with drop shadows, Blinn shading, and billboarding.

Figure 1: This work examined how various design decisions affect user depth perceptions of virtual objects in augmented reality applications. We conducted two studies with a perceptual mapping task in which participants aligned a floating virtual object with one of eight real world targets. Independent variables consisted of various design decisions regarding virtual object rendering, including (A) aerial perspective, (B) billboarding, cast shadows (either (C) simple drop shadows or (G) via ray tracing), (D) dimensionality (2D versus 3D shape), (E) shading, and (F) texture. Both experiments were full factorial designs, enabling us to examine interactions among cue combinations (H).

Abstract

Augmented reality technologies allow people to view and interact with virtual objects that appear alongside physical objects in the real world. For augmented reality applications to be effective, users must be able to accurately perceive the intended real world location of virtual objects. However, when creating augmented reality applications, developers are faced with a variety of design decisions that may affect user perceptions regarding the real world depth of virtual objects. In this paper, we conducted two experiments using a perceptual matching task to understand how shading, cast shadows, aerial perspective, texture, dimensionality (i.e., 2D vs. 3D shapes) and billboarding affected participant perceptions of virtual object depth relative to real world targets. The results of these studies quantify trade-offs across virtual object designs to inform the development of applications that take advantage of users' visual abilities to better blend the physical and virtual world.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – *artificial, augmented, and virtual realities, evaluation/methodology*—; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – *color, shading, shadowing, and texture, virtual reality.*

1 Introduction

Augmented Reality (AR) technologies are rapidly maturing and hold great promise for improving human efforts across a variety of domains. For instance, researchers have recently designed AR applications that enhance architecture and construction projects [10], help doctors and surgeons in healthcare and medical settings [76], improve human-robot interaction [22], and aid in design and manufacturing processes [51]. Within the space of AR technologies, see-through augmented reality head-mounted displays (ARHMDs) appear particularly promising in providing hands-free, low-latency AR interactions. While primitive ARHMD devices that display virtual objects on top of the real world have existed since Sutherland's "Sword of Damocles" [67], recent advances in technology quality, availability, and cost mean that designers have an increasing interest in developing novel consumer ARHMD applications. In order to best take advantage of the capabilities provided by ARHMDs, developers need to understand how the *design* of AR applications may influence their effectiveness.

One of the most critical issues for AR application designers is ensuring that virtual objects appear in the correct places in the real world and are perceived accurately relative to other virtual and physical objects in a scene. Properly positioning virtual objects is a challenging task that may require visual markers (e.g., [36]) or computer vision techniques (e.g., [76]) to build understandings of the physical environment. This task is complicated by the fact that, in an ARHMD paradigm, *all positioning of virtual objects is simulated*—while a virtual object may appear to be in the real world, it is physically displayed on a two-dimensional screen near the viewer's eyes. It is only the graphics techniques used to render virtual objects, such as stereo rendering, perspective projection, or choice of lighting model, that create the illusion of these objects being physically present at various locations in the real world. As a result, understanding the relationship between how virtual objects are rendered and how users will perceive the spatial position of those objects represents a crucial design challenge.

Designers make many decisions when crafting virtual objects for AR applications, such as selecting the material properties of an object, the shading or shadows used, and the kinds of lighting

*Department of Computer Science | catherine.diaz@colorado.edu
†Department of Computer Science | Michael.Walker-1@colorado.edu
‡Department of Information Science | danielle.szafir@colorado.edu
§Department of Computer Science & ATLAS Institute daniel.szafir@colorado.edu

that will be simulated. This design space is complicated by the fact that different decisions may alter user perceptions of where a virtual object is located. For instance, research has found that users consistently underestimate distance in fully virtual environments [38, 45, 73, 77–79]. Augmented reality research has found similar effects for ARHMDs [43, 69], although a great deal of this research has examined factors such as convergence [20] and latency [48], which are typically intrinsic properties of AR hardware and may be outside of a designer's control. There has been comparatively less research examining the perceptual influence of factors that designers can control, such as choices of surface shading model or shadow design. Our research provides a systematic study of how these design factors can influence depth perceptions in the context of see-through ARHMDs, in which virtual and physical objects are intermixed.

The goal of this research is to provide AR developers insight into how design trade-offs influence the perceived real-world depth of virtual objects in AR applications, which we explore in the context of depth alignment with real-world objects. Specifically, we conducted two experiments to examine how shadows, aerial perspective, dimensionality (i.e., 2D vs. 3D shapes), billboarding, texture, and shading models affect AR depth perceptions (Figure 1). Our experiments explore these factors across different implementations, shapes, and object states (e.g., size, position, rotation) to help contextualize our findings in realistic practices where a variety of environmental considerations may influence virtual object perceptions and alignments. Overall, we find that cast shadows provide a highly useful cue in improving depth judgments across all conditions. In addition, our results show that these factors combine in complex ways, with certain combinations cumulatively supplementing each other to provide additional depth refinements. Our results provide empirical guidance for increasing the accuracy and speed of user depth judgments in AR applications that designers can readily integrate into existing development practices.

2 Related Work

In this work, we build on prior research in visual perception and mixed reality. Below, we review related work examining the interaction between visual design and depth perceptions and discuss perceptual research in virtual environments.

2.1 Depth Perception

Perceptual psychology provides a rich literature on the visual mechanisms involved in depth perception (see Howard [29] for a survey). This literature enumerates specific visual properties of a scene that cue the viewer to the spatial position of different objects. Augmented reality and other computer graphics technologies can take advantage of these *depth cues* in order to make objects appear to be in their intended spatial positions. Many of these depth cues are *monocular*—they can be detected in any two-dimensional image. For example, if one object occludes another, the first object is perceived as closer to the viewer. Other cues are *binocular*—they rely on fused inputs from the left and right eyes to create a sense of depth. While conventional graphics applications rely heavily on monocular cues due to constraints of conventional displays (i.e., 2D monitors), many modern AR technologies allow designers to leverage binocular cues by rendering two separate images—one for the right eye and a second for the left—that mimic the left and right eye images of the virtual object as positioned within the real world.

Studies in perceptual psychology have compared the effectiveness of these cues in communicating depth for physical objects in the real world. For example, Gilliam synthesizes a series of findings to identify what depth cues most accurately communicate the spatial configuration of real-world objects at different viewing distances [25]. The physical world may provide a strong baseline for understanding perceptions of virtual objects; however, it is critical to note that virtual objects are inherently simulations—they *approximate* aspects of the real world. As a result, perceptions can vary between physical and virtual objects [27] and their interplay can lead to depth discrepancies [17, 19, 37]. Additionally, virtual objects are designed with respect to available graphics techniques, not isolated depth cues. These techniques might influence multiple perceptual cues simultaneously. For example, perspective projection manipulates both object size and height above the plane. Due to these considerations, we use examples from computer graphics to inform our evaluation (see Thompson et al. [72] for a survey).

Many studies in graphics have evaluated depth interactions in purely virtual scenes. For example, research has shown that lighting direction [39] and shading model [40] influence perceived depth along virtual surfaces. Berbaum et al. [4] found that interactions between shadow and materials in simulated lighting affect depth perceptions. Cipiloglu et al. [12] and Zheng et al. [82] provide functions for layering different graphics techniques to improve apparent depth. In this work, we build on these approaches to understand how the techniques used to render virtual objects influence their perceived position in the real world.

2.2 Depth Perception in Virtual and Augmented Environments

We draw additional inspiration for our research from perceptual studies in virtual environments (VEs) and prior work in AR. Researchers have devoted a great deal of study to depth perceptions in VEs, recognizing that "in order for a user to act in a virtual world as if present in the physical world being simulated, he or she must perceive spatial relations the same way they would be perceived if the user were actually in the physical world" [73]. The majority of prior work in VE depth perception has investigated absolute egocentric depth judgments, meaning real-world estimates of how far away an object is from a viewer. Such studies (e.g., [38, 45, 73, 77–79]) typically have a user walk to or toward a previously seen target (e.g., *blind walking*, *imagined blind walking*, or *triangulation by walking*) and have consistently shown that people underestimate depth in purely virtual environments.

There has been comparatively less research performed on modern AR systems (see Swan & Gabbard [68], Dünser et al. [18], and Livingston et al. [43] for surveys), although past research has found similar underestimation of depth for virtual objects (e.g., [69]). To gauge depth perceptions, AR research has used *walking tasks* similar to those used in VEs (sometimes adapted as reaching tasks, rather than walking) [71], *perceptual mapping* tasks that involve aligning virtual and physical objects [5, 21, 48], *verbal reports* of object depth [35], *forced choice* tasks in which participants must select from a given set of possible depth judgments (e.g., choose whether or not an object is closer than another object) [24, 44, 57], or combinations of these methods [31, 58, 63, 69, 70].

Such past research has begun to build a preliminary understanding of depth perceptions in AR. For example, early work hinted at a potential relationship between object size and depth estimations in AR [57], a finding confirmed by later research [58]. Other work has explored occlusion as a depth cue in AR [15, 44, 69], especially for x-ray augmented reality, in which AR enables users to "see through" occluding or partially occluding surfaces (e.g., [59]). Ellis and Menges [21] show that physical surfaces near virtual objects affect viewer depth judgments, indicating that the interplay between real objects and virtual objects may complicate the AR design space beyond that of purely virtual environments. Perhaps most similar to our current work, Cidota et al. [11] explore the impact of visual fade and blur effects on perceived virtual object depth, while Sugano et al. [65] investigate using a marker-based system to estimate a shading model for virtual objects, finding that shading cues increase virtual object presence and aid in ordinal depth estimations. While these studies have been invaluable in building initial knowledge depth perceptions in AR, they are limited in their ability to inform modern AR designers: several use custom-built HMDs in artificial laboratory environments (e.g., with fixed-in-place headsets that restrict participant mobility), while others lack rigorous evaluations of multiple cues with a sufficiently large population size. Moreover, many prior studies have evaluated strict perceptual cues or properties of HMD hardware (e.g., [14]), rather than those cues arising from conventional design decisions, making it difficult to generalize findings to designer practices. In our study, we replicate several existing findings and offer new evidence of how these features might interact with one another in practice. To our knowledge, this work is the first to characterize the main and interaction effects of such a large variety of different AR design decisions (that can be influenced by application developers) on user depth perceptions using a modern, commercially available ARHMD.

3 Potential Factors Influencing Depth in AR

To investigate user depth perceptions of virtual objects in augmented reality, we developed an AR application that enabled strict control over the design of virtual objects. We used this application, which is detailed in Section 4, in two laboratory experiments that evaluated the effects of several common design techniques developers might use to communicate virtual object position in AR applications.

While our experimental infrastructure is extensible to any desired design configuration, we drew our candidate stimuli from a survey of literature on perceptual psychology, depth cue interaction, and conventional graphics design techniques. From this survey, we selected six primary factors representing commonly used design choices relevant to a variety of AR use cases: aerial perspective, shadows, surface shading models, billboarding, dimensionality, and surface texture. We hypothesized that each of these factors may have a significant impact on depth perceptions. Moreover, unlike cues such as stereo viewing or accommodation, these cues reflect rendering choices AR developers can readily control that are not typically considered to be defaults or constrained by hardware. Three of these properties (billboarding, dimensionality, and surface texture) reflect properties of the virtual object, whereas the other three (aerial perspective, shadows, and shading model) simulate interactions between the object and environment. Below, we outline how each factor might influence depth perceptions based on observations from prior literature. The specific implementation of each factor in our experimental application is described in Sections 5 & 6, with examples shown in Figure 1.

Aerial Perspective: Also known as atmospheric attenuation, aerial perspective results from colors becoming more hazy as the distance between the object and observer increases. This haziness reduces the contrast between the object and the environment, which can result in significant changes to perceived depth [56]. In the real world, this effect is generally only noticeable for objects far in the distance, such as the bluish hue cast over mountains along a horizon. However, it is a commonly used cue in computer graphics at both long and short ranges [12, 34]. While prior work has simulated atmospheric attenuation in AR by adjusting virtual object opacity [41, 43], we are not aware of any research in AR that uses traditional graphics simulations of aerial perspective. We chose to examine aerial perspective as it is currently poorly understood in AR, has shown utility at close range in traditional graphics, offers an opportunity to explore perceptions when virtual and physical cues may not match, and may help build our knowledge regarding the role of color shift and other typically far-field cues in AR applications.

Cast Shadows: In the real world, when lighting interacts with physical objects, it creates shadows. In virtual scenes, we can simulate this interaction using cast shadows. Cast shadows are commonly used in graphics applications, such as video games, to increase realism and help localize objects in a scene. While there are several methods for implementing shadows (e.g., shadow mapping, baked lighting, etc.), creating realistic shadows can be an extensive and time-consuming effort for designers and may impact simulation or game quality (e.g., rendering speed). AR applications have the added challenge that the simulated lighting model used to render virtual objects may not precisely align with the physical lighting in the room. There is evidence both for and against the utility of cast shadows for depth judgments. While several studies provide evidence that shadows help people determine object position in virtual scenes (both monocular and stereoscopic) [5,33,40,46,72,81], evidence by Hartzell et al. [26] and Sugano et al. [65] suggests that shadows may be of limited importance in stereo scenes.

We seek to disambiguate the mixed results of past work by conducting a systematic study to determine the role of cast shadows in perceiving virtual object depth for AR. By understanding the importance of shadows for relative positioning, we can provide empirical guidance for future research efforts and help designers determine trade-offs in explicitly designing for shadows and simulated lighting models that may not precisely align with the physical lighting in the room. However, implementing cast shadows in AR is complicated by the lack of a virtual ground plane on which to render shadows. One approach to rendering shadows without a virtual ground is to create a semitransparent dark virtual plane of the shape and size of a shadow aligned with the physical ground plane. For example, Berning et al [5] looked at perceptual effects from virtual *drop shadows* (planes projected immediately below objects), while Zollmann et al. used virtual drop shadows to support spatial perceptions of real micro aerial vehicles [83]. Alternatively, ray tracing can be used to approximate shadow position, bounds, and intensity, rendering shadows that better correspond to actual scene lighting. While planar approximations may not precisely mirror all attributes of shadows in the real world (e.g., softness), the visual system is quite tolerant of shadow imperfections [30], which can be used to accelerate rendering for interactive applications in practice [60].

Our first study looks at performance for drop shadows, whereas the second evaluates ray traced cast shadows with aligned simulated and physical light sources. While the virtual objects in our experiment cast shadows on the real world, our experimental framework did not consider shadows real world objects might cast on virtual objects. To mitigate any potential conflicts from this decision, no real world objects were located between the light sources and virtual stimuli in our environment; therefore no real-to-virtual shadows would naturally occur. However, virtual objects did cast shadows upon themselves based on their physical structure, the tested shading model, and lighting position and intensity.

Shading: Surface shading models define the material and reflectance properties of a virtual object. Research in computer graphics [40] and psychology [1,9,16,52] suggest that the shading models used to render a virtual object influences its perceived spatial properties. These studies point to the importance of specular highlights for communicating depth information both along the surface of a virtual object and about the relative positioning of objects [1, 4, 52, 74]. However, misalignments between virtual highlights and the real world, which are common in AR due to the separation of real and virtual lighting, may have an adverse impact on these perceptions [1]. In this work, we explore three different shading models that apply varying levels of diffuse and specular shading to virtual objects: Lambertian (diffuse with no highlight), Blinn (diffuse highlight), and Phong (strong highlight). To better understand the effects of design in practice, our implementations apply surface shading through fragment shaders extracted directly from Maya[1], a popular 3D modeling system. Examining these three models allows us to explore how different properties of shading models commonly used in design tools might influence a virtual object's perceived depth.

[1] http://www.autodesk.com/products/maya/overview

Billboarding: Billboarding occurs when a two-dimensional virtual object continually updates its orientation to face the camera at a fixed angle. In AR, this means that the object rotates to constantly face the viewer, which may be useful for certain user interface elements, such as creating "holographic" menu systems. However, billboarding may also degrade perceptions of virtual object depth by reducing volumetric visual cues: the object always appears exactly the same to viewers regardless of their viewing position, limiting viewers' abilities to gather additional depth information from head motion. We tested the influence of billboarding on depth estimation as we currently know very little about this factor, despite the importance of menu systems and other labeling schemes in AR applications.

Dimensionality: Virtual objects in AR applications may take a variety of shapes, such as complex 3D objects or 2D menu systems. The dimensionality of the object itself might improve depth perceptions by providing viewers with additional pictorial depth cues, such as volume, curvature, or linear perspective [17]. Prior studies have provided preliminary evidence of this effect, identifying systematic underestimation for 2D objects relative to their 3D equivalents [5]. We hypothesize that these added cues will help people more accurately position virtual objects in the real world, leading to improved performance for depth estimations.

Texture: Texture density is a well-established pictorial depth cue: as objects get further away, their surface textures appear more dense [17]. Prior studies have shown that texturing objects enhances depth perceptions (e.g., [5, 28, 66]). As a result, past AR systems have synthetically added textures to objects to improve depth perceptions [6, 75]. However, many of these studies measure texture in isolation. Studies exploring texture in conjunction with other cues found that textures had little effect on perceptions relative to the other cues [50]. Given the use of supplemental textures in AR systems, we sought to further explore the impact of added textures on user depth judgments.

Other Cues: Other object attributes might also influence apparent position in AR or otherwise affect user abilities to estimate depth. We considered three such factors in our study: object size, position, and rotation. Object size might influence user abilities to discern depth cues associated with the virtual object [8]; for example, smaller objects might show less discernible shading [64]. Cutting & Vishton [13] showed that the effectiveness of different cues may change based on the position of an object relative to the viewer (e.g., binocular disparity is a stronger cue when objects are close), and physical displacement between objects may hinder depth alignments. Similarly, Al-Kalbani et al. show that virtual object position can influence perceived depth and affect freehand grasping accuracy [2], although such effects might be mitigated with dual view visual feedback [3]. Finally, the rotation of an object may change viewers' access to depth cues. For example, if the flat face of a three-dimensional cube faces a viewer, it obscures information about the physical depth of the cube and may also provide less information about surface shading. We treated size, initial position, and rotation as random effects to understand the impact of these attributes on depth perceptions and to control for potential experimental confounds.

Cue Interactions: When designing virtual objects, developers may often use multiple techniques at once. Layering different depth channels can improve depth perceptions beyond any single baseline [47, 53]. However, how this integration happens and to what degree it remains effective is still a matter of debate [12]. As a result, we sought to test not only the influence of each technique in isolation, but also in combination to identify effects from cue layering.

3.1 Hypotheses

Building on findings from prior research, we developed three hypotheses for depth perceptions in AR environments:

H1: Virtual object depth will be underestimated using an ARHMD.
H2: The design of a virtual object will influence the perceived depth of that object. Specifically:

(a) *Aerial Perspective* will improve depth judgments as demonstrated in prior experiments [43]; although a cue mismatch at the distances evaluated, it has proven useful in depth estimation at similar distances in traditional graphics applications [12, 34].
(b) *Cast shadows* will improve depth judgments as they provide visual information about how the virtual object acts on the physical world.
(c) *Shading models* with specular highlights (Blinn and Phong) will improve depth judgments by communicating the virtual object's position relative to the lighting in the room.
(d) *Billboarding* will degrade depth judgments by reducing visual information about object volume.
(e) *Dimensionality* will improve depth judgments with 3D shapes enhancing overall depth cueing.
(f) *Surface textures* will improve depth judgments through cues provided by texture density changes.

H3: Cues will exhibit interaction effects with cue layering improving depth judgments beyond applying cues individually.

4 General Experimental Design

We conducted two experiments to evaluate our hypotheses. Each of our experiments followed the same general procedure, which we describe below. Individual variations accounting for the specific factors considered in each study are described in the respective experiment sections.

4.1 Experimental Apparatus

For each experiment, we used a Microsoft HoloLens[2] as an experimental apparatus. The HoloLens is a wireless, optical see-through stereographic augmented reality HMD with a 30 x 17 degree field of view (FOV). Sensors utilized by the HoloLens include an integrated inertial measurement unit (IMU), depth sensor, ambient light sensor, and several cameras and microphones that support voice input, gesture recognition, and head tracking. While research is ongoing in comparing optical ARHMDs such as the HoloLens or Meta Glasses with video-based ARHMDs (e.g., [49]), the HoloLens was chosen due to its emerging popularity, ease of access (i.e., no hardware modifications are required, unlike in most video-based systems), and high potential as a model for future consumer ARHMD systems.

We built a custom experimental framework for the HoloLens using Unity,[3] a popular game engine for designing and developing virtual and augmented reality applications. This extensible framework allows designers to specify the design conditions to be tested, target positions, and real-world lighting to generate a set of interactive trials based on these conditions that can be ported directly to the HoloLens. To promote replicability and further investigation of depth cues, our framework is available at `http://iron-lab.org/research/ar-depth`. Our application ran at a constant 30fps with no visible latency. In this framework, overhead lighting was approximated by point and directional lights placed at the same locations as physical light sources (see Figure 2). Virtual light color and intensity approximated the real-world lights.

4.2 Experimental Task

Our application presented users with a perceptual mapping task inspired by Swan et al. [69] in which participants viewed a virtual shape that they could move forward and backward (positive and negative along the z axis). Participants used this system to align the center of the virtual object with one of eight real world targets

[2]https://www.microsoft.com/microsoft-HoloLens/en-us
[3]https://unity3d.com/unity

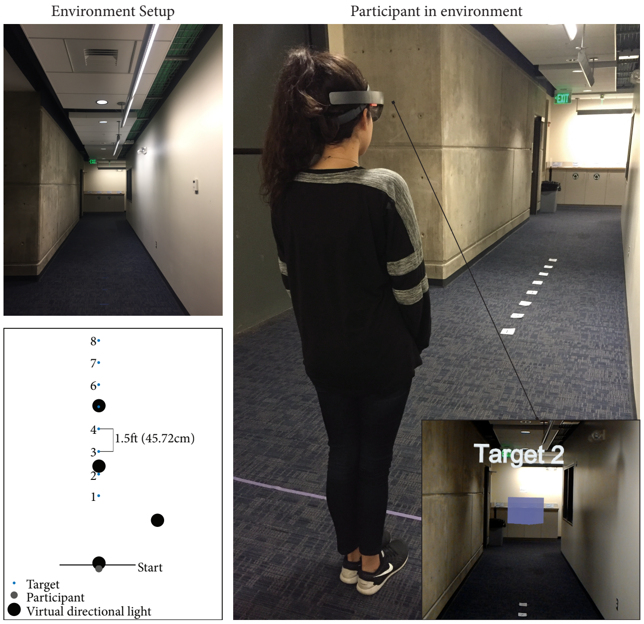


Figure 2: Participants used a custom-built AR application to align virtual objects with real-world targets (white cards) using the Microsoft HoloLens. We measured the corresponding depth accuracy by computing the distance between virtual objects rendered using different graphics techniques and the corresponding physical targets.

lined up in front of them in a hallway setup analogous to that used in Livingston et al. [42] and Swan et al. [71]. The eight targets were arrayed along the ground plane, which has been shown to enable people to accurately judge object distances up to 20*m* [80]. In our experiments, the targets were located throughout *action* space [73], arrayed at even intervals from 8*ft*–18.5*ft* (2.44*m*–5.64*m*) away from the viewer. Each target was separated by 1.5*ft* (45.7*cm*). According to developer guidelines, this target range is within the comfort zone of the HoloLens device.[4] Prior to the experiments, eight virtual targets that were invisible to participants were precisely aligned with the physical targets on the floor. This enabled the framework to record the distance between the aligned virtual object and the real-world targets and also allowed for calibration with the physical lighting in the environment.

In each experiment, every participant completed a fixed number of trials. Each trial presented the participant with a single virtual object rendered with various depth cues present or absent corresponding to the experimental factors under investigation. Text indicating the desired depth target was displayed just above the participants' calibrated line of sight (i.e., participants would look up to see the directions such that the text would not interfere with object alignment, Figure 2). Participants only had control over moving the object in the z direction. Participants indicated when they were satisfied with the object's position, after which the application would load another trial or conclude the experiment if all trials had been completed.

4.3 Experimental Design & Measures

Each experiment was designed as a full factorial within-participants study, with participants completing repeated experimental trials that corresponded to all possible combinations of depth cues. Independent variables consisted of a specific subset of the depth cues described in §3. Experiment 1 evaluated aerial perspective, drop shadows, billboarding, and shading; Experiment 2 evaluated dimensionality, texture, cast shadows, billboarding, and shading.

In addition to these fixed effects factors, the starting x and z location of the virtual object, the initial virtual object yaw rotation ψ, and the size of the virtual object were treated as random factors with values drawn from uniform distributions for each trial.

Two objective measures captured the primary outcomes of the experimental manipulations: *error*, the distance in *cm* between the center of the virtual shape and the physical target, and *completion time*, the time taken in seconds for users to align the virtual cube with the physical target. We primarily examined *signed error*, which helps gauge systematic depth under- or overestimation (underestimation: *signederror* < 0, overestimation: *signederror* > 0). Completion time served as a proxy for participants' abilities to rapidly estimate the position of a virtual object, a desirable property for AR applications.

In addition to these measures, we also recorded trial orderings (randomized per participant) and collected data on participant gender (previously shown to potentially affect AR depth estimation [31]) and use of corrective lenses. We analyzed this data for both experiments but found no evidence that our within-participants design induced biases due to trial ordering or transfer effects (e.g., learning or fatigue), nor did we find significant effects of gender or participant use of corrective lenses on depth judgments.

4.4 Experimental Procedure

Each study followed the same experimental protocol (lasting ~30-45 minutes/participant) consisting of six phases: (1) introduction, (2) screening, (3) calibration, (4) main task, (5) survey, and (6) debrief.

First, the experimenter greeted the participant, obtained informed consent, and gave the participant a brief overview of their task. In Phase 2, the experimenter administered a basic test for stereoblindness to screen for normal stereo vision. In Phase 3, the experimenter calibrated the augmented reality application by measuring participant height to align the virtual targets (used to measure participant error) with the real-world depth targets. The application was then initialized with a starting position relative to a fixed participant physical location, centered along a set horizontal calibration line drawn on the ground that was perpendicular to the targets and maintained a minimum distance between the participant and the first target. In Phase 4, participants completed a series of randomized trials, each requiring participants to carefully align the center of a virtual shape with the depth of a given real-world target. The physical targets consisted of white notecards, numbered 1 through 8, uniformly spaced along the z-axis as described in §4.2. Participants were instructed that they could move freely in horizontal space, but were not allowed to cross the fixed calibration line (c.f., Figure 2). While optional movement may provide for some added cues, we believe this contributes to the generalizability of our findings by reflecting the idea that ARHMDs are intended to provide dynamic, interactive experiences. Also, our within-participants design enabled participants to act as their own control, minimizing the potential impact of any additional participant movement on our results. After completing all trials, the experimenter collected demographic information (Phase 5) and debriefed participants (Phase 6), who were compensated with a $5.00 Amazon gift card. All experiments were conducted under the approval of the CU Boulder Institutional Review Board (IRB).

5 Experiment One

In our first experiment, we designed and conducted a $2 \times 2 \times 2 \times 3$ within-participants study evaluating aerial perspective (present or absent), simplified cast shadows (drop shadows present or absent), billboarding (present or absent), and three shading models (Lambertian, Blinn, or Phong, c.f. Figure 1).

[4] https://https://developer.microsoft.com/en-us/windows/holographic/hologram_stability

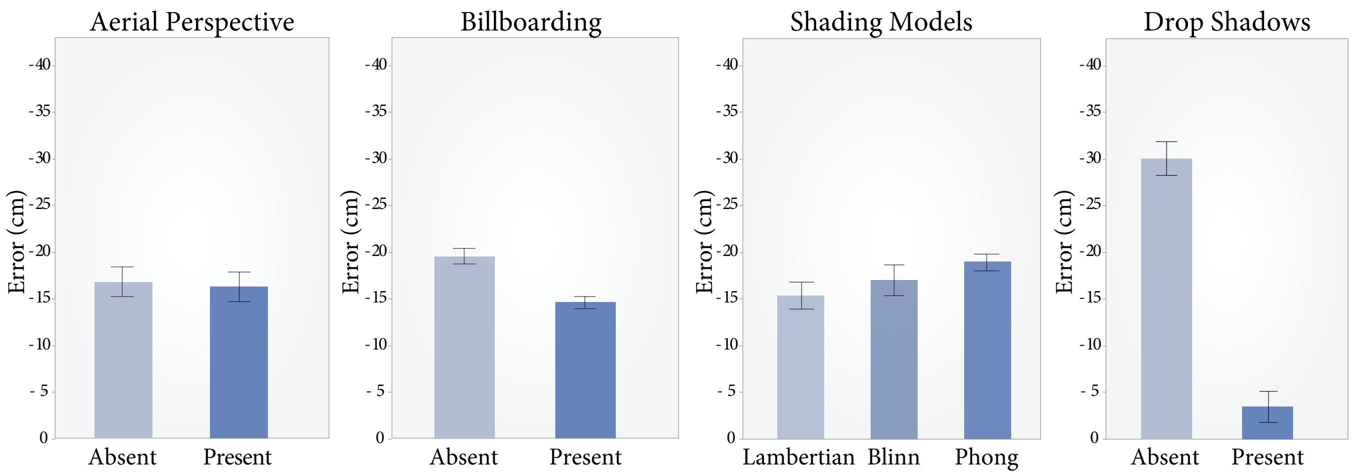


Figure 3: In Experiment 1, our analysis found support for H1 and partial support for H2, the notion that the design of a virtual object has direct ramifications on its perceived depth. Principally, drop shadows and billboarding increased participant depth judgment accuracy. Error bars in bar charts encode standard error.

5.1 Experimental Implementation

Our first experiment presented users with a single virtual cube that participants could move forward and backward using voice commands of "Forward" and "Backward" (each utterance moved the cube $10cm$) as suggested in Livingston et al. [43]. Once participants felt they had aligned the virtual cube with the real-world target, they used the voice command "Done" to confirm the virtual cube placement and start the next trial. Each trial presented participants with a virtual cube rendered using combinations of the four independent variables, implemented as follows:

Aerial perspective used the fog function from Cipiloglu et al. [12], which adjusts fragment colors as a function of the computed distance between the viewer (camera) and the virtual surface. While an exponential factor may mimic how atmospheric particles scatter light [54], we used a linear fog to mimic how increasing distance attenuates contrast in the real world, which provides reasonable results for small distances [61], such as those tested in our study (Figure 1B). Fog intensity was determined in piloting.

Cast shadows are not currently supported by Unity for see-through ARHMD applications. In Experiment 1, we approximated shadows using a simple drop shadow implementation, similar to the virtual object drop shadows tested in Berning et al. [5]. We rendered shadows as a semitransparent dark plane immediately below the virtual object, aligned with the real-world floor (c.f., Figure 1A). Shadow intensity was determined through piloting that compared real and virtual shadow intensity in the experimental environment.

Billboarding adjusts an object's orientation to consistently face a target, giving the illusion of dimensionality while reducing an object's complexity. Our application implemented a standard billboarding algorithm [23], rotating the cube along its center such that the flat face of the virtual cube always faced the participant (Figure 1D). This approach allowed us to remove virtual object volumetric information by presenting a single constant view of the cube while retaining all other rendering effects.

Shading models used one of three three implementations: Lambertian (diffuse reflection) [62], Blinn (a diffuse highlight) [7], and Phong (diffuse reflection with specular highlights) [55] (Figure 1C). Blinn and Phong can produce similar results depending on properties of the surface (e.g., spectral reflectance); however, Blinn models are more computationally efficient, a valuable property for AR applications. Shader parameters were computed using Maya defaults.

Each of these four techniques can be turned on or off independently (or, for shading, switched between one of the three shading models), resulting in 24 possible stimuli. We additionally treated object size ($15cm$ to $50cm$), initial x and z position ($\pm 40cm$ from the center of the room and $1.36m$ to $7.5m$ from the participant, respectively), and yaw orientation ψ ($0°$ to $90°$) as random factors. z position was randomized to remove bias for time taken to align the cube with targets (i.e., to prevent closer target alignment times from always being lower than farther targets) as in Swan et al. [69]. Ranges for these factors were determined through piloting, based on the field of view of the HoloLens. As visual perspective affects perceived object height as a function of distance to the observer, we treated the cube's initial y position as a constant factor, aligning it with each user's height (i.e., at $y = 0$ such that the cube appeared at eye level for each user). Similarly, cube pitch (θ) and roll (ϕ) rotations were treated as a constant factors ($\theta = 0, \phi = 0$). Participants viewed each of the 24 different stimuli four times–twice for near targets (1-4) and twice for far targets (5-8)–resulting in 96 trials per-participant. Each trial use a random cube size, initial x position, initial z position, and yaw rotation ψ.

5.2 Participants

A total of 24 participants (14 male, 10 female) took part in this experiment (7 wore corrective lenses while using the HoloLens). All participants were native English speakers recruited from CU Boulder. Participants were between 18 and 31 years of age ($M = 21.7$, $SD = 3.00$). On a seven-point scale, participants reported low prior familiarity with VR ($M = 2.38$, $SD = 1.69$) and AR ($M = 1.46$, $SD = 0.83$). Each participant completed 96 trials during the experiment, yielding a total of $2,304$ trials. Of these, 6 trials were excluded from data analysis due to speech recognition errors during the experiment. Our data analysis focused on the remaining $2,298$ trials.

5.3 Analysis

Our main analysis of error and time used a four-way repeated measures analysis of covariance (ANCOVA). This model treated the main experimental factors as fixed effects, included target as a covariate (to account for potential variance if depth judgment difficulty was affected by target distance), and participant ID as a random variable to account for variance caused by individual participant behaviors. Tukey's Honest Significant Difference (HSD) test with an $\alpha = .05$ controlled for Type I errors in planned comparisons across experimental conditions. We also performed regression analysis to understand relationships between factors considered random effects (cube size, horizontal position, and rotation) and depth alignment.

5.4 Results

H1 Depth Underestimation: We anticipated that participants would underestimate the depth of virtual objects in AR, as in prior experiments in purely virtual environments. Although it is not possible from our experimental design to evaluate this hypothesis using inferential statistics (there is no equivalent non-AR condition to compare against), we believe our signed error measurement provides initial support for this hypothesis. We found participant average error across all conditions to be $M = -16.42cm, 95\%CI = \pm 1.7cm$, meaning that the cube appeared further away than it actually was. This led participants to regularly place the virtual cube *in front* of the physical target, rather at the target. This systematic underestimation is consistent with prior findings in virtual and augmented reality.

H2 Main Effects: We predicted that virtual object design would influence depth judgments, specifically that cast shadows, shading models with specular highlights (Blinn and Phong), and aerial perspective would improve user depth accuracy, while billboarding would degrade depth perceptions. Our results partially supported H2: we found evidence that design decisions significantly affected depth perceptions. However, observed effects did not always conform to our predictions for each design technique (c.f., Figure 3).

We found a significant main effect of cast shadows on depth perceptions, $F(1, 2244) = 315.86$, $p < .0001$, with drop shadows increasing depth alignment accuracy by 90.06% (error without shadows: $M = -29.90cm \pm 2.44cm$; with shadows: $M = -2.99cm \pm 2.12cm$). Drop shadows also had a significant main effect on completion time, $F(1, 2244) = 23.02$, $p < .0001$, increasing user performance by 9.07% (time without: $M = 17.30s \pm 0.5s$, time with: $M = 15.73s \pm 0.5s$).

Billboarding also had a significant main effect on depth error, $F(1, 2244) = 12.99$, $p < .001$, but not completion time, $F(1, 2244) = .002$, $p = .967$. Interestingly, the direction of the effect of billboarding on error was contrary to our prediction from H2: the use of billboarding actually increased user depth perception accuracy (error without: $M = -19.14cm \pm 2.43cm$; error with: $M = -13.71cm \pm 2.39cm$). Overall, we found no significant main effect of shading model or aerial perspective on depth perception accuracy or completion time.

H3 Interactions: Cue integration theory suggests that using multiple techniques might improve depth judgments beyond any single technique. We examined our data for interaction effects from our main factors to evaluate this hypothesis and found three significant interaction effects. The first was between cast shadows and aerial perspective, $F(1, 2244) = 6.62$, $p = .010$, where the highest depth estimation accuracy was produced using both cast shadows and aerial perspective ($M = -0.79cm$), while the lowest accuracy was found with aerial perspective and no shadows ($M = -31.62cm$). Tukey's HSD found a significant difference between these conditions indicating that aerial perspective could enhance depth judgments when used in conjunction with drop shadows, but was not helpful by itself. We also found a significant interaction effect between drop shadows and billboarding, $F(1, 2244) = 7.20$, $p = .007$, with the highest depth estimation accuracy when using both shadows and billboarding ($M = -2.27cm$), while the lowest accuracies were found when both cues were absent ($M = 34.67cm$). Finally, we observed a three-way interaction effect between cast shadows, billboarding, and shading model, $F(1, 2244) = 3.68$, $p = .025$, with highest accuracy using cast shadows, billboarding, and Lambertian shading ($M = -.55cm$) and worst when shadows and billboarding were absent while using Blinn shading ($M = -36.10$). This finding is interesting to note as it runs contrary to our expectation that specular highlights would improve depth accuracy.

Regression Analysis: While we randomized cube size, horizontal position, and rotation to focus on design properties of a virtual object,

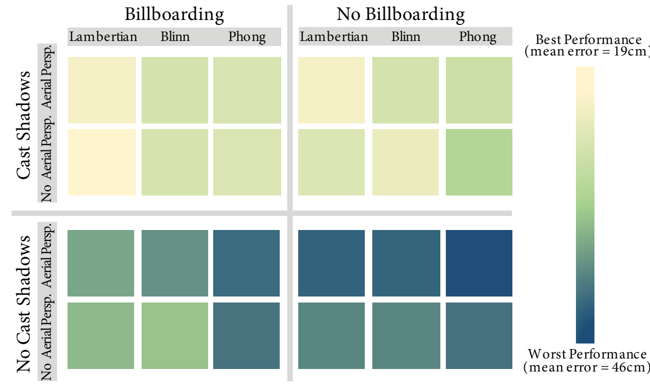


Figure 4: A visual summary of cue interaction on user depth estimation performance across all conditions in Experiment 1; lighter colors correspond to improved performance among various cue combinations.

we still wanted to understand if these attributes might influence the perceived spatial position of an object. We performed a regression analysis comparing performance across these factors to explore this interaction. We found a significant linear relationship between error and size: $depth error = -39.59 + 71.30 \times cubesize$; $F(1,2296) = 72.21$, $R^2 = 0.03$, $p < .0001$. As the virtual object got larger, participant performance improved. Regression analysis did not yield a significant relationship between cube horizontal position or rotation and depth estimation accuracy.

5.5 Discussion

Cast shadows proved to be the most important design decision for improving spatial perceptions in Experiment 1. This result runs contrary to prior findings (e.g., [26, 65]) that suggest cast shadows may be of only limited importance in stereo scenes. Instead, our findings reveal that shadows may play a crucial role in spatial perceptions between real and virtual objects. Shadow mapping algorithms can be computationally expensive, especially if they involve close simulations of real-world lighting; however, the strength of this effect suggests that better understanding the role of cast shadows for spatial perceptions in AR may greatly improve the perceptual effectiveness of augmented reality applications. Experiment 1 used a simple drop shadow approach, mimicking that found in Berning et al. [5], which only loosely approximated the actual lighting in the experimental environment. To better understand the perceptual effects of cast shadows in AR, we implemented a more robust cast shadow algorithm using ray tracing for Experiment 2.

Billboarding had an unexpected positive effect on depth alignment, even though billboarding removed spatial cues provided by the object's 3D structure since viewers only saw the front face of the cube. We believe there are three possible explanations for the observed effect: (a) removing such cues from the scene reduced the amount of (potentially redundant or conflicting) information that viewers have to cognitively process to make depth judgments, (b) viewers are more familiar with two-dimensional virtual objects, which are commonly seen on traditional monitors, and thus are better able to place virtual objects with the same visual properties, and/or (c) billboarding provided a consistent visual appearance from different perspectives that may act as a reference point when performing depth alignment tasks. Understanding factors that affect the perceived depth of billboards is critical for effectively designing interactive AR applications as the technique may commonly be used for interactive widgets such as menus, projections, and labels. The positive effect of billboarding on depth estimation indicates that further studies are needed to better understand spatial perceptions of two-dimensional objects for AR environments in practice, which led us to explore both billboarding and dimensionality in Experiment 2.

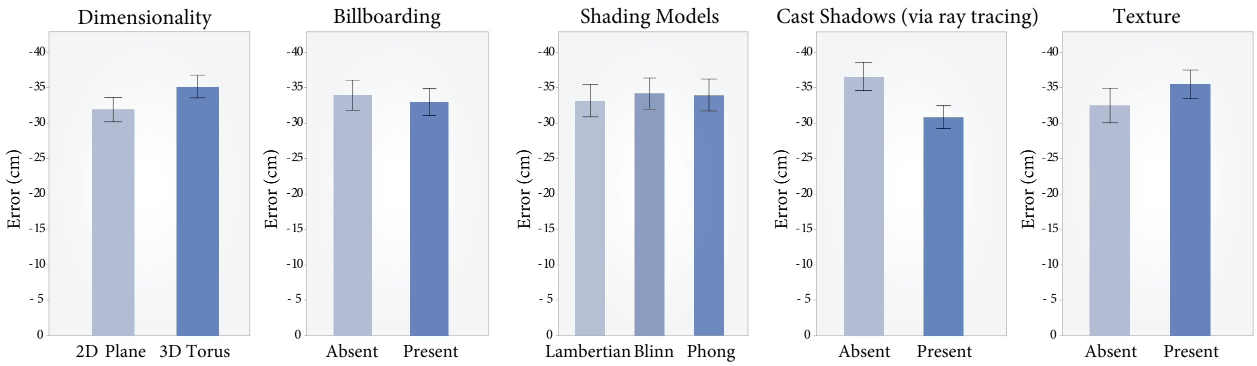


Figure 5: In Experiment 2, cast shadows rendered via ray tracing provided the greatest increase in participant abilities to align virtual objects to physical targets. Error bars encode 95% confidence intervals.

We found no significant main effects of aerial perspective or shading, although they did interact with the other factors. This result suggests that perceptions of the design techniques we studied might combine in different ways. While methods of cue combination for strict depth cues are still unknown (see [12, 47, 53] for discussions), our findings suggest that depth cues from graphics techniques combine in a complex manner, with certain combinations improving user spatial perceptions (Figure 4).

6 Experiment Two

Our first experiment provided support for H1 and mixed support for H2 and H3, with certain cues aiding depth perceptions and others only contributing via interaction effects. To further explore the relationship between virtual object design and perceived depth, we conducted a second experiment in the same environment considering more realistic shadows, dimensionality, billboarding, shading, and texture to further our understandings of how various design decisions might affect user depth judgments when using an ARHMD.

6.1 Experimental Conditions & Implementation

To provide more precise alignment capabilities, participants used a wireless Xbox One controller as an input device instead of verbal commands for the perceptual matching task. Participants positioned the object along the z axis using a joystick and confirmed the position using the 'A' button, which provided continuous control over the object's position. Object size, starting z position, and yaw rotation ψ were randomized as in Experiment 1.

This experiment took the form of a $2 \times 2 \times 2 \times 2 \times 3$ within-participants study evaluating dimensionality (2D plane or 3D object), texture (present or absent), ray-traced cast shadows (present or absent), billboarding (present or absent), and shading (Lambertian, Blinn, or Phong) leading to 48 distinct stimuli. Billboarding and shading models were implemented identically to Experiment 1. We introduced dimensionality as a design factor (2D versus 3D shape) to further examine the billboarding effect found in Experiment 1. Dimensionality also allowed us to approximate perceived positioning in virtual menu systems, generally rendered as 2D planes that may optionally be billboarded to consistently face the user. We used two shapes to explore dimensionality: a 2D plane and a 3D torus. These shapes allowed us to test increased geometric complexity (torus) and a closer approximation of a menu system (plane).

Our first study suggested that cast shadows provide critical cues for positioning virtual objects in the real world, but used only a basic drop shadow approximation. In Experiment 2, we increased the fidelity of our shadow mapping by using ray tracing to identify the shape, size, and position of the object's shadows based on the position of the light sources in the room. We then projected a semi-transparent dark plane onto the floor at the corresponding locations computed from each of the overhead light sources. Shadow opacity was directly proportional to the magnitude of the vector between the light source and location of the furthest point of the shadow, with the opacity parameters determined in piloting. This approach closely approximated the default cast shadow algorithms used in Unity, which we could not use directly since objects must be opaque to receive shadows whereas our virtual ground plane in the tested scenes was transparent.

Experiment 1 revealed some benefit to aerial perspective when used in conjunction with shadows. However, the fog function exaggerated real-world effects and did not appear to be a strong cue on its own. In our second study, we removed aerial perspective to instead explore a more realistic near-field object-based depth cue: texture density. Shapes were either a solid color (texture absent) or mapped with a checkerboard texture.

Although we removed aerial perspective, adding both dimensionality and texture increased the complexity of our factorial design. To account for the larger number of conditions (48 distinct stimuli) while keeping the experiment of a similar length to first experiment, each combination of conditions was tested twice (once each with a randomly selected near/far target), rather than four times as in Experiment 1. This sampling resulted in 96 total trials/participant.

6.2 Participants

We recruited 24 participants (14 male, 9 female, 1 prefer not to identify) for this experiment, all native-English speakers from CU Boulder. Seven participants wore corrective lenses during the study. Participant age ranged from 18 to 33 ($M = 23.0$, $SD = 4.12$). On a seven-point scale participants reported low prior familiarity with VR ($M = 2.92$, $SD = 1.86$) and AR ($M = 2.13$, $SD = 1.30$). The wireless controller disconnected for one participant, leaving her unable to complete the experiment. We excluded her partial data from analysis, leaving full data from 23 participants. Each participant completed two trials per stimuli (96 total per participant), yielding 2,208 trials. We examined this dataset for misclicks (trials with response time under $2s$ and error greater than $100cm$) and excluded one trial for this reason. We analyzed the remaining 2,207 trials.

6.3 Results

Our analysis procedure was identical to that in Experiment 1.

H1 Depth Underestimation: Consistent with Experiment 1, we found that participants systematically underestimated virtual object depth (average error across all conditions $M = -31.33cm, 95\%CI = \pm 1.22cm$).

H2 Main Effects: We again found partial support for our hypothesis of how various design decisions might impact user depth judgments. As in Experiment 1, we found a significant main effect of cast shadows on depth judgment error, $F(1, 2130) = 30.21$, $p < .0001$, with cast shadows increasing depth perception accuracy by 17.96% (error without cast shadows: $M = -34.42cm \pm 1.59cm$; with cast shadows: $M = -28.24cm \pm 1.93cm$). We also found a significant main effect of cast shadows on completion time, $F(1, 2130) = 8.19$, $p = .004$, with cast shadows improving response time by 6.53% (time without: $M = 15.16s \pm 0.51s$, time with: $M = 14.17s \pm 0.53s$).

We did not find a significant effect of dimensionality on depth perception error, but did find a marginal effect on completion time, $F(1, 2130) = 3.29$, $p = .070$, with the effect opposite that predicted by H2: users aligned the 2D plane ($M = 14.38s \pm 0.39s$) slightly faster than the 3D torus ($M = 14.95s \pm 0.55s$). We did not find significant or marginal main effects of billboarding, texture, or shading on depth error or time.

H3 Interactions: We found two significant interaction effects when examining cue combinations. First, we found a two-way effect of dimensionality and cast shadows, $F(1, 2130) = 4.41$, $p = .036$, where cast shadows improved depth judgments for the 3D torus (mean improvement with shadows $M = 8.63cm$) to a greater extant than that of the 2D plane (mean improvement with shadows $M = 3.86cm$). Second, we found a four-way interaction effect of cast shadows, billboarding, texture, and shading, $F(1, 2130) = 3.09$, $p = .046$, with most accurate depth judgments of stimuli with cast shadows and billboarding present, texture absent, and Phong shading ($M = -24.49cm$) and least accurate judgments with cast shadows absent, billboarding and texture present, and Blinn shading ($M = -37.76cm$). We also found two marginal interaction effects on response time, with a marginal effect of billboarding and shading on time, $F(1, 2130) = 2.49$, $p = .083$, where billboarding led to faster response times with Lambertian shading but slower response times with Blinn and Phong shading, and a marginal effect of dimensionality, billboarding, and shading on time, $F(1, 2130) = 2.43$, $p = .088$, with fastest response times from a 2D plane without billboarding using Phong shading ($M = 13.37s$) and slowest response times from a 3D torus with billboarding and a Phong shader ($M = 15.91s$). Together, this reveals mixed results for H2 and H3: shaders with specular highlights improved completion time for a 2D shape over a billboarded 3D shape, but did not help across 2D billboards and 3D billboards.

Regression Analysis: As in Experiment 1, we found a significant linear relationship between depth error and virtual object size, where depth estimation accuracy improved with larger virtual objects: $deptherror = -46.44 + 54.53 \times cubesize$; $F(1,2205) = 39.69$, $R^2 = 0.02$, $p < .0001$. Regression analysis did not yield a significant relationship between object rotation and depth estimation accuracy.

6.4 Discussion

We again found evidence of systematic depth underestimation overall (H1) and results that support the use of cast shadows in improving depth judgments (H2), although we found that the effect size of ray traced shadows was less than that of the simple drop shadows from Experiment 1. Three possible explanations for this difference are: (1) ray traced shadows better matched the real scene lighting compared to the simple drop shadows from Experiment 1, but may have been less salient as intensity was proportional to distance from the light source, (2) ray tracing could lead to an object casting multiple shadows (due to multiple light sources), which may be more cognitively demanding to interpret than a single drop shadow, (3) drop shadows always fell immediately beneath the target whereas ray traced shadows were displaced from the target by lighting angles, which may also be more challenging to interpret.

We also identified some surprising findings that were opposite H2 predictions, such as that users were slightly faster placing 2D objects than 3D objects while maintaining equivalent accuracies. This finding, along with the unanticipated billboarding results from Experiment 1, indicates users will readily perceive the location of AR menu systems and other 2D objects, possibly due to prior familiarity with 2D objects and/or removal of conflicting cues. However, it also suggests that more study is needed to better understand effects of object dimensionality in AR. We also found no significant effects of texture, replicating previous findings where texture was less effective relative to other cues [50].

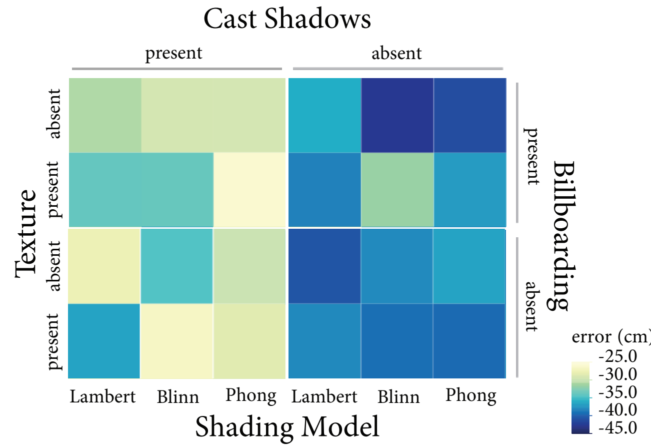


Figure 6: A visual summary of cue interaction on user depth estimation performance across four conditions in Experiment 2; lighter colors correspond to improved performance among various cue combinations.

As in Experiment 1, we found a variety of interaction effects that overall support the notion that cue integration affects user depth estimation (Figure 6), although it appears to be a complex phenomena that does not always significantly improve depth judgment accuracy. In particular, we found possible conflicts between cues of dimensionality, billboarding, and shading model, indicating that more study is needed to understand the relative importance of these designs from the perspective of cue integration theory.

7 Synthesis and Conclusions

Our experiments measured spatial perceptions in augmented reality as a function of virtual object design. In these studies, we found:

1. On average, viewers underestimated virtual object depth (H1).
2. Certain design decisions regarding virtual object rendering directly influenced its perceived spatial position (H2). In particular, cast shadows appear highly beneficial for depth estimation, improving accuracy by 90% in Experiment 1 and 18% in Experiment 2.
3. Design combinations interacted in complex ways, potentially improving the perceived depth of virtual objects in the real world over individual design choices (H3).

Across all experiments, cast shadows were an important cue for aligning virtual and physical objects. While both drop shadows and cast shadows via ray tracing improved depth judgments, open-ended comments suggested that participants directly understood the utility of drop shadows (e.g., *Exp1 P78* stated: "Placing the cube became a lot harder when there were no shadows"), but did not perceive ray traced shadows as useful (*Exp2 P120*: "Shadows were difficult to use as gauge of depth"). Participants expressed a preference for drop shadows (which also had a larger effect size), even though cast shadows from ray tracing provided a closer approximation of the real world. This contrast between preference and simulation fidelity supports prior findings in visual perception that suggest people are affectively tolerant of imperfections in shadows [30, 60] and may help designers carefully reason about usability trade-offs among performance, perception, and realism.

While cast shadows provided a strong beneficial cue, as predicted by H2, other cues such as shading, aerial perspective, and texture did not appear to have a significant impact on their own. However, these cues did exhibit interaction effects that could improve perceptions when used in combination with multiple secondary cues and/or cast shadows, providing preliminary support for H3. These interaction effects were often complex and did not always align with our predictions, providing opportunities for more study.

We hypothesize that the primacy of cast shadows indicates the importance of virtual objects acting on the physical world for precise positioning. Purely virtual cues (texture & billboarding) and physical-to-virtual cues (aerial perspective & shading model) primarily helped when used in conjunction with virtual-to-physical interactions (drop and cast shadows). These results may offer insight into other important considerations for perceived placement of virtual objects. For instance, we speculate that prioritizing other virtual-to-physical interactions, such as caustics (e.g., [32]), occlusion, and physicality (e.g., weight-based deformations) may also assist rapid and precise localization of virtual objects in the real world in the same manner we observed with cast shadows.

Recent advances in consumer-ready augmented reality head-mounted display technologies is creating an exciting design space for new AR applications. While ARHMDs seem well poised to enhance human efforts in a variety of consumer and industry domains, we still lack fundamental knowledge regarding how various design decisions will impact AR application effectiveness. This research explored how common graphics techniques affect user perception of virtual object depth in AR across two experiments in a realistic test environment. Our results point to a specific design implication for improving user depth perceptions of virtual objects in AR: cast shadows can significantly reduce depth estimation time while improving accuracy. We found simple drop shadows and robust ray traced shadows to be a highly effective cue for both traditional 3D objects as well as 2D or billboarded representations, suggesting cast shadows should also be used in designing menu systems or overlays. In addition, we found several interaction effects where cue layering produced complex results, with certain interactions increasing depth judgment performance beyond that of single cues, pointing to the need for more research in this complicated perceptual space. While our experiment only evaluated a subset of possible graphics techniques, we believe that our experimental framework is well-suited to additional explorations within this space. By further quantifying the relationship between design and perception, we can explore tailoring designs to optimize perceptual effectiveness, as in Cipiloglu at al. [12], which presents an algorithm for automatically prioritizing various depth cues. Overall, our findings demonstrate the importance of considering design trade-offs for AR applications and offer hardware-independent findings that can readily improve human interactions with AR in practice.

8 Acknowledgments

This work was supported by an IGP grant from the University of Colorado Boulder and an Early Career Faculty grant from NASA's Space Technology Research Grants Program under award NNX16AR58G.

References

[1] W. J. Adams and J. H. Elder. Effects of specular highlights on perceived surface convexity. *PLoS Computational Biology*, 10(5):e1003576, 2014.

[2] M. Al-Kalbani, I. Williams, and M. Frutos-Pascual. Analysis of Medium Wrap Freehand Virtual Object Grasping in Exocentric Mixed Reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2016*, pages 84–93, 2016.

[3] M. Al-Kalbani, I. Williams, and M. Frutos-Pascual. Improving Freehand Placement for Grasping Virtual Objects via Dual View Visual Feedback in Mixed Reality. In *ACM Conference on Virtual Reality Software and Technology*, pages 279–282, 2016.

[4] K. Berbaum, D. Tharp, and K. Mroczek. Depth perception of surfaces in pictures: Looking for conventions of depiction in pandora's box. *Perception*, 12(1):5–20, 1983.

[5] M. Berning, D. Kleinert, T. Riedel, and M. Beigl. A study of depth perception in hand-held augmented reality using autostereoscopic displays. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2014*, pages 93–98. IEEE, 2014.

[6] C. Bichlmeier, T. Sielhorst, S. M. Heining, and N. Navab. Improving depth perception in medical ar. In *Bildverarbeitung für die Medizin 2007*, pages 217–221. Springer, 2007.

[7] J. F. Blinn. Models of Light Reflection for Computer Synthesized Pictures. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques 1977*, volume 11, pages 192–198. ACM, 1977.

[8] M. F. Bradshaw, A. Glennerster, and B. J. Rogers. The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, 36(9):1255–1264, 1996.

[9] H. H. Bülthoff and H. A. Mallot. Integration of depth modules: stereo and shading. *Josa a*, 5(10):1749–1758, 1988.

[10] H.-L. Chi, S.-C. Kang, and X. Wang. Research Trends and Opportunities of Augmented Reality Applications in Architecture, Engineering, and Construction. *Automation in Construction*, 33:116–122, 2013.

[11] M. A. Cidota, R. M. Clifford, S. G. Lukosch, and M. Billinghurst. Using Visual Effects to Facilitate Depth Perception for Spatial Tasks in Virtual and Augmented Reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2016*, pages 172–177, 2016.

[12] Z. Cipiloglu, A. Bulbul, and T. Capin. A framework for enhancing depth perception in computer graphics. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 141–148. ACM, 2010.

[13] J. Cutting and P. Vishton. Perception of space and motion. handbook of perception and cognition, 1995.

[14] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr. Tablet versus Phone: Depth Perception in Handheld Augmented Reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2012*, pages 187–196, 2012.

[15] A. Dey and C. Sandor. Lessons learned: Evaluating Visualizations for Occluded Objects in Handheld Augmented Reality. *International Journal of Human-Computer Studies*, 72(10):704–716, 2014.

[16] 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. *The Visual Computer*, pages 1–15, 2015.

[17] D. Drascic and P. Milgram. Perceptual Issues in Augmented Reality. In *Electronic Imaging: Science & Technology*, pages 123–134. International Society for Optics and Photonics, 1996.

[18] A. Dünser, R. Grasset, and M. Billinghurst. *A Survey of Evaluation Techniques used in Augmented Reality Studies*. Human Interface Technology Laboratory New Zealand, 2008.

[19] S. R. Ellis and U. J. Bucher. Distance perception of stereoscopically presented virtual objects optically superimposed on physical objects by a head-mounted see-through display. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 38, pages 1300–1304. SAGE Publications, 1994.

[20] S. R. Ellis, U. J. Bucher, and B. M. Menges. The Relationship of Binocular Convergence and Errors in Judged Distance to Virtual Objects. In *Proceedings of the International Federation of Automatic Control*, volume 95, pages 297–301, 1995.

[21] S. R. Ellis and B. M. Menges. Localization of Virtual Objects in the Near Visual Field. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):415–431, 1998.

[22] H. Fang, S. Ong, and A. Nee. Interactive Robot Trajectory Planning and Simulation using Augmented Reality. *Robotics and Computer-Integrated Manufacturing*, 28(2):227–237, 2012.

[23] A. R. Fernandes. Billboarding Tutorial, 2001. `http://lighthouse3d.com/opengl/billboarding/billboardingtut.pdf`.

[24] C. Furmanski, R. Azuma, and M. Daily. Augmented-reality Visualizations Guided by Cognition: Perceptual Heuristics for Combining Visible and Obscured Information. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2002*, pages 215–320. IEEE, 2002.

[25] B. Gillam. The perception of spatial layout from static optical information. *Perception of space and motion*, pages 23–67, 1995.
[26] T. Hartzell, M. Thompson, and B. Kim. The influence of graphics effects on perceiving depth in minimized virtual environments. In *Computer Applications for Communication, Networking, and Digital Contents*, pages 235–242. Springer, 2012.
[27] M. Hedrich, M. Bloj, and A. I. Ruppertsberg. Color constancy improves for real 3d objects. *Journal of Vision*, 9(4):16–16, 2009.
[28] M. Hou. User experience with alignment of real and virtual objects in a stereoscopic augmented reality interface. In *Proceedings of the 2001 Conference of the Centre for Advanced Studies on Collaborative Research*, page 6. IBM Press, 2001.
[29] I. P. Howard and B. J. Rogers. Seeing in depth, volume 2: Depth perception. *Ontario, Canada: I. Porteous*, 2002.
[30] J. Jacobson and S. Werner. Why Cast Shadows are Expendable: Insensitivity of Human Observers and the Inherent Ambiguity of Cast Shadows in Pictorial Art. *Perception*, 33(11):1369–1383, 2004.
[31] C. Jerome and B. Witmer. The Perception and Estimation of Egocentric Distance in Real and Augmented Reality Environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 2249–2252, 2005.
[32] P. Kán and H. Kaufmann. High-quality Reflections, Refractions, and Caustics in Augmented Reality and their Contribution to Visual Coherence. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2012*, pages 99–108. IEEE, 2012.
[33] D. Kersten. Inverse 3-d graphics: A metaphor for visual perception. *Behavior Research Methods, Instruments, & Computers*, 29(1):37–46, 1997.
[34] M. Kersten, J. Stewart, N. Troje, and R. Ellis. Enhancing depth perception in translucent volumes. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006.
[35] S. E. H. Kirkley Jr. Augmented Reality Performance Assessment Battery (ARPAB): Object Recognition, Distance Estimation and Size Estimation using Optical See-through Head-worn Displays. 2003.
[36] C. Koch, M. Neges, M. König, and M. Abramovici. Natural Markers for Augmented Reality-based Indoor Navigation and Facility Maintenance. *Automation in Construction*, 48:18–30, 2014.
[37] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual Issues in Augmented Reality Revisited. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2010*, pages 3–12. IEEE, 2010.
[38] D. R. Lampton, D. P. McDonald, M. Singer, and J. P. Bliss. Distance Estimation in Virtual Environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meetings*, volume 39, pages 1268–1272, 1995.
[39] M. S. Langer and H. H. Bülthoff. A prior for global convexity in local shape-from-shading. *Perception*, 30(4):403–410, 2001.
[40] F. Lindemann and T. Ropinski. About the influence of illumination models on image comprehension in direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1922–1931, 2011.
[41] M. A. Livingston, Z. Ai, K. Karsch, and G. O. Gibson. User Interface Design for Military AR Applications. *Virtual Reality*, 15(2-3):175–184, 2011.
[42] M. A. Livingston, Z. Ai, J. E. Swan, and H. S. Smallman. Indoor vs. Outdoor Depth Perception for Mobile Augmented Reality. In *IEEE Virtual Reality Conference*, pages 55–62, 2009.
[43] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas. Pursuit of "X-ray vision" for Augmented Reality. In *Human Factors in Augmented Reality Environments*, pages 67–107. 2013.
[44] M. A. Livingston, J. E. Swan II, J. L. Gabbard, T. H. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving Multiple Occluded Layers in Augmented Reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2016*, page 56. IEEE, 2003.
[45] J. M. Loomis and J. M. Knapp. Visual Perception of Egocentric Distance in Real and Virtual Environments. *Virtual and Adaptive Environments*, 11:21–46, 2003.
[46] P. Mamassian, D. C. Knill, and D. Kersten. The perception of cast shadows. *Trends in Cognitive Sciences*, 2(8):288–295, 1998.
[47] G. Mather and D. R. Smith. Combining depth cues: effects upon accuracy and speed of performance in a depth-ordering task. *Vision Research*, 44(6):557–562, 2004.
[48] J. W. McCandless, S. R. Ellis, and B. D. Adelstein. Localization of a Time-delayed, Monocular Virtual Object Superimposed on a Real Environment. *Presence*, 9(1):15–24, 2000.
[49] D. Medeiros, M. Sousa, D. Mendes, A. Raposo, and J. Jorge. Perceiving Depth: Optical Versus Video See-Through. In *ACM Conference on Virtual Reality Software and Technology*, pages 237–240, 2016.
[50] M. Mikkola, A. Boev, and A. Gotchev. Relative importance of depth cues on portable autostereoscopic display. In *Proceedings of the 3rd Workshop on Mobile Video Delivery*, pages 63–68. ACM, 2010.
[51] A. Nee, S. Ong, G. Chryssolouris, and D. Mourtzis. Augmented Reality Applications in Design and Manufacturing. *CIRP Annals-Manufacturing Technology*, 61(2):657–679, 2012.
[52] J. F. Norman, J. T. Todd, and G. A. Orban. Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science*, 15(8):565–570, 2004.
[53] I. Oruç, L. T. Maloney, and M. S. Landy. Weighted linear cue combination with possibly correlated error. *Vision Research*, 43(23):2451–2468, 2003.
[54] R. P. O'Shea, S. G. Blackburn, and H. Ono. Contrast as a Depth Cue. *Vision Research*, 34(12):1595–1604, 1994.
[55] B. T. Phong. Illumination for Computer Generated Pictures. *Communications of the ACM*, 18(6):311–317, 1975.
[56] A. G. Rempel, W. Heidrich, and R. Mantiuk. The role of contrast in the perceived depth of monocular imagery. In *Proceedings of the Symposium on Applied Perception in Graphics and Visualization*, pages 115–115. ACM, 2011.
[57] J. P. Rolland, W. Gibson, and D. Ariely. Towards Quantifying Depth and Size Perception in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 4(1):24–49, 1995.
[58] 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: Teleoperators and Virtual Environments*, 11(6):610–625, 2002.
[59] M. E. C. Santos, I. de Souza Almeida, G. Yamamoto, T. Taketomi, C. Sandor, and H. Kato. Exploring Legibility of Augmented Reality X-ray. *Multimedia Tools and Applications*, 75(16):9563–9585, 2016.
[60] M. Sattler, R. Sarlette, T. Mücken, and R. Klein. Exploitation of Human Shadow Perception for Fast Shadow Rendering. In *Proceedings of the Symposium on Applied Perception in Graphics and Visualization*, pages 131–134. ACM, 2005.
[61] M. Segal and K. Akeley. The OpenGL Graphics Interface. *California: Silicon Graphics Systems*, 1994.
[62] P. Shirley, M. Ashikhmin, and S. Marschner. *Fundamentals of Computer Graphics*. CRC Press, 2015.
[63] G. Singh, J. E. Swan II, J. A. Jones, and S. R. Ellis. Depth Judgment Measures and Occluding Surfaces in Near-field Augmented Reality. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 149–156. ACM, 2010.
[64] M. Stone, D. A. Szafir, and V. Setlur. An engineering model for color difference as a function of size. In *Color and Imaging Conference*, volume 2014, pages 253–258. Society for Imaging Science and Technology, 2014.
[65] N. Sugano, H. Kato, and K. Tachibana. The Effects of Shadow Representation of Virtual Objects in Augmented Reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2003*, pages 76–83. IEEE, 2003.
[66] R. T. Surdick, E. T. Davis, R. A. King, and L. F. Hodges. The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence: Teleoperators and Virtual Environments*, 6(5):513–531, 1997.
[67] I. E. Sutherland. The Ultimate Display. *Proceedings of the International Information Security and Privacy Conference Congress*, 65(2):506–508, 1965.
[68] J. E. Swan and J. L. Gabbard. Survey of User-based Experimentation in Augmented Reality. In *Proceedings of the International Conference*

on Virtual Rehabilitation 2005, pages 1–9, 2005.

[69] 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.

[70] 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*, 21(11):1289–1298, 2015.

[71] J. E. Swan 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*, pages 1–16, 2017.

[72] W. Thompson, R. Fleming, S. Creem-Regehr, and J. K. Stefanucci. *Visual perception from a computer graphics perspective*. CRC Press, 2011.

[73] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the Quality of the Computer Graphics Matter when Judging Distances in Visually Immersive Environments? *Presence*, 13(5):560–571, 2004.

[74] J. T. Todd, J. F. Norman, J. J. Koenderink, and A. M. Kappers. Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception*, 26(7):807–822, 1997.

[75] K. Uratani, T. Machida, K. Kiyokawa, and H. Takemura. A study of depth visualization techniques for virtual annotations in augmented reality. In *Virtual Reality*, pages 295–296. IEEE, 2005.

[76] J. Wang, H. Suenaga, K. Hoshi, L. Yang, E. Kobayashi, I. Sakuma, and H. Liao. Augmented Reality Navigation with Automatic Marker-free Image Registration using 3-D Image Overlay for Dental Surgery. *IEEE Transactions on Biomedical Engineering*, 61(4):1295–1304, 2014.

[77] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The Effects of Head-mounted Display Mechanics on Distance Judgments in Virtual Environments. In *Proceedings of the Symposium for Applied Perception in Graphics and Visualization*, pages 35–38, New York, NY, USA, 2004. ACM.

[78] B. G. Witmer and P. B. Kline. Judging Perceived and Traversed Distance in Virtual Environments. *Presence*, 7(2):144–167, 1998.

[79] B. G. Witmer and W. J. Sadowski. Nonvisually Guided Locomotion to a Previously Viewed Target in Real and Virtual Environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):478–488, 1998.

[80] B. Wu, T. L. Ooi, and Z. J. He. Perceiving Distance Accurately by a Directional Process of Integrating Ground Information. *Nature*, 428(6978):73–77, 2004.

[81] A. Yonas, L. T. Goldsmith, and J. L. Hallstrom. Development of sensitivity to information provided by cast shadows in pictures. *Perception*, 7(3):333–341, 1978.

[82] L. Zheng, Y. Wu, and K.-L. Ma. Perceptually-based depth-ordering enhancement for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):446–459, 2013.

[83] S. Zollmann, C. Hoppe, T. Langlotz, and G. Reitmayr. FlyAR: Augmented Reality Supported Micro Aerial Vehicle Navigation. *IEEE Transactions on Visualization and Computer Graphics*, 20(4):560–568, 2014.