Egocentric Depth Judgments in Optical, See-Through Augmented Reality

J. Edward Swan II, *Member*, *IEEE*, Adam Jones, Eric Kolstad,
Mark A. Livingston, *Member*, *IEEE*, and Harvey S. Smallman

Abstract—A fundamental problem in optical, see-through augmented reality (AR) is characterizing how it affects the perception of spatial layout and depth. This problem is important because AR system developers need to both place graphics in arbitrary spatial relationships with real-world objects, and to know that users will perceive them in the same relationships. Furthermore, AR makes possible enhanced perceptual techniques that have no real-world equivalent, such as *x-ray vision*, where AR users are supposed to perceive graphics as being located behind opaque surfaces. This paper reviews and discusses protocols for measuring egocentric depth judgments in both virtual and augmented environments, and discusses the well-known problem of depth underestimation in virtual environments. It then describes two experiments that measured egocentric depth judgments in AR. Experiment I used a perceptual matching protocol to measure AR depth judgments at medium and far-field distances of 5 to 45 meters. The experiment studied the effects of upper versus lower visual field location, the x-ray vision condition, and practice on the task. The experimental findings include evidence for a switch in bias, from underestimating to overestimating the distance of AR-presented graphics, at $\sim$ 23 meters, as well as a quantification of how much more difficult the x-ray vision condition makes the task. Experiment II used blind walking and verbal report protocols to measure AR depth judgments at distances of 3 to 7 meters. The experiment examined real-world objects, real-world objects seen through the AR display, virtual objects, and combined real and virtual objects. The results give evidence that the egocentric depth of AR objects is underestimated at these distances, but to a lesser degree than has previously been found for most virtual reality environments. The results are consistent with previous studies that have implicated a restricted field-of-view, combined with an inability for observers to scan the ground plane in a near-to-far direction, as explanations for the observed depth underestimation.

Index Terms—Artificial, augmented, and virtual realities, ergonomics, evaluation/methodology, screen design, experimentation, measurement, performance, depth perception, optical see-through augmented reality.

1 Introduction

Optical, see-through augmented reality (AR) is the variant of AR where graphics are superimposed on a user's view of the real world with optical, as opposed to video, combiners. Because optical, see-through AR (simply referred to as "AR" for the rest of this paper) provides direct, heads-up access to information that is correlated with a user's view of the real world, it has the potential to revolutionize the way many tasks are performed. In addition, AR makes possible enhanced perceptual techniques that have no real-world equivalent. One such technique is *x-ray vision*, where the intent is for AR users to accurately perceive objects which are located behind opaque surfaces.

The AR community is applying AR technology to a number of unique and useful applications [1]. The application that motivated the work described here is mobile, outdoor AR for situational awareness in urban settings (the Battlefield Augmented Reality System (BARS) [19]). This is a very difficult application domain for AR; the biggest challenges are outdoor tracking and registration, outdoor display hardware, and developing appropriate AR display and interaction techniques.

In this paper, we focus on AR display techniques, in particular, how to correctly display and accurately convey depth. This is a hard problem for several reasons. Current head-mounted displays are compromised in their ability to display depth, because they often dictate a fixed accommodative focal depth, and they restrict the field of view. Furthermore, it is well known that distances are consistently underestimated in VR scenes depicted in head-mounted displays [5], [16], [21], [23], [34], [36], but the reasons for this phenomenon are not yet clear. In addition, unlike virtual reality, with AR users see the real world, and therefore graphics need to appear to be at the same depth as colocated real-world objects, even though the graphics are physically drawn directly in front of the eyes. Furthermore, there is no real-world equivalent to x-ray vision, and it is not yet understood how the human visual system reacts to information displayed with purposely conflicting depth cues, where the depth conflict itself communicates useful information.

- *J.E. Swan II, A. Jones, and E. Kolstad are with the Department of Computer Science and Engineering and the Institute for Neurocognitive Science and Technology, Mississippi State University, 300 Butler Hall, PO Box 9637, Mississippi State, MS 39762. E-mail: swan@acm.org.*
- *M.A. Livingston is with the 3D Virtual and Mixed Environments Laboratory, Code 5580, Naval Research Laboratory, Washington DC, 20375. E-mail: markl@ait.nrl.navy.mil.*
- *H.S. Smallman is with the Pacific Science & Engineering Group, San Diego, CA, 92121. E-mail: Smallman@pacific-science.com.*

Manuscript received 1 Aug. 2006; revised 26 Oct. 2006; accepted 8 Nov. 2006; published online 2 Feb. 2007.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-0119-0806.
Digital Object Identifier no. 10.1109/TVCG.2007.1035.

2 Background and Related Work

2.1 Depth Cues and Cue Theory

Human depth perception delivers a vivid three-dimensional perceptual world from flat, two-dimensional, ambiguous retinal images of the scene. Current thinking on how the human visual system is able to achieve this performance emphasizes the use of multiple *depth cues*, available in the

scene, that are able to resolve and disambiguate depth relationships into reliable, stable percepts. *Cue theory* describes how and in which circumstances multiple depth cues interact and combine. Generally, 10 depth cues are recognized (Howard and Rogers [11]):

1. binocular disparity,
2. binocular convergence,
3. accommodative focus,
4. atmospheric haze,
5. motion parallax,
6. linear perspective and foreshortening,
7. occlusion,
8. height in the visual field,
9. shading, and
10. texture gradient.

Real-world scenes combine some or all of these cues, with the structure and lighting of the scene determining the relative salience of each cue. Although *depth cue interaction* models exist (Landy et al. [18]), these were largely developed to account for how stable percepts could arise from a variety of cues with differing salience. The central challenge in understanding human depth perception in AR is determining how stable percepts can arise from inconsistent, sparse, or purposely conflicting depth cues, which arise either from imperfect AR displays, or from novel AR perceptual situations such as x-ray vision. Therefore, models of AR depth perception will likely inform both applied AR technology as well as basic depth cue interaction models.

2.2 Near, Medium, and Far-Field Distances

Depth cues vary both in their salience across real-world scenes, and in their effectiveness by distance. Cutting [6] has provided a useful taxonomy and formulation of depth cue effectiveness by distances that relate to human action. He divided perceptual space into three distinct regions, which we term near-field, medium-field, and far-field. The *near field* extends to about 1.5 meters: It extends slightly beyond arm's reach, it is the distance within which the hands can easily manipulate objects, and within this distance, depth perception operates almost veridically. The *medium field* extends from about 1.5 meters to about 30 meters: It is the distance within which conversations can be held and objects thrown with reasonable accuracy; within this distance, depth perception for stationary observers becomes somewhat *compressed* (items appear closer than they really are). The *far field* extends from about 30 meters to infinity, and as distance increases, depth perception becomes increasingly compressed. Within each of these regions, depth cues vary in their availability, salience, and potency.

2.3 Egocentric Distance Judgment Techniques

Researchers have long been interested in measuring the perception of distance, but, faced with the classic problem that perception is an invisible cognitive state, have had to find measurable quantities that can be related to the perception of distance. Therefore, they have devised experiments where distance perception can be inferred from distance *judgments*. The most general categorization of distance judgments is egocentric or exocentric: *egocentric* distances are measured from an observer's own view point, while *exocentric* distances are measured between different objects in a scene. Loomis and Knapp [21] and Foley [10] review and discuss the methods that have been developed to measure judged egocentric distances.

There have been three primary methods: *verbal report*, *perceptual matching*, and *open-loop action-based tasks*. With *verbal report* [10], [16], [21], [23], observers verbally estimate the distance to an object, typically using whatever units they are most familiar with (e.g., feet, meters, or multiples of some given referent distance). Observers have also verbally estimated the size of familiar objects [21], which are then used to compute perceived distance. *Perceptual matching* tasks [9], [10], [22], [30], [37] involve the observer adjusting the position of a target object until it perceptually matches the distance to a referent object. Perceptual matching is an example of an *action-based task*; these tasks involve a physical action on the part of the observer that indicates perceived distance. Action-based tasks can be further categorized into open-loop and closed-loop tasks. In an *open-loop* task, observers do not receive any visual feedback as they perform the action, while in a *closed-loop* task they do receive feedback. By definition, perceptual matching tasks are closed-loop action-based tasks.

A wide variety of *open-loop action-based tasks* have been employed. For all of these tasks, observers perceive the egocentric distance to an object, and then perform the task without visual feedback. The most common open-loop action-based task has been *blind walking* [5], [16], [21], [23], [36], [37], where observers perceive an object at a certain distance, and then cover their eyes and walk until they believe they are at the object's location. Blind walking has been found to be very accurate for distances up to 20 meters, and there is compelling evidence that blind walking accurately measures the percept of egocentric distance (Loomis and Knapp [21]). Because of these benefits, blind walking has been widely used to study egocentric depth perception at medium and far-field distances, in both real-world and VR settings. A closely related technique is *imagined blind walking* [7], [26], where observers close their eyes and imagine walking to an object while starting and stopping a stopwatch; the distance is then computed by multiplying the time by the observers' normal walking speed. Yet another variant is *triangulation by walking* [21], [34], [36], where observers view an object, cover their eyes, walk a certain distance in a direction oblique to the original line of sight, and then indicate the direction of the remembered object location; their perception of the object's distance can then be recovered by simple trigonometric calculations. Near-field distances have been studied by *open-loop pointing* tasks [10], [25], where observers indicate distance with a finger or manipulated slider that is hidden from view.

In addition, some researchers have used *forced-choice* tasks [20], [29], [30] to study egocentric depth perception. In forced-choice tasks, observers make one of a small number of discrete depth judgment choices, such as whether one object is closer or farther than another; or at the same or a different depth; or at a near, medium, or far depth, etc. These tasks tend

to use a large number of repetitions for a small number of observers, and can employ psychophysical techniques to measure and analyze the judged depth [29], [30].

Finally, although depth judgment tasks are considered the best method available for measuring the egocentric percept of distance and have been widely used, researchers have determined that they can be influenced by cognitive factors that are unrelated to actual egocentric distance. For example, Decety et al. [7] and Proffitt [27] have argued that distance judgments are influenced by the amount of energy observers anticipate expending to traverse the distance. Proffitt [27] and collaborators have further observed that distance judgments are influenced by the possibility of injury, by the observer's current emotional state, and even by social factors such as whether or not the observer owns the item to which distances are judged.

2.4 The Virtual Reality Depth Underestimation Problem

Over the past several years, many studies have examined egocentric depth perception in VR environments. A consistent finding has been that *egocentric depth is underestimated* when objects are viewed on the ground plane, at near to medium-field distances, and the VR environment is presented in a head-mounted display (HMD) [5], [16], [21], [23], [28], [34], [36]. As discussed above, most of these studies have utilized open-loop action-based tasks, although the effect has been observed with perceptual matching tasks as well [37]. These studies have examined various theories as to why egocentric depth is underestimated, and have found evidence that underestimation is caused by an HMD's limited field-of-view [37]; that underestimation is *not* caused by an HMD's limited field-of-view [5], [16]; that the weight of the HMD itself might contribute to the phenomenon [36]; that monocular versus stereo viewing does not cause it [5]; that the quality of the rendered graphics does not cause it [34]; that the effect persists even when observers see live video of the real world in an HMD [23]; that the effect might exist when VR is displayed on a large-format display screen as well [26]; that the effect might disappear when observers know that the VR room is an accurate model of the physical room in which they are located [13]; that the amount of underestimation is significantly reduced by as little as 5 to 7 minutes of practice with feedback [24], [28]; and that the underestimation effect can be compensated by modifying the way the graphics are rendered [17]. In summary, the egocentric distance underestimation effect is real, and although its parameters are being explored, it is not yet fully understood.

2.5 Previous AR Depth Judgment Studies

There have been a small number of studies that have examined depth judgments with optical, see-through AR displays. Ellis and Menges [9] summarize a series of AR depth judgment experiments, which used a perceptual matching task to examine near-field distances of 0.4 to 1.0 meters, and studied the effects of an occluding surface (the x-ray vision condition), convergence, accommodation, observer age, and monocular, biocular, and stereo AR displays. They found that monocular viewing degraded the depth judgment, and that the x-ray vision condition caused a change in vergence angle which resulted in depth judgments being biased toward the observer. They also found that cutting a hole in the occluding surface, which made the depth of the virtual object physically plausible, reduced the depth judgment bias. McCandless et al. [22] used the same experimental setup and task to additionally study motion parallax and AR system latency in monocular viewing conditions; they found that depth judgment errors increased systematically with increasing distance and latency. Rolland et al. [29], in addition to a substantial treatment of AR calibration issues, discuss a pilot study at near-field distances of 0.8 to 1.2 meters, which examined depth judgments of real and virtual objects using a forced-choice task. They found that the depth of virtual objects was overestimated at the tested distances. Rolland et al. [30] then ran additional experiments with an improved AR display, which further examined the 0.8 meter distance, and compared forced-choice and perceptual matching tasks. They found improved depth accuracy and no consistent depth judgment biases. Jerome and Witmer [14] used a perceptual matching task as well as verbal report to examine distances from 1.5 to 25 meters. They found that the depth of real-world objects were judged more accurately than virtual objects, but their dependent measure does not allow the error to be categorized as underestimation or overestimation. They also found a very interesting interaction between error and gender. Kirkely [15] used verbal report to study the effect of the x-ray vision condition, the ground plane, and object type (real objects, realistic virtual objects (e.g., a chair), and abstract virtual objects (e.g., a sphere)), on monocularly-viewed objects at distances from 3 to 33.5 meters. He found that the x-ray vision condition reduced performance, placing objects on the ground plane improved performance, and that real objects resulted in the best performance, realistic virtual objects resulted in intermediate performance, and abstract virtual objects resulted in the worst performance. Livingston et al. [20] used a forced-choice task to examine graphical parameters such as drawing style, intensity, and opacity on occluded AR objects at far-field distances of 60 to 500 meters. They found that certain parameter settings were more effective for their task.

Taken together, these studies have just begun to explore how depth perception operates in AR displays. In particular, only two previous studies have examined AR depth perception in the medium-field to far-field, which is an important range of distances for many imagined outdoor AR applications. In this paper, we describe two AR egocentric depth judgment experiments that have studied this range of distances. Experiment I used a perceptual matching task, and Experiment II used verbal report and blind walking tasks. Furthermore, Experiment II is the first reported AR depth study to use the open-loop action-based task of blind walking, and as discussed above, in VR open-loop action-based tasks have been the most wildly used task category.

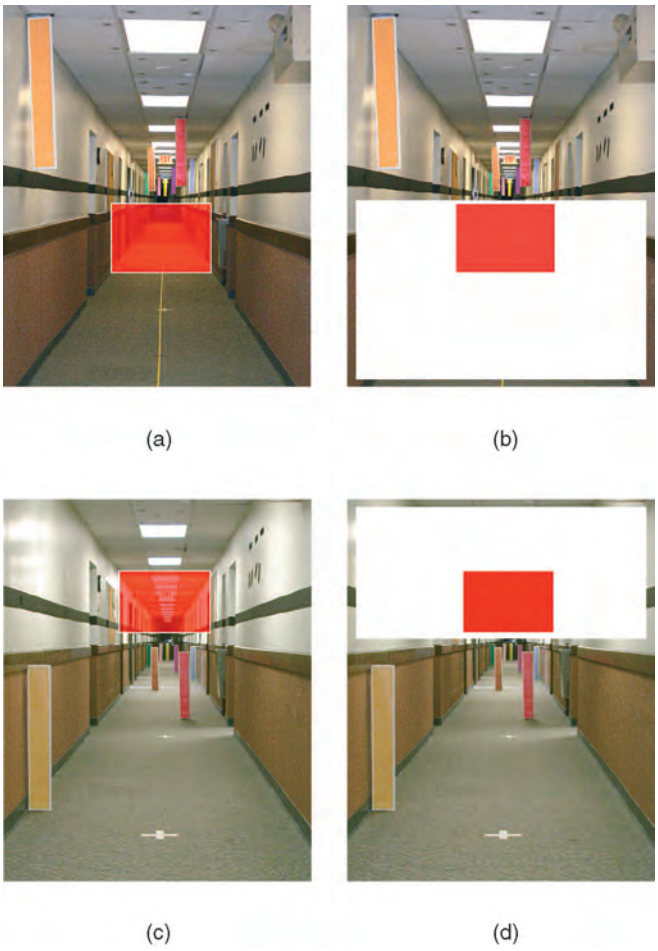

Fig. 1. The experimental setting and layout of the real-world referents and the virtual target rectangle. Observers manipulated the depth of the target rectangle to match the depth of the real-world referent with the same color (red in this example). Note that these images are not photographs taken through the actual AR display, but instead are accurate illustrations of what observers saw. (a) Referents on *ceiling*, occluder *absent*. (b) Referents on *ceiling*, occluder *present*. (c) Referents on *floor*, occluder *absent*. (d) Referents on *floor*, occluder *present*.

3 Experiment I: Perceptual Matching Protocol

3.1 Experimental Task and Setting

In Experiment I,[1] we used a perceptual matching task to study depth judgments of medium-field to far-field distances of 5.25 to 44.31 meters. Fig. 1 shows the experimental setting. Observers sat on a stool at one end of a long hallway, and looked through an optical, see-through AR display mounted on a frame. Observers saw a series of eight real-world *referents*, approximately positioned evenly down the hallway (Fig. 1). Each referent was a different color. The AR display showed a virtual *target*, which we drew as a semitransparent rectangle that horizontally filled the hallway, and vertically extended about half of the hallway's height. Our target and task was motivated by our initial problem domain, outdoor augmented reality in urban settings [19], which required users to visualize the spatial layout of rectangular building components, such as walls, floors, doors, etc., within a radius of one to several blocks. The visualized rectangular building components typically abutted other parts of the building, such as the hallway in our experimental setting.

1. This experiment has been previously described by Swan et al. [32]; this section summarizes the experiment and its most interesting results.

TABLE 1
Independent Variables and Levels, and Dependent Variables, for Experiment I

Independent Variables				
observer	8	(random variable)		
height in visual field	2	ceiling, floor		
occluder	2	present, absent		
distance	8	Distance (Meters)	Angular Size (° Visual Angle)	Color
		5.25	1.75	orange
		11.34	.808	red
		17.42	.526	brown
		22.26	.412	blue
		27.69	.331	purple
		33.34	.275	green
		38.93	.235	pink
		44.31	.206	yellow
repetition	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10		
Dependent Variables				
judged distance	measured by perceptual matching, *meters*			
absolute error	| *judged distance – distance* |, *meters*			
error	*judged distance – distance*, *meters*			

Observers adjusted the target's depth position in the hallway with a trackball. For each trial, our software drew the target rectangle at a random initial depth position; it drew the target rectangle with a white border, and colored the target interior to match the color of one of the referents (Fig. 1). The observer's task was to adjust the target's depth position until it matched the depth of the referent with the same color. When the observer believed the target depth matched the referent depth, they pressed a mouse button on the side of the trackball. This made the target disappear; the display then remained blank for approximately one second, and then the next trial began. For the display device we used a Sony Glasstron LDI-D100B stereo optical see-through display. It displays 800×600 (horizontal by vertical) pixels in a transparent window which subtends $27^\circ \times 20^\circ$[2] and, thus, each pixel subtends approximately $.033^\circ \times .033^\circ$.

3.2 Variables and Design

3.2.1 Independent Variables

The independent variables are summarized in Table 1. We recruited eight *observers* from a local population of scientists and engineers. As shown in Fig. 1, we placed the referents at two different *heights in the visual field*: we mounted the referents either on the *ceiling* or the *floor*. Our experimental control program rendered the target in the opposite field of view as the referents. As discussed above, we were interested in understanding AR depth perception in the x-ray vision condition, so we varied the presence of an occluding surface. When the *occluder* was *absent* (Figs. 1a and 1c), observers could see the hallway behind the target. When the *occluder* was *present* (Figs. 1b and 1d), we mounted a heavy rectangle of

2. Angular measures in this paper are in degrees of visual arc.

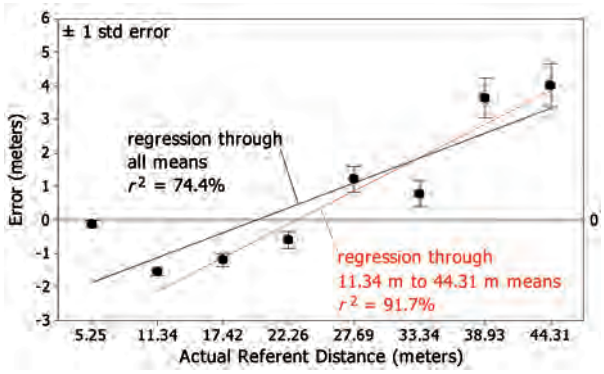

Fig. 2. The effect of distance on error $(N = 2,560)$, which exhibits a strong linear regression beginning at 11.34 meters. This reveals a switch in bias from underestimating to overestimating target distance at ~ 23 meters.

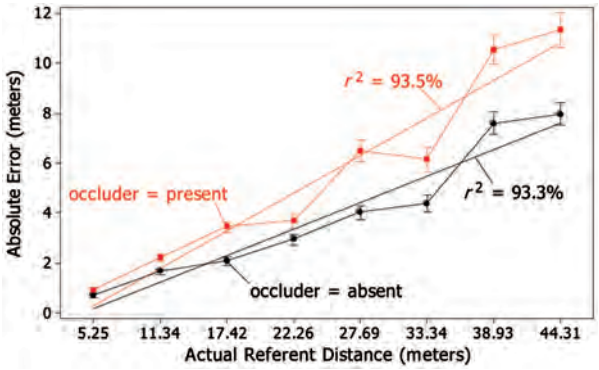

Fig. 3. Effect of occluder by distance on absolute error $(N = 2,560)$. Observers had more error in the occluded (x-ray vision) condition (red line and points) than in the nonoccluded condition (black and points), and the difference between the occluded and nonoccluded conditions increased with increasing distance.

foamcore posterboard across the observer's field-of-view, which occluded the view of the hallway behind the target. We placed the eight referents at the *distances* from the observer indicated in Table 1. We built the referents out of triangular shipping boxes, which measured 15.3 cm wide by 96.7 cm tall. We covered the boxes with the colors listed in Table 1. We created the colors by printing single-colored sheets of paper with a color printer. To increase the contrast of the referents against the hallway background, we created a border around each color with white gaffer's tape. We affixed the referents to the ceiling and floor with velcro. We presented each *repetition* of the other independent variables 10 times.

3.2.2 Dependent Variables

For each trial, observers manipulated a trackball to place the target at their desired depth down the hallway, and pressed the trackball's button when they were satisfied. The trackball produced 2D cursor coordinates, and we converted the y-coordinate into a depth value with the perspective transform of our graphics pipeline; we used this depth value to render the target rectangle. When an observer pressed the mouse button, we recorded this depth value as the observer's *judged distance*. As indicated in Table 1, we used the judged distance to calculate two dependent variables, *absolute error* and *error*. An *absolute error* or *error* close to 0 indicates an accurately judged distance. An $error > 0$ indicates an overestimated judged distance, while an $error < 0$ indicates an underestimated judged distance.

3.2.3 Experimental Design and Procedure

We used a factorial nesting of independent variables for our experimental design, which varied in the order they are listed in Table 1, from slowest (observer) to fastest (repetition). We collected a total of 2,560 data points (eight observers $\times$ two fields of view $\times$ two occluder states $\times$ eight distances $\times$ 10 repetitions). We counterbalanced presentation order with a combination of Latin squares and random permutations. Each observer saw all levels of each independent variable, so all variables were within-subject.

3.3 Results and Discussion

Here, we discuss the main results qualitatively; full statistical details are given in Swan et al. [32]. Fig. 2[3] shows that error increased linearly with increasing distance ($r^2 = 74.4\%$; black line in Fig. 2). However, the 5.25 meter referent weakens the linear relationship; it is likely close enough that near-field distance cues are still operating. The linear relationship between error and distance increases when analyzed for referents 2-8 ($r^2 = 91.7\%$; red line in Fig. 2). Even more interesting is a shift in bias from underestimating (referents 2-4) to overestimating (referents 5-8) distance. The bias shift occurs at around 23 meters, which is where the red line in Fig. 2 crosses zero meters of error. Foley [10] found a similar bias shift, from underestimating to overestimating distance, when studying binocular disparity in isolation from all other depth cues. He found that the shift occurred in a variety of perceptual matching tasks, and although its magnitude changed between observers, it was reliably found. However, in Foley's tasks, the point of veridical performance was typically found at closer distances of 1-4 meters. The similarity of this finding to Foley's suggests that stereo disparity may be an important depth cue in this experimental setting, although the strength of stereo disparity weakens throughout the medium-field range. It seems likely that linear perspective is also an important depth cue here.

Fig. 3 shows an occluder by distance interaction effect on absolute error. When an occluder was present (the x-ray vision condition), observers had more error than when the occluder was absent, and the difference between the occluder present and occluder absent conditions increased with increasing distance. Fig. 3 shows a linear modeling of the occluder present condition (red line), which explains $r^2 = 93.5\%$ of the observed variance, and a linear modeling of the occluder absent condition (black line), which explains $r^2 = 93.3\%$ of the observed variance. These two linear models allow us to estimate the magnitude of the occluder effect according to distance:

$$y_{present} - y_{absent} = .08x - .33,$$

where y_{present} is the occluder present (red) line, y_{absent} is the occluder absent (black) line, and x is distance. This equation says that for every additional meter of distance, observers made 8 cm of additional error in the occluder present versus the occluder absent condition.

3. In this and future graphs, N is the number of data points that the graph summarizes.

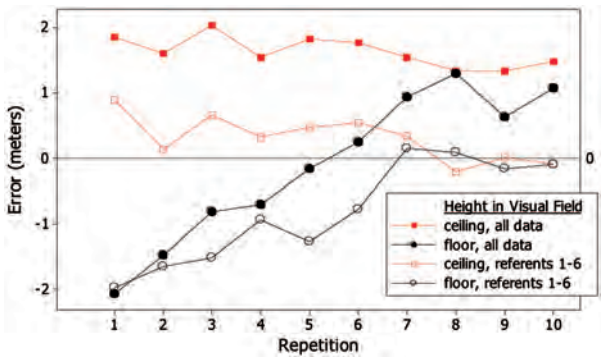

Fig. 4. Effect of height in the visual field by repetition on error $(N = 2,560)$. Solid shapes (■, •) are means for all the data; hollow shapes (□, ○) are means for the first six referents. Squares (■, □) are referents mounted on the ceiling; circles (•, ○) are referents mounted on the floor. For clarity, standard error bars are not shown.

Fig. 4 shows an interesting interaction between height in the visual field and repetition. The solid shapes (■, •) show the interaction for all of the data. When the referents were mounted on the ceiling (■), observers overestimated their distance by about 1.5 meters, and when the referents were mounted on the floor (•), observers began with an underestimation (low repetitions), and with practice, by repetition 8 matched the overestimation of the ceiling-mounted referents. The general bias toward overestimation can be explained by the overestimation of the last two referents, as seen in Fig. 2. In Fig. 4, the hollow shapes (□, ○) show the same interaction when the last two referents are removed. When the referents were mounted on the ceiling (□), observers did not show a bias, and by repetition 7 were quite accurate. For referents mounted on the floor (○), observers initially demonstrated the same underestimation as they did for the full data set, and with practice, by repetition 7 matched the veridical performance of the ceiling-mounted referents (□).

This interaction is puzzling. We hypothesize that the underestimation of the first two or three floor-mounted referents (○) is similar to the underestimation that has been demonstrated in VR environments, and that the underestimation's disappearance is a practice effect, which has not been seen in previous experiments because open-loop action-based tasks such as blind walking typically only have 1-3 repetitions. This hypothesis is consistent with the findings of Mohler et al. [24] and Richardson and Waller [28], who found that as little as three additional repetitions of blind walking (but with feedback) significantly reduced the amount of underestimation. On the other hand, the ceiling-mounted referents (□), which are hanging at eye level, do not show underestimation. Among the very few studies to examine the egocentric distance of ceiling-mounted referents is Dilda et al. [8], who used a perceptual matching task that is very similar to the one we used, and found that the distance was overestimated by 10 percent. Interestingly, in Fig. 4, for the first three repetitions the difference between the ceiling (□) and floor (○) referents is also roughly 10 percent.

4 Experiment II: Blind Walking and Verbal Estimation Protocol

Our experiences conducting Experiment I motivated us to design and conduct an experiment which replicated the type of depth judgment task and medium-field setting that has been most often studied in VR. Experiment II utilized the depth judgment protocols of 1) blind walking and 2) verbal report to measure egocentric distance perception of ground-based objects in an AR head-mounted display (HMD). We again studied medium-field distances, this time from 3 to 7 meters. As discussed previously, the VR egocentric depth perception literature describes a number of studies utilizing blind walking [5], [16], [21], [23], [36] and verbal report [10], [16], [21], [23], at distances ranging from ~ 2 to ~ 25 meters. Therefore, Experiment II is more directly comparable to the VR depth perception literature —the main difference being the use of a see-through AR display as opposed to an opaque VR display. Our motivation was to further characterize the depth underestimation phenomena in AR, as well as to study depth judgments of 1) virtual objects and 2) virtual objects that augment the appearance of real objects. As a control condition, we also studied depth judgments of 3) real objects seen with an unencumbered view, and 4) real objects seen through the AR HMD display.

4.1 Experimental Setup and Task

Observers judged the distance to both a physical referent object (Fig. 5a), as well as a virtual model of the referent object. Our referent object was a wooden pyramid, 23.5 cm tall, with a square base of 23.5 cm. Our display device was a Sony Glasstron LDI-100B monoscopic (biocular), optical see-through HMD. Our HMD displays 800×600 (horizontal by vertical) pixels in a transparent window which subtends $27^\circ \times 20^\circ$, and thus each pixel subtends approximately $.033^\circ \times .033^\circ$. This window is approximately centered in a larger semitransparent frame, which is tinted like sunglasses and so attenuates the brightness of the real world. The outer edge of this frame subtends $66^\circ \times 38^\circ$. Because our HMD is monoscopic, we used an anaglyphic stereo technique to give observers a stereo disparity depth cue. We presented the virtual referent in blue to the left eye and red to the right eye (Fig. 5a), and we attached appropriately colored red and blue plastic filters to the inside of the HMD. We ordered the filters from a supplier of 3D anaglyphic stereo equipment; their colors matched the red and blue produced from common monitors. For each eye, there was negligible ghosting through the other eye's filter. The resulting virtual object appeared neither red nor blue, but instead a shade of white. There was also a subtle shimmering effect, which did not disrupt the sense that the virtual referent object was located in a definite position in space. We rendered the back line of the virtual object with a dashed appearance, to graphically suggest that it was behind the front lines.

Attaching the red and blue filters to the HMD further attenuated the brightness of the real world. Although we set the display opacity to its most transparent setting, it was difficult to see the real world, and the physical referent object,

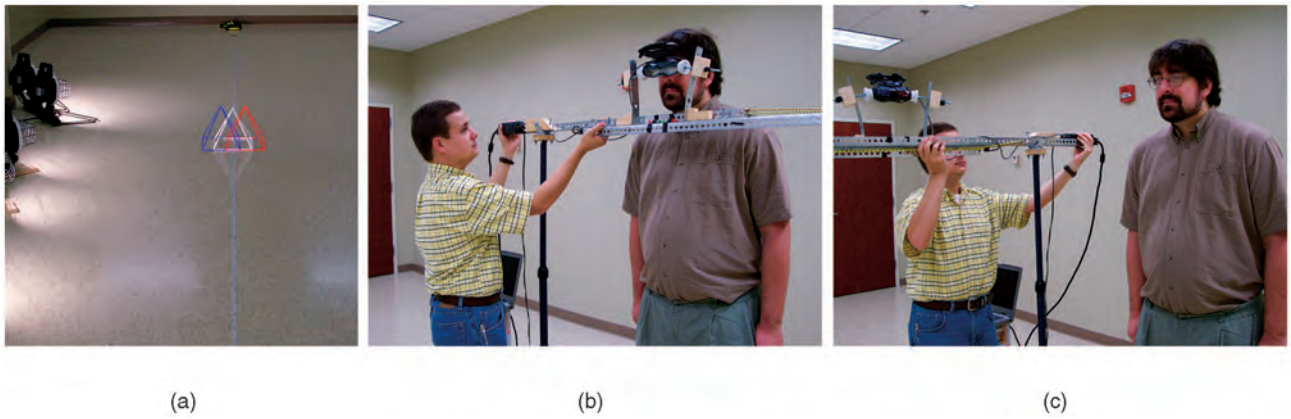

Fig. 5. (a) Observer's view of the real-world referent object, illuminated by the halogen lights, and the virtual referent object (the $real + virtual + HMD$ environment). Observers viewed the virtual object in red/blue anaglyphic stereo. We rendered the backmost line of the virtual object with a dashed appearance, which further enhanced the sense that the virtual and real objects were merged. Note that we created this image using video see-through AR, while observers used optical see-through AR. (b) Observer looking through the frame-mounted AR HMD during a blind walking trial. An experimenter is prepared to swing the frame out of the way. (c) The experimenter has swung the frame out of the way, and the observer is now free to walk forward.

under normal indoor illumination conditions. Therefore, like other studies that have utilized Glasstron displays [14], we illuminated the referent object with six 600-watt halogen lamps (Fig. 5a), which provided enough illumination so that the object could be readily perceived through the display. In addition, we painted the physical referent object white, both to match the virtual pyramid, and to better reflect the illumination of the halogen lamps. We adjusted the HMD's brightness setting so that the virtual object matched the brightness of the real object. We corrected the display for an optical barrel distortion effect using the 2D polygonal grid-based texture mapping technique initially described by Watson and Hodges [35] and refined by Bax [2]; we separately calibrated a 16×12 cell grid for the left and right display channels. Our display had a nonadjustable interpupilary separation, so we measured observers' interpupilary distance and eye height, and modeled these parameters in software. Our display also had a nonadjustable accommodative demand of 1.2 meters.

As mentioned above, we wanted to study the condition where the virtual referent augmented the appearance of the physical referent. This meant that we needed to achieve a very precise alignment between the virtual and physical referents—more precise than is possible with current 6 degree-of-freedom tracking technology. Therefore, similar to Experiment I, we mounted the AR HMD on a rigid frame, supported by two tripods. We adjusted the height of the tripods so that each observer could comfortably look through the HMD at their normal standing eye height.

The blind walking protocol requires subjects to observe a referent object, close (or cover) their eyes, and walk forward. This meant that it was necessary to engineer the HMD frame so that it could swing out of the way (Fig. 5). The frame was attached to one tripod with a caster wheel mount that allowed 360° of rotation, while the other side of the frame rested in an "L" shaped holder. We engineered this apparatus to be stable enough so that, when the HMD was swung out of the way and then back into position, the alignment was preserved as much as possible. During the experiment, we typically only had to make minor adjustments to restore the alignment. We stereo calibrated the display by stereo-aligning a virtual wireframe model of the experimental room to the actual room, and as discussed below, we tested and recalibrated the alignment between the virtual and real referent objects as often as every trial.

We conducted the experiment in two different buildings[4] on the Mississippi State University campus. Location 1 was a 2.28×30.4 meter hallway; observers stood 8.83 meters from one end, and walked down the center of the hallway. Location 2 was an 11.35×7.26 meter empty room in a different building; observers stood 1.7 meters from one wall and faced the long axis of the room. Observers walked down a path that was approximately centered between one wall of the room and a folding wall that extends 2.77 meters into the room. In both locations, we attached a long, flexible measuring tape down the center of the pathway; we used this tape to place the physical referent object at precise distances, and to measure the observer's position during the blind walking trials. The numbers on the tape were much too small to be legible to observers during experimental trials.

We ran the experiment on a Pentium M 1.80 GHz laptop computer with an NVIDIA GeForce FX Go5200 graphics card, which outputs frame-sequential stereo. We monitored the experiment's progress on the laptop screen. We implemented our experimental control code in C++, using the OpenGL library, and Perl.

4.2 Variables and Design

4.2.1 Independent Variables

Observers: We recruited 16 observers from a population of university students (undergraduate and graduate), and staff. Nine of the observers were male, seven were female;

4. Although it was not our desire to change locations during the experiment, we were forced to by two factors: 1) the halogen lights, a lack of air conditioning, and the onset of summer resulted in uncomfortable conditions in Location 1, and 2) the Institute for Neurocognitive Science and Technology, where we conducted this experiment, moved into a new building (Location 2), which meant we had to move our equipment as well. In Section 4.2.3, we discuss where this location change fell in the experimental design.

TABLE 2
Independent Variables and Levels, and Dependent Variables, for Experiment II

INDEPENDENT VARIABLES		
observer	16	(random variable)
environment	4	real world, real + HMD real + virtual + HMD virtual + HMD
protocol	2	blind walking verbal report
distance	3	3, 5, 7 meters
repetition	4	1, 2, 3, 4
DEPENDENT VARIABLES		
judged distance	measured from each *protocol*, *meters*	
error	*judged distance – distance*, *meters*	

they ranged in age from 20 to 33, with a mean age of 25.4. We screened the observers, via self-reporting, for color blindness and visual acuity. All observers volunteered, and were compensated $10 per hour for their time. Observers spent an average of 2.25 hours completing the experiment.

Environment: As shown in Table 2 and Fig. 5, observers judged the depth of referents presented in four different environments. In the *real-world* environment, observers saw the real-world referent object, and did not look through the HMD. We included this as a control condition, as it duplicates the setup of distance perception studies with real-world referents [21]. In the $real + HMD$ environment, observers saw the real-world referent object, but this time regarded the referent object through the HMD. In the $real + virtual + HMD$ environment, observers saw the real-world referent object and the virtual referent object at the same time. As discussed below, we carefully calibrated the display so that the two aligned with a high degree of precision. In the $virtual + HMD$ environment, observers saw only the virtual referent object.

Protocol: Observers used two different protocols to judge the depth of referent objects. When using the *blind walking* protocol, observers regarded the referent object for as long as they wished (typically a few seconds), closed their eyes, and then verbally notified the experimenter that they were ready to respond. An experimenter swung the HMD out of the way and said "walk forward"; this operation typically took ~ 2 seconds. After hearing "walk forward," observers walked, with their eyes closed, to their remembered location of the referent object. For environments where a physical referent object was present, a second experimenter removed the object before the observer reached the location. After stopping, observers stood and looked ahead (not down), while the two experimenters silently recorded their distance from the floor-mounted tape. When this was recorded, observers walked to an isolation area, which was a room off of the hallway (Location 1), or an area separated by a folding wall (Location 2). In the isolation area, observers could not see the experimental room. While the observer was gone, the experimenters reset the HMD, set the physical referent to the next distance, and checked and adjusted the HMD calibration. When all was ready, the experimenters asked the observer to return to the starting position without looking at the room, and begin the next trial. During real world environment trials, observers did not look through the HMD. Instead, after the observer closed their eyes, the experimenter waited ~ 2 seconds, and then said "walk forward."

When using the *verbal report* protocol, observers regarded the referent object for as long as they wished (typically a few seconds), and then reported the distance, in whatever units the observer desired. Observers then moved to the isolation area while the experimenters readied everything for the next trial. When all was ready, the experimenters asked the observer to return to the starting position without looking at the room, and begin the next trial. Although the calibration was checked every trial, because the HMD was not swung out of the way, it was generally only necessary to adjust it at the beginning of each block of verbal report trials.

Distance: For *experimental trials*, observers saw referent objects placed at distances of 3, 5, and 7 meters. Because observers may notice the repetition in such a small set of distances, and this can influence their distance judgments (especially verbal reports), 25 percent of the distance judgments were *noise trials*. For these trials, distances were randomly chosen from 0.25-meter increments in the 3 to 7 meter range; the experimenters recorded the data from the noise trials using the same procedures that were used for the experimental trials. The noise trials are not analyzed in this paper.

Repetition: Observers saw four repetitions of each combination of the other independent variables.

4.2.2 Dependent Variables

As shown in Table 2, the primary dependent variable was *judged distance*, which was either measured from the observer's foot position (blind walking), or verbally reported by the observer. We also calculated *error*, which has the same meaning as it did in Experiment I: an *error* close to 0 indicates an accurately judged distance, an $error > 0$ indicates an overestimated judged distance, and an $error < 0$ indicates an underestimated judged distance.

4.2.3 Experimental Design

We used a factorial nesting of independent variables in our within-subjects experimental design. Table 3 shows the loop that our experimental control program used to present the

TABLE 3
Stimulus Presentation Loop and Counterbalancing

PRESENTATION LOOP	LEVELS	ORDER CONTROL
for each *environment*	4	4×4 Latin Square
for each *protocol*	2	2×2 Latin Square
for *distance* ⊗ *repetition* + *noise*	(3 × 4) + 4	Restricted random permutation
present trial		

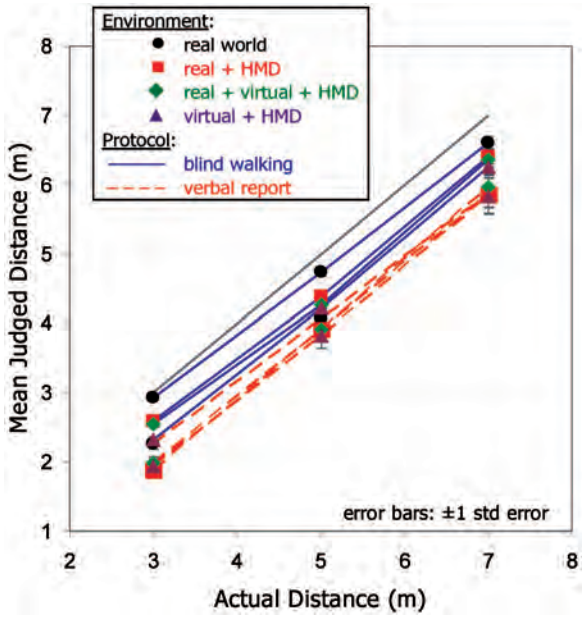

Fig. 6. The main results, plotted as judged distance versus actual referent distance $(N = 1,536)$. The light gray line indicates veridical performance.

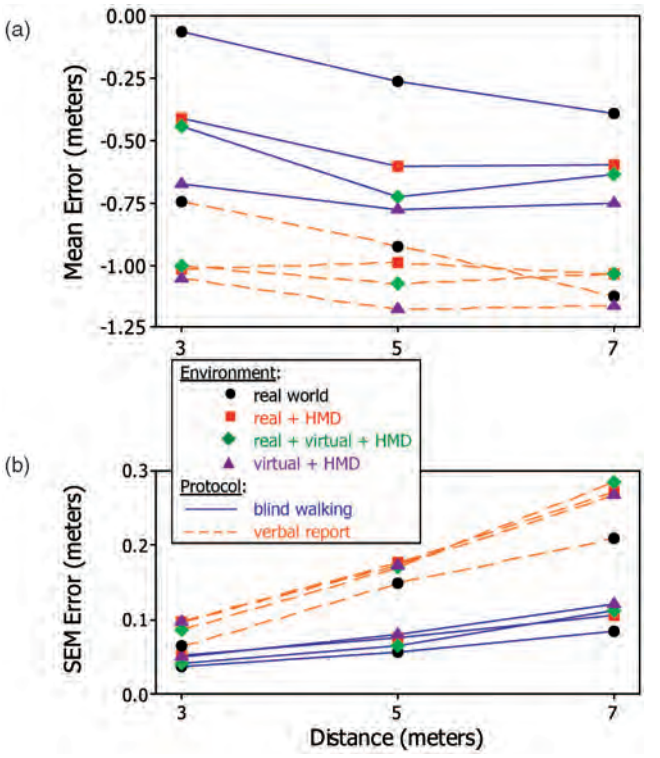

Fig. 7. The main results, plotted as (a) mean error $(N = 1,536)$, and (b) standard error of the mean (SEM) error $(N = 1,536)$, for each referent distance.

independent variables to the observers. *Environment* varied the slowest; within each environment observers saw each *protocol*. The presentation order of *environment* was controlled by a 4×4 between-subjects Latin Square, while the presentation order of *protocol* was controlled by a 2×2 between-subjects Latin Square; when combined, these two Latin Squares resulted in a presentation order design that repeated modulo eight subjects. Within each $environment \otimes protocol$ block, our control program generated a list of 3 $(distance) \times$ 4 $(repetition) = 12$ experimental distances, and then added four random *noise* distances. The program then randomly permuted the presentation order of the resulting 16 distances, with the restriction that the same distance could not show up twice in a row. We collected a total of 1,536 data points (16 observers $\times$ four environments $\times$ two protocols $\times$ three distances $\times$ four repetitions). As discussed above, the 16 observers participated in two different locations. Observers 1-8 participated in Location 1, while observers 9-16 participated in Location 2. Therefore, the experiment was counterbalanced with respect to the presentation order of the data collected in each location.

4.3 Results and Discussion

4.3.1 Descriptive Results

Fig. 6 shows the main results from the study, which by the convention established in much of the recent VR depth perception literature, is displayed as a correlation between the actual distance and the judged distance. This shows that, like virtual environments presented in opaque HMDs, there is a general trend of egocentric distance underestimation for virtual objects presented in transparent, AR HMDs. The judged distances fell into three main groups, which are listed here along with their mean percentages of actual distance $(percentage = judged\ distance/actual\ distance)$: 1) blind walking in the real-world environment: 96 percent, 2) blind walking in the HMD environments, which includes the real-world seen through the HMD: 86 percent, and 3) verbal report: 77 percent. These results can be compared to the percentages from six studies of virtual environment distance perception that examined a similar range of distances with open-loop action-based protocols, as reported by Thompson et al. [34]. These studies reported real-world judgments that ranged from 92-100 percent of actual distances, and virtual environment judgments that ranged from 42-85 percent of actual distances. Our control condition (blind walking in the real-world) had results (96 percent) that are similar to what has been reported across these studies (92-100 percent), and we interpret this as some assurance that our implementation of the blind walking protocol was essentially correct. However, others have achieved results very close to 100 percent [33], and it seems likely that further improvements are possible. More interestingly, we found that the degree of underestimation for the HMD environments (86 percent) is on the low end of what has been observed for virtual environments (42-85 percent).

The rest of the graphs in this paper show results in terms of *error* (Table 2); this metric allows differences in judged distances to be more clearly plotted. Fig. 7a gives the main results in terms of mean *error*. As discussed above, these indicate that all blind walking conditions had less underestimation than verbal report conditions, and that blind walking in the real world was the most accurate of all. In Section 4.3.3 below, we analyze the blind walking results in more detail. Fig. 7b gives the variability of the main results, expressed in terms of the standard error of *error*. These results indicate that as the degree of underestimation increases, so does the variability and, thus, the verbal report results are more variable than the blind walking results. In addition, similar to Experiment I, variability

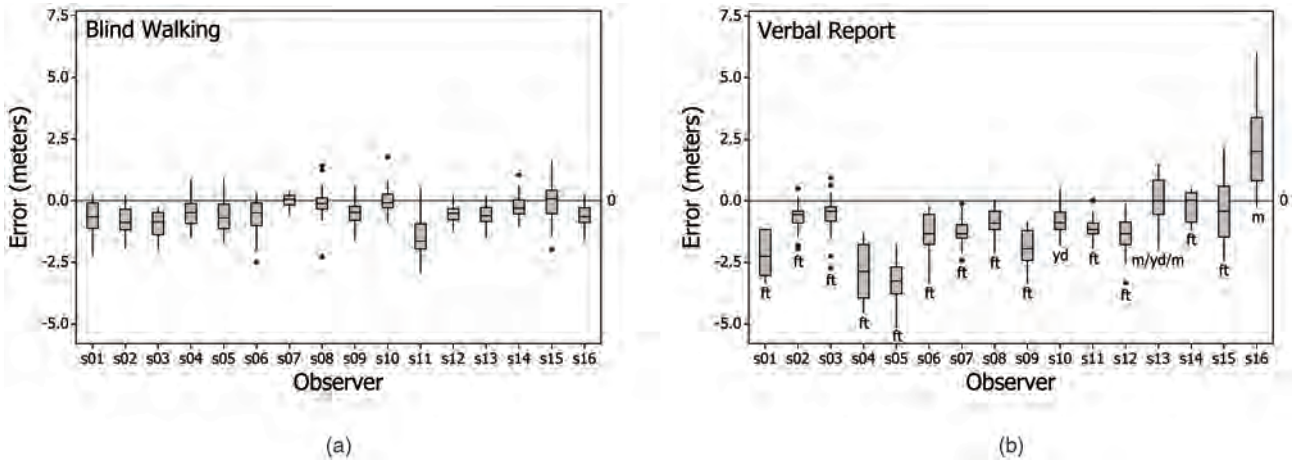

Fig. 8. Boxplots showing the error results for each observer. (a) The blind walking results $(N = 768)$. (b) The verbal report results $(N = 768)$. These are labeled with the units that the observers used: *ft*: feet, *yd*: yards, and *m*: meters. Observer *s*13 began using meters, then switched to yards, and then back to meters. Asterisks "$*$" indicate single outlying data points.

increased with increasing distance, which we generally expect because observer responses are based on depth cues of linearly decreasing effectiveness (i.e., observers are following Weber's law [31]). Finally, there appears to be an increase in gain as well as a bias shift for verbal report, relative to blind walking.

Fig. 8 shows the results for each observer, separated according to protocol. Observers were consistent with blind walking (Fig. 8a), as compared to verbal estimation (Fig. 8b). Observer *s*07 gave extremely consistent blind walking results; this subject reported walking and running on a treadmill with their eyes closed on a regular basis. Observer *s*11, who gave the most underestimated blind walking results, reported being quite fatigued. As indicated in Fig. 8b, observers displayed much more variability with verbal estimation. This variability is also reflected in Fig. 7b, but Fig. 8b shows that most of the extra variability of verbal estimation comes from between-subject differences. When drawing graphs in the style of Fig. 7a, we found that dropping individual observers with high verbal estimation variability (such as *s*05, *s*16, etc.) substantially changed the verbal estimation lines (dotted orange), while the blind walking means (solid blue) were relatively stable. Because of this variability, we do not have much faith in the verbal estimation results, and we do not inferentially analyze them below.

Therefore, in this experiment, the verbal report protocol did not prove itself to be very useful. While some researchers have reached the same conclusion (Jerome and Witmer [14]), others have found a high correlation between open-loop action-based tasks and verbal report (e.g., Loomis and Knapp [21]). It is possible that we could modify the protocol to reduce the noise; for example, we could have used a modified magnitude estimation procedure where observers state their unit of preference (feet, yards, meters, etc.) ahead of time, and then present a 1-unit example stimulus in their field of view, such as a one-foot ruler, or yardstick, or meterstick.

4.3.2 Analysis Techniques

In this section, we describe how we statistically analyzed our results. In addition to the typical ANOVA analysis, we also subjected the results to a *power analysis*, and the techniques for doing this are described in some detail here. Although some of this material is tutorial in nature, the power analysis discussion has two benefits: 1) it shows how to compute *standardized effect sizes* for most of the previously reported studies in the depth perception literature, and 2) it illustrates how to compute a *null hypothesis confidence interval*, which is the statistically proper technique for arguing the truth of a null hypothesis. To date, we have not encountered a discussion of these techniques in the depth perception literature.

We analyzed our results with univariate analysis of variance (ANOVA); these results are given in Table 4. With ANOVA, we modeled our experiment as a repeated-measures design that considers *observer* a random variable and all other independent variables as fixed (Table 2). The distributions on which ANOVA analysis is based assume that, for each tested effect, the data is normally distributed and the variance is homogenous. For repeated-measures designs such as the ones we report here, these two assumptions are jointly referred to as *sphericity* of the variance/covariance matrix. Sphericity is usually violated [3], [12], and Fig. 7b indicates that it is likely violated in this study, at least across protocol and distance. Therefore, following the recommendations of Howell [12, p. 486] and Buchner et al. [3], for each tested effect we applied the Huynh and Feldt correction ε (Table 4). Instead of the standard F-test on n, d degrees of freedom, where n is the numerator and d the denominator of the F ratio, under this correction we calculate the F-test on εn, εd degrees of freedom. This results in a more conservative test, which corrects for the degree to which sphericity is violated.

In addition to significance testing, in this analysis, we also performed two types of *power analysis* (Cohen [4]): 1) *post-hoc power analysis* and 2) establishing *null hypothesis confidence intervals*. Standard significance testing is based on comparing the calculated p value to α, and rejecting the null hypothesis when $p < \alpha$. Typically, and in this study, $\alpha = 0.05$. α is the probability of committing a Type I error (finding an effect when no effect is present in the data [12]); minimizing this error is why α is set to a small number.

TABLE 4
ANOVA Results for Experiment II

Effect	On	N	ε	n	d	F	p	f^2	r	λ	power
1 Environment ***	all data	1536	0.98	3	45	5.89	0.002	0.39	0.65	49.4	1.00
2 Repetition ***	all data	1536	0.89	3	45	18.75	< .000	1.25	0.74	192.8	1.00
3 Environment ***	blind	768	1.00	3	45	12.54	< .000	0.84	0.38	61.1	1.00
4 Environment ***	blind, not real world, 3 meters	192	1.00	2	30	9.38	0.001	0.63	0.51	38.5	1.00
5 Environment (null)	blind, real+HMD and real+virtual+HMD, 3 meters	128	1.00	1	15	0.28	0.604	0.02	0.51	0.6	0.11
6	Null hypothesis confidence interval		1.00	1	15			0.30	0.50	9.0	0.80
7 Environment (null)	blind, not real world, 5 meters	192	0.78	2	30	1.69	0.208	0.11	0.46	4.9	0.45
8 Environment (null)	blind, not real world, 7 meters	192	1.00	2	30	0.69	0.510	0.05	0.59	3.4	0.33
9	Null hypothesis confidence interval		0.78	2	30			0.26	0.46	11.3	0.81

N is the number of data points analyzed; ε is the Huynh and Feldt correction; n, d are the numerator, denominator degrees of freedom; F is the value of the ANOVA F-Test; p is the conditional probability of the ANOVA F-Test; f^2 is Cohen's effect size; r is the averaged pair-wise correlation; λ is the noncentrality parameter, and power is post-hoc power.

Power analysis calculates a number typically called *power*; 1-*power* is the probability of committing a Type II error (failing to find an effect when one is actually present). Cohen [4] recommends, and we adopt, a goal of achieving $power \geq 0.80$.

Post-hoc power analysis calculates the power of statistically significant findings. Power is a function of three numbers: n, d, and λ, where n is the numerator and d the denominator of the F ratio, and λ is called the *noncentrality parameter*. For a repeated-measures design such the one in this paper,

$$\lambda = \frac{\varepsilon(S-1)nf^2}{1-r}, \tag{1}$$

where ε is the Huynh and Feldt correction factor described above, S is the number of observers in the study, and r is the averaged pair-wise correlation between the levels of the independent variable of the statistically significant finding. f^2 is a standardized measure of effect size for factorial ANOVA designs. As discussed by Cohen [4],

$$f^2 = \frac{\eta^2}{1-\eta^2}, \tag{2}$$

where η^2 (partial eta-squared) is calculated

$$\eta^2 = \frac{nF}{nF+d}, \tag{3}$$

and n, d, and F are the numerator, denominator, and F value of the F-test.

The value of (2) and (3) is that they allow the standardized effect size f^2 to be calculated from the commonly-reported F-test parameters n, d, and F. For example, the effect in Table 4, line 1, would typically be reported $F(3, 45) = 5.89$, $p = .002$; here, $n = 3$, $d = 45$, $F = 5.89$ and (2) and (3) give $f^2 = 0.39$. This allows effect sizes to be computed and compared with previous studies that do not directly report f^2, and most of the studies reported in the depth perception literature give F-tests for important findings. However, (1) shows that λ is a function of ε, S, n, f^2, and r, and while the number of observers S is typically reported, values for ε and r are typically not. Therefore, it is generally not possible to directly compute the power of previously reported repeated-measures designs. Most of the previous studies in the depth perception literature are repeated-measures designs, because the tested distances are usually measured multiple times for each observer, although other variables often vary between observers. For Experiment II, Table 4 gives the values of all of these parameters, as well as the resulting post-hoc power, for each significant effect discussed in the next section. We used G*Power [3] and SPSS to calculate power.

When a finding is not statistically significant (e.g., when $p \geq 0.05$), power analysis can be used to establish a *null hypothesis confidence interval*. In general, a large p value cannot establish the truth of the null hypothesis, because the null hypothesis is a point result (Howell [12]). However, power analysis can bound the possible effect size f^2 to lie within a confidence interval. If the resulting interval is small enough, then the null hypothesis has effectively been argued. Establishing such an interval requires assuming values for the parameters ε, n, d, f^2, and r. In Table 4, lines 6 and 9 list the parameter values that we assumed to establish null hypothesis confidence intervals. In all cases, we chose our parameters to be conservative population estimates, based on the parameter values in the rest of Table 4.

4.3.3 Inferential Results

In this section, when we discuss hypothesis tests, we also give the Table 4 line number that lists the additional parameters. There was a main effect over all of the data ($N = 1,536$ data points) of environment ($F(3, 45) = 5.89$, $p = .002$, line 1), which is explored in more detail below. There was also an effect of repetition ($F(3, 45) = 18.75$, $p < .000$, line 2); observers increased their accuracy with repeated exposure to each condition. This repetition effect also appeared in most of the ANOVAs of subsets of the data that are reported below, but we do not further consider it.

Fig. 9 shows the blind walking error means and standard errors from Figs. 7a and 7b. Within the blind walking data ($N = 768$), there was an effect of environment ($F(3, 45) = 12.54$, $p < .000$, line 3). The standard error bars in Fig. 9 indicate that this is due to a separation between the real world condition and the HMD conditions; unsurprisingly, it was easier to judge the distance of the real-world referent. Interestingly, for the nonreal-world conditions $\text{real} + \text{HMD}$, $\text{real} + \text{virtual} + \text{HMD}$, and $\text{virtual} + \text{HMD}$, the overlap in the error bars suggests that the HMD conditions were equally difficult at 5 and 7 meters. We investigated this possibility by performing separate ANOVAs on the nonreal world

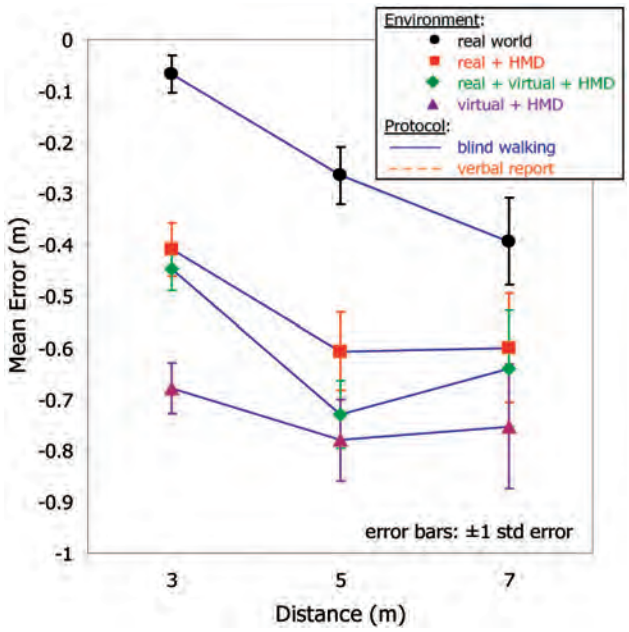

Fig. 9. The mean error results for blind walking $(N = 768)$.

conditions at 3 meters, 5 meters, and 7 meters ($N = 192$ for each test). At 3 meters, as suggested by the separation between the virtual + HMD condition and the other two conditions (real + HMD, real + virtual + HMD), there was still an effect of environment ($F(2, 30) = 9.38$, $p = .001$, line 4). However, a test on the remaining two conditions ($N = 128$) indicated no effect of environment ($F(1, 15) = .28$, $p = .604$, line 5). Furthermore, our experiment could detect effects as small as $f^2 = .30$ with $power = .80$ (line 6), and .30 is small compared to the f^2 sizes of the significant effects just discussed (lines 1-5). At 5 meters, there was no effect of environment for the nonreal-world conditions ($F(2, 30) = 1.69$, $p = .208$, line 7), nor was there an effect at 7 meters ($F(2, 30) = .69$, $p = .510$, line 8). For either of these distances, our experiment could reliably detect effects as small as $f^2 = .26$ with $power = .80$ (line 9).

The relative accuracy of the real-world (control) condition is not surprising; this has been found by many researchers who have compared real-world referents to virtual environment referents (e.g., Thompson et al. [34]). The interesting aspect of these findings, which is implied by the null confidence intervals just presented, is that the real + HMD environment exhibits the same degree of underestimation as both the real + virtual + HMD and virtual + HMD environments (with the exception of the virtual + HMD environment at 3 meters). We hypothesize that the most likely explanation is a combination of the framing effect of our display's narrow field-of-view, as well as the fact that observers were not free to rotate their heads when looking through the HMD. Although some researchers have hypothesized that a limited HMD field-of-view does not cause distance underestimation (Creem-Regehr et al. [5], Knapp and Loomis [16]), Wu et al. [37] found evidence that it does cause underestimation. However, the field-of-view studied for the negative results was $42^\circ \times 32^\circ$ (horizontal $\times$ vertical) (Creem-Regehr et al.) and $47^\circ \times 36^\circ$ (Knapp and Loomis), while Wu et al. only found underestimation when the field of view was restricted to at least $21.2^\circ \times 21.2^\circ$. Our field-of-view was $27^\circ \times 20^\circ$, which compares to Wu et al.'s vertical dimension. Furthermore, Creem-Regehr et al. found that distances were underestimated when head rotations were prevented, and Wu et al. found that distances were not underestimated with a narrow field-of-view when observers were allowed to scan the ground plane in the near-to-far direction (from their feet to the object). Given the size of our HMD's field-of-view and the fact that our HMD's mounting prevented head rotations, our results are consistent with the findings of both Creem-Regehr et al. and Wu et al.

We noticed that when we looked through the display in the real + virtual + HMD environment, and the real object was pulled away, the virtual object seemed to float up from the ground and move closer to us. We hypothesize that the floating upward effect is caused by a lack of cues suggesting that the virtual objects are attached to the ground, and the movement closer is caused by an inward change in vergence angle,[5] driven by accommodative/vergence mismatch. When the accommodative demand (1.2 meters for our HMD) is closer than the fixation distance (3 to 7 meters in this experiment), the resting vergence angle of the eyes shifts inward, causing objects to be perceived as closer than their actual location (Mon-Williams and Tresilian [25]). In the situation described here, when the real and the virtual object are seen together, the eyes accommodate to the real object, and there is no accommodative/vergence mismatch, but when the real object is pulled away, the mismatch occurs. The greater underestimation of the virtual + HMD environment at 3 meters, relative to the real + virtual + HMD and real + HMD environments, is consistent with this hypothesis.

5 Conclusions

AR has many compelling applications, but many will not be realized until we understand how to place graphical objects in depth relative to real-world objects. This is difficult because imperfect AR displays and novel AR perceptual situations such as x-ray vision result in conflicting depth cues. Egocentric distance perception in the real world is not yet completely understood (Loomis and Knapp [21]), and its operation in VR is currently an active research area. Even less is known about how egocentric distance perception operates in AR settings; the comprehensive survey in Section 2 found only seven previously published papers describing unique experiments.

To our knowledge, along with Jerome and Witmer [14] and Kirkley [15], we have conducted the first experiments that have measured AR depth judgments at medium and far-field distances, which are important distances for a number of compelling AR applications. Experiment I used a perceptual matching protocol, and studied distances of 5 to 45 meters. It provides evidence for a switch in bias, from underestimating to overestimating distance, at ~ 23 meters (Fig. 2), and provides an initial quantification of how much more difficult the depth judgment task is in the x-ray vision condition (Fig. 3). It also found an effect of height in the

5. Postexperiment, the first three authors used nonius lines to test for changes in vergence angle for this situation, using a technique similar to the one reported by Ellis and Menges [9]. For all three authors, the test indicated an inward change in vergence angle.

visual field in the form of an interaction with repetition (Fig. 4). We suggest that part of this interaction replicates the VR depth underestimation problem, and further suggest that the effect of practice on VR depth underestimation should be explored. Experiment II used blind walking and verbal report protocols, and studied distances of 3 to 7 meters. Experiment II provides evidence that the egocentric depth of AR objects is underestimated at these distances, but to a lesser degree than has previously been found for most virtual reality environments. Furthermore, the results are consistent with previous studies that have implicated a restricted field-of-view, combined with an inability for observers to scan the ground plane in a near-to-far direction, as explanations for the observed depth underestimation.

The perceptual matching protocol used in Experiment I is generally representative of the types of depth estimation tasks we can imagine users performing in an AR-based situational awareness system such as BARS [19]; such tasks might involve estimating or specifying the distance to urban objects such as buildings, personnel, or vehicles, even if the objects are hidden from sight. While we can also imagine users giving a verbal estimate of depth, we cannot imagine BARS users blind walking. However, as Loomis and Knapp [21] discuss, there are compelling theoretical arguments and substantial empirical evidence that depth judgments from open-loop action-based protocols such as blind walking are driven by a relatively pure percept of egocentric distance. However, to achieve this purity, the protocols must be carefully implemented, in order to counteract cognitive techniques such as footstep counting. In contrast, the depth judgments from the perceptual matching protocol are likely primarily driven by minimizing the exocentric distance between the referent and the target objects, although some percept of egocentric depth of the referent may also be involved. So while there is substantial theoretical value in the blind walking protocol, there is also practical value in studying protocols, such as perceptual matching, that are closer to the real-world tasks we imagine AR users actually performing.

ACKNOWLEDGMENTS

Experiment I was supported by the Advanced Information Technology Branch of the US Naval Research Laboratory, the US Office of Naval Research, and Mississippi State University. Experiment I was conducted at the US Naval Research Laboratory, when the first author was employed by the US Naval Research Laboratory. Experiment II was supported by The Institute for Neurocognitive Science and Technology through a seed grant provided by the US Office of Naval Research, a grant from the US National Aeronautics and Space Administration (NASA), and Mississippi State University. Experiment II was conducted at The Institute for Neurocognitive Science and Technology at Mississippi State University. The authors gratefully acknowledge several very helpful conversations with Stephen R. Ellis of NASA Ames Research Center (Experiment I), and William B. Thompson of the University of Utah (Experiment II). Finally, the detailed feedback of several anonymous reviews has substantially improved this paper.

REFERENCES

[1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S.J. Julier, and B. MacIntyre, "Recent Advances in Augmented Reality," *IEEE Computer Graphics and Applications,* vol. 21, no. 6, pp. 34-47, Nov./Dec. 2001.

[2] M.R. Bax, "Real-Time Lens Distortion Correction: 3D Video Graphics Cards Are Good for More than Games," *Stanford Electrical Eng. and Computer Science Research J.,* Spring, 2004, http://ieee.stanford.edu/ecj/spring04.html.

[3] A. Buchner, F. Faul, and E. Erdfelder, "G*Power: A General Power Analysis Program," http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/, July 2006.

[4] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences,* second ed. Academic Press, 1988.

[5] 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 Environments," *Perception,* vol. 34, no. 2, pp. 191-204, 2005.

[6] J.E. Cutting, "How the Eye Measures Reality and Virtual Reality," *Behavior Research Methods, Instrumentation and Computers,* vol. 29, pp. 29-36, 1997.

[7] J. Decety, M. Jeannerod, and C. Prablanc, "The Timing of Mentally Represented Actions," *Behavioural Brain Research,* vol. 34, pp. 35-42, 1989.

[8] V. Dilda, S.H. Creem-Regehr, and W.B. Thompson, "Perceiving Distances to Targets on the Floor and Ceiling: A Comparison of Walking and Matching Measures [Abstract]," *J. Vision,* vol. 5, no. 8, p. 196a, 2005.

[9] S.R. Ellis and B.M. Menges, "Localization of Virtual Objects in the Near Visual Field," *Human Factors,* vol. 40, no. 3, pp. 415-431, Sept. 1998.

[10] J.M. Foley, "Stereoscopic Distance Perception," *Pictorial Comm. in Virtual and Real Environments,* second ed., S.R. Ellis, M.K. Kaiser, and A.J. Grunwald, eds., pp. 558-566, Taylor & Francis, 1993.

[11] I.P. Howard and B.J. Rogers, "Depth Perception," *Seeing in Depth,* vol. 2, I. Porteus, Ontario, Canada, 2002.

[12] D.C. Howell, *Statistical Methods for Psychology,* fifth ed. Duxbury, 2002.

[13] V. Interrante, L. Anderson, and B. Ries, "Distance Perception in Immersive Virtual Environments, Revisited," *Proc. IEEE Virtual Reality Conf. '06,* 2006.

[14] C.J. Jerome and B.G. Witmer, "The Perception and Estimation of Egocentric Distance in Real and Augmented Reality Environments," submitted manuscript, US Army Research Inst., 2005.

[15] S. Kirkley, "Augmented Reality Performance Assessment Battery (ARPAB)," PhD dissertation, Instructional Systems Technology, Indiana Univ., 2003.

[16] J.M. Knapp and J.M. Loomis, "Limited Field of View of Head-Mounted Displays Is Not the Cause of Distance Underestimation in Virtual Environments," *Presence: Teleoperators and Virtual Environments,* vol. 13, no. 5, pp. 572-577, Oct. 2004.

[17] S.A. Kuhl, W.B. Thompson, and S.H. Creem-Regehr, "Minification Influences Spatial Judgments in Virtual Environments," *Proc. Symp. Applied Perception in Graphics and Visualization (APGV '06),* pp. 15-19, 2006.

[18] M.S. Landy, L.T. Maloney, E.B. Johnston, and M. Young, "Measurement and Modeling of Depth Cue Combination: In Defense of Weak Fusion," *Vision Research,* vol. 35, no. 3, pp. 389-412, 1995.

[19] M.A. Livingston, L.J. Rosenblum, S.J. Julier, D. Brown, Y. Baillot, J.E. Swan II, J.L. Gabbard, and D. Hix, "An Augmented Reality System for Military Operations in Urban Terrain," *Proc. Interservice/Industry Training, Simulation, & Education Conf. (I/ITSEC '02),* 2002.

[20] 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," *Proc. Second Int'l Symp. Mixed and Augmented Reality (ISMAR '03),* pp. 56-65, 2003.

[21] J.M. Loomis and J.M. Knapp, "Visual Perception of Egocentric Distance in Real and Virtual Environments," *Virtual and Adaptive Environments: Applications, Implications and Human Performance Issues,* L.J. Hettinger and J.W. Haas, eds., pp. 21-46, Lawrence Erlbaum Assoc., 2003.

[22] J.W. McCandless, S.R. Ellis, and B.D. Adelstein, "Localization of a Time-Delayed, Monocular Virtual Object Superimposed on a Real Environment," *Presence: Teleoperators and Virtual Environments,* vol. 9, no. 1, pp. 15-24, Feb. 2000.

[23] R. Messing and F.H. Durgin, "Distance Perception and the Visual Horizon in Head-Mounted Displays," *ACM Trans. Applied Perception,* vol. 2, no. 3, pp. 234-250, July 2005.
[24] B.J. Mohler, S.H. Creem-Regehr, and W.B. Thompson, "The Influence of Feedback on Egocentric Distance Judgments in Real and Virtual Environments," *Proc. Symp. Applied Perception in Graphics and Visualization (APGV '06),* pp. 9-14, 2006.
[25] M. Mon-Williams and J.R. Tresilian, "Ordinal Depth Information from Accommodation," *Ergonomics,* vol. 43, no. 3, pp. 391-404, Mar. 2000.
[26] J.M. Plumert, J.K. Kearney, J.F. Cremer, and K. Recker, "Distance Perception in Real and Virtual Environments," *ACM Trans. Applied Perception,* vol. 2, no. 3, pp. 216-233, July 2005.
[27] D.R. Proffitt, "Embodied Perception and the Economy of Action," *Perspectives on Psychological Science,* vol. 1, no. 2, pp. 110-122, 2006.
[28] A.R. Richardson and D. Waller, "The Effect of Feedback Training on Distance Estimation in Virtual Environments," *Applied Cognitive Psychology,* vol. 19, pp. 1089-1108, 2005.
[29] J.P. Rolland, W. Gibson, and D. Ariely, "Towards Quantifying Depth and Size Perception in Virtual Environments," *Presence: Teleoperators and Virtual Environments,* vol. 4, no. 1, pp. 24-49, Winter 1995.
[30] J.P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci, "Method of Adjustment Versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Helmet-Mounted Displays," *Presence: Teleoperators and Virtual Environments,* vol. 11, no. 6, pp. 610-625, Dec. 2002.
[31] R. Sekuler and R. Blake, *Perception,* fourth ed. McGraw-Hill, 2001.
[32] J.E. Swan II, M.A. Livingston, H.S. Smallman, D. Brown, Y. Baillot, J.L. Gabbard, and D. Hix, "A Perceptual Matching Technique for Depth Judgements in Optical, See-Through Augmented Reality," *Proc. IEEE Virtual Reality Conf. '06,* pp. 19-26, 2006.
[33] W.B. Thompson, personal comm., July 2006.
[34] 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: Teleoperators and Virtual Environments,* vol. 13, no. 5, pp. 560-571, Oct. 2004.
[35] B.A. Watson and L.F. Hodges, "Using Texture Maps to Correct for Optical Distortion in Head-Mounted Displays," *Proc. Virtual Reality Ann. Symp. (VRAIS '95),* pp. 172-178, 1995.
[36] 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," *Proc. First Symp. Applied Perception in Graphics and Visualization,* pp. 35-38, 2004.
[37] B. Wu, T.L. Ooi, and Z.J. He, "Perceiving Distance Accurately by a Directional Process of Integrating Ground Information," *Nature,* vol. 428, pp. 73-77, Mar. 2004.

J. Edward Swan II received the PhD degree in computer science, with specializations in computer graphics and human-computer interaction, from Ohio State University in 1997. From 1997 through 2004, he was a scientist with the Virtual Reality Lab at the US Naval Research Laboratory. Since 2004, he has been an associate professor in the Department of Computer Science and Engineering and a Research Fellow in the Institute for Neurocognitive Science and Technology at Mississippi State University. His research centers on perceptual and cognitive aspects of augmented and virtual reality technology and visualization techniques. He is a member of the IEEE and the IEEE Computer Society.

Adam Jones received the BS degree from Mississippi State University in 2004. He is a graduate student in the Department of Computer Science and Engineering at Mississippi State University, and he is affiliated with the Institute for Neurocognitive Science and Technology. His research interests include virtual and augmented reality, medical and scientific visualization, visual perception, cognitive science, and human-computer interaction.

Eric Kolstad received the BS degree in computer science and the MS degree in environmental monitoring from the University of Wisconsin in 1991 and 1994. He is a graduate student in the Computational Engineering PhD program at Mississippi State University, and he is affiliated with the Institute for Neurocognitive Science and Technology. His research interests include geospatial data visualization and feature characterization, 3D simulation, terrain modeling, and augmented reality.

Mark A. Livingston received the AB degree in computer science and mathematics from Duke University in 1993 and the MS and PhD degrees in computer science from the University of North Carolina at Chapel Hill in 1996 and 1998. He is a research scientist in advanced information technology at the Naval Research Laboratory. He directs and conducts research on interactive graphics, including AR, visualization metaphors, mathematical representations, perceptual and cognitive factors, and applications. He is a program cochair for the IEEE/ACM International Symposium on Mixed and Augmented Reality 2007, on the conference committee of IEEE Virtual Reality, and a member of ACM SIGGRAPH and the IEEE and IEEE Computer Society.

Harvey S. Smallman received the PhD degree in experimental psychology from the University of California at San Diego in 1993. He has been a senior scientist at Pacific Science & Engineering Group in San Diego since 1998. He is interested in the mechanisms of visual perception and in how they constrain information visualization. He is a two-time winner of the Jerome Ely Award of the Human Factors and Ergonomics society for best paper in its flagship journal *Human Factors*.