Proceedings of IEEE Virtual Reality 2006, March 25–29, pages 19–26, Alexandria, Virginia, USA, © 2006 IEEE
Awarded ***Honorable Mention*** *as one of 3 best papers at VR 2006*

A Perceptual Matching Technique for Depth Judgments in Optical, See-Through Augmented Reality

J. Edward Swan II[1]
Computer Science and
Engineering Department
Mississippi State University

Mark A. Livingston[2]
3D Virtual and Mixed
Environments Laboratory
Naval Research Laboratory

Harvey S. Smallman[3]
Pacific Science &
Engineering Group

Dennis Brown[4]
3D Virtual and Mixed
Environments Laboratory
Naval Research Laboratory

Yohan Baillot[5]
Advanced Engineering
and Science
ITT Industries

Joseph L. Gabbard[6]
Systems Research Center
Virginia Tech

Deborah Hix[7]
Systems Research Center
Virginia Tech

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 techniques for measuring egocentric depth judgments in both virtual and augmented environments. It then describes a perceptual matching task and experimental design for measuring egocentric AR depth judgments at medium- and far-field distances of 5 to 45 meters. The experiment studied the effect of field of view, the x-ray vision condition, multiple distances, and practice on the task. The paper relates some of the findings to the well-known problem of depth underestimation in virtual environments, and further reports evidence for a switch in bias, from underestimating to overestimating the distance of AR-presented graphics, at ~23 meters. It also gives a quantification of how much more difficult the x-ray vision condition makes the task, and then concludes with ideas for improving the experimental methodology.

CR Categories: H.5 [**Information Interfaces and Presentation**]: H.5.1: Multimedia Information Systems — Artificial, Augmented, and Virtual Realities; H.5.2: User Interfaces — Ergonomics, Evaluation / Methodology, Screen Design

Keywords: 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 AR users 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 [10]. 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 are focused 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 — for example, they often dictate a fixed accommodative focal depth. Furthermore, it is well known that distances are persistently underestimated in VR scenes depicted in head-mounted displays [3, 8, 12, 14, 17, 22, 25, 26], 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 co-located 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. In the work reported in this paper, our larger goal was to study AR depth perception, and our specific goal was to develop an experimental methodology for measuring AR depth judgments at medium- and far-field distances.

2. BACKGROUND AND RELATED WORK

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 [9]. Generally, ten depth cues are recognized [7]: (1) binocular disparity, (2) binocular convergence, (3) accommodative fo-

---

[1] swan@acm.org (corresponding email)
[2] markl@ait.nrl.navy.mil
[3] smallman@pacific-science.com
[4] dbrown@ait.nrl.navy.mil
[5] baillot@simulation3d.biz
[6] jgabbard@vt.edu
[7] hix@vt.edu

cus, (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 of the scene determining the salience of each cue. Although *depth cue interaction* models exist, 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 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 AR technology, as well as depth cue interaction models.

NEAR-, MEDIUM-, AND FAR-FIELD DISTANCES: Depth cues vary both in their salience across real-world scenes, and in their effectiveness by distance. Cutting [2] 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, different combinations of depth cues are available.

EGOCENTRIC DISTANCE JUDGMENT TECHNIQUES: In the development of AR (and VR) environments, we are interested in measuring the perception of distance, but we suffer from the classic problem that perception is an invisible cognitive state, and so we have to find something measurable which can be theoretically related to the perception of distance. Therefore, we devise experiments where we measure distance *judgments*, and then infer distance perception from these judgments. The most general categorization of the judgments we can measure is ego- 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 [12] and Foley [5] review and discuss the methods that have been developed to measure judged egocentric distances.

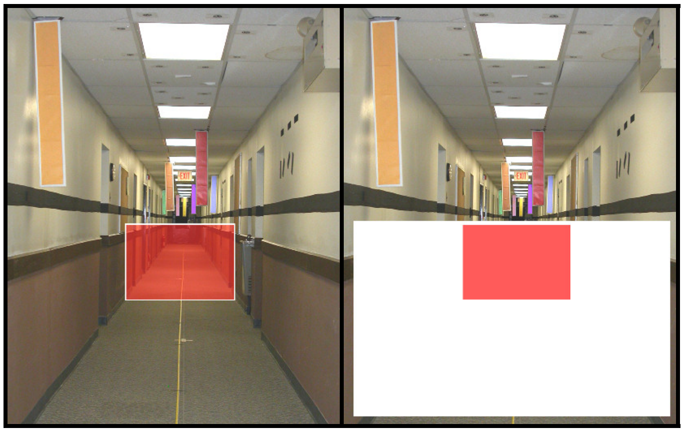
There have been three primary methods: *verbal report*, *perceptual matching*, and *open-loop action-based tasks*. With *verbal report* [5, 8, 12, 14] observers verbally estimate the distance to an object, typically using whatever units they are most familiar with (e.g., feet, meters, or multiples of some given referent distance). Observers have also verbally estimated the size of familiar objects [12], which are then used to compute perceived distance. *Perceptual matching* tasks [4, 5, 13, 19, 26] 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- 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. A common open-loop action-based task has been *visually directed walking* [3, 8, 12, 14, 25, 26], 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. Visually directed walking has been found to be very accurate for distances up to 20 meters [12], and 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 visually directed walking* [17], 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* [12, 22, 25], 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 [5, 15], where observers indicate distance with a finger or manipulated slider that is hidden from view.

In addition, some researchers have used *forced-choice* tasks [11, 18, 19] 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 [18, 19].

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) [3, 8, 12, 14, 17, 22, 25, 26]. 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 [26]. 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 [26]; that underestimation is *not* caused by an HMD's limited field-of-view [3, 8]; that the weight of the HMD itself might contribute to the phenomenon [25]; that monocular versus stereo viewing does not cause it [3]; that the quality of the rendered graphics does not cause it [22]; that the effect persists even when observers see live video of the real world in an HMD [14]; and that the effect might exist when VR is displayed on a large-format display screen as well [17]. In summary, the egocentric distance underestimation effect is real, and although its parameters are being explored, it is not yet fully understood.

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 [4] 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 an occluding surface (the x-ray vision



(a) referents in *upper field of view*, occluder *absent*

(b) referents in *upper field of view*, occluder *present*

(c) referents in *lower field of view*, occluder *absent*

(d) referents in *lower field of view*, occluder *present*

Figure 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.

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 towards 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. [13] 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. [18], 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. [19] 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. Livingston et al. [11] discuss an experiment that 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.

3. AR DEPTH EXPERIMENT

We developed a perceptual matching technique for measuring AR depth judgments. As we developed our experimental protocol, setting, and task, we pursued the following design goals:

- Study medium- and far-field distances, which interest us because they have not been well-studied in AR, different depth cues operate at these distances, and these distances are meaningful in our application domain [10]. We studied distances between 5.25 and 44.31 meters.
- Compare the occluded (x-ray vision) condition to the non-occluded condition.
- Require observers to simultaneously attend to the real world and virtual objects in order to correctly perform the task. This addresses a criticism of some previous AR studies [6, 11], where observers could essentially ignore the real world and yet still perform the task.
- Ensure that our task is not 2D solvable, but requires a depth judgment to correctly perform. A *2D solvable* task can be solved by only attending to 2D geometry. For example, if we used height in the visual field to encode the depth of two virtual objects, and then asked observers which one was farther, they could correctly answer by simply noting which had the greater 2D y-coordinate.
- Control the ratio of environmental illumination to AR display brightness. Even though our application domain calls for using AR outdoors, we needed to control this ratio because our AR system and display cannot adjust to or match outdoor illuminance values [6]. Therefore, we found an indoor space (a hallway) that was large enough to study medium- and far-field distances, and we covered the windows with thick black felt.

3.1 Experimental Task

We measured depth judgments with a perceptual matching task. Figure 1 shows the experimental setting. We seated observers on a tall stool 3.4 meters from one end of a 50.1-meter long hallway. Observers looked down the hallway, through an optical, see-through AR display mounted on a frame. We mounted the display so the center of each lens was 147.3 cm above the floor, and we adjusted the height of the stool so that observers could comfortably look through the display. Because the display was rigidly mounted, each observer saw exactly the same field-of-view. Observers saw a series of eight real-world *referents*, approximately positioned evenly down the hallway (Figure 1). Each referent was a different color. The AR display showed a virtual *target*, which we drew as a semi-transparent rectangle that horizontally filled the hallway, and vertically extended about half of the hallway's height. We utilized a rectangular target because our application domain [10] involves the AR presentation of rectangular building elements, such as hallways and doorways. Observers placed their right hand on a trackball; by rolling the trackball forwards and backwards, they moved the target in depth up and down the hallway.

For each trial, our software drew the target rectangle at a random initial depth position; it drew the target rectangle with a

Table 1: Independent and dependent variables.

INDEPENDENT VARIABLES				
observer	8	(random variable)		
(referent) *field of view*	2	upper, lower		
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				
absolute error	| *judged distance – actual distance* |, *meters*			
signed error	*judged distance – actual distance, meters* +: observer overestimated target distance –: observer underestimated target distance			

white border, and colored the target interior to match the color of one of the referents (Figure 1). The software smoothly modulated the opacity of the color according to distance: close to the observer the color was more opaque, and it grew progressively more transparent with increasing distance. This was in addition to the transparency of the graphics induced by the AR display; Livingston et al. [11] previously determined this to be an effective graphical technique for distance encoding, which approximates the depth cue of atmospheric haze. The software also printed a text label that named the color at the bottom of the display screen. The observer's task was to adjust the target's depth position until it matched the depth of the referent with the same color (Figure 1). 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–100B stereo optical see-through display. We increased the display's transparency by removing the LCD opacity filter, and we set the display brightness to its maximum setting. Our Glasstron displays 800 × 600 (horizontal by vertical) pixels in a transparent window which subtends 28.0° × 21.3°[1], and thus each pixel subtends approximately .033° × .033°. This window is approximately centered in a larger semi-transparent frame, which is tinted like sunglasses and so attenuates the brightness of the real world. The outer edge of this frame subtends 63.3° × 39.7°. We stereo calibrated the display by stereo-aligning a rectangle that matched a rectangular window at the far end of the hallway; in Figure 1 this window is covered by heavy black felt and so is not visible. We had to slightly rotate (yaw and pitch) the scene in each eye in order to horizontally and vertically stereo-align the stimuli; we perform this rotation in software. Because the display was rigidly mounted and not tracked, we only had to calibrate the display once; it was not recalibrated on a per-observer basis. We ran the experiment on a Pentium IV 3.06 GHz computer with an Nvidia Quadro4 graphics card, which outputs frame-sequential stereo. We split the video signal, sending one signal to the AR display, and one to a monitor, so we could see the observers' progress. We implemented our experimental control code in Java.

3.2 Variables and Design

3.2.1 Independent Variables

OBSERVERS: We recruited eight observers from a population of scientists and engineers. Seven of the observers were male, one was female; they ranged in age from 21 to 47. We screened the observers, via self-reporting, for color blindness and visual acuity. All observers volunteered and received no compensation.

FIELD OF VIEW: As shown in Figure 1, we placed the referents in the observer's *upper* and *lower* field of view, by mounting 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. We manipulated field of view in this experiment because we earlier ran a four-observer pilot experiment with the same task, but with the referents exclusively in the lower field of view. The pilot data suggested that observers consistently underestimated target depth, similar to the results that have been found in virtual environments [3, 8, 12, 14, 17, 22, 25, 26]. Wu, Ooi, and He [26] have argued that this effect is caused by a patch of far ground surface, which is actually flat, being perceived as tilted towards the vertical. Tyler [23] found that objects slightly closer in the lower field of view were judged equidistant to objects in the upper field of view. Because our experimental setup (Figure 1) has a ceiling with rich perspective depth cues, we decided to test referents mounted on both the ceiling and the floor. All of the studies which show distance underestimation in virtual environments cited in this paragraph studied referent objects on the ground plane, and hence (using the terminology of this paper) in the observer's lower field of view.

OCCLUDER: As discussed above, we are interested in understanding AR depth perception in the x-ray vision condition. When the *occluder* was *absent* (Figure 1, (a) and (c)), observers could see the hallway behind the target. When the *occluder* was *present* (Figure 1, (b) and (d)), we mounted a heavy rectangle of foamcore posterboard across the observer's field-of-view, which occluded the view of the hallway behind the target. We carefully positioned the occluder so that it did not cut off the observer's view of the bottom (top) of the referents, and yet so it fully occluded the target throughout the entire possible depth range.

Because the hallway's linear perspective becomes quite compressed at 50 meters, we had to calibrate the position of the occluder and the display. In fact, the tightness of this positioning was our original motivation for rigidly mounting the display: without it, observers could easily look over (or under) the occluder to see an unoccluded view of the target, by moving their head up or down only a few centimeters. In addition, our hallway contains a dark, wooden molding between the brown-colored lower walls and the cream-colored upper walls (Figure 1). In the occluded condition, when the referents were in the lower field of view (Figure 1 (d)), this molding formed a strong linear perspective cue that was missing when the field of view was reversed (Figure 1 (b)). Therefore, we carefully positioned and applied black gaffer's tape to the upper walls, which yielded a comparable linear perspective cue in both field of view conditions.

---

[1] Angular measures in this paper are in degrees of visual arc.

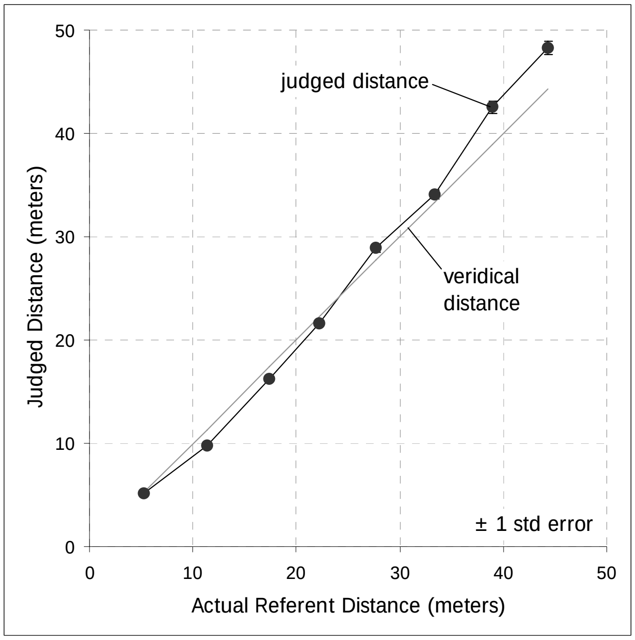


Figure 2: The main result, plotted as judged distance versus actual referent distance. The light grey line indicates veridical performance. For this and all figures, absent error bars indicate the standard error is smaller than the symbol size.

REFERENT DISTANCE: We placed the eight referents at the distances from the observer indicated in Table 1; these distances are measured from the front of the Glasstron AR display. We positioned the referents left and right in the visual field so that they were all visible from the observer's position. As indicated in Figure 1, we placed three of the referents adjacent to a wall and the last referent in the very center; we slightly offset the remaining four referents from the center. The width of the referents subtended from 1.75° to .206°; the farthest referent was over 12 times wider than the standard limit of visual acuity of about 1 minute of visual arc [20]. In person, it was easier to perceive the far referents than it is to see them in Figure 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; these are the eight chromatic colors from the eleven *basic color terms*, which are the colors with one-word English names that Smallman and Boynton [21] have shown to be maximally discriminable and unambiguously named, even cross-culturally (the remaining color terms are 'white', 'black', and 'grey'). We created the colors by printing single-colored sheets of paper with a color printer. To increase the contrast of the referents, we created a border around each color with white gaffer's tape. We affixed the referents to the ceiling and floor with Velcro.

REPETITION: We presented each combination 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 *signed error*.

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 2560 data points (8 observers * 2 fields of view * 2 occluder states * 8 distances * 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.

Each observer first read and signed a consent form, and then took a stereo acuity test, which all observers passed. The observer next completed 5 practice trials, which used a clear, colorless target rectangle that was only perceptible because of its white border; we verbally asked the observer to place the target on random referents until we felt that the observer understood the task. The observer next completed four blocks of 80 trials each. Between blocks the observer rested for as long as they desired, but at least long enough for us to either mount or dismount the occluder, and to move all of the referents from the floor to the ceiling or vice versa. The entire procedure took from 60 to 90 minutes to complete.

3.3 Results

We analyzed our results with analysis of variance (ANOVA) and regression analysis. 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 1). This type of design factors out between-subject differences; it allows greater sensitivity for detecting experimental effects with fewer observers, but at the cost of not allowing us to examine individual differences. Eight observers allowed us to detect main effects as small as 1.04 meters for signed error ($N = 1280$, *power* = 95%, $\alpha = 5\%$, $\sigma = 7.27$ meters) and .79 meters for absolute error ($N = 1280$, *power* = 95%, $\alpha = 5\%$, $\sigma = 5.55$ meters), and these effect sizes are small compared to the effects discussed in this section. Therefore, eight observers was an adequate number of subjects for this study.

When deciding which results to report, in addition to considering the p value, the standard measure of *effect significance*, we also considered η^2 (eta-squared), a standard measure of *effect size*. η^2 is an approximate measure of the percentage of the observed variance that can be explained by a particular effect, and is an appropriate effect size measure for a non-additive repeated-measures design [24].

Figure 2 summarizes the main experimental results, which by convention is given as a correlation between the actual referent distances and the judged distances. Theoretically perfect (veridical) performance is indicated by the diagonal line. The data indicate distance underestimation for referents 2, 3, and 4, followed by increasing distance overestimation. This trend is analyzed in more detail below.

Figure 3(a) shows that the variability (expressed as the standard error of the mean) of the judged target distance grew linearly (r^2 = 96.5%) with increasing referent distance, and Figure 3(b) shows that absolute error also grew linearly (r^2 = 93.7%) with increasing referent distance; Figure 3(b) also shows a main effect

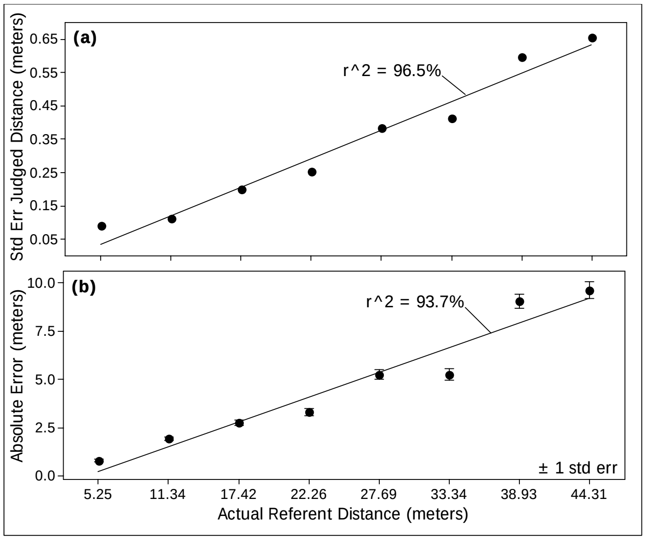


Figure 3: As the referent distance increased, (a) the variability and (b) the absolute error of the estimated target distance grew in a linear fashion. Both regressions indicate decreasing depth cue effectiveness with distance.

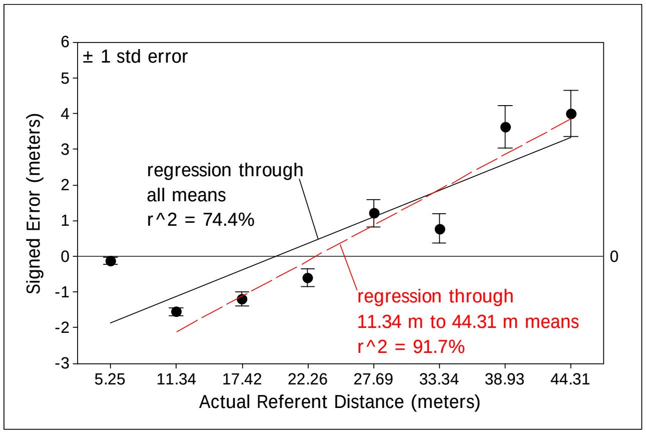


Figure 4: The effect of distance on signed error. Signed error exhibits a strong linear regression beginning at 11.34 meters, which reveals a switch in bias from underestimating to overestimating target distance at ~23 meters.

of distance on absolute error ($F(7,49) = 30.5$, $p < .000$, $\eta^2 = 29.4\%$). Both regressions demonstrate that our experimental task is not 2D solvable, and is in fact measuring a depth judgment, because the linear relationship with distance indicates judgments based on depth cues of linearly decreasing effectiveness (e.g., observer responses are following a Weber's law [20]). In this experiment, observers made depth judgments with virtual targets, and therefore the experiment lacks the "ground truth" that comes from tasks where observers manipulate a real-world target to match a virtual referent, such as Ellis and Menges [4] and McCandless et al. [13]. Therefore, the correlations in Figure 3 are an important validation of the experimental methodology. Loomis and Knapp [12] use a similar line of reasoning, which relates errors to depth cue availability, to validate open-loop action-based tasks, and McCandless et al. [13] found monotonic increases in both variation and error with increasing distance.

Figure 4 shows the effect of distance on signed error ($F(7,49) = 3.20$, $p = .007$, $\eta^2 = 7.31\%$). Like unsigned error (Figure 3), signed error shows a linear relationship with increasing distance ($r^2 = 74.4\%$; solid line in Figure 4). 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 signed error and distance increases when analyzed for referents 2–8 ($r^2 = 91.7\%$; dashed line in Figure 4). Even more interesting is a shift in bias from underestimating (referents 2–4) to overestimating (referents 5–8) distance; this bias shift is also seen in Figure 2. The bias shift occurs at around 23 meters, which is where the dashed line in Figure 4 crosses zero meters of signed error. Foley [5] 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 is an important depth cue in this experimental setting.

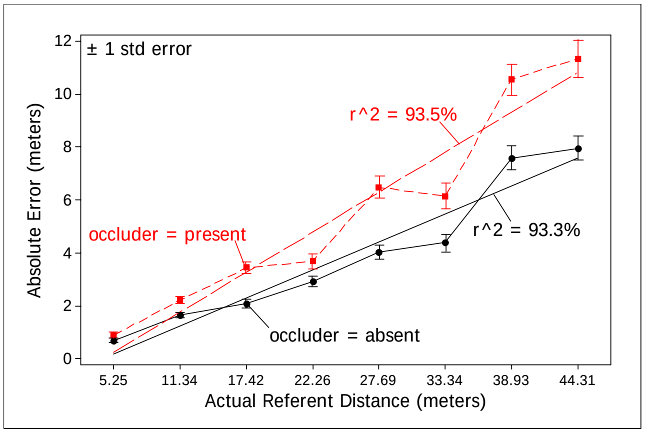


Figure 5: Effect of occluder by distance on absolute error. Observers had more error in the occluded (x-ray vision) condition (dashed line and points) than in the non-occluded condition (solid line and points), and the difference between the occluded and non-occluded conditions increased with increasing distance.

We found a main effect of occluder on absolute error ($F(1,7) = 5.78$, $p = .047$, $\eta^2 = 2.28\%$); when the occluder was absent, observers made an average depth judgment error of 3.91 meters, versus 5.59 meters when the occluder was present. This effect was expected because fewer depth cues are available when the occluder is present. We also found an occluder by distance interaction on absolute error (Figure 5, $F(7,49) = 2.06$, $p = .066$; $\eta^2 = .97\%$). When an occluder was present (the x-ray vision condition), observers had more error then when the occluder was absent, and the difference between the occluder present and occluder absent conditions increased with increasing distance. Figure 5 shows a linear modeling of the occluder present condition (dashed line), which explains $r^2 = 93.5\%$ of the observed variance, and a linear modeling of the occluder absent condition (solid 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:

$$y_{\text{present}} - y_{\text{absent}} = .08x - .33, \tag{1}$$

where y_{present} is the occluder present (dashed) line, y_{absent} is the occluder absent (solid) line, and x is distance. This equation says that for every additional meter of distance, observers made 8 cm

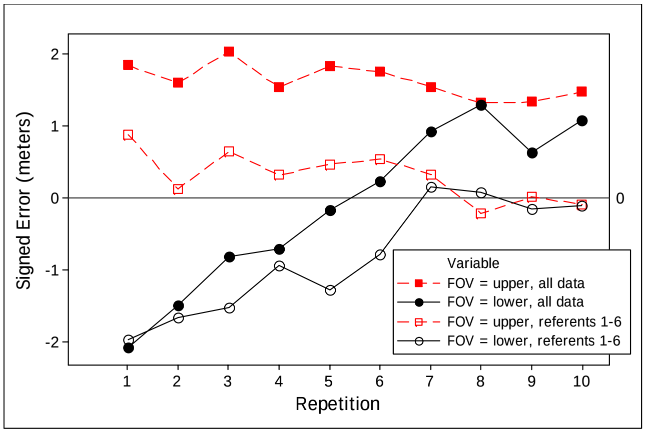


Figure 6: Effect of field of view (FOV) by repetition on signed error. Solid shapes (■,●) are means for all the data; hollow shapes (□,○) are means for the first six referents. Squares (■,□) are referents in the upper field of view; circles (●,○) are referents in the lower field of view. For clarity, standard error bars are not shown.

of additional error in the occluder present versus the occluder absent condition.

We found a field of view by repetition interaction on signed error ($F(9,63) = 3.24$, $p = .003$; $\eta^2 = .72\%$). This is shown by the solid shapes (■,●) in Figure 6. When the referents were in the upper field of view (■, mounted on the ceiling), observers overestimated their distance by about 1.5 meters, and when the referents were in the lower field of view (●, mounted on the floor), observers began with an underestimation (low repetitions), and with practice, by repetition 8 matched the overestimation of the upper field of view. The general bias towards overestimation can be explained by the overestimation of the last two referents, as seen in Figures 2 and 4. In Figure 6 the hollow shapes (□,○) show the field of view by repetition interaction when the last two referents are removed; the interaction is still significant for this reduced data set ($F(9,63) = 2.44$, $p = .019$; $\eta^2 = 1.02\%$). When the referents were in the upper field of view (□), observers did not show a bias, and by repetition 7 were quite accurate. For referents in the lower field of view (○), 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 upper field of view (□) referents.

These results raise the question as to why distance judgments of referents in the lower field of view were initially underestimated. We propose that the results for repetitions 1–3 of the lower field of view referents (○) demonstrate the same distance underestimation that has been demonstrated by VR environment studies [3, 8, 12, 14, 17, 22, 25, 26]. All of these studies share the following properties: (1) they demonstrated distance underestimation for virtual environments presented in HMDs; (2) they measured distance judgments to referent objects in the lower field of view (placed on the ground plane); (3) they used open-loop action-based tasks (primarily visually directed walking and triangulation by walking); and (4) observers completed 1–3 repetitions of each experimental condition[1]. The results for repetitions 1–3 of the lower field of view referents (○) share all of these properties except for property 3: here the underestimation is demonstrated with a perceptual matching task (although Wu, Ooi, and He [26] found the underestimation for both a perceptual matching and a visually directed walking task).

Wu, Ooi, and He [26] also found that, with 2 repetitions and when observers cannot look around, a vertical view subtending 29.6° is adequate for accurate depth judgments, but a vertical view subtending 21.1° causes distance underestimation. This compares to the transparent window of our display, which allows a 21.3° vertical view. It is possible that this explains the distance underestimation for the first several repetitions of the lower field of view referents (○). But regardless of the explanation, the facts that (1) with practice observers became more accurate when placing lower field of view referents and (2) the methodologies of this study and the VR depth underestimation studies [3, 8, 12, 14, 17, 22, 25, 26] were very similar, suggest that the general VR distance underestimation effect might be transitory, and could disappear with practice.

4. DISCUSSION

As mentioned in the Introduction, 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 [12]), 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 five previously published papers describing unique experiments. The current study contributes to the important task of understanding AR depth perception.

To our knowledge, we have conducted the first experiment that has measured AR depth judgments at medium- and far-field distances, which are important distances for a number of compelling AR applications. We have demonstrated a perceptual matching task, and found a linear relationship between distance and depth judgment variability and error (Figure 3), which argues for the validity of our results. We have also detected evidence for a switch in bias, from underestimating to overestimating distance, at ~23 meters (Figure 4), and we have made an initial quantification of how much more difficult the depth judgment task is in the x-ray vision condition (Figure 5). Finally, we found an effect of field of view in the form of an interaction with repetition (Figure 6). 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.

The finding of a bias switch at ~23 meters (Figure 4) immediately suggests distorting the graphics so that depth is judged veridically regardless of distance. However, before pursuing this goal, the reliability of the bias switch needs to be verified by additional studies, especially ones which utilize open-loop action-based depth judgment tasks such as visually directed walking or triangulation by walking. In addition to verifying the bias switch, such studies would allow us to more closely compare our results to the VR depth perception literature. If the bias switch proves to be reliable, an important theoretical goal would be to explain, in the language of cue theory, precisely why it occurs. Such a description would likely indicate the most efficient way to counteract the bias switch.

---

[1] Messing and Durgin [14] point out that the small number of repetitions is part of the visually directed walking methodology; this is done so that observers do not develop strategies (such as counting footsteps) which do not depend on egocentric distance perception.

5. METHODOLOGICAL IMPROVEMENTS

In hindsight, we have determined at least two areas where the reported experimental methodology needs improvement:

- In our study observers' eyes were all at the same height, but there is ample evidence that the human visual system uses the angular declination below the horizon as an absolute egocentric distance cue for objects on the ground plane [12, 16]. Because this is calibrated by an individual's eye height, future studies should place observers at their standing eye height.
- Our targets used a high-contrast white border around a featureless interior (Figure 1). This high-contrast border is a very salient cue for stereo disparity judgments; and it is known that stereo disparity is more sensitive in the center of the visual field [7]. In our study design the target became smaller as the distance increased, and this could have made stereo disparity a more salient cue with increasing distance. Future studies should consider and perhaps control for this potentially confounding depth cue.

ACKNOWLEDGEMENTS

This work was supported in part by the Advanced Information Technology Branch of the Naval Research Laboratory, the Office of Naval Research, and Mississippi State University. The experiment was conducted at the Naval Research Laboratory, when the first author was employed by the Naval Research Laboratory. We gratefully acknowledge several very helpful conversations with Stephen R. Ellis of NASA Ames Research Center.

REFERENCES

[1] AZUMA, R, BAILLOT, Y, BEHRINGER, R, FEINER, S, JULIER, S, MACINTYRE, B, "Recent Advances in Augmented Reality", *IEEE Computer Graphics & Applications*, 21(6), Nov / Dec 2001, pages 34–47.

[2] CUTTING, JE, "Reconceiving Perceptual Space", in M ATHERTON, H HECHT, R SCHWARTZ (EDS.) *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space*, Cambridge, MA, MIT Press, 2003, pages 215–238.

[3] CREEM-REGEHR, SH, WILLEMSEN, P, GOOCH, AA, THOMPSON, WB, "The Influence of Restricted Viewing Conditions on Egocentric Distance Perception: Implications for Real and Virtual Environments", *Perception*, 34(2), 2005, pages 191–204.

[4] ELLIS, SR, MENGES, BM, "Localization of Virtual Objects in the Near Visual Field", *Human Factors*, 40(3), September 1998, pages 415–431.

[5] FOLEY, JM, "Stereoscopic Distance Perception", in SR ELLIS, M KAISER, AJ GRUNWALD (EDS), *Pictorial Communication in Virtual and Real Environments (2nd Ed)*, Taylor & Francis, 1993, pages 558–566.

[6] GABBARD, JL, SWAN II, JE, HIX, D, SCHULMAN, RS, LUCAS, J, GUPTA, D, "An Empirical User-Based Study of Text Drawing Styles and Outdoor Background Textures for Augmented Reality", *Technical Papers, IEEE Virtual Reality 2005*, pages 11–18.

[7] HOWARD, IP, ROGERS, BJ, *Depth Perception*, Volume 2 of *Seeing in Depth*, I. Porteus, Ontario, Canada, 2002.

[8] KNAPP, JM, LOOMIS, JM, "Limited Field of View of Head-Mounted Displays Is Not the Cause of Distance Underestimation in Virtual Environments," *Presence*, 13(5), October 2004, pages 572–577.

[9] LANDY, MS, MALONEY, LT, JOHNSTON, EB, YOUNG, M, "Measurement and Modeling of Depth Cue Combinations: In Defense of Weak Fusion", *Vision Research*, 35, 1995, pages 389–412.

[10] LIVINGSTON, MA, ROSENBLUM, L, JULIER, SJ, BROWN, D, BAILLOT, Y, SWAN, J.E. II, GABBARD, JL, HIX, D, "An Augmented Reality System for Military Operations in Urban Terrain", In *Proceedings of the Interservice / Industry Training, Simulation, & Education Conference (I/ITSEC '02)*, Orlando, FL, Dec 2–5, 2002.

[11] LIVINGSTON, MA, SWAN II, JE, GABBARD, JL, HÖLLERER, TH, HIX, D, JULIER, SJ, BAILLOT, Y, BROWN, D, "Resolving Multiple Occluded Layers in Augmented Reality", *The 2nd International Symposium on Mixed and Augmented Reality (ISMAR '03)*, October 7–10, 2003, Tokyo, Japan, pages 56–65.

[12] LOOMIS, JM, KNAPP, JM, Visual Perception of Egocentric Distance in Real and Virtual Environments, in LJ HETTINGER, JW HAAS (EDS.), *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues*, Lawrence Erlbaum Associates, 2003, pages 21–46.

[13] MCCANDLESS, JW, ELLIS, SR, ADELSTEIN, BD, "Localization of a Time-Delayed, Monocular Virtual Object Superimposed on a Real Environment", *Presence*, 9(1), Feb 2000, pages 15–24.

[14] MESSING, R, DURGIN, FH, "Distance Perception and the Visual Horizon in Head-Mounted Displays", *ACM Transactions on Applied Perception*, 2(3), July 2005, pages 234–250.

[15] MON-WILLIAMS, M, TRESILIAN, JR, "Ordinal depth information from accommodation?", *Ergonomics*, 43(3), March 2000, pages 391–404.

[16] OOI, TL, WU, B, HE, ZJ, "Distance Determined by the Angular Declination below the Horizon", *Nature*, Vol. 414, November 8th, 2001, pages 197–200.

[17] PLUMERT, JM, KEARNEY, JK, CREMER, JF, RECKER, K, "Distance Perception in Real and Virtual Environments", *ACM Transactions on Applied Perception*, 2(3), July 2005, pages 216–233.

[18] ROLLAND, JP, GIBSON, W, ARIELY, D, "Towards Quantifying Depth and Size Perception in Virtual Environments", *Presence*, 4(1), Winter 1995, pages 24–49.

[19] ROLLAND, JP, MEYER, C, ARTHUR, K, RINALDUCCI, E, "Method of Adjustment versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Helmet-Mounted Displays", *Presence*, 11(6), Dec 2002, pages 610–625.

[20] SEKULER, R, BLAKE, R, *Perception, 4th ed.*, McGraw-Hill, 2001.

[21] SMALLMAN, HS, BOYNTON, RM, "On the Usefulness of Basic Colour Coding in an Information Display", *Displays*, 14, 1993, pages 158–165.

[22] THOMPSON, WB, WILLEMSEN, P, GOOCH, AA, CREEM-REGEHR, SH, LOOMIS, JM, BEALL, AC, "Does the Quality of the Computer Graphics Matter when Judging Distances in Visually Immersive Environments?", *Presence: Teleoperators & Virtual Environments*, 13(5), October 2004, pages 560–571.

[23] TYLER, CW, "The Horopter and Binocular Fusion", In D REGAN (ED.), *Binocular Vision*, Boca Raton, FL, CRC Press, 1991, chapter 9, pages 19–37.

[24] VAUGHAN, G, CORBALLIS, MC, "Beyond Tests of Significance: Estimating Strength of Effects in Selected ANOVA Designs", *Psychological Bulletin*, 72(3), 1969, pages 204–213.

[25] WILLEMSEN, P, COLTON, MB, CREEM-REGEHR, SH, THOMPSON, WB, "The Effects of Head-Mounted Display Mechanics on Distance Judgments in Virtual Environments", Proceedings of the *1st Symposium on Applied Perception in Graphics and Visualization*, ACM International Conference Proceedings Series, Vol. 73, 2004, pages 35–38.

[26] WU, B, OOI, TL, HE, ZJ, "Perceiving Distance Accurately by a Directional Process of Integrating Ground Information", *Nature*, Vol. 428, March 4th, 2004, pages 73–77.