Technical Report #MSU-050531, Department of Computer Science and Engineering, Mississippi State University, 2005.

A Methodology for Quantifying Medium- and Far-Field Depth Perception in Optical, See-Through Augmented Reality

J. Edward Swan II[2*] Mark A. Livingston[1†] Harvey S. Smallman[3] Joseph L. Gabbard[4] Dennis Brown[1]
Yohan Baillot[1] Simon J. Julier[1] Greg S. Schmidt[1] Catherine Zanbaka[5] Deborah Hix[4]
Lawrence Rosenblum[1]

[1]Naval Research Laboratory
[2]Mississippi State University
[3]Pacific Science & Engineering Group
[4]Virginia Tech
[5]University of North Carolina at Charlotte

Abstract

A fundamental problem in optical, see-through augmented reality (AR) is characterizing how it affects human depth perception. 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. However, achieving this is difficult, because the graphics are physically drawn directly in front of the eyes. Furthermore, AR makes possible enhanced perceptual techniques that have no real-world equivalent, such as *x-ray vision*, where AR users perceive that graphics are located behind opaque surfaces. Also, to date AR depth perception research has examined near-field distances, yet many compelling AR applications operate at longer distances, and human depth perception itself operates differently at medium-field and far-field distances.

This paper describes the first medium- and far-field AR depth perception experiment that provides metric results. We describe a task and experimental design that measures AR depth perception, with strong linear perspective depth cues, and matches results found in the general depth perception literature. Our experiment quantifies how depth estimation error grows with increasing distance across a range of medium- to far-field distances, and we also find evidence for a switch in bias from underestimating to overestimating depth at ~19.4 meters. Our experiment also examined the x-ray vision condition, and found initial evidence of how depth estimation error grows for occluded versus non-occluded graphics.

Keywords: Augmented Reality Depth Perception, Optical See-Through Augmented Reality

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

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.

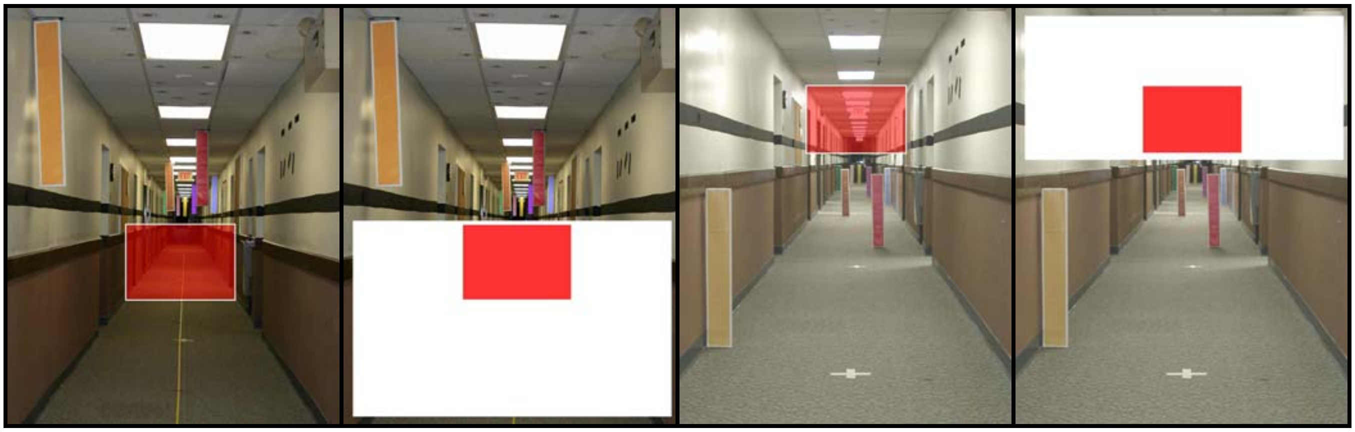
The AR community is applying AR technology to a number of unique and useful applications [Azuma et al. 2001]. The application that motivated the work described here is mobile, outdoor AR for situational awareness in urban settings [Livingston et al. 2002]. This is a very difficult application domain 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. 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. Yet current AR displays are compromised in their ability to display depth (for example, they often dictate a fixed focal depth), and it is not yet known if this is simply due to engineering limitations, or if the limits are more fundamental. Furthermore, there is no real-world equivalent to x-ray vision, and how the human visual system processes x-ray visual information is not yet understood, much less the depth accuracy limitations for applications such as the ones mentioned above.

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 [Landy et al. 1995]. Generally, ten depth cues are recognized [Howard and Rogers 2002]: (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 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 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, AR depth perception will likely inform both AR technology, as well as depth cue interaction models.

---

[*] swan@acm.org
[†] mark.livingston@nrl.navy.mil



(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. Subjects manipulated the depth of the target rectangle to match the depth of the real-world referent with the same color (red in this example). (b) and (d) show the x-ray vision condition.

2 Related Work

Depth cues vary both in their salience across real-world scenes, and in their effectiveness by distance. Cutting [2003] 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.

There have been a small number of studies that have examined depth perception with optical, see-through AR displays. Ellis and Menges [1998] summarize a series of AR depth experiments, which examined near-field distances of 0.4 to 1.0 meters, and studied an occluding surface (the x-ray vision condition), convergence, accommodation, subject age, and monocular, biocular, and stereo AR displays. McCandless et al. [2000] used the same experimental setup and task to additionally study motion parallax, AR system latency, and the effect of cutting a hole in the occluding surface. In all of these experiments, subjects used a *method of adjustment* technique: they manipulated the depth of a real object to match the depth of a virtual object. Rolland et al. [1995] discuss a pilot study at near-field distances of 0.8 to 1.2 meters, which examined depth perception of real and virtual objects. The study used a *forced choice* technique, where subjects must choose one object as "closer" or "farther" than a reference object. Rolland et al. [2002] ran further experiments that examined these topics, but used an improved AR display, and compared forced-choice to method of adjustment techniques. Livingston et al. [2003] discuss an experiment that examined graphical techniques such as drawing style, intensity, and opacity on occluded AR objects at far-field distances of 60 to 500 meters; they used a forced-choice technique.

In addition to the experiments reported above, a large number of visualization tools and interactive techniques have been proposed for viewing and manipulating objects in depth in virtual and augmented reality systems, including hidden or occluded (x-ray vision) conditions. Bane and Höllerer [2004] describe one current effort, which gives a set of far-field, x-ray vision techniques for visualizing the interior structure of buildings. Their paper also contains an extensive review of the work in this area.

3 AR Depth Experiment

When developing 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 [Livingston et al. 2002]. We studied distances between 5.25 and 44.31 meters.
- Determine the fidelity (ordinal or metric) of AR depth perception at these distances. *Ordinal* fidelity means subjects could only make judgments such as "in front of" or "behind", while *metric* implies a continuous sense of depth. We therefore used a method-of-adjustment technique, which allows metric measurements, as opposed to a forced-choice technique, which would only allow ordinal measurements.
- Compare the occluded (x-ray vision) condition to the non-occluded condition.
- Require subjects to simultaneously attend to the real world and virtual objects in order to correctly perform the task. This addresses a criticism of some previous work [Livingston et al. 2003; Gabbard et al. 2005], where subjects could essentially ignore the real world and yet still perform the task.
- Ensure that our task is not 2D solvable, but requires depth perception to correctly perform. A *2D solvable* task can be solved by attending to only 2D geometry. For example, if we used only height in the visual field to encode the depth of two virtual objects, then subjects could correctly determine which was farther by noting which had the greater 2D *y*-coordinate.

Table 1: Independent and dependent variables.

INDEPENDENT VARIABLES				
subject	8	(random variable)		
(referent) *field of view*	2	upper, lower		
occluder	2	present, absent		
distance	8	**DISTANCE FROM SUBJECT**	**COLOR**	**POSITION**
		5.25 meters	orange	wall left
		11.34 meters	red	center right
		17.42 meters	brown	center left
		22.26 meters	blue	wall right
		27.69 meters	purple	center right
		33.34 meters	green	wall left
		38.93 meters	pink	center left
		44.31 meters	yellow	center
repetition	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10		
DEPENDENT VARIABLES				
absolute error	| *estimated_distance – actual_distance* |, *meters*			
signed error	*estimated_distance – actual_distance*, *meters* +: subject overestimated target distance –: subject underestimated target distance			

- Control the ratio of environmental illumination to AR display brightness. Even though our application domain of mobile AR calls for outdoor use, we needed to control this ratio because our AR system and display cannot adjust to or match outdoor illuminance values. 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

Figure 1 shows the experimental setting. We seated subjects 3.4 meters from one end of a 50.1-meter long hallway. Subjects looked down the hallway, through an optical, see-through AR display mounted on a frame. Because the display was rigidly mounted, each subject saw exactly the same field of view. Subjects saw a series of eight real-world *referents*, approximately positioned every 5.6 meters 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 filled approximately half of the hallway. Subjects 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 in the hallway. The software drew the target rectangle with a 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 subject 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. [2003] previously determined this to be an effective graphical technique for distance encoding. The software also printed a text label that named the color at the bottom of the display screen.

The subject'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 subject 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, with SVGA resolution and a 28° horizontal field of view in each eye. We increased the display's transparency by removing the LCD opacity filter, and we set the display brightness to its maximum setting. 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 observe subjects' progress. We implemented our experimental control code in Java.

3.2 Variables and Design

3.2.1 Independent Variables

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

Field of View: As shown in Figure 1, we placed the referents in the subject'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-subject pilot experiment with the same task, but with the referents exclusively in the lower field of view. The pilot data suggested that subjects consistently underestimated the target depth, and we hypothesized that this might be due to an implicitly tilted visual reference plane, called a *horopter*, against which matches are made. In the depth perception community it is well-known that the vertical horopter is tilted, with objects lying physically slightly closer in the lower field of view appearing equidistant to objects in the upper field of view [Tyler 1991]. If subjects in our experiment made depth judgments by matching against a tilted vertical horopter, it should show up as a main effect or interaction with 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)), subjects 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 subject'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 subject'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, subjects 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 gaf-

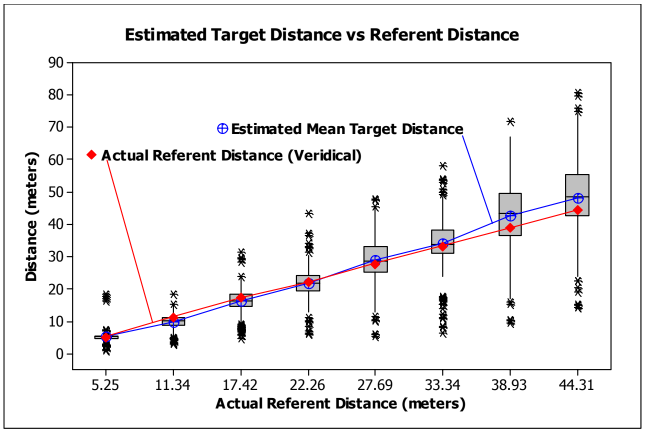


Figure 2: The general results from the first study, indicating where subjects placed the targets (blue line), versus the actual referent locations (red line).

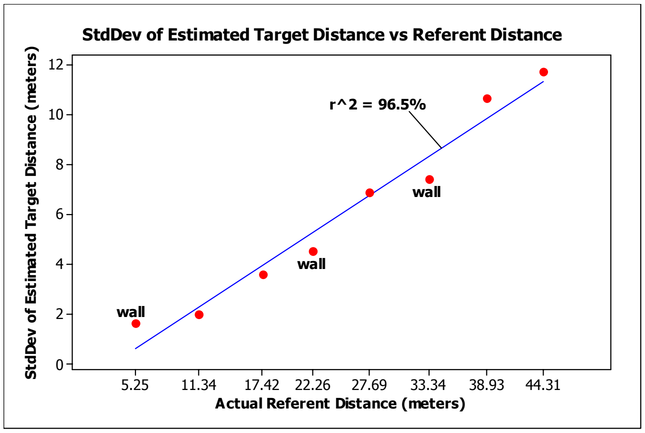


Figure 3: As the referent distance increased, the variability of the estimated target distance grew in a linear fashion, which clearly indicates decreasing depth cue effectiveness. "wall" marks referents placed next to a wall.

fer's tape to the upper walls, which yielded a comparable linear perspective cue in both field of view conditions.

Referent Distance: We placed the eight referents at the distances from the subject 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 subject's position. As indicated in Table 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. 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 [1993] 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, subjects 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 an estimated target distance, which we used to render the target rectangle. When a subject pressed the mouse button, we recorded the estimated target distance, and used this to calculate and record *absolute error* and *signed error*, using the formulas shown in Table 1.

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 (subject) to fastest (repetition). We collected a total of 2560 data points (8 subjects * 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 subject saw all levels of each independent variable, so all variables were within-subject.

Each subject first read and signed a consent form, and then took a stereo acuity test, which all subjects passed. The subject 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 subject to place the target on random referents until we felt that the subject understood the task. The subject next completed four blocks of 80 trials each. Between blocks the subject 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 *subject* a random variable and all other independent variables as fixed (Table 1). 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 [Vaughan and Corballis 1969]. In general, when $\eta^2 < 1\%$, we were hesitant to consider an effect.

Figure 2 summarizes the actual referent distances (red line) versus subjects' estimated target distances. Note that although in Figure 2 the data is labeled "estimated target distance", because depth perception compresses at medium- and far-field distances, subjects perceived both target and referent distances as being closer than veridical [Smallman et al. 2002].

Figure 3 shows that the variability (expressed as the standard deviation) of the estimated target distance grew linearly (r^2 = 96.5%) with increasing referent distance. This demonstrates that our experimental task and design are actually measuring depth perception, because the linear variability growth with distance clearly indicates that the depth matches are being made with depth cues of linearly decreasing effectiveness, and because depth stud-

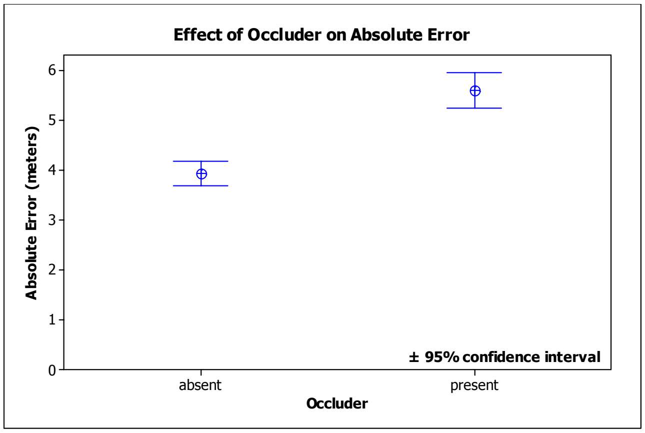


Figure 4: Subjects made ~44% greater errors when an occluder was present.

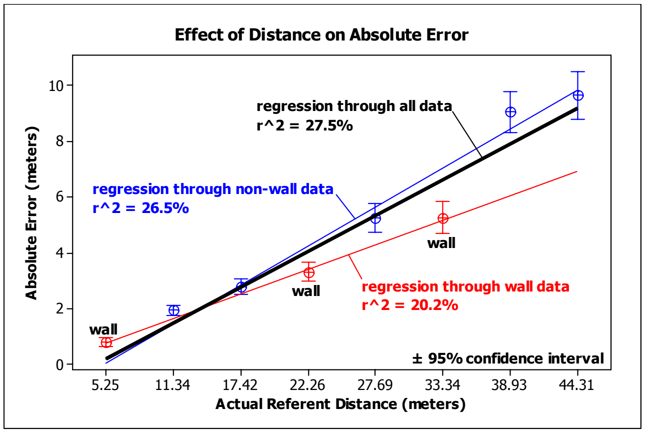


Figure 5: Absolute error increased with increasing distance, but the rate of increase was less for those referents that were next to a wall.

ies with physical stimuli have found a similar linear relationship between variability and depth [Cutting and Vishton 1995]. Figure 3 also indicates that, of the three referents that were located next to a wall, the 33.34 meter referent appears to have less variance than would be expected from the linear growth in variability. The effects of wall proximity are further explored below.

There was no main effect of field of view on either absolute position error or signed position error, nor were there any interactions. Therefore, in contrast to what our pilot data suggested, we can not conclude that subjects are making depth judgments relative to a tilted vertical horopter.

Figure 4 shows the effect of occluder on absolute error ($F(1,7) = 5.78$, $p = .047$, $\eta^2 = 2.28\%$); subjects made ~44% more error when the occluder was present. Although this finding has practical importance for designers of AR x-ray vision techniques, it was expected because fewer depth cues are available when the occluder was present. We did not, however, find an effect of occluder on signed error ($F < 1$), nor did we find any interactions. This result diverges from the findings of Ellis and Menges [1998]; in a near-field experiment, they found that an occluder pushed estimated target distance closer to the subject. However, Ellis and Menges determined that this occurred because the occluder changed subjects' convergence, and convergence is not an effective depth cue beyond about 10 meters [Cutting 2003].

Figure 5 shows the effect of distance on absolute error ($F(7,49) = 30.5$, $p < .000$, $\eta^2 = 29.4\%$); absolute error increased with distance. This is another indication that the task became more difficult as distance increased. An overall linear model of this difficulty (black line) explains $r^2 = 27.5\%$ of the observed variance. Figure 5 also shows the effect of the referents that were positioned next to a wall; these resulted in reduced error, as well as reduced rate of error growth with increasing distance, because they afforded easier use of linear perspective. Note that a linear model applied only to the wall data (red line) has a shallower slope than the overall linear model (black line); in contrast a linear model applied only to the non-wall data (blue line) runs relatively close to the overall linear model.

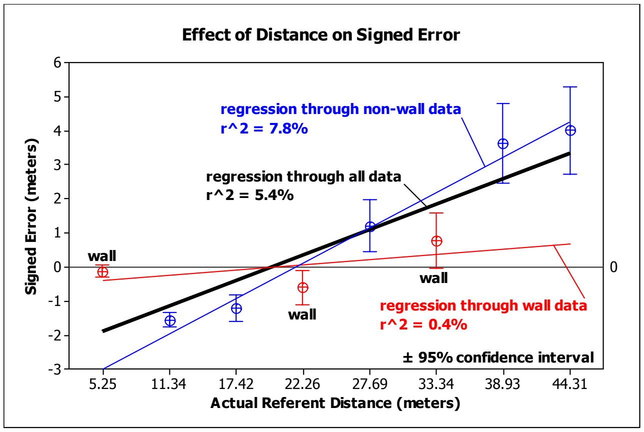


Figure 6: Similar to absolute error (Figure 5), signed error increased with increasing distance, and the rate of increase was less for those referents that were next to a wall. In addition, signed error revealed a switch in bias from underestimating to overestimating target distance at ~19.4 meters.

We also found a distance by repetition interaction for absolute error ($F(63,441) = 1.53$, $p = .008$; $\eta^2 = 1.43\%$). For the first four referents, there was very little absolute error variation with repetition, but for the second four referents, there was considerable absolute error variation with repetition. This interaction is just another manifestation of the absolute error increasing with distance (Figure 5).

Figure 6 shows the effect of distance on signed error ($F(7,49) = 3.20$, $p = .007$, $\eta^2 = 7.31\%$). Signed error generally displayed the same effects as absolute error (Figure 5): signed error increased with distance, and linear modeling of all of the data (black line), and the data split into wall data (red line), and non-wall data (blue line), indicates that error and rate of error growth were reduced when referents were next to a wall. However, the r^2 values indicate that linear models do not explain as much variance as they did for absolute error (Figure 5); this is particularly true for the wall data. Comparing the relative magnitude of the confidence intervals between Figures 5 and 6 indicates there is more variability in signed error, because with absolute error positive and negative values with nearly the same magnitude are folded over into values that are nearly equal.

The most interesting finding from signed error, which is not seen in the absolute error results, is evidence of a shift in bias from underestimating to overestimating target distances (Figure 6). This begins at the 11.34 meter referent — the 5.25 meter referent is close enough that stereopsis is still available as a depth cue, but by 11.34 meters subjects have transitioned from using stereopsis to using linear perspective. The bias shift occurs at around 19.4 meters, which is where the black line in Figure 6 crosses zero meters of signed error. Before this point, subjects underestimated target distances (negative signed error); after this

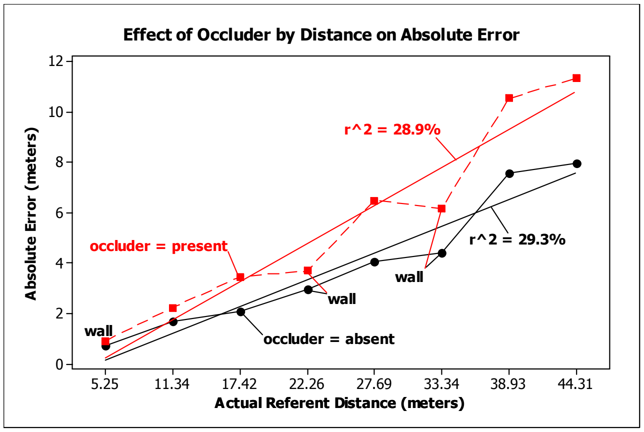


Figure 7: Subjects had more error in the occluded (x-ray vision) condition (red line and points) than in the non-occluded condition (black line and points), and the difference between the occluded and non-occluded conditions increased with increasing distance.

point, subjects increasingly overestimated target distances (positive signed error). This bias shift has not been found by previous AR depth studies that examined near-field distances [Ellis and Menges 1998; McCandless et al. 2000; Rolland et al. 1995; Rolland et al. 2002].

Because of the interesting wall effects noted in Figures 5 and 6, we conducted an ANOVA analysis which followed the repeated-measures design described in Table 1, except that we replaced *referent distance* with the independent variable *referent position*, which had the values *wall* and *not wall*. We found a main effect of referent position on absolute error ($F(1,7) = 28.5$, $p = .001$, $\eta^2 = 5.17\%$); subjects made an average error of 3.12 meters for the wall condition versus 5.73 meters for the not wall condition. In addition, we found several other significant main effects and interactions for both absolute error and signed error, however we do not consider them important because of low η^2 values (η^2 was $\leq .64\%$).

Figure 7 shows an occluder by distance interaction on absolute error ($F(7,49) = 2.06$, $p = .066$; $\eta^2 = .97\%$). When an occluder was present (the x-ray vision condition), subjects 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 7 also shows a linear modeling of the occluder present condition (red line), which explains $r^2 = 28.9\%$ of the observed variance, and a linear modeling of the occluder absent condition (black line), which explains $r^2 = 29.3\%$ of the observed variance. The slope of the occluder present (red) line is larger than the occluder absent (black) line; the slopes (1) indicates that the occluded condition becomes increasingly more difficult than the non-occluded condition with increasing distance, and (2) estimate the magnitude of this effect. Figure 7 also shows an interesting effect for the two wall referents at 22.26 and 33.34 meters. When the occluder is present, these wall referents are the only two that lie below the red regression line; this pattern is not repeated when the occluder is absent. This indicates that in the occluded condition, where most linear perspective cues from the hallway are missing, subjects can still gain some linear perspective by aligning the target with a wall (Figure 1); but this is only helpful when the referent is against a wall. In the non-occluded condition there are enough perspective cues that subjects do not attend as closely to the wall. Thus, Figure 7 shows that most of the wall effect from Figures 5 and 6 comes from the occluded condition.

4 Discussion and Future Work

As discussed in the Introduction, AR has many compelling applications, but some will not be realized until we understand how to place graphics 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. Our study contributes to the important task of understanding AR depth perception.

To our knowledge, we have conducted the first experiment that metrically examines AR depth perception at medium- and far-field distances, which are important distances for a number of compelling AR applications. We have demonstrated an experimental task and design that measures depth perception, finding a linear relationship between estimated depth variability and increasing distance which is similar to what has been found in a large body of depth perception literature (Figure 3). Our experiment has quantified how depth estimation error grows with increasing distance (Figure 5) across a range of interesting medium- to far-field distances. We have also detected evidence for a switch in bias, from underestimating to overestimating distance, at ~19.4 meters (Figure 6), and finally, we quantified how depth error grows for occluded versus non-occluded graphics (Figure 7).

However, like most controlled user studies, this one had many limitations that restrict the generality of our findings. We list a few limitations here; all of them suggest future experiments, some of which we plan to conduct:

- We only examined a subset of the depth cues discussed in the Introduction, and features in the hallway, such as the ceiling lights and molding, gave very strong linear perspective cues (Figure 1). This is the likely the reason why there was not an even higher cost for the occluded condition (Figure 7): subjects were able to use linear perspective cues from the non-occluded referents to help place the occluded graphics. However, there are AR applications, such as visualizing building interiors, which can use strong linear perspective cues [Bane and Höllerer 2004]. We would like to run a similar experiment in a large room instead of a hallway, where we could control the strength of linear perspective cues with appropriately placed props.
- Like most optical see-through AR user-based studies to date, a large limitation is the optical quality of our AR display itself: although they have been widely used in AR research, the Sony Glasstron was originally designed for personal use and general desktop applications. We are interested in potentially building an AR display out of off-the-shelf optical components, similar to the one built by Rolland et al. [2002]. This is especially important for near-field depth experiments, where cues that are strongly influenced by display optics, such as binocular disparity and accommodative focus, are dominant.
- In our task subjects only manipulated the depth of a virtual target to match the depth of a real referent. We might find different results if subjects matched a real target to a virtual referent, like Ellis and Menges [1998] and McCandless et al. [2000]. We have already conducted a study where we compared matching both real and virtual objects to real referents [Livingston et al. 2005].
- A challenging AR visualization related to x-ray vision involves understanding how depth perception operates when users perceive multiple, semi-transparent layers of occluded information. Livingston et al. [2003] studied this issue, but only measured ordinal depth perception, and did not require subjects to attend to the real world. We would like to conduct a similar

study, based on the methods and task of the experiment reported here, that measures depth sorting ability metrically.

Acknowledgements

This work was supported by the Advanced Information Technology Branch of the Naval Research Laboratory, the Office of Naval Research (ONR), and Mississippi State University. This experiment was conducted at the Naval Research Laboratory.

References

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

BANE, R, HÖLLERER, T, 2004, "Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality", *Proceedings of the Third IEEE / ACM International Symposium on Mixed and Augmented Reality (ISMAR)*.

CUTTING, JE, VISHTON, PM, 1995, "Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth", in W EPSTEIN, S ROGERS (EDS), *Handbook of Perception and Cognition, Vol 5: Perception of Space and Motion*, Academic Press, pages 69–117.

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

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

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

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

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

LIVINGSTON, MA, ROSENBLUM, L, JULIER, SJ, BROWN, D, BAILLOT, Y, SWAN, J.E. II, GABBARD, JL, HIX, D, 2002, "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.

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

LIVINGSTON, MA, ZANBAKA, C, SWAN II, JE, SMALLMAN, HS, 2005, "Objective Measures for the Effectiveness of Augmented Reality", *Poster Papers, IEEE Virtual Reality 2005*.

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

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

ROLLAND, JP, MEYER, C, ARTHUR, K, RINALDUCCI, E, 2002, "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, pages 610–625.

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

SMALLMAN, HS, ST. JOHN, M, COWEN, MB, 2002, "Use and Misuse of Linear Perspective in the Perceptual Reconstruction of 3D Perspective View Displays", *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*, HFES, Santa Monica, CA, pages 1560–1564.

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

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