Resolving Multiple Occluded Layers in Augmented Reality

Mark A. Livingston* J. Edward Swan II* Joseph L. Gabbard† Tobias H. Höllerer‡
Deborah Hix† Simon J. Julier§ Yohan Baillot§ Dennis Brown*
Naval Research Laboratory
Washington D.C.

Abstract

A useful function of augmented reality (AR) systems is their ability to visualize occluded infrastructure directly in a user's view of the environment. This is especially important for our application context, which utilizes mobile AR for navigation and other operations in an urban environment. A key problem in the AR field is how to best depict occluded objects in such a way that the viewer can correctly infer the depth relationships between different physical and virtual objects. Showing a single occluded object with no depth context presents an ambiguous picture to the user. But showing all occluded objects in the environments leads to the "Superman's X-ray vision" problem, in which the user sees too much information to make sense of the depth relationships of objects.

Our efforts differ qualitatively from previous work in AR occlusion, because our application domain involves far-field occluded objects, which are tens of meters distant from the user. Previous work has focused on near-field occluded objects, which are within or just beyond arm's reach, and which use different perceptual cues. We designed and evaluated a number of sets of display attributes. We then conducted a user study to determine which representations best express occlusion relationships among far-field objects. We identify a drawing style and opacity settings that enable the user to accurately interpret three layers of occluded objects, even in the absence of perspective constraints.

1 Introduction

Augmented reality (AR) refers to the mixing of virtual cues into the user's perception of the real three-dimensional environment. In this work, AR denotes the merging of synthetic imagery into the user's natural view of the surrounding world, using an optical, see-through, head-worn display. Figure 1 is an example from our AR system.

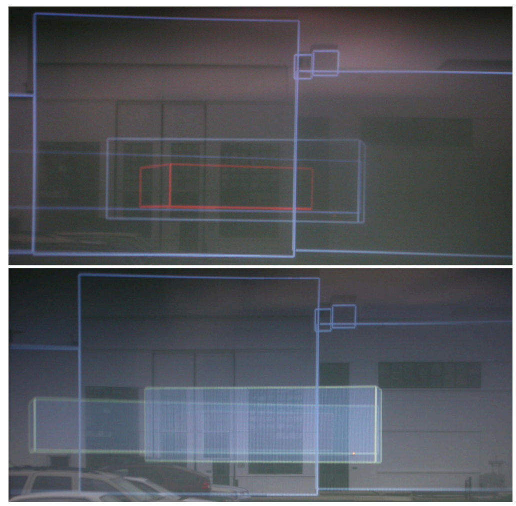

Figure 1. Before-and-after pictures of one of our visualization techniques. The occluded target lies behind the physically visible building (always in wireframe) and the two other occluded buildings. The bottom picture–with a filled, partly opaque drawing style–vastly improves the ability of users to discern this depth ordering.

Through the ability to present direct information overlays, integrated into the user's environment, AR has the potential to provide significant benefits in many application areas. Many of these benefits arise from the fact that the virtual cues presented by an AR system can go beyond what is physically visible. Visuals include textual annotations, directions, instructions, or "X-ray vision," which shows objects that are physically present, but occluded

*Virtual Reality Laboratory, Naval Research Laboratory. Corresponding email: livingston@ait.nrl.navy.mil
†Systems Research Center, Virginia Polytechnic Inst. and State Univ.
‡Dept. of Computer Science, University of California, Santa Barbara
§ITT Advanced Engineering and Sciences

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 IEEE

from view. Potential application domains include manufacturing [4], architecture [26], mechanical design and repair [10], medical applications [7, 23], military applications [17], tourism [9], and interactive entertainment [25].

1.1 Context for Our Work

This study is set in the larger context of research and development of mobile, outdoor AR. Our system supports information gathering and human navigation for situation awareness in an urban setting [17]. A critical aspect of our project is that it equally addresses both technical and human factors issues in fielding mobile AR. Technical challenges on which we are focusing include tracking and registration and display design. To address human factors issues, we are systematically incorporating usability engineering activities [14] at every phase of development, to ensure that our AR system meets its human users' needs.

We determined one such user need by performing a task analysis with domain experts [13], who identified a strong need to visualize the spatial locations of personnel, structures, and vehicles occluded by buildings and other urban structures. While we can provide an overhead map view to view these relationships, using the map requires a context switch. We hope to design visualization methods that enable the user to understand these relationships when directly viewing, in a heads-up manner, the augmented world in front of them. In our application domain, typically only the first layer of objects is physically visible.

1.2 Visualization of Occluded Objects

Giving the user the ability to discern the correct depth ordering among several physical and virtual objects that partially or completely occlude one another is complicated by the "Superman's X-ray vision" problem. If the user sees all depth layers of a complex environment, there will be too much information to understand the depth ordering. But if only the objects of interest are presented, there may not be sufficient context to grasp the depth of these objects.

The complexity can be partially managed by information filtering methods [16], which use rules and reasoning to reduce the set of objects displayed to the user to the "important" ones. Our goal in this work is to discover a set of graphical cues that addresses the depth ordering problem—that is, provides sufficient cues that the user can understand the depth relationships of virtual objects that overlap in screen space. In order to achieve this, we designed a number of sets of display attributes for the various layers of occluded virtual objects. Figure 1 shows an example from the experiment.

2 Related Work

2.1 Viewing Occluded Objects in AR

The KARMA system [10] built on earlier work in computer-generated illustrations to create an AR system that used ghosting (represented, for example, with partial transparency or dashed lines) and cutaway views to express depth ordering between real and virtual objects. The cutaway view provides a context for the 3D relationships. The apparent conflict created by a virtual object overlapping a real object that should occlude the virtual object is thus resolved by surrounding the virtual object with a "virtual hole" in the real object [22].

Furmanski et al. [12] utilized a similar approach in their pilot experiment. Using video AR, they showed users a stimulus which was either behind or at the same distance as an obstructing surface. They then asked users to identify whether the stimulus was behind, at the same distance as, or closer than the obstruction. Only a single occluded object was present in the test. The parameters in the pilot test were the presence of a cutaway in the obstruction and motion parallax. The presence of the cutaway significantly improved users' perceptions of the correct location when the stimulus was behind the obstruction. The authors offered three possible locations to the users, even though only two locations were used. Users consistently believed that the stimulus was in front of the obstruction, despite the fact that it was never there. The authors also discuss issues related to depth perception in AR, including system issues, such tracker noise and visual display complexity, and traditional perceptual cues such as transparency, occlusion, apparent size, shading gradients, motion parallax, and stereopsis.

Other AR systems have used similar techniques as well. The Architectural Anatomy project [26] used overlays to denote the location of hidden objects. These were understood to be one layer behind the visible surface. A similar approach was taken by Neumann and Majoros [19] in an aircraft maintenance prototype application.

The perceptual community has studied depth and layout perception for many years. Cutting [5] divides the visual field into three areas based on distance from the observer: near-field (within arms reach), medium-field (within approximately 30 meters), and far-field (beyond 30 meters). He then points out which depth cues are more or less effective in each field. Occlusion is the primary cue in all three spaces, but with the AR metaphor and the optical see-through, this cue is diminished. Perspective cues are also important for far-field objects, but this assumes that they are physically visible. The question for an AR system is which cues work when the user is being shown virtual representations of objects integrated into a real scene.

2.2 Perceptual Issues in Augmented Reality

The issue of correctly understanding depth ordering of virtual and real objects is one piece of the larger puzzle of perception in AR. Ellis and Menges [8] found that the presence of a visible (real) surface near a virtual object significantly influences the user's perception of the depth of the virtual object. For most users, the virtual object appeared to be nearer than it really was. This varied widely with the user's age and ability to use accommodation, even to the point of some users being influenced to think that the virtual object was further away than it really was. Adding virtual backgrounds with texture reduced the errors, as did the introduction of virtual holes, similar to those described above.

Drasic and Milgram [6] list a number of cues that a user may use to interpret depth, including image resolution and clarity, contrast and luminance, occlusion, depth of field (e.g. blur), accommodation, and shadows. AR uses one of two technologies to see the real world, optical see-through and video see-through. Both technologies can present occluded objects, and each has a variety of challenges [21].

Several authors observe that providing correct occlusion of real objects by virtual objects requires a scene model. As demonstrated by many previous applications, correct occlusion relationships do not necessarily need to be displayed at all pixels; the purpose of many applications is to see through real objects. Even among occluded objects, some may have higher semantic importance, such as a destination in a tourism application. Studies found that occlusion of the real object by the virtual object gave the incorrect impression that the virtual object was in front, despite the object being located behind the real object and other perceptual cues denoting this relationship [21]. Blurring can help compensate for depth perception errors [11].

3 Experiment

3.1 Design Methodology

We used a systematic approach to determine factors for this study. Our AR team performed six cycles of structured expert evaluation on a series of mockups representing occluded objects in a variety of ways. Results from one cycle informed redesign of mockups for the next cycle of evaluation; more than 100 mockups were created. Parameters that varied during the mockups included line width, line style, number of levels of occlusion, shading, hidden lines/surfaces, shadows, color, and stereopsis. Iteratively evaluating the mockups, our team collectively found that intensity was the most powerful graphical encoding for occlusion (i.e., it was the most consistently discriminable). Drawing style and opacity were also key discriminators.

From these findings, **drawing style**, **opacity**, and **intensity** comprised a critical yet tenable set of parameters for our study. Also based on our expert evaluations, we chose to use three different positions for the target, giving us a total of four levels of occlusion (three buildings plus the target). This introduced the question of whether the ground plane (i.e. perspective) would provide the only cue that users would actually use. Because our application may require users to visualize objects that are not on the ground or are at a great distance across hilly terrain, we added the use of a consistent, flat ground plane for all objects as a parameter.

3.2 Hardware

The hardware for our AR platform consisted of three components. For the image generator, we used a Pentium IV 1.7 GHz computer with an ATI FireGL2 graphics card (outputting frame-sequential stereo). For the display device, we used a Sony Glasstron LDI–100B stereo optical see-through display (SVGA resolution). The user was seated indoors for the experiment and was allowed to move and turn the head and upper body freely while viewing the scene, which was visible through an open doorway to the outdoors. We used an InterSense IS-900 6-DOF ultrasonic and inertial tracking system to track the user's head motion to provide a consistent 3D location for the objects as the user viewed the world.

The user entered a choice for each trial on a standard extended keyboard, which was placed on a stand in front of the seat at a comfortable distance. The display device, whose transparency can be adjusted in hardware, was set for maximum opacity of the LCD, to counteract the bright sunlight that was present for most trials. Some trials did experience a mix of sunshine and cloudiness, but the opacity setting was not altered. The display brightness was set to the maximum. The display unfortunately does not permit adjustment of the inter-pupillary distance for each user. If IPD is too small, then the user will be seeing slightly cross-eyed and tend to believe objects are closer than they are. The display also does not permit adjusting the focal distance of the graphics. The focal distance of the virtual objects is therefore closer than the real object that we used as the closest obstruction. This would tend to lead users to believe the virtual objects were closer than they really were.

3.3 Experimental Design

3.3.1 Independent Variables

From our heuristic evaluation and from previous work, we identified the following independent variables for our experiment. These were all *within-subject* variables: every user saw every level of each variable.

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**

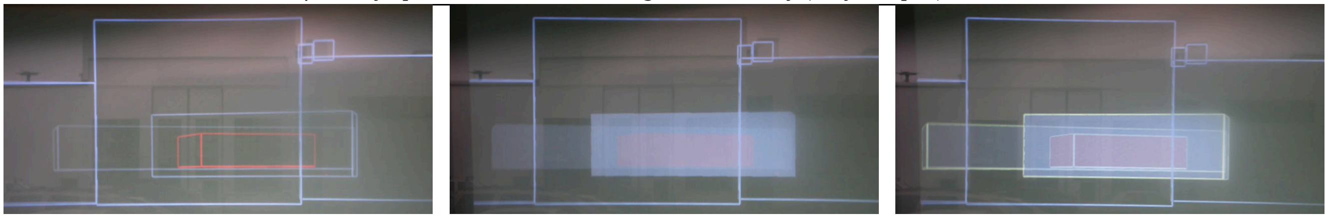

Figure 2. User's view of the stimuli. *Left*: **"wire" drawing style.** *Center*: **"fill" drawing style.** *Right*: **"wire+fill" drawing style. The target (smallest, most central box) is between (position "middle") obstructions 2 and 3 in all three pictures. These pictures were acquired by placing a camera to the eyepiece of the HMD, which accounts for the poor image quality. The vignetting and distortion are due to the camera lens and the fact that it does not quite fit in the exit pupil of the HMD's optics.**

Drawing Style ("wire", "fill", "wire+fill"): Although the same geometry was visible in each stimulus (except for which target was shown), the representation of that geometry was changed to determine what effect it had on depth perception. We used three drawing styles (Figure 2). In the first, all objects are drawn as wireframe outlines. In the second, the first (physically visible) object is drawn as a wireframe outline, and all other objects are drawn with solid fill (with no wireframe outline). In the third style, the first object is in wireframe, and all other layers are drawn with solid fill with a white wireframe outline. Backface culling was on for all drawing styles, so that the user saw only two faces of any occluded building.

Opacity (constant, decreasing): We designed two sets of values for the α channel based on the number of occluding objects. In the "constant" style, the first layer (visible with registered wireframe outline) is completely opaque, and all other layers have the same opacity ($\alpha = 0.5$). In the "decreasing" style, opacity changes for each layer. The first (physically visible, wireframe) layer is completely opaque. The successive layers are not opaque; the α values were 0.6, 0.5, and 0.4 for the successively more distant layers.

Intensity (constant, decreasing): We used two sets of intensity modulation values. The modulation value was applied to the object color (in each color channel, but not in the opacity or α channel) for the object in the layer for which it was specified. In the "constant" style, the first layer (visible with registered wireframe outline) has full intensity (modulator=1.0) and all other layers have intensity modulator=0.5. In the "decreasing" style, the first layer has its full native intensity, but successive layers are modulated as a function of occluding layers: 0.75 for the first, 0.50 for the second, and 0.25 for the third (final) layer.

Target Position (close, middle, far): As shown in the overhead map view (Figure 3), there were three possible locations for the target.

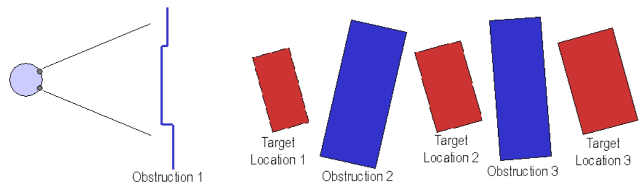

Figure 3. The experimental design (not to scale) shows the user position at the left. Obstruction 1 denotes the visible surfaces of the physically visible building. The distance from the user to obstruction 1 is approximately 60 meters. The distance from the user to target location 3 is approximately 500 meters, with the obstructions and target locations roughly equally spaced.

Ground Plane (on, off): From the literature and everyday experience, we know that the perspective effects of the ground plane rising to meet the horizon and apparent object size are a strong depth cues. In order to test the representations as an aide to depth ordering, we removed the ground plane constraint in half of the trials. The building sizes were chosen to have the same apparent size from the users' location for all trials. When the ground plane constraint was not present in the stimulus, the silhouette of each target was fixed for a given pose of the user. In other words, targets two and three were not only scaled (to yield the same apparent size) but also *positioned vertically* such that all three targets would occupy the same pixels on the 2D screen for the same viewing position and orientation. No variation in position with respect to the two horizontal dimensions was necessary when changing from using the ground plane to not using it. The obstructions were always presented with the same ground plane. We informed the users for which

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**

half of the session the ground plane would be consistent between targets and obstructions.

We did this because we wanted to remove the effects of perspective from the study. Our application requires that we be able to visualize objects that may not be on the ground, may be at a distance and size that realistic apparent size would be too small to discern, and may be viewed over hilly terrain. Since our users may not be able to rely on these effects, we attempted to remove them from the study.

Stereo (on, off): The Sony Glasstron display takes left and right eye images. The inter-pupillary distance and vergence angle are not adjustable, so we can not provide a true stereo image for all users. However, we can present images with disparity (which we shall call "stereo" for the experiment) or present two identical images ("biocular").

Repetition (1, 2, 3): Each user saw three repetitions of each combination of the other independent variables.

3.3.2 Dependent Variables

For each trial, we recorded the user's (three-alternative forced) choice for the target location and the time the user took to enter the response after the software presented the stimulus. All combinations of these parameters were encountered by each user; however, the order in which these were presented was also randomly permuted. Thus each user viewed 432 trials. The users ranged in time from twenty to forty minutes for the complete set of trials. The users were told to make their best guess upon viewing the trial and not to linger; however, no time limit per trial was enforced. The users were instructed to aim for a balance of accuracy and speed, rather than favoring one.

3.3.3 Counterbalancing

Figure 4 describes how we counterbalanced the stimuli. We observed (in conjunction with many previous authors) that the most noticeable variable was ground plane [5, 24]. In order to minimize potentially confusing large-scale visual changes, we gave ground plane and stereo the slowest variation. Following this logic, we next varied the parameters which controlled the scene's visual appearance (drawing style, alpha, and intensity), and within the resulting blocks, we created nine trials by varying target position and repetition.

3.4 Experimental Task

We designed a small virtual world that consisted of six buildings (Figure 3). The first building was an obstruction that corresponded (to the limit of our modeling accuracy) to a building that was physically visible during the experiment. The remaining five buildings consisted of three targets, only one of which was shown at a time, and two obstructions. The obstructions were always drawn in blue; the target that was drawn always appeared in red. The three targets were scaled such that their apparent 2D sizes were equal, regardless of their locations. Obstructions 2 and 3 roughly corresponded to real buildings. The three possible target locations did not correspond to real buildings.

The task for each trial was to determine the location of the target that was drawn. The user was shown the overhead view before beginning the experiment. This helped them visualize their choices and would be an aide available in a working application of our system. The experimenter explained that only one target would appear at a time. Thus in all of the stimulus pictures, four objects were visible: three obstructions and one target. For the trials, users were instructed to use the number pad of a standard extended keyboard and press a key in the bottom row of numbers (1–3) if the target were closer than obstructions 2 and 3, a key in the middle row (4–6) if the target were between obstructions 2 and 3, or a key in the top row (7–9) if the target were further than obstructions 2 and 3. A one-second delay was introduced between trials within sets, and a rest period was allowed between sets for as long as the user wished. We showed the user 48 sets of nine trials each. The users reported no difficulties with the primitive interface after their respective practice sessions. The users did not try to use head motion to provide parallax, which is not surprising for a far-field visualization task.

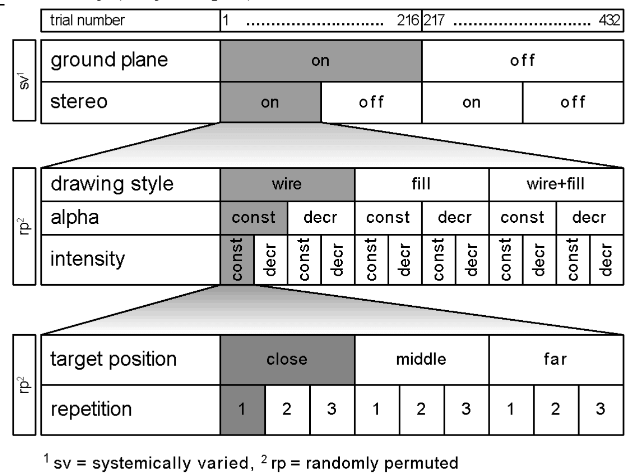

Figure 4. Experimental design and counterbalancing for one user. Systematically varied parameters were counterbalanced between subjects.

3.5 Subjects

Eight users participated. All subjects were male and ranged in age from 20 to 48. All volunteered and received no compensation. Our subjects reported being heavy computer users. Two were familiar with computer graphics, but none had seen our representations. Subjects did not have difficulty learning or completing the experiment.

Before the experiment, we asked users to complete a stereo acuity test, in case stereo had produced an effect. The test pattern consisted of nine shapes containing four circles each. For each set of four circles, the user was asked to identify which circle was closer than the other three. Seven users answered all nine test questions correctly, while the other user answered eight correctly.

4 Hypotheses

We made the following hypotheses about our independent variables.

1. The ground plane would have a strong positive effect on the user's perception of the relative depth.

2. The wireframe representation (our system's only option before this study) would have a strong negative effect on the user's perception.

3. Stereo imagery would not yield different results than biocular imagery, since all objects are in the far-field [5].

4. Decreasing intensity would have a strong positive effect on the user's perception for all representations.

5. Decreasing opacity would have a strong positive effect on the user's perception of the "fill" and "wire+fill" representations. In the case of wireframe representation the effect would be similar to decreasing intensity. Apart from the few pixels where lines actually cross, decreasing opacity would let more and more of the background scene shine through, thereby indirectly leading to decreased intensity.

5 Results

Figure 5 categorizes the user responses. Subjects made 79% correct choices and 21% erroneous choices. We found that subjects favored the far position, choosing it 39% of the time, followed by the middle position (34%), and then by the close position (27%). We also found that subjects were the most accurate in the far position: 89% of their choices were correct when the target was in the far position,

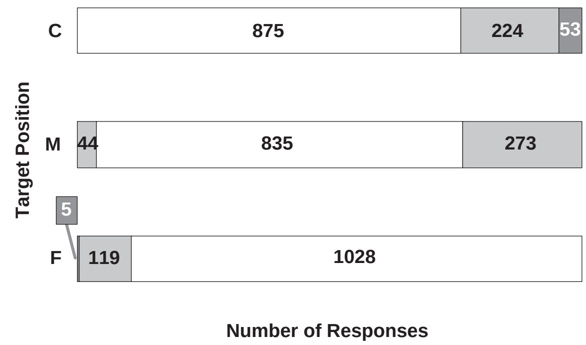

Figure 5. User responses by target position. For each target position, the bars show the number of times subjects chose the (C)lose, (M)iddle, and (F)ar positions. Subjects were either correct when their choice matched the target position (white), off by one position (light gray), or off by two positions (dark gray).

as compared to 76% correct in the close position, and 72% correct in the middle position.

As discussed above, we measured two dependent variables: ***user response time***, and ***user error***. For user response time, the system measured the time in milliseconds (ms) between when it drew the scene and when the user responded. For user error, we calculated the metric $e = |a - u|$, were a is the actual target position (between 1 and 3), and u is the target position chosen by the user (also between 1 and 3). Thus, if $e = 0$ the user has chosen the correct target; if $e = 1$ the user is off by one position, and if $e = 2$ the user is off by two positions. We conducted significance testing for both response time and user error with a standard analysis of variance (ANOVA) procedure. In the summary below, we report user errors in positions (pos).

5.1 Main Effects

There was a main effect of ground plane $(F(1,7) = 51.50,\ p < .01)$ on absolute error; as we expected, subjects were more accurate when a ground plane was present (.1435 pos) then when it was absent (.3056 pos). Interestingly, there was no effect on response time $(F < 1)$. This indicates that subjects did not learn to just look at the ground plane and immediately respond from that cue alone, but were in fact also attending to the graphics.

There was a main effect of drawing style on response time $(F(2,14) = 8.844,\ p < .01)$, and a main effect on absolute error $(F(2,14) = 12.35,\ p < .01)$. As shown in Figure 6, for response time, subjects were slower with the

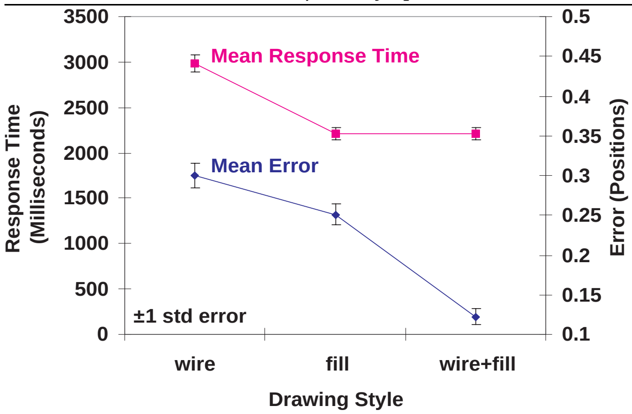

Figure 6. Main effect of drawing style on response time (□) and error (◇).

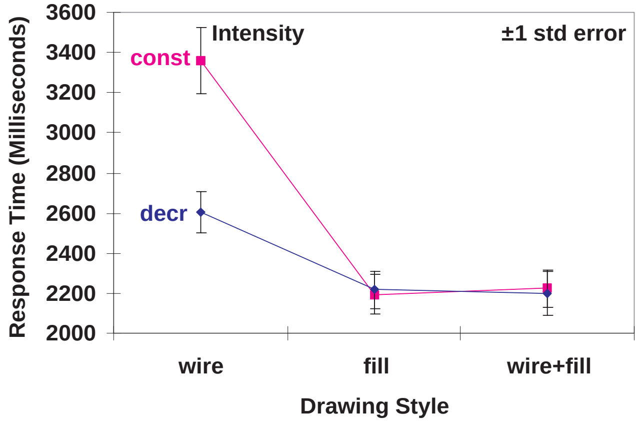

Figure 7. Drawing style by intensity (constant (□), decreasing (◇)) interaction on response time.

"wire" style, while they had comparable times for the "fill" and "wire+fill" styles. For error, subjects had the fewest errors with the "wire+fill" style. These results verified our expectations that the "wire" style would not be very effective, and the "wire+fill" style would be the most effective, since it combines the occlusion properties of the "fill" style with the wireframe outlines, which help convey the targets' shapes.

There was no main effect of stereo on response time $(F < 1)$, and there was no main effect on absolute error $(F < 1)$. This supports our hypothesis that stereo would have minimal effect on a far-field task.

There was a main effect of opacity on absolute error $(F(1,7) = 7.029,\ p < .05)$. Subjects were more accurate with decreasing opacity (.1962 pos) than with constant opacity (.2529 pos). This makes sense because the decreasing opacity setting made the difference between the layers more salient. However, there was no effect of opacity on response time $(F < 1)$; the weakness of this effect $(p = .960)$ is interesting compared to intensity, which was effective for response time at the .01 level.

There was a main effect of intensity on response time $(F(1,7) = 13.16,\ p < .01)$, and a main effect on absolute error $(F(1,7) = 18.04,\ p < .01)$. Subjects were both faster (2340 versus 2592 ms), and more accurate (.1811 versus .2679 pos), with decreasing intensity. This result was expected, as decreasing intensity did a better job of differentiating the different layers. However, this effect can be explained by the interaction between drawing style and intensity. (See Section 5.2.)

There was a main effect of target position on absolute error $(F(2,14) = 4.689,\ p < .05)$, but no effect on response time $(F(2,14) = 2.175,\ p = .15)$. Subjects were most accurate when the target was in the far position, while the close and middle positions were comparable. The effect on error is shown as the "mean" line in Figure 11.

There was a main effect of repetition on response time $(F(2,14) = 20.78,\ p < .01)$. As expected from training effects, subjects became faster with practice. However, repetition had no effect on absolute error $(F < 1)$, so although subjects became faster, they did not become more accurate. This can be taken as a sign that the presented visuals were understandable for the subjects right from the outset. No learning effect took place regarding accuracy. Subjects became faster, though, which is a sign that their level of confidence increased.

5.2 Interactions

There was an interaction between drawing style and intensity on response time $(F(2,14) = 9.38,\ p < .01)$ and on absolute error $(F(2,14) = 8.778,\ p < .01)$. Figure 7 shows that the effect on response time is due to the difference between constant and decreasing intensity when the target is drawn in the "wire" style. Here, subjects were faster when the wireframe targets were drawn with decreasing intensity, which indicates that decreasing intensity was salient enough to be perceptual when the stimuli were just lines. Figure 8 shows the effect on absolute error again comes primarily from the difference for the "wire" style, where subjects were more accurate with decreasing intensity. Thus, this analysis shows that the improvement in speed and accuracy ascribed to decreasing intensity in Section 5.1 is due to decreasing intensity's effect on the wireframe renderings. This appears to refute our hypothesis that decreasing intensity would have a strong positive effect.

Figure 9 shows a target position by drawing style interac-

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 IEEE

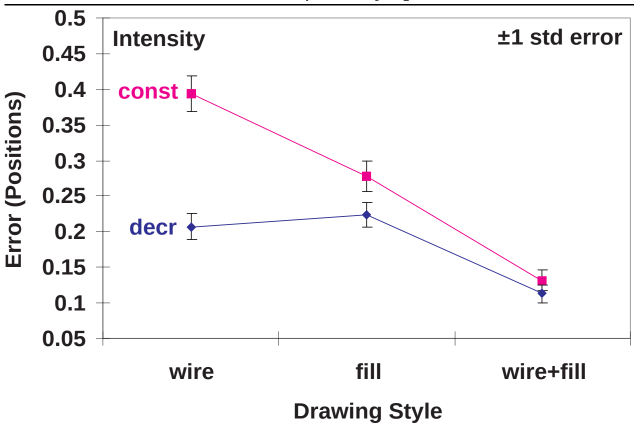

Figure 8. Drawing style by intensity (constant (□), decreasing (⋄)) interaction on absolute error.

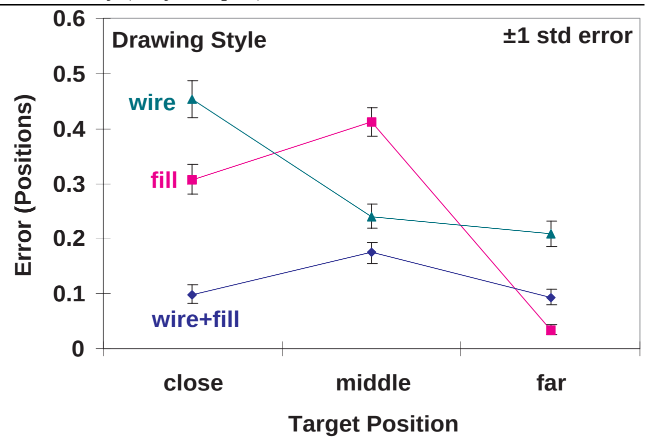

Figure 9. Target position by drawing style (fill (□), wire+fill (⋄), wire (△)) interaction.

tion for absolute error $(F(4,28) = 11.42,\ p < .01)$. Considering the "wire" and "wire+fill" styles, the trend is similar for the middle and far positions, but the "wire" style was particularly difficult in the close position. The "fill" style, which only facilitated layering comparisons using hue and intensity without the 3D structure given by the wireframe lines, was particularly difficult in the middle position, when the target was of intermediate saliency. However, it was quite effective in the far position, when the target saliency was very low. This indicates that subjects used low target saliency as a cue that the target was in the far position.

Figure 10 shows a stereo by opacity interaction for absolute error $(F(1,7) = 8.923,\ p < .05)$. This effect is primarily due to the poor performance of constant opacity in the stereo off condition. Although we do not yet have a theory as to why stereo and opacity would exhibit this interaction, this effect again argues for the global effectiveness of decreasing opacity, as this setting is able to counteract the deleterious effect of the stereo off condition.

Figure 11 shows a target position by ground plane interaction for absolute error $(F(2,14) = 4.722,\ p < .05)$. With no ground plane, this interaction shows an almost linearly decreasing effect as the target position moves farther out. When the ground plane is present, the interaction shows that subjects had the most difficulty in the middle position, but were able to use the extremal ground plane positions to accurately judge the close and far target positions.

6 Discussion

We knew a priori that we could improve upon our previous visualization: "wire" drawing style with all objects drawn at full intensity and opacity. We note that our inde-

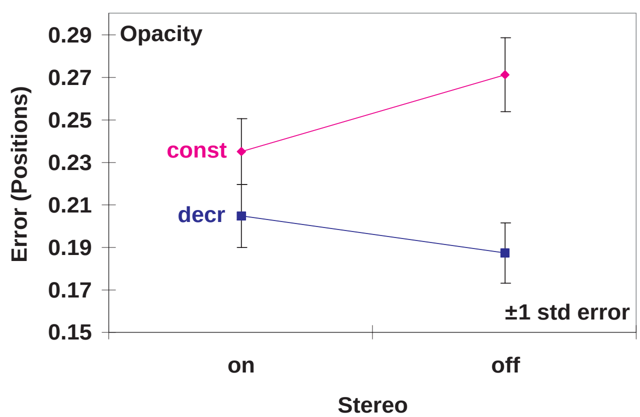

Figure 10. Stereo by opacity (decreasing (□), constant (⋄), interaction on absolute error.

pendent variables had several positive main effects on accuracy and no negative effects on response time. Thus it would appear that, to a first approximation, we have found representations that convey more information about relative depth to the user than our standard wireframe representation, without sacrificing speed in reaching that understanding.

It is well-known that a consistent ground plane is a powerful depth cue. However, we can now provide statistical backing for our fundamental hypothesis that graphical parameters can provide strong depth cues, albeit not physically realistic cues. We found that with the ground plane on the average error was .144 pos, whereas the with the ground plane off and the following settings:

- drawing style: "wire+fill"

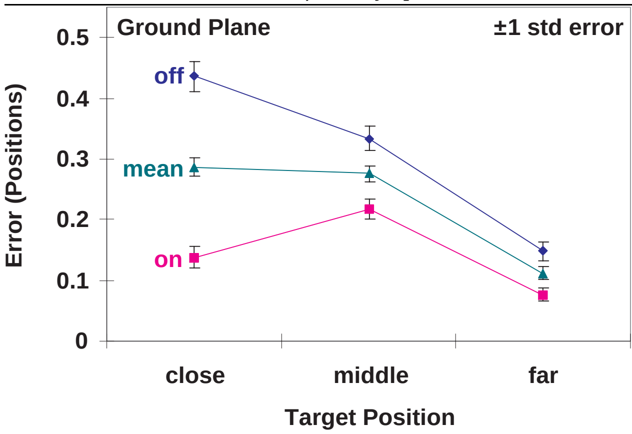

Figure 11. Target position by ground plane (on (□), off (◇)) interaction on absolute error. In addition, this graph shows the main effect of target position (mean (△)).

- opacity: decreasing
- intensity: decreasing

the average error was .111 pos. The data thus suggest that we did find a set of graphical parameters as powerful as the presence of the ground plane constraint. This would indeed be a powerful statement, but requires further testing before we can say for sure whether this is our finding. The fact that there was a main effect of repetition on response time but not on accuracy indicates that the subjects could quickly understand the semantic meaning of the encodings.

The "wire+fill" drawing style yielded the best accuracy. This is consistent with the HCI literature that supports using redundant encodings to convey information [15]. We believe the wireframe portion of the representation helps convey the object shape, whereas the filled portion helps convey the depth ordering. Clearly, however, the two are more powerful together than either is separately.

It is curious to note that the users showed a tendency to pick the far target position and were (thus) more accurate when the target was in the far position. But there was no effect on response time, so the bias towards the third position does not seem very strong.

The main effects of opacity and intensity modulation seem to support the psychophysical literature that dimmer objects appear to be more distant. But, the main effect of intensity can be completely explained by its effect on the wireframe representations, as indicated by the interactions noted in Figures 7 and 8. Thus we can not accept our hypothesis that decreasing intensity would provide a strong cue. However, the main effect of opacity cannot similarly be explained by any interactions, which means that this effect remains across all the other independent variables. This argues for accepting the hypothesis that opacity is a globally effective layering and ordering cue. In addition, during our heuristic evaluation sessions, we discovered that expert evaluators could learn to accurately discern depth ordering with an *increasing* opacity per layer. Since the closer layers are more transparent with such a scheme, this allows users to visualize a greater number of layers. So it remains to be seen whether the number of layers can be increased without sacrificing accuracy or speed, with any scheme of opacity settings: decreasing, constant, or perhaps even increasing.

7 Future Work

In future studies, we hope to overcome confounding factors that were beyond our control, such as the limitations of the display (no inter-pupillary distance, vergence, or focal distance adjustment). As noted, we believe that any errors in the current settings of these conditions are likely to make users believe that objects are closer than they are, which would appear to conflict with the favoritism our users showed for believing the target to be in the furthest position. Similarly, the brightness of the environment from the sun affects the display usability in ways that we have not yet tested. We hope to devise a test in which we can at least measure the influence the sun may have on our visualizations. Video see-through AR would help overcome the brightness difference, but is neither something we have studied nor a popular methodology with our intended users. Finally, an obvious criticism of our current task, which we intend to address in future studies, is that it did not require any interaction between the user's view of the real and virtual worlds, and yet this interaction is at the heart of AR.

An important next step is to draw design recommendations from our results. It appears that filled representations with wireframe outlines, decreasing opacity, and decreasing intensity are sufficient to convey three layers of far-field occluded objects to the user. As we continue this work, we hope to enable AR system developers to create more usable user interfaces. We are excited by the results of this first study, and while there are clearly interactions that we do not yet understand, we are currently planning future studies to improve our understanding of these results and to build on them. We are confident that we have begun to solve the "Superman's X-ray vision" problem for augmented reality.

References

[1] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. In E. E. Catmull, editor, *Computer Graphics (SIGGRAPH '92 Proceedings)*, volume 26, pages 203–210, July 1992.

[2] B. Bell, S. K. Feiner, and T. Höllerer. View management for virtual and augmented reality. In *Proceedings of ACM Symposium on User Interface Software and Technology*, pages 101–110, Nov. 2001.

[3] M. Billinghurst, J. Bowskill, N. Dyer, and J. Morphett. Spatial information displays on a wearable computer. *IEEE Computer Graphics and Applications*, 18(6):24–31, November/December 1998.

[4] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads up display technology to manual manufacturing processes. In *Proceedings of Hawaii International Conference on System Sciences*, volume II, pages 659–669. IEEE Computer Society Press, Jan. 1992.

[5] J. E. Cutting. How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments, and Computers*, 29(1):29–36, 1997.

[6] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In M. T. Bolas, S. S. Fisher, and J. O. Merritt, editors, *SPIE Volume 2653: Stereoscopic Displays and Virtual Reality Systems III*, pages 123–134, January/February 1996.

[7] P. Edwards, D. Hawkes, D. Hill, D. Jewell, R. Spink, A. Strong, and M. Gleeson. Augmented reality in the stereo operating microscope for otolaryngology and neurological guidance. In *Medical Robotics and Computer Assisted Surgery*, Sept. 1995.

[8] S. R. Ellis and B. M. Menges. Localization of object in the near visual field. *Human Factors*, 40(3):415–431, Sept. 1998.

[9] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In *International Symposium on Wearable Computing (ISWC)*, pages 74–81, Oct. 1997.

[10] S. Feiner, B. MacIntyre, and D. Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):52–62, July 1993.

[11] A. Fuhrmann, G. Hesina, F. Faure, and M. Gervautz. Occlusion in collaborative augmented environments. *Computers and Graphics*, 23(6):809–819, 1999.

[12] C. Furmanski, R. Azuma, and M. Daily. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, pages 215–224, Sept. 2002.

[13] J. L. Gabbard, J. E. Swan II, D. Hix, M. Lanzagorta, M. A. Livingston, D. Brown, and S. Julier. Usability engineering: Domain analysis activities for augmented reality systems. In *Proceedings of SPIE (International Society for Optical Engineering)*, The Engineering Reality of Virtual Reality 2002, Jan. 2002.

[14] D. Hix and J. L. Gabbard. *Usability Engineering of Virtual Environments*, pages 681–699. Lawrence Erlbaum Associates, 2002.

[15] D. Hix and H. R. Hartson. *Developing User Interfaces: Ensuring Usability through Product and Process*. John Wiley and Sons, New York, 1993.

[16] S. Julier, Y. Baillot, D. Brown, and M. Lanzagorta. Information filtering for mobile augmented reality. *IEEE Computer Graphics and Applications*, 22(5):12–15, September/October 2002.

[17] 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. In *Interservice/Industry Training, Simulation, and Education Conference*, page 89, Dec. 2002.

[18] C. Loscos, G. Drettakis, and L. Robert. Interactive virtual relighting of real scenes. *IEEE Transactions on Visualization and Computer Graphics*, 6(4):289–305, October/December 2000.

[19] U. Neumann and A. Majoros. Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In *Proceedings of IEEE Virtual Reality Annual International Symposium*, pages 4–11, 1998.

[20] J. Nielsen. *Heuristic Evaluation*, pages 25–62. John Wiley and Sons, New York, 1994.

[21] J. P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287–309, June 2000.

[22] A. State, D. T. Chen, C. Tector, A. Brandt, H. Chen, R. Ohbuchi, M. Bajura, and H. Fuchs. Case study: Observing a volume-rendered fetus within a pregnant patient. In *Proceedings of IEEE Visualization '94*, pages 364–368, 1994.

[23] A. State, M. A. Livingston, G. Hirota, W. F. Garrett, M. C. Whitton, E. D. Pisano MD, and H. Fuchs. Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies. In *SIGGRAPH 96 Conference Proceedings*, Annual Conference Series, pages 439–446. ACM SIGGRAPH, Addison Wesley, Aug. 1996.

[24] R. T. Surdick, E. T. Davis, R. A. King, and L. F. Hodges. The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence: Teleoperators and Virtual Environments*, 6(5):513–531, Oct. 1997.

[25] B. Thomas, B. Close, J. Donoghue, J. Squires, P. D. Bondi, M. Morris, and W. Piekarski. ARQuake: An outdoor/indoor augmented reality first person application. In *International Symposium on Wearable Computers*, pages 139–146, Oct. 2000.

[26] A. Webster, S. Feiner, B. MacIntyre, W. Massie, and T. Krueger. Augmented reality in architectural construction, inspection, and renovation. In *Proceedings of the Third ASCE Congress for Computing in Civil Engineering*, June 1996.

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**