*Office of Naval Research Workshop on Attention, Perception, and Modeling for Complex Displays,
Troy, New York, USA, June 4–6, 2003.*

Far-Field Occlusion and Distance Perception in Augmented Reality

J. Edward Swan II*

The Naval Research Laboratory, Washington, DC

Augmented reality (AR) is an emerging technology, which allows computer-generated graphical information to be correlated with a person's view of the real world (see Figure 1). These graphics can be placed in three positions relative to the real world: (1) in front of the view of the real world, (2) on the view of the real world, and (3) beyond the view of the real world. For decades researchers have studied perception using artificial stimuli (both computer-generated and non-computer-generated), as well as natural scenery. While many of the results will doubtless be found to apply to AR, AR itself provides a display capability which is qualitatively different from those which have been previously studied. This is particularly true for graphics drawn on the view of the real world (which look like computer-drawn signs or labels), and drawn beyond the view of the real world (which provide an "x-ray vision" capability).

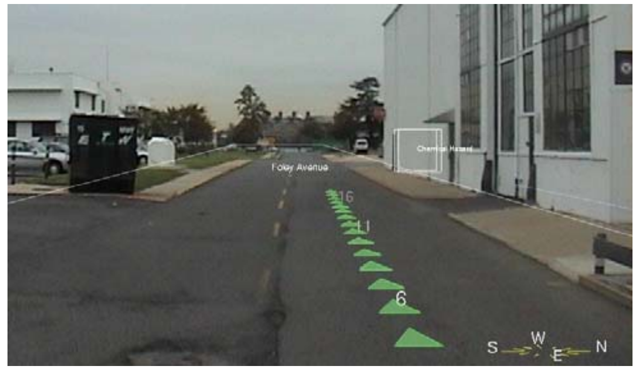

Figure 1: An example of augmented reality (AR), where graphical information overlays the user's view of the real world. In this example, a compass shows which direction the user is facing, the triangles indicate a path the user is following, the numbers on the path indicate distances in meters in front of the user, a hidden chemical hazard is annotated, and the name of the street is given. The graphics are registered with the world, so for example the triangles appear to be painted onto the road surface. The result is an integrated display which allows *heads-up* viewing of the graphical information.

Literature Survey

Not much is known about perception in AR displays. In February 2003 we conducted a survey of the 7 primary publishing venues for AR research, comprising a corpus of 880 papers, as well as the HCI Bibliography [6], comprising 24,000 records. This survey [7] revealed only 14 reported user-based studies of AR systems — and only about half of these deal with perceptual issues. This is not surprising, because only very recently have AR display and tracking technology matured to the point where AR can be used to study perceptual and other ergonomic issues.

The "X-Ray Vision" Problem in Mobile AR

For the past 4 years, we have been developing a mobile AR system, termed BARS, to support urban operations [5]. Throughout the development of this system, we have conducted extensive domain analysis activities with subject matter experts [2]. These, combined with the expected applications of BARS, have yielded the following list of mobile AR affordances, which require perceptual and ergonomic study:

- Heads-up display of objects occluded by urban structures.
- Heads-up display of object distance.
- Simultaneous display of overlapping objects.
- Graphical clutter inherent in 'x-ray vision'.
- User tolerance of tracking and registration errors.
- Display of object importance.
- Heads-up textual information layout.
- Hands-free, heads-up system control.

Because there are no commercially available mobile AR systems, we have had to time systematically studying these affordances with the system's emerging technical capabilities. Within the past year, the BARS system matured to the point where we could study issues related to object occlusion and distance perception.

Our task analysis [2] identified a user 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 for these relationships, using the map requires a context switch. We hope to design an AR visualization method that enables 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 — the buildings across the street — is physically visible. The AR field has termed this the "x-ray vision" problem: if the user sees all depth layers of a complex environment, there will be too much information

*Naval Research Laboratory, Code 5580, 4555 Overlook Ave SW, Washington, DC 20375. Email: swan@acm.org

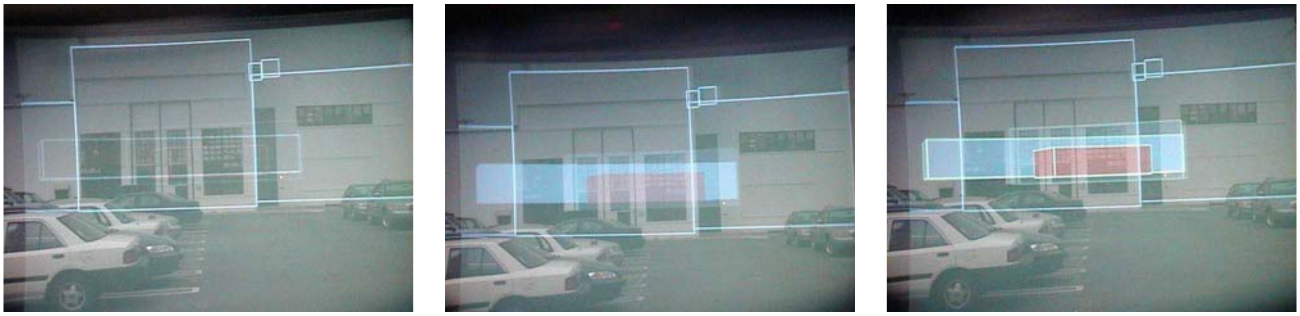

Figure 2: User's view of the stimuli. *Left*: wireframe drawing style. *Center*: filled drawing style. *Right*: "both" drawing style. The target (red) is between obstructions 2 and 3 (blue) in all three pictures. These pictures were acquired by placing a camera to the eyepiece of the head-mounted display.

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.

There is a very compelling finding from comparing the urban x-ray vision problem to our literature review [7] discussed above. To date, all of the reported work is for tasks in the near visual field. Such near-field tasks are natural when a user employs their hands, and these are the types of tasks which motivated much of the work reported to date. However, viewing occluded objects in an urban setting, as well as other urban military operations, require looking at least as far as across a street, and thus use far-field perception. While the small number of results discussed above could hardly be considered a complete study of near-field AR perception, to date we could not find even one reported study (other than our own [4]) of a far-field task. Perception researchers have pointed out the very large perceptual differences between near-field and far-field perception [1], and we cannot expect near-field results to apply to far-field tasks. Furthermore, while it is true that far-field perception has been studied with VR and other optical stimuli [1] (and the same is certainly true for near-field perception), with AR tasks the view of the real world behind the graphical annotations, and the interaction between the graphics and the real world, make far-field AR perception qualitatively different from anything previously studied.

Experiment

To date, we have conduced one experiment in far-field occlusion and distance perception, using our mobile AR system. This section is an extremely brief summary of our paper describing the experiment [4].

Design Methodology

Rather than simply guessing at critical factors in AR occlusion, we used a systematic approach based on expert heuristic evaluation [2]. This type of user interface evaluation employs usability experts (not domain experts) to assess an evolving user interface against relevant design guidelines (heuristics), for the purpose of determining usability problems (which, in this case, would inform our selection of factors). Results from the several experts are then combined and ranked to prioritize iterative redesign of each usability issue discovered.

Our 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; during the six cycles, 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 consistently discriminable graphical encoding for occlusion. Drawing style and opacity were also key discriminators. Further, only three or four levels of occlusion were discriminable under almost any combination of varying factors.

Experimental Task

We designed a small virtual world that consisted of representations of four buildings (see Figure 2). The first representation corresponded to the building that was physically visible during the experiment — see the wireframe outline in Figure 2 above. The remaining three buildings consisted of a target (drawn in red), and two obstructions (drawn in blue). The target was drawn in one of three locations: in front of both obstructions, between the obstructions, or behind both obstructions. In each case the target was scaled so its apparent 2D size was approximately equal. The task for each trial was to determine the position of the target — either close, middle, or far. We recorded the user's three-alternative forced choice, and the time elapsed from the introduction of the stimulus until the user's response.

Experimental Design

Our experiment varied the following within-subject independent variables:

- *Drawing Style*: wireframe, filled, wireframe and filled
- *Opacity*: constant, decreasing
- *Intensity*: constant, decreasing
- *Target Position*: close, middle, far

- *Ground Plane*: on, off
- *Stereo*: on, off
- *Repetition*: 1, 2, 3

We used the three drawing styles shown in Figure 2. In the wireframe style, all objects were drawn as wireframe outlines. In the filled style, all objects were drawn with solid fill, and in the wireframe and filled style, all objects were drawn with a solid fill and with a white wireframe outline. For opacity and intensity, in the constant style, all the layers had the same opacity (intensity), while in the decreasing style, the opacity (intensity) decreased with farther layers. From both the literature and from everyday experience, we know that the perspective effect of the ground plane rising to meet the horizon is a strong depth cue. In order to test the other variables, we removed the ground plane constraint in half of the trials, by drawing the target so it occupied the same screen pixels regardless of position.

We counterbalanced the above variables using Latin squares and random permutations. Each of our eight subjects completed 432 trials. Subjects took between 20 to 40 minutes to complete the experiment. Subjects did not have difficultly learning or completing the task.

We made the following hypotheses about our independent variables:

- The ground plane would have a strong positive effect on the user's task performance.
- The wireframe representation (which our system had been using before this work) would have a strong negative effect on the user's task performance.
- Stereo imagery would not yield different results than biocular imagery, since all objects are in the far-field [1].
- Decreasing intensity would have a strong positive effect on the user's perception for all representations.
- Decreasing opacity would have a strong positive effect on the user's perception of the filled and wireframe and filled 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.

Results

Our dependent variables were user response time, and user error. We calculated user error as the absolute number of positions between the correct choice and the user's choice: 0 if correct, 1 if off by one position, and 2 if off by two positions.

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

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 ($F(2,14) = 12.35$, $p < .01$). Subjects were the slowest and had the most errors with the wireframe style; subjects had the fewest errors with the wireframe and filled style. This verified our expectation that the wireframe style would not be very effective, and the wireframe and filled style would be the most effective, since it combines the occlusion properties of the filled style with the wireframe outlines, which help give a sense of depth to the targets.

There was a main effect of opacity on absolute error ($F(1,7) = 7.029$, $p < .05$). Subjects were more accurate with decreasing opacity than with constant opacity. This makes sense because the decreasing opacity setting made the level differences more salient. However, there was no effect of opacity on response time ($F < 1$).

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 and more accurate 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 ($F(2,14) = 9.38$, $p < .01$ for response time, $F(2,14) = 8.778$, $p < .01$ for absolute error). The main effects of intensity, for both response time and absolute error, were only significant for the wireframe style; there were no differences for the filled or wireframe and filled styles.

There was a main effect of target position on absolute error ($F(2,14) = 4.689$, $p < .05$), but no effect on response time. Subjects were most accurate when the target was in the far position, while the close and middle positions were comparable.

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.

Additional results and interactions are discussed in our publication [4].

Discussion

Before this study, the BARS system utilized only the wireframe drawing style, with no intensity or opacity modulation. We knew a priori that we could improve upon this visualization method. We note that our independent 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 initial 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 the mean error with the ground plane on was 0.144 positions, whereas the mean error with the ground plane off and the following settings was 0.111 positions: wireframe and filled drawing style, decreasing opacity, decreasing intensity. The data thus suggests that we did find a set of graphical parameters as powerful as the presence of the perspective constraint. This would indeed be a powerful statement, but requires further testing before we can say for sure whether this is our finding. Even if this is not the case, we believe this result is a powerful validation of our usability engineering design methodology, discussed above. This methodology yielded both the visual parameters and the specific parameter settings which we tested.

The wireframe and filled drawing style yielded the best accuracy. This is consistent with the HCI literature that supports using redundant encodings to convey information. 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.

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. 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.

Future Work

Our future plans include further experiments studying occlusion and distance perception. In addition, we want to research perceptual models which can help us understand these findings. Finally, an obvious criticism of our current task 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. We want to address this shortcoming in future studies.

Acknowledgements

I would like to acknowledge my collaborators on this work, and co-authors on many of the papers listed below: Joseph Gabbard and Deborah Hix at Virginia Tech; Simon Julier, Mark Livingston, Yohan Baillot, and Dennis Brown at the Naval Research Laboratory, and Tobias Höllerer at the University of California, Santa Barbara. This work was funded by the Office of Naval Research, and the Naval Research Laboratory's Base Program.

References

[1] Cutting, JE, "*How the Eye Measures Reality and Virtual Reality*", Behavior Research Methods, Instrumentation, and Computers, 29, 1997, page(s): 29–36.

[2] Joseph L. Gabbard, J. Edward Swan II, Deborah Hix, Marco Lanzagorta, Mark Livingston, Dennis Brown, Simon Julier, "*Usability Engineering: Domain Analysis Activities for Augmented Reality Systems*", The Engineering Reality of Virtual Reality 2002, Proceedings of SPIE (International Society for Optical Engineering), January 24, 2002.

[3] Deborah Hix, Joseph L. Gabbard, J. Edward Swan II, Mark A. Livingston, Tobias H. Höllerer, Simon J. Julier, Yohan Baillot, Dennis Brown, "*A Cost-Effective Usability Evaluation Progression for Novel Interactive Systems*", (in submission to) 37th Hawaii International Conference on System Sciences (HICSS-37), 2004.

[4] Mark A. Livingston, J. Edward Swan II, Joseph L. Gabbard, Tobias Höllerer, Deborah Hix, Simon J. Julier, Yohan Baillot, Dennis Brown, "*Resolving Multiple Occluded Layers in Augmented Reality*", (in submission to) IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), 2003.

[5] Mark A. Livingston, Lawrence J. Rosenblum, Simon J. Julier, Dennis Brown, Yohan Baillot, J. Edward Swan II, Joseph L. Gabbard, Deborah Hix, "*An Augmented Reality System for Military Operations in Urban Terrain*", Proceedings of the Interservice / Industry Training, Simulation, & Education Conference (I/ITSEC '02), Orlando, FL, December 2–5, 2002.

[6] Gary Perlman, "*The HCI Bibliography: Human-Computer Interaction Resources*", web page www.hcibib.org, accessed February 2003.

[7] J. Edward Swan II, Deborah Hix, Joseph Gabbard, "*Perceptual and Ergonomic Issues in Mobile Augmented Reality for Urban Operations*", unpublished white paper (available from the author), 2003.