

Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information

Chris Furmanski, Ronald Azuma, Mike Daily

HRL Laboratories, LLC

3011 Malibu Canyon Rd.; Malibu, CA 90265; USA

{chris, azuma, mjdaily}@HRL.com

Abstract

One of the unique applications of Mixed and Augmented Reality (MR / AR) systems is that hidden and occluded objects can be readily visualized. We call this specialized use of MR/AR, Obscured Information Visualization (OIV). In this paper, we describe the beginning of a research program designed to develop such visualizations through the use of principles derived from perceptual psychology and cognitive science. In this paper we surveyed the cognitive science literature as it applies to such visualization tasks, described experimental questions derived from these cognitive principles, and generated general guidelines that can be used in designing future OIV systems (as well improving AR displays more generally). Here we also report the results from an experiment that utilized a functioning AR-OIV system: we found that in a relative depth judgment, subjects reported rendered objects as being in front of real-world objects, except when additional occlusion and motion cues were presented together.

Keywords: augmented and mixed reality, cognition, human-computer interaction, motion, perception, occlusion

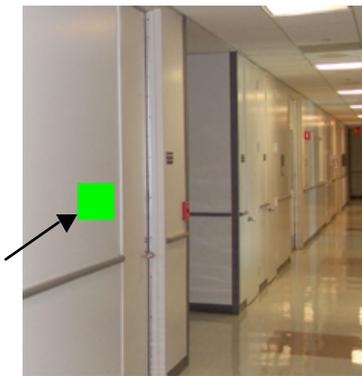


Figure 1: An Example of depth ambiguity in an OIV/AR mockup. In this display, the location (in depth) of the rendered square (marked with an arrow) is ambiguous. Is it on the surface in the hallway, to right of the near room? In the near room? In the far room? There is no definite way to tell because 2D-planar projections of 3D space can be ambiguous.

1. Motivation

In Mixed Reality (MR) and Augmented Reality (AR) systems, virtual objects are combined with real images at interactive rates in 3D. Such displays can enhance the user's perception of the real environment by showing information the user cannot directly sense when unaided. For example, in many AR applications we wish to endow the user with "X-ray vision," enabling the user to see through objects to view a fetus inside a womb, to see pipes and conduits behind walls, or to spot the location of a hidden enemy soldier. Being able to see occluded objects is a useful capability in a variety of medical, architectural, inspection, and military applications. This technology is especially useful in urban environments, where broad, angular surfaces (e.g., walls in hallways, buildings on a street, etc.) limit one's field of view of nearby visually-obscured locations.

However, displaying such hidden objects in a manner that a user intuitively understands is not always trivial. Take the example of a soldier in one room of a building using an AR system to spot the known location of a

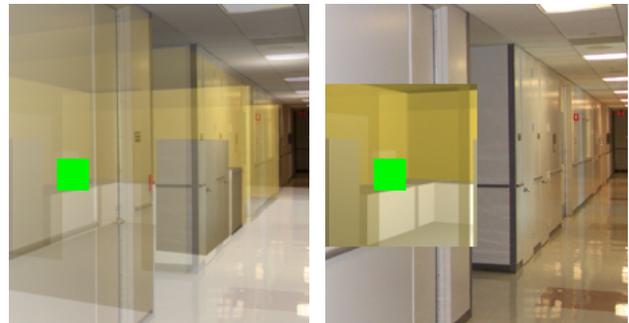


Figure 2: Sample solutions to depth ambiguity in an OIV/AR mockup. In these displays, the location of the rendered square is communicated more clearly by the use of transparency in these visualizations (compare to Figure 1). The use of transparent overlays (LEFT) conveys depth by letting the viewer see structure not otherwise visible, but while still perceiving the real-world structure. A similar approach (RIGHT) presents normally unseen structure by over-rendering a virtual "cut-away" of the occluding surfaces. This approach more clearly depicts the inside of the room, but at the cost of occluding real-world surfaces.

hostile soldier, several rooms away. How does the AR display show the enemy's location? It is not obvious how to visualize this because there may be many walls in between the user and the enemy, and it must be made clear where the enemy is in relation to the user and the other rooms. The AR display must make this relationship intuitively obvious or it may be more confusing than it is helpful. For example, in the mock-up displays presented in Figure 1, it is not clear where the occluded virtual object exists in the real environment.

The goal of this research, then, is to develop new concepts and guidelines for developing effective visualizations of occluded information in MR/AR applications. We call this approach OIV: Computer-Assisted Visualization Augmentation. Unlike previous efforts, our research is driven by the application of cognitive science principles. The contribution in this paper consists of several parts. First, we draw upon a wealth of cognitive science knowledge in relevant areas, such as visual perception, attention, and visual-spatial memory. From this basic knowledge, we outline *a priori* guidelines and hypotheses for designing visualizations. These guidelines are demonstrated in several visualization concept images. The mock-up images in Figure 2 (that utilize exactly the same spatial configuration used in Figure 1) are examples of possible solutions to depth ambiguity. While it may seem obvious that Figure 2 is better at communicating the spatial location of the virtual object, the guidelines and hypotheses behind such concepts must be validated through experimental techniques. We list several key experimental questions and present results from a preliminary experiment that represent our initial steps toward answering these questions.

2. Previous work

The application of findings from cognitive and perceptual psychology is not a new idea, as cognitive studies and analyses have been applied to AR for manufacturing and maintenance tasks [22]. Also, Drascic and Milgram discussed perceptual issues involving depth cues in stereoscopic MR displays [6]. However, these did not specifically address visualization of occluded objects.

Similarly, the visualization community has made use of cognitive and perceptual guidelines to drive the design of visualizations. For example, Interrante [15] uses the principal directions and curvatures in isosurfaces of volume datasets to define textures that follow those directions and curvatures, better illustrating the shape of the isosurface. However, we are not aware of this approach being explicitly taken previously for MR/AR visualization problems.

Many papers have focused on the problem of recovering the tracking information needed to support occlusion (i.e., the location of the user and the depth map of the objects in the real environment). This paper assumes that capability already exists, and instead focuses

on designing visualizations to present the occluded information to the user.

There are a handful of papers that focused on visualization design in AR, including some that specifically addressed occlusion. KARMA [8] used a rule-based approach to determine which objects to highlight and label in a maintenance application. Feiner et. al. [9] developed an application showing pipes, support beams, and other architectural objects hidden behind walls. Julier et. al. [17] developed a means for filtering data, based upon importance metrics, to reduce clutter in the display. MacIntyre and Coelho [19] described techniques to adjust the visualization based upon the expected error in the tracking system. Fuhrmann et. al. [10] built a virtual model of a real user to allow gradual blending of the real and virtual along the edges of the real user, creating a smooth transition between virtual and real at the points of occlusion. Bell et. al. [2] built a view management system to prevent labels from occluding inappropriate objects and each other.

AR technology has also been utilized for medical applications in order to visualize co-registered medical imaging data as a surgical aid. Examples include rendering a model of a pit around occluded tumors inside a breast to aid in ultrasound-guided biopsies [24], as well as the volume-rendering of a fetus in a pregnant woman [23].

Stephen Ellis and his group at NASA Ames have conducted many experiments exploring issues of perception in AR displays. For example, a real, physical surface in the proximity of a virtual object can markedly alter the user's perception of the distance to the virtual object [7].

The difference in our work compared to previous works is in the cognitive-science-based approach of developing guidelines for visualizing occluded objects in MR/AR applications. We seek design principles that are scientifically grounded and justified. This paper represents the beginning of this research program, and the contribution lies in 1) surveying existing cognitive knowledge, 2) listing general guidelines, 3) illustrating sample visualization concepts based on these guidelines, 4) describing the key experimental questions, and 5) conducting an AR-based perceptual pilot experiment.

3. OIV issues

While providing human viewers with extra-sensory information has a wide range of applications, adding visual information also presents the viewer (and system designer) with several perceptual and engineering hurdles. Two of the primary perceptual issues involve (1) conveying the difference between what is normally perceptible and what is extra-sensory in a way (2) that is algorithmically and perceptually easy to visualize in a cluttered and complex environment. Some other important issues for the development of a OIV system,

though not covered in depth here, include: managing and controlling the quantity of extra-sensory information displayed, integrating multimodal cues (haptic and auditory sensations) from obscured locations, and interacting with information across multiple depth planes. This paper focuses on the perceptual aspects of OIV systems, but other aspects of human cognition are also important, including conceptual metaphors, working memory, and training/perceptual learning.

3.1. Depth ambiguity

One of the primary problems with displaying rendered information in an OIV is conveying the difference between what is obscured (not visible) and what would be in plain sight (visible). This is especially problematic for AR/VR displays that are most often a planar projection (2D) view of a mixed-reality 3D environment; here, the position of objects in depth can become ambiguous when 3D information is presented on a 2D planar surface (i.e., information about depth becomes lost and thus is ambiguous).

In such a planar view, the exact location (especially in depth) of computer-rendered symbols is inherently ambiguous (see Figure 1). Thus, in an extended vision environment, where a computer-rendered target could be on the wall next to you or in the adjoining room, it is important to properly convey the depth information of the target. Such a problem is part of a larger of class problems often confronted in AR/VR visualization called depth ambiguities.

3.2. Visual complexity

While conveying information about the distance from the viewer to a visually-rendered target is an important factor in developing an OIV system, one critical constraint is that rendered information must remain clearly discernable to the viewer, especially in complex visual scenes.

To a large degree, OIV faces many of the same challenges faced when developing AR displays. For example, whenever information is added to an AR display, visual displays can become cluttered, limiting their effectiveness.

Drastic changes in luminance and color are particularly problematic in OIV where gradient information about depth and distance must be conveyed in addition to the rendered AR information. In typical AR displays, the augmented information is linked to items that are normally visible.

However, in OIV, augmented information that is linked to visible items must be differentiated from information that is linked to items that are occluded. Thus, additional information about distance (or information that conveys occlusion) must be added to the display. There are a host of approaches by which occlusion and distance information can be conveyed

(which will be addressed in more detail later in Section 6, Design solutions). Nevertheless, the development of image-processing algorithms that allow for rapid changes in contrast, brightness, and transparency of the visual display will be a key component of OIV development.

4. Image processing

Three approaches to mixing computer-generated imagery with video are possible: virtual (replace reality), augmented (enhance reality), and mediated (change reality). Processing methods can be classified into image enhancement and image understanding techniques. With image enhancement, qualities such as contrast, brightness, and transparency are manipulated to improve visibility of important features or highlights. Image understanding attempts to recognize structures and features with the aim of automatically describing the contents of an image.

4.1. Image enhancement

Alpha blending displays additional information over video through multiple channels of graphic overlays. Each overlay consists of colored pixels where each pixel may have different levels of transparency, or be opaque.

Traditional image enhancement techniques, such as histogram equalization, produce improved contrast across an image, but without taking into account possibly relevant local features.

More advanced techniques such as homomorphic filtering combine multiple images of different exposures (due to automatic gain control) to enhance the dynamic range of the resulting image. In this approach, regions of greater homometric certainty correspond to regions of the image that are midtones. Highlights or shadows have lesser homometric certainty [20].

The medical imaging world uses a variety of techniques to highlight information present in imagery obtained from three-dimensional sensors such as CT and NMRI. Physics-based quantitative models have been used for enhancing imagery from x-rays in applications such as mammography [12]. In this approach, the entire process of acquiring an x-ray image is modeled quantitatively to understand the degrading factors (e.g. scattered radiation and beam hardening) and counter them. High pass filtering in the Fourier domain is also useful in highlighting portions of images.

4.2. Image understanding

Visual saliency in imagery refers to the regions of the image that have special interest or draw attention due to unique or significant features (visual conspicuity). Saliency can be measured using a variety of techniques ranging from information theoretic scale space approaches [10], bottom up feature based approaches [18], and saliency networks for extracting salient curves [25], among others.

Higher level processing enables regions of images to be characterized in many ways such as through segmentation of figure and ground and perceptual grouping (bottom up grouping of structures into objects). There are typically four types of image segmentation approaches: those based on threshold techniques, edges, regions, or connectivity preserving relaxation [5]. In addition, techniques that operate on range imagery obtained from stereo or active ranging systems are useful to identify regions with common depth characteristics [9].

5. Visual perception

There are a host of perceptual phenomena that can help disambiguate depth ambiguity that results from displaying a 3D world in a planar view (2D). Outside of having stereo-depth information displayed to viewers, monocular depth cues (innate perceptual cues that the human visual system can use that do not take advantage of disparity information gained from binocular viewing) can provide many useful clues about depth that may otherwise be ambiguous. These cues can be integrated into the development of OIV displays to provide perceptually salient depth and distance information without the use of stereo.

5.1. Depth-dependent perceptual cues

One applicable class of perceptually-salient visual cues that carries inherent depth information are monocular depth cues. These cues are robust, are just as informative when presented to one eye as two, and are effective for static, as well as dynamic images. Some relevant examples include:

- *Transparency* – Transparency, or the use of clear or translucent surfaces, is one of the most common and most intuitive types of visual representations used to visualize depth. One basic property of transparency is that surfaces should be additive; overlapping regions of transparent surfaces that overlap should become darker in order to relate a sense of depth. Although it should be noted that humans tend to become confused from the depth ambiguities that arise from a large number of overlapping transparent surfaces.
- *Occlusion* – The interposition of objects in depth is another intuitive monocular cue commonly used to convey depth. Surfaces that exist between the viewer and another object will obscure distant objects. While this is a good cue for relating a relative sense of depth and distance, the use of opaque occluding surfaces might defeat the purpose of OIV by concealing occluded information.
- *Size-scaling gradients & texture* – Another common and intuitive conveyance of depth/distance involves the varying of size as a function of distance. (This takes advantage of a perceptual bias to conceptualize size that changes with distance, referred to as size

constancy). Size-scaling is commonly used in 3D perspective displays, which provides many obvious clues about depth (e.g., items known to have a constant texture and/or series of parallel lines, such as in a grid) as patterns and textures become finer in the distance.

- *Shading gradients* – Along with size dependent changes, changes in an object's shading also relates information of depth and distance. The perceived contrast of objects in the distance decreases as a function of distance, so that realistic displays could reduce the effective contrast of rendered objects to facilitate effective distance and depth information.
- *Cross-referenced depth* - Other cues, such as shadows on a ground plane (e.g., drop shadows for floating objects), can help disambiguate the location of certain objects in depth, especially in relation to other objects (e.g., walls or a ground plane) [14]. Other obvious manufactured, visual clues relating distance might include the use of virtual yardsticks [19] or distance markers can be used, as well.

5.2. Perceptual motion

Another important class of perceptual cues that provide depth/distance information are motion-related percepts. While depth-dependent gradients are effective for static displays, how objects appear to move, especially in relation to the viewer and viewer's movements, can convey very accurate and very meaningful representations to human viewers. Two of the most relevant motion-related cues are:

- *Motion parallax* - Objects that are closer move farther and faster than objects in the distance. This provides important innate depth cues, and the use of a dynamically changing video display will provide a wealth of innate depth clues. Violations of these cues can lead to unnecessary confusion and localization errors. Thus proper modeling of the virtual world will be an important part of OIV displays.
- *Structure-from-motion (SFM)* - A related cue that ties shape and motion together is SFM; here, visible bits of an object move in such a way as to give the viewer a sense of structure even when none is apparent in static displays. Implementing SFM into OIV systems would be useful if displays were found to have too much visual complexity---thus non-relevant information could be reduced from rendered models to simple outlines or vertices, which, when moved could give the sense of a solid structure without actually having to render it. This would have the benefit of reducing clutter by providing information about the shape defined only by simple vertices and movement of the viewer.

5.3. Binocular cues: stereopsis

Besides using information about depth and distance through the perceptual system's innate monocular-processing techniques (such as size constancy and motion parallax), one obvious AR design approach is to use binocular information. When perceiving visual stimuli with two eyes, humans automatically compare the visual signals that arrive to both eyes in order to make estimates about depth and distance. Other less salient visual-motor cues include ocular convergence (eye position information) and accommodation (changes of eye shape for focus).

6. Design solutions

A wide range of possible design solutions exist which can overcome the problems of depth ambiguity and visual complexity in an OIV environment. The optimal design will likely unite a combination of approaches, integrating image-processing techniques that make use of a variety of perceptually salient cues. In fact, studies have shown that different visual cues provide varying amounts of accuracy when judging depth and distance [26]. Binocular cues are most accurate only for short distances (<3 m), while people tend to utilize monocular cues at longer distances (motion parallax is used at medium distances (3-10m) while size-constancy is used at longer distances (>10m)) [4].

Some practical approaches that use perceptual cues to convey distance include:

- *Additive transparency*: Rendered surfaces that use additive properties of transparency to convey depth/distance (see Figure 2 (left)).
- *Size scaling of rendered surfaces*: Utilize rendered objects with a known, fixed size, and scale the size as a function of distance to convey depth/distance.
- *Over-rendered transparency*: rendering virtual cut-aways of existing solid objects to portray what is on the other side (see Figure 2 (right)).

Other useful visualizations in OIV might include a blend of perceptual and metaphorical information:

- *Distance markers*: Virtual yard sticks or text information conveying distance could act as literal distance markers between the viewer and an object, similar to virtual tape measures described in AR [21].
- *Temporal distance coding*: AR information could be presented such that all items of a similar distance are displayed at the same time, but AR information for objects at other distances would appear at different times (different temporal phases) or different temporal rates (2 Hz vs. 0.5 Hz).
- *Ground-plane grids*: Incorporate rendered grids as a ground plane for relating relative and absolute distances.
- *Marker fore-shortening*: The width of lines that connect rendered AR information (e.g., text tags)

with real-world objects could vary as a function of distance; line widths in the foreground would be wider than line widths at greater distances.

- *Alternate perspective*: Distance information can also be accurately conveyed through the use of multiple perspectives, such as a top-down exocentric view.
- *Symbolic representation*: A novel system/language of symbols could be developed and used that specified depth, distance, and/or specific spatial location information.

6.1. General guidelines

We have generated some *a priori* guidelines based on the cognitive principles outlined above. These guidelines address some of the major perceptual issues involved with OIV systems but may also generalize to other types of AR and VR displays.

Important design guidelines include:

- *Distance conveyance* – In OIV environments, distance and absolute location can be confusing, so AR renderings should disambiguate information about distance or position.
- *Proper motion physics* – For dynamic displays, motion parallax is an important cue to human observers, so it is important that the information in depth move in such a way as to convey its proper position. This can be achieved with properly defined geometries and metrically-accurate models of the environment that rotate and move in realistic ways.
- *Eliminate unneeded AR motion* – Because the human visual system is so sensitive to motion, unneeded motion of rendered material (e.g., the slowly moving self-organization of rendered AR tags) as well as mis-registration should be eliminated or minimized as much as possible.
- *Selective or multiple cues* – Because the accuracy of different perceptual cues vary with distance, specific perceptual cues (e.g., motion parallax or size-constancy/scaling) should be selected if displays are operating in an environment of a limited range of depths/distances. Multiple perceptual cues should be integrated if the ranges of display depth/distances are variable.

There are many other factors that do not deal directly with perceptual influence on OIV, but are important features to cover, including:

- *Define rule space* – Another important factor for OIV system development is to define the conditions under which augmented information should be displayed. For example, if the viewer knows that augmented information will only be presented on hallway surfaces or only within the confines of the building, then these rules can help disambiguate otherwise vague location information.

Carefully defining a series of rules, such as

having different color or symbols to designate particular locations, will go a long way towards building an unambiguous OIV system. Even implicit rules or context clues, such as augmented messages only appear in one's office, also reduce positional uncertainty. So, while more detailed rules increase the cognitive complexity for the human user, more sophisticated rule systems can drastically reduce the ambiguity in OIV displays.

- *Effectiveness testing* – While these general guidelines can be used as a starting point, some form of experimental testing (whether pilot testing or more involved experimental procedures) should be incorporated through the design development. Such an empirical approach can improve overall system design as certain specific perceptual cues may be better suited for certain applications. Multiple cues may interact in a way that could be quantified and subsequently compared to other designs. The instructions that participants get in experiments should be created very carefully so as not to deliberately bias subjects' responses. Finally, the design development of similar systems could be improved using other formalized usability-engineering processes that provide a structured evaluation technique [11].

7. OIV testbed

We have begun to implement an OIV environment on an actual AR setup. The AR system is a video see-through design, where the camera is tracked with a HiBall 3000 optical tracker, made by 3rdTech. The rendering code is written in OpenGL and the video capturing is done through DirectShow. The rendering is done on a dual processor Xeon 1.7 GHz PC running Windows XP, with an NVIDIA Quadro 2 graphics board. This OIV testbed was used to generate the stimuli for the pilot experiment (Section 9). This OIV system could be implemented on a biocular head-worn system, but, for the purpose of this experiment, the AR output of the system was recorded on video tape and then saved out as movie files to be played back as stimuli for the experiment participants.

8. Alphanumeric information

While the types of information presented in an OIV system can vary (text, objects, avatars, etc.), one obvious direction of OIV development is the integration of alphanumeric information. The practical application of alphanumeric information in OIV will face the same general issues listed above (Section 3. OIV issues) as well as a unique subset of problems.

The presentation of rendered text in OIVs must not only convey distance and depth information, but must also convey the augmented information in a way that maintains the readability of the text. Thus, when

algorithms are used that convey distance information using variations of transparency, contrast, size, occlusion, and color/saturation, care must be taken to ensure that text labels are still readable.

An additional guideline that should be considered when using alphanumeric information is:

- *Readability testing* – The readability of text in OIV displays should be tested across a range of distances, light-levels, and scene complexities to ensure that the algorithms for text presentations are robust.

9. Preliminary experiment

The use of pilot experiments can be instrumental in guiding the design of successful human-centered systems. And since the focus of this discourse is on the use of cognitive principles to guide the design of an OIV display, application of experimental methods can reveal which specific features (or combination of features) are best suited for effectively relaying information.

One key to successful experimental design is pinpointing a precise question that will be the focus of the experiment. Many potential lines of empirical testing exist, including determining:

- Which combination of monocular perceptual cues provide the most accurate distance information for text labels in an OIV environment?
- What are the best ways to toggle the amount of transparency in a display? What are the best controls for a user to vary transparency information?
- How does the addition of concurrent exocentric maps improve localization performance?
- Are literal distance markers (such as quantifying distance) more effective at conveying distance than implied or relative markers (i.e., the by-product of perceptual cues or rendered ground plane grids)?
- How are people's depth perception and distance judgments affected when using an OIV/AR display as compared to unaided judgments?
- How does practice /training with an OIV device improve perceptual performance accuracy and speed? What role does previous spatial knowledge / expertise play in successfully disambiguating uncertain OIV displays?

We chose to focus on two questions in particular:

- (1) Whether people actually suffer from depth ambiguity in an OIV environment?
- (2) How well do monocular perceptual cues provide accurate distance information in an OIV environment?

9.1. Goals

The goals of this pilot experiment were three fold. First, this experiment aimed to validate that, for even the simplest of displays, rendered information (a non-shaded, colored square) that does not literally convey distance or depth, does produce depth ambiguities that impair

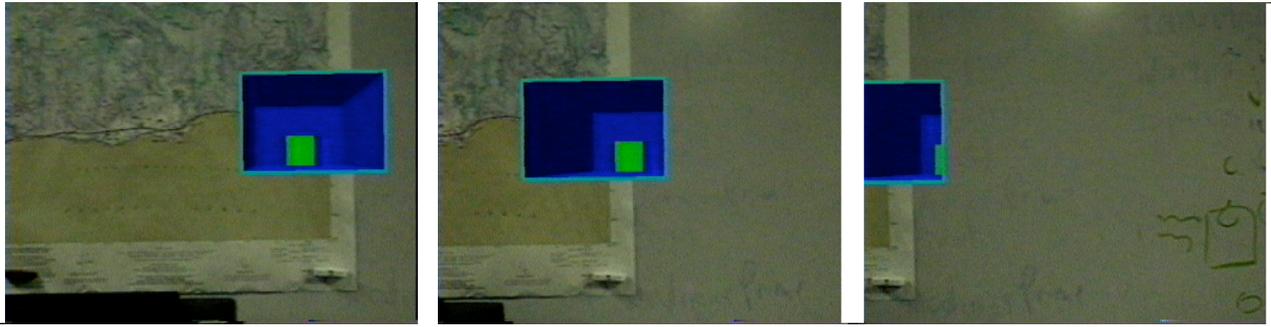


Figure 3: Three frames from a stimulus video capture generated with an AR-OIV system. In this particular example of a dynamic display (panning right), the target square is rendered to a location 1m behind the real-world map and is located within a rendered cut-away box. These frames illustrate 2 major perceptual cues: motion parallax (the small rendered target square moves relatively less than the map because the map is closer to the viewer than the rendered target) and occlusion (the target square is obscured by the boundaries of the cut away).

location judgments in static OIV displays. Second, we aimed to test if a subset of perceptual cues could be used to more accurately convey depth information, thus overcoming depth ambiguities. Third, and finally, we wanted to develop and test our OIV system.

9.2. Methods

This experiment typifies the kind of study that could be run on our OIV system. The methods for this pilot experiment were as follows:

- *Stimuli*: Images were generated using our functioning OIV testbed (see Section 7 for details). The actual stimuli were OIV displays that were digitally captured and presented in a series of 6 edited, uncompressed digital-video clips, each lasting about 10 seconds. In each display, a rendered bright-green 2-D square acted as the target of the subjects' depth judgments (see *Task*, below). Three frames from a dynamic display are presented in Figure 3. Note: all of the elements of displays (rendered graphics and video) used in the experiment were presented to subjects in color (even though they appear here as gray-scale images).
- *Conditions*: This experiment had 6 different conditions (2 camera motions x 3 target conditions). The rendered target square could appear at a fixed location on the map (in the same depth plane as the map), fixed approximately 1m behind the wall, or fixed about 1m behind the wall with the addition of a rendered cut-away which began at the wall and extended back 1m to the location of the square.

Because we didn't want subjects to use size as cue (closer targets would normally appear larger), we adjusted the size of target in the far position so that the projected sizes of the target in all conditions were approximately equal.

For half the trials, the camera remained still for the length of the presentation, while for the other half of trials, the camera panned left and right, while rotating in order to keep the target square in view.

While the camera moved (intended to simulate a person moving) yielding the dynamic display, the physical location of the rendered square remained fixed.

- *Task*: The subjects' primary task was to judge the location (in depth) of the rendered target square. Subjects were given a 3-alternative forced choice of target locations (1 m behind the map, on the map, or 1 m in front of the map) even though the target actually only appeared behind or on the map. The entire procedure lasted about 5 minutes.
- *Design*: The experimental design was a standard within-subjects design, in which each subject was exposed to all conditions. The order of the conditions was presented in a pseudo-randomized manner using Latin-square counterbalancing.
- *Participants*: Subjects were 8 HRL employee volunteers.
- *Apparatus*: Stimuli were presented to subjects on an Apple dual processor 1Ghz G4 with an NVIDIA GeForce4 MX graphics card and a 17" flat panel display. Subjects responded via a pen-and-paper questionnaire.
- *Instructions*: Subjects were instructed that they would be viewing 6 video clips from an AR setup. Subjects were told that the rendered object was at fixed spatial location and could appear at one of 3 different depth planes (1 m behind, on, or in 1m in front of the map wall), and were also given a schematic diagram illustrating the different locations in depth. They were told they were to make a judgment on the depth of the target, as well as to rate their confidence on their judgment on a 1-3 scale. They were also told not to explicitly use size or reflectance as the basis for their depth judgment.

9.3. Results

The results from this experiment are presented in Figure 4. In general, subjects tended to perceive the target square in front of the map, even when the target appeared at a location behind the map (e.g., Figure 4, top-middle cell (Static camera; target rendered behind the map)). The only condition in which subjects reported perceiving the target in the correct location was for dynamic displays, when the cut-away was added to the target in the back position.

9.4. Discussion

The data from this experiment reveal several interesting patterns. First, the data presented in Figure 4 (top-left and top-middle cells) are consistent with a depth-ambiguous percept. In fact, the two static displays (with the target rendered behind and on the wall) were perceptually identical because we made sure the projected sizes and perceived locations of the targets were identical. While it may seem redundant to present subjects with two perceptually identical stimuli, it was a necessary control; we wanted to experimentally validate that even though the target stimuli were rendered in different spatial locations (on the wall or 1m behind the wall) and at different sizes (to eliminate any size differences as a cue for distance), subjects could not perceptually distinguish between the two (which they couldn't). The fact that subjects responded more or less equally poorly across these conditions is consistent with an ability to properly identify the position of the target in depth (thus demonstrating depth ambiguity).

Subjects also showed a strong tendency to perceive the stimulus as in front of the wall, even in the face of strong perceptual cues indicating otherwise. Cues from motion parallax, often regarded as one of the more salient perceptual cues, by themselves, failed to overcome subjects' perception of seeing the rendered target as in front of the wall (see Figure 4, bottom-middle cell). This finding is consistent with other reports that find that occlusion is the dominant perceptual cue for depth judgments [3].

However, other subtle percepts revealed in a post-experiment debriefing suggest that some of the bias of "in front" may be due to technical limitations of our AR implementation. In this information debriefing, subjects were asked to explain why they made their decision and were to comment once they were told the correct location of the target. Subjects often pointed out small vertical and horizontal jitters (due to mis-registration) in the position of the target (especially during the dynamic displays). This jitter was due to noise in the optical tracker. Subjects said they used this jitter as evidence that the target was not part of the wall, but was in front of the wall. This points out the importance of proper AR registration, and how technical constraints may have a profound effect on the desired perception of the rendered

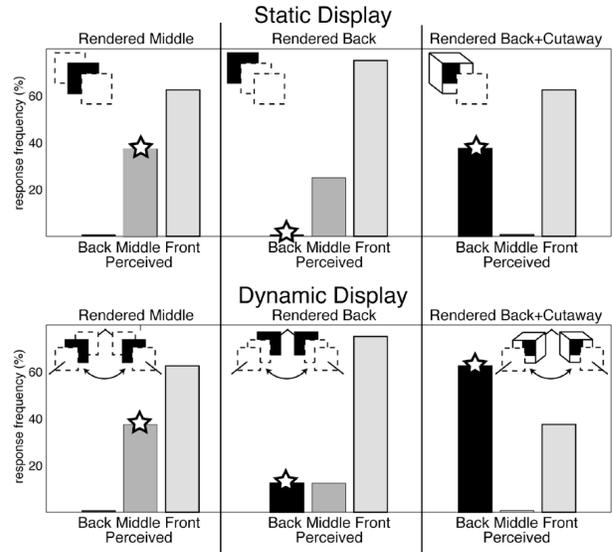


Figure 4: Experimental results for both static and dynamic AR-OIV displays. Bars are the response frequency (in percent of responses) plotted as a function perceived location. Each of the 6 triplets (cells) of bars represents 1 of the 6 experimental conditions; cell rows are static and dynamic displays and cell columns are the actual rendered locations of the target. Stars depict the rendered location of the target (which are also represented as shaded squares in the schematic graphic in the corner of each cell).

images. For example, reducing jitter through a "closed loop" tracking system that additionally observes fiducial markers at known locations might significantly affect the results. Thus additional guidelines and future parametric experiments that describe the tolerance of the human perceptual system to mis-registration, especially in a AR-OIV environment would be extremely beneficial for future system designs.

Yet, even when the target moved with the map (see Figure 1, bottom-left cell), subjects reported the target as being in front of the wall more often (62.5%) than on the wall (37.5%). This is consistent with the idea that subjects cannot suspend their innate perceptual knowledge of occlusion (the target square did occlude the wall), even when other cues (such as motion) convey the more relevant information (though contradict the percept that occlusion means in front). Further, the importance of occlusion is supported by fact the subjects were much more inclined to correctly identify the position of the target as behind the map (62.5%) when the target was occluded by the dynamic motion of the cut-away box (see Figure 4, bottom-right cell for data, and Figure 3 for gray-scale versions of frames from the actual dynamic cut-away stimulus).

The main goal of this experiment was to test if a subset of perceptual cues could be used to convey depth information. Four of the most relevant perceptual cues in OIV design that could affect the localization of augmented

information are size, transparency, occlusion, color, and motion. The ideal experimental design would test effectiveness of each of the cues separately, and then quantify how the use of multiple cues together interacted to improve (or diminish) performance. However, here, we simply tested if a certain subset of cues (i.e., motion, occlusion, and the image superposition (occluded items presented in front of occluding object)) could produce more accurate localization of AR targets.

10. Conclusions and future work

The goals of this paper were threefold: First, it was to outline some of the important issues involved in developing an AR-based system for visualizing obscured information. Second this paper aimed to review perceptual factors that could lead to effective designs of OIV systems. And third, this paper intended to empirically validate the effectiveness of implementing perceptual cues in an OIV mockup system through the use of a pilot experiment.

This paper serves to elucidate many of the important research questions involved with OIV. This paper also serves as the starting point for several potentially fruitful lines of future research, including the development of efficient ways to overcome people's seemingly innate tendency to use occlusion as the dominant cue in an OIV setting. Many of the research questions and important guidelines outlined in Sections 6 and 9 are currently being investigated through additional experimentation.

11. Acknowledgements

Howard Neely helped with video recording and capturing the sequences used in the experiment.

12. References

- [1] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent Advances in Augmented Reality. *IEEE Comp. Graph. & App*, vol. 21, no. 6 (Nov/Dec 2001), pp. 34-47.
- [2] B. Bell, S. Feiner, and T. Höllerer. View Management for Virtual and Augmented Reality. *Proc. ACM Symp. on User Interface Software and Technology* (Orlando, FL, 11-14 Nov. 2001), pp. 101-110.
- [3] Braunstein, M. L. Andersen, G. J., Rouse, M. W., & Tittle, J. S. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception & Psychophysics*, 40, pp. 216-224.
- [4] J.E. Cutting. (1997). How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments & Computers*, 29(1), pp. 27-36.
- [5] M.J. Daily. Color Image Segmentation. In *Advances in Image Analysis*, R. Gonzalez and Y. Mahdavih, eds., SPIE Press, 1993, pp. 552-562.
- [6] D. Drascic and P. Milgram. Perceptual Issues in Augmented Reality. *Proc. SPIE vol. 2653 Stereoscopic Displays and Virtual Reality Systems III* (San Jose, CA, Feb. 1996), pp. 123-134.
- [7] S.R. Ellis and B.M. Menges. Localization of Objects in the Near Visual Field. *Human Factors*, vol. 40, no. 3 (Sept. 1998), pp. 415-431.
- [8] S. Feiner, B. MacIntyre, and D. Seligmann. Knowledge Based Augmented Reality. *Comm. ACM*, vol. 36, no. 7 (July 1993), pp. 52-62.
- [9] S.K. Feiner, A.C. Webster, T.E. Krueger III, B. MacIntyre, and E.J. Keller. Architectural Anatomy. *Presence: Teleoperators and Virtual Environments*, vol. 4, no. 3 (Summer 1995), pp. 318-325.
- [10] A. Fuhrmann, G. Hesina, F. Faure, and M. Gervautz. Occlusion in collaborative augmented environments. *Computers & Graphics*, vol. 23, no. 6, (Dec. 1999), pp. 809-819.
- [11] J.L. Gabbard, J.E. Swan II, D. Hix, M. Lanzagorta, M. Livingston, D. Brown, and S. Julier, Usability Engineering: Domain Analysis Activities for Augmented Reality Systems. *Proceedings of the Conference on The Engineering Reality of Virtual Reality 2002*, SPIE (International Society for Optical Engineering) and IS&T (Society for Imaging Science and Technology) Electronic Imaging 2002, January 24, 2002.
- [12] R. Highnam and M. Brady. Model-Based Image Enhancement. *Proc. 12th Eurographics UK* (22-24 March 1994), pp. 410-415.
- [13] A. Hoover, G. Jean-Baptiste, X. Jiang, P. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D. Eggert, A. Fitzgibbon, & R. Fisher. An Experimental Comparison of Range Image Segmentation Algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 7 (July 1996), pp. 673-689.
- [14] G.S. Hubona, P. Wheeler, G. Shirah, and M. Brandt. The Relative Contributions of Stereo, Lighting, and Background Scenes in Promoting 3D Depth Visualization. *ACM Transactions on Computer-Human Interaction*, Vol. 6 (3), September 1999, pp. 214-242.
- [15] V. Interrante. Illustrating Surface Shape in Volume Data via Principal Direction-Driven 3D Line Integral Convolution. *Proc. SIGGRAPH '97* (Los Angeles, 3-8 Aug. 1997), pp. 109-116.

[16] M. Jägersand. Saliency Maps and Attention Selection in Scale and Spatial Coordinates: An Information Theoretic Approach. *Proc. 5th Int'l Conf. Computer Vision* (Boston, 20-23 June 1995), pp. 195-202.

[17] S. Julier, M. Lanzagorta, Y. Baillet, L. Rosenblum, S. Feiner, T. Höllerer, and S. Sestito. Information Filtering for Mobile Augmented Reality. *Proc. Int'l Symp. Augmented Reality 2000 (ISAR'00)*. (Munich, 5-6 Oct. 2000), pp. 3-11.

[18] C. Koch and S. Ullman. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. *Human Neurobiology*, vol. 4 (1985), pp. 219-227.

[19] B. MacIntyre and E.M. Coelho. Adapting to Dynamic Registration Errors Using Level of Error (LOE) Filtering. *Proc. Int'l Symp. Augmented Reality 2000 (ISAR'00)*. (Munich, 5-6 Oct. 2000), pp. 85-88.

[20] S. Mann. Humanistic Intelligence: 'WearComp' as a new framework and application for intelligent signal processing. *Proc. IEEE* vol. 86, no. 11 (Nov. 1998), pp. 2123-2151.

[21] P. Milgram, S. Zhai, D. Drascic, and J. Grodski. Applications of Augmented Reality for Human-Robot Communication. *Proceedings of IROS'93* (Yokohama, Japan, July 1993), pp. 1467-1476.

[22] U. Neumann, and A. Majoros. Cognitive, Performance, and Systems Issues for Augmented Reality Applications in Manufacturing and Maintenance. *Proc. IEEE Virtual Reality Ann. Int'l Symp. '98. (VRAIS '98)*. (Atlanta, 14-18 Mar. 1998), pp. 4-11.

[23] A. State, D. 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. *Proc. IEEE Visualization '94* (Washington, DC, 17-21 Oct. 1994), pp. 364-368.

[24] A. State, M.A. Livingston, G. Hirota, W.F. Garrett, M.C. Whitton, H. Fuchs and E.D. Pisano. Techniques for Augmented-Reality Systems: Realizing Ultrasound-Guided Needle Biopsies. *Proc. SIGGRAPH '96* (New Orleans, LA, 4-9 Aug. 1996), pp. 439-446.

[25] A. Sha'ashua and S. Ullman. Structural Saliency: The Detection of Globally Salient Structure Using a Locally Connected Network. *Proc. 2nd Int'l Conf. Computer Vision* (Tarpon Springs, FL, 5-8 Dec. 1988), pp. 321-327.

[26] Surdick, R.T., Davis, E.T., King, R.A., Hodges, L.F., 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) 1997, pp. 513-531.