Pursuit of "X-ray Vision" for Augmented Reality

Mark A. Livingston, Arindam Dey, Christian Sandor, and Bruce H. Thomas

Abstract The ability to visualize occluded objects or people offers tremendous potential to users of augmented reality (AR). This is especially true in mobile applications for urban environments, in which navigation and other operations are hindered by the urban infrastructure. This "X-ray vision" feature has intrigued and challenged AR system designers for many years, with only modest progress in demonstrating a useful and usable capability. The most obvious challenge is to the human visual system, which is being asked to interpret a very unnatural perceptual metaphor. We review the perceptual background to understand how the visual system infers depth and how these assumptions are or are not met by augmented reality displays. In response to these challenges, several visualization metaphors have been proposed; we survey these in light of the perceptual background. Because augmented reality systems are user-centered, it is important to evaluate how well these visualization metaphors enable users to perform tasks that benefit from X-ray vision. We summarize studies reported in the literature. Drawing upon these analyses, we offer suggestions for future research on this tantalizing capability.

Mark A. Livingston
U.S. Naval Research Laboratory, Washington, DC, 20375, USA
e-mail: mark.livingston@nrl.navy.mil

Arindam Dey
Univ. of South Australia, Adelaide, South Australia 5001
e-mail: Arindam.Dey@unisa.edu.au

Christian Sandor
Univ. of South Australia, Adelaide, South Australia 5001
e-mail: chris.sandor@gmail.com

Bruce H. Thomas
Univ. of South Australia, Adelaide, South Australia 5001
e-mail: Bruce.Thomas@unisa.edu.au

1 Introduction

One of the most intriguing capabilities envisioned for augmented reality (AR) systems is the notion of “X-ray vision,” or the ability to virtually “see through” one surface to what in reality is hidden from view. Many AR systems have been premised on this feature for the primary value provided by the application. Furness’ pioneering work [15, 16] on head-up displays for pilots was motivated in large part by the ability to virtually see through the solid cockpit walls and floor to a virtual copy of the real world hidden by the aircraft infrastructure. The ultrasound AR system [2] provided the ability to understand the data acquired by the ultrasound probe from inside the body; to do this requires a metaphor that ensures the user will understand the data to reside behind the visible skin surface, and in fact this was a problem noted early in the development of the system. We believe that the metaphor “X-ray vision” was first applied to this work [45].

There are two sides to this unique perceptual capability. There is the issue of the absolute distance of graphical entities within the coordinate system of the real environment. There is also the relative order of real and virtual surfaces within the merged environment. Neither of these perceptual capabilities seem to come naturally for AR users, and thus numerous visual metaphors have been conceived in order to give the impression of relative and absolute depth. Another issue that challenges the designer of X-ray visualization metaphors is the potential to overload the user with information [20, 28]. If the entire database of known objects is shown (generously assumed to be accurately registered to the real world), then the user could easily be overwhelmed with information and be unable to understand the depths of any object.

Developing X-ray vision AR systems is a difficult problem from a number of perspectives. First, X-ray vision is truly an unnatural act. This is not an ability people perform without the aid of perceptual and cognitive mechanisms, and as such there are few *metaphors* to which people can relate. Therefore, careful consideration of depth cues and their usability, as well as the introduction of new cues, are required. Second, the presentation of X-ray vision information to the user is more difficult than presenting depth in traditional three-dimensional (3D) graphics or even virtual environments (VE). In the latter, there is a rich history of of experiments showing that depth is underestimated. This is due to the fact that the system is not in complete control of the visual information received by the user. This chapter will explore the perceptual background, visual metaphors, and empirical investigations into overcoming this challenging and important topic in AR research. We conclude with an analysis of the research and suggestions for further research.

2 Perceptual Background

Our perception of depth draws on a number of cues. These cues interact with the distance from us; they also interact with the properties of AR displays in unique ways.

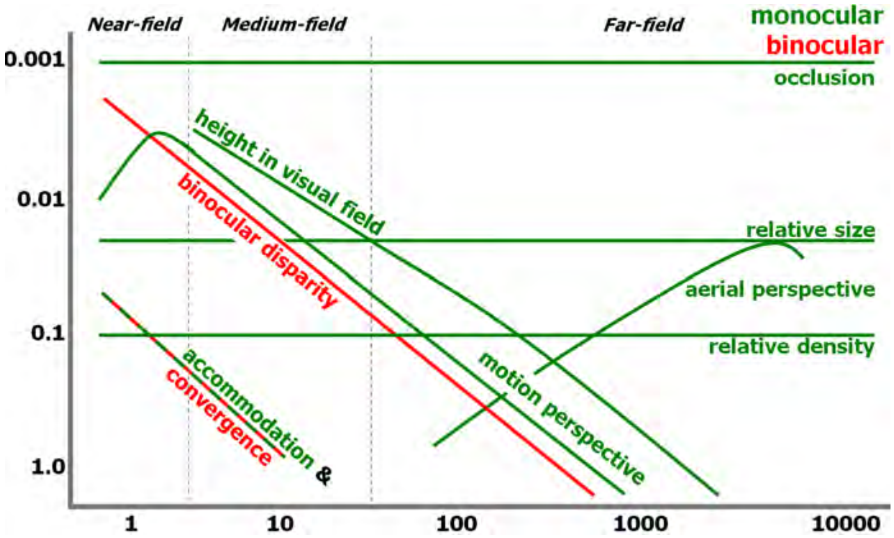


Fig. 1 Depth cues and the depth discrimination threshold (where smaller numbers indicate more potent cues) in the three fields of depth. In Heiko Hecht, Robert Schwartz, and Margaret Atherton, eds., Looking into Pictures: An Interdisciplinary Approach to Pictorial Space, top half of Figure 11.2, page 223, ©2003 Massachusetts Institute of Technology, by permission of The MIT Press. This chart is modeled on one drawn by Cutting [6] using data from Nagata [34] and Cutting.

Cutting divided the environment into three regions: personal space (generally, within arm’s length), action space (a distance at which one can reliably and accurately interact with other entities), and vista space (anything beyond action space)[1]. Fig. 1 shows his ordering of the cues within each of these regions. This section consists of a brief review of each depth cue and how they can be affected by AR displays. For more complete reviews of depth perception, we refer to reader to textbooks on perception [5, 42] or chapters [6, 7, 34] which detail depth cues, their relative strengths, and variations noted from the general patterns depicted in Fig. 1. Readers will also note that some authors delineate and categorize cues differently; we use the categories of Cutting [6] to match Fig. 1. We also note that depth cues do not, in general, function in isolation; *cue combination* leads the human visual system to understanding of depth. Our analysis here is limited to individual cues.

A few terms will be helpful in the following discussion. The *human visual system* should be understood as consisting of the complete pathway that begins with the retina, travels up the optic nerve, passes through the lateral geniculate nucleus, and ends in the visual cortex. Interested readers may consult a recommended textbook for details, but the important concept is that all of these components play a role in perception, and in depth perception in particular. Measurements taken in the retina will be compared along the pathway and processed in the cortex to interpret

[1] Personal space, action space, and vista space are commonly termed as near-field, medium-field, and far-field distances, respectively. In this chapter, we will use the latter set of terms.

depth. *Ordinal depth* refers to identification of the relative depth of objects: which is closest, which is next after that, etc. *Metric depth* indicates that a measurement (e.g. in meters) can be ascertained by the observer, at whatever level of accuracy the observer is capable. Most of the cues are *monocular*; they apply to a single eye. Two cues are *binocular*; they require two eyes to be applied to a scene or a visual system. Fig. 1 shows monocular cues in green and binocular cues in red. When considering distances in the environment, we can classify them as *egocentric* distances, indicating a distance from the observer's viewpoint, or *exocentric*, indicating a distance between two objects within the field of regard. All of the work in AR has studied egocentric distance for the AR user.

2.1 Occlusion

When a solid object prevents light rays from another object from reaching the observer directly, we define this as a case of **occlusion**. If the closer object, known as the *occluder*, blocks only some of the rays from the more distant, or *occluded*, object, we call this **partial occlusion**. This cue provides information about ordinal depth amongst the objects, but not about the metric depth. As depicted in Fig. 1, occlusion is generally the most powerful depth cue in any region of space. That this cue provides such powerful depth information derives from the fact that we are accustomed to a world populated by solid objects that are opaque. Of course, we do encounter transparent or translucent objects regularly, and situations such as windows have become sufficiently common that we incorporate them smoothly into our understanding. Other translucent media will be discussed under the aerial perspective cue.

Using occlusion as a depth cue also relies on an assumption often known as Helmholtz's rule: the contour of the occluder does not change direction where it intersects the contour of the occluded object. Such a coincidence would likely cause a misinterpretation of the depth order and/or the shapes of the respective objects. (The phenomenon of false contours gives another example of a possible misinterpretation.) One can imagine the difficulty of interpreting shape in a world made up of objects that are composed of wireframe outlines, with no solid surfaces. In such a world, the ambiguity of occlusion ordering would be high. Another assumption should be obvious but is worth noting for the discussion below. One must also be able to differentiate the two objects – i.e. there must be a difference in the brightness or color of the occluder and occluded object so that the contour is detected by the visual system.

Optical see-through (OST) AR displays often prevent the use of the occlusion cue by the human visual system. The optical combination of real and virtual is generally achieved through a partially-silvered mirror. Thus for every pixel in the virtual image, there is a real solid angle behind it that will show through the combiner. Wherever graphical objects are visible, the graphics do reduce the salience of the real world. Thus the graphics cannot be ordered in depth relative to the real objects merely with the occlusion cue. As noted above, for certain translucent objects we experience everyday, this situation may become sufficiently familiar to understand the geometric relationships. However, for most AR users of optical displays, the situation has yet to become familiar. A few research prototypes of optical displays have been demonstrated [23] with the capability (usually via a "masking" display in the optical path to the real world) to occlude the real world at pixels (or blocks of pixels) that are part of virtual objects. At this time, we are unaware of commercially-available optical displays with this feature. On the other hand, video-overlay AR displays can overwrite the video pixels completely and thus occlude the real world with the virtual objects. This is true both for those that use a form of chromakey replacement and those that use a graphical frame buffer and render directly onto the image. The occlusion cue is thus available in AR systems that use video displays. Also, since this cue is completely reliant on the 2D projection of the world onto the retina, there is no difference between head-worn and hand-held displays for this cue. All AR applications will suffer some loss of fidelity of this cue if they are unable to accurately align the graphics to the real environment, an error in *registration*.

2.2 Binocular Disparity

Because of the displacement between the two eyes, each eye sees a slightly different image of an object; in particular, there is a shift in the position on each retina. This translation is known as **binocular disparity**, and from it the visual system can compute the distance to the object. In terms borrowed from AR (or more precisely, from computer vision), this cue relies on a calibration of the distance between the eyes. Our visual system learns this calibration through everyday experience of the world. As also known in computer vision, the visual system must identify corresponding points on the object in order to compute the disparity. This cue interacts with convergence, discussed below, since that also can affect the apparent translation on the retina. As shown in Fig. 1, binocular disparity is a very strong cue at close distances; however, because the amount of translation decays with increasing distance (due to perspective foreshortening), its strength diminishes as an object moves farther away.

Providing correct binocular disparity provides a challenge for head-worn AR displays, but recent commercial offerings are able to meet this requirement. Many older systems did not permit the adjustment of the inter-pupillary distance (IPD) of the display hardware. Thus, the AR user would experience an incorrect binocular disparity without careful calibration of the device's fit for each user. For optical displays, measurement of the IPD and setting the hardware and software rendering to match provides correct binocular disparity. For video displays, the camera mountings on the display must be included on the display hardware mounting and moved in concert. For hand-held displays, there is little choice but to experience the disparity that is dependent on the distance to the display surface, which is generally held within the near-field (especially if one accepts the limitation to arm's length). Some mobile phones are now using autostereoscopic displays, which is an example of the possibility for a hand-held display to enable the user to experience binocular disparity that is dependent on the virtual distance to a virtual object rather than the real distance to the display surface. In this case, a calibration of the IPD and the distance to the display will permit proper binocular disparity for virtual objects. Another fundamental issue for depicting binocular disparity is the limited resolution of the display elements used for the virtual imagery in optical displays and both real and virtual imagery in video displays. Given that the visual angle occupied by each pixel barely supports normal human vision (cf. Chapter 3), it stands to reason that the retinal measurements of disparity suffer from this limited resolution as a consequence of the display hardware limitation.

2.3 Motion Parallax

The relative movement on the retina of stationary objects caused by egomotion of the observer is known as **motion parallax**. The similar term **motion perspective** is sometimes used as a synonym and sometimes applied to depth perceived for objects moving near or around an observer (whether parallel or perpendicular to the central view ray, or at any angle). In general, one may infer the relative motion and depths of objects moving relative to the observer. This cue is quite powerful in the near-field, and – in an analogous fashion to the decay in strength of binocular disparity – decays with increasing distance to the objects due to the relatively small motion on the retina that a distant object's relative motion will cause. Fig. 1 shows another curious effect for motion parallax (and motion perspective, if considered separately), but one which has a simple explanation. The cue decreases in effectiveness at extremely close distances, because relatively large translations on the retina are difficult to track. Additionally, rapid motions on the retina do not lend themselves detection of differences, so interpreting relative depth of two nearby moving objects through motion perspective is challenging. Thus the graph of the sensitivity for this cue (Fig. 1) is not monotonic.

Since motion parallax is a purely geometric cue, both optical and video AR displays are capable of depicting the necessary information to make use of this cue. Of course, this statement assumes that the problems described under binocular disparity do not prevent understanding of the relative motion on the display surface. The limited resolution of the display hardware – and the camera of video displays – can present a significant problem for both nearby and distant objects, but most AR applications avoid displaying extremely close objects. Given the observation of the loss of fidelity at extremely close distances, this avoidance is probably advisable for proper depth perception. The loss of resolution for distant objects makes this cue ineffective for some of the studies described below. Hand-held displays convey this cue in theory, but the limited angular field of view (FOV) may limit the ability to apply this cue. Noise in the tracking that creates dynamic registration error will interfere with interpretation of smooth motion and thus with this cue.

2.4 Height in Visual Field

We are generally accustomed to seeing object sit on the ground. Imagine a desktop with objects sitting on the planar surface and the observer's eye above the plane. Then the **height in the visual field** would be a simple function of the egocentric distance from the observer to the objects and the perspective projection parameters of the human visual system. Just as with the innate calibration of the IPD, humans have an understanding of their height from the ground plane. However, this knowledge can be clouded by the situation (e.g. uneven ground). In addition, visual acuity decreases the precision of the measurements with increasing distance needed to turn height in the visual field into metric depth. Thus while ordinal depth may be computed via height in the visual field, metric depth presents more of a challenge, and the utility of this cue decreases with increasing distance to the object. Further, the sharp observer will note that Fig. 1 omits this cue from the near-field. This is based on the assumption that an observer's central view ray is parallel to the ground. Given the vertical limits of the human FOV (in particular, approximately 45° below horizontal), this means that an object must have a minimum distance equal to the height of the observer's eye in order to be in the FOV. Clearly, the observer can simply look down to remedy this situation for a nearby object (at the obvious cost of limiting the distance objects can be). In unusual situations such as looking down from a balcony, this cue has analogies (e.g. "height" – distance – away from the side of the building that are likely of little practical value.

Height in the visual field is also a purely geometric cue, so AR displays – both hand-held and head-worn – generally handle this accurately. Again, limited display resolution will limit the utility of this cue in providing metric depth information for graphics in all AR displays and for real objects in video AR displays. Also, the limited FOV may move the effective region of this cue even farther from the user than the beginning of action space. Many head-worn displays do not offer a large vertical FOV. Hand-held displays, inasmuch as they are held farther from the eye than head-worn displays rest, suffer even greater loss of the effective distance of this cue. Registration error, especially vertical misalignment to the correct real-world location, will further reduce the usefulness of this as a cue in AR.

2.5 Relative Size

Artists have long understood the basic relationships of linear perspective: distant objects occupy a smaller area on the retina than nearby objects of the same size. Thus the **relative size** of the nearby object on the retina is larger. If the objects are recognized (e.g. another person, or a standard door frame) or their absolute size is otherwise known, then ordinal and metric depth may be ascertained from this cue. The key concept is that this is a *relative* cue; there must be a basis for comparison, either within the scene or from experience. Thus a tree, which may be quite recognizable but have an uncertainty associated with its height or with the breadth of its branches, provides a less accurate (but still useful) cue through relative size. The threshold to see a difference from relative size is higher than to discern occlusion, so while relative size is effective throughout the visual fields, it is less effective than occlusion, and in the near-field and medium-field, less effective than the other cues discussed above (Fig. 1).

Relative size is also a purely geometric cue, and as such is in theory achievable by both hand-held and head-worn AR displays. But like the other geometric cues, display and camera resolution as well as registration error can reduce its utility.

2.6 Relative Density

We turn our attention to a monocular geometric cue that is quite similar to relative size. A cluster of objects or features in a texture will have a characteristic spacing on the retina which is called **relative density**. This is perhaps most easily conceived as analogous to a checkerboard pattern used in computer vision calibration techniques. If such an object were of known size and placed in an AR environment, then the observer would be able to infer its distance by the perspective foreshortening effect on the texture. Generalize this to any pattern of objects for which a comparison can be made – again, either to another object (or group of objects) in the scene or to something from experience – and one has successfully constructed this cue. Relative density can also yield both ordinal and metric depth perception. It is less effective than relative size, perhaps by almost an order of magnitude (Fig. 1).

As relative density is a purely geometric cue, the only observation necessary for AR displays is that, owing to its lower effectiveness, it would seem likely to be even more sensitive to display (and camera) resolution and, potentially, dynamic registration error than relative size.

2.7 Convergence

The other binocular cue is the angle between the central view ray of the two eyes, a measure of **convergence**. By convention, large angles are those needed to fix nearby objects in both retinal images, whereas a focus on the horizon reduces the angle to 0° (parallel view rays for the two eyes). This cue interacts with binocular disparity, since both change the retinal position of objects. The human visual system is calibrated by experience to understand the difference, since convergence is an **oculomotor** cue – i.e. dependent on the muscles of the eye – whereas binocular disparity is not. As Fig. 1 indicates, this cue is effective only in the near-field and slightly into the medium-field. An exercise in trigonometry should convince the reader why this cue's effectiveness drops so dramatically with distance, but can be used to provide metric depth for nearby objects. The angular resolution of the oculomotor system would need to be very high.

This cue is typically completely lost with AR displays. Hand-held displays force the user to converge on a single display surface; thus any information about distance is to the display, not to the virtual objects depicted. Head-worn displays that provide separate display elements for each eye could in theory automatically adjust this angle or change the imagery in order to compensate appropriately for the geometric adjustment, but in practice, no displays do this. Because it is an oculomotor cue, it is nearly impossible for an AR display to correctly stimulate the visual system with this cue. For a video AR display, the cameras would have to move in concert with the oculomotor changes of the eyes. McCandless and Ellis (2000) [32] presented a quantitative analysis of eye position during changes in convergence, concluding that accurate depiction of a fixed stimulus distance in a binocular display requires real-time compensation for the observer’s eye movements.

2.8 Accommodation

The eye naturally focuses on an object of interest; the change in the depth of the focus distance is called **accommodation**. A static focus distance may not provide much information, but changes do, albeit only in the near-field and some of the medium-field (Fig. 1), exactly equal to the effectiveness of convergence (hence the graph of their effectiveness looks to be dashed). This cue’s effectiveness also decreases with increasing age. The information provides metric depth in the near-field.

This cue is also typically completely lost with AR displays. Most graphics (AR and other) are rendered with a pinhole camera model, thus creating infinite depth of field; thus, all objects in the FOV are in focus. While rendering methods exist for out-of-focus rendering, they can be computationally expensive and are thus rarely used in interactive systems. Post-processing methods can compute and add the proper amount of blur [24], but these are also rarely used. While recent research has demonstrated an optical display with multiple focal planes [25], this seems to be a long time from entering commercial markets. Video AR displays rely on the camera to provide the accommodation cue (i.e. focus) for the real environment, and generally cannot change this setting in real time. In theory, a real-time focus mechanism and an eye tracker could be coupled together via software control, but we cannot cite an example of this being implemented.

2.9 Aerial Perspective

We began our discussion with the strongest depth cue, the occlusion of distance objects by intervening objects and noted that translucent objects produced variation in the interpretations of depth. However, if the human visual system can attribute the appearance of an object to intervening translucent media, such as atmospheric haze, fog, or even “clear” air, then it invokes the cue of **aerial perspective**. The uniqueness of this cue is seen in two ways in Fig. 1. First, it is generally ineffective in the near-field and medium-field; second, the dominant trend is for its effectiveness *increase* with distance, until at great distances near the limits of the human visual system, it suffers a slight reversal of this trend. Both of these features may be explained by the amount of intervening media. Until there is a sufficient amount, the cue is ineffective, and as more intervening media accumulates, the effectiveness increases until the object is no longer discerned and the cue can no longer be applied. It is exceptionally difficult to characterize all instances of this effect; for example, an extremely thick fog may provide significant information in the near-field and block information beyond a few meters into the medium-field. The more typical case is that which is graphed in Fig. 1.

Graphics hardware has offered simulations of this effect for many years, although its use in AR is undocumented. Moreover, for this cue to be effective, it should match the real environment, which would require measures of the atmospheric media at and up to the application-critical distances from the user. While in theory such measurements are obtainable, in practice they would seem unworthy of the attention given that the maximum strength of this cue is less than relative size, and only approaches that level at great distances, beyond where most AR applications have required (or demonstrated) capability of accurately registering graphics. The nature of this effect would also seem to work against optical AR; this cue, when accurately depicted, will make objects' apparent colors mix with the intervening media. This may be difficult to differentiate from the optical mixing of colors of virtual objects with the background colors. Video AR can overcome this mixing, but still may need to account for the color context to properly convey the subtle color shift of virtual objects due to this cue. Hand-held and head-worn display would appear to have little difference, although if the application wants to depict a thick fog in the near-field, the distance of the display from the user's eyes may also be critical to proper depiction of this cue.

3 Depth Estimation Protocols

In this section, we briefly review the tasks that have been used to study the perception of egocentric distance. A more complete review may be found in the perception literature [31], which will include protocols for measurement of exocentric distances. We note the applicability to AR for each measurement protocol.

3.1 Verbal Estimation

Perhaps the most straightforward way to measure an observer's perception of distance is to have the observer report the distance in some standard measurement metric, most commonly feet or meters. This protocol should be understood to include non-verbal methods of entering the data, such as typing a number into a keyboard. The key aspect is that the stimulus is constant, and the observer’s response is (in theory) a continuous value (although the observer will discretize the response). The advantage of this protocol is that it can be applied in nearly any situation; it was used in several of the experiments discussed in Sec. 5. An inverse method of this protocol could be implemented by instructing an observer to place an object at a given egocentric distance. These methods are *open-loop*, since there is no feedback available to the user to determine whether an answer is accurate while the judgment is being made. This method requires that the user have a strong sense of the distance metric, which is a skill that an observer may not have acquired or practiced.

3.2 Forced-Choice

Another simple and widely-applicable method to measure perception is to ask the observer to choose between a small number of alternatives, such as which of two (or some small number) of objects is closest. This is, of course, a widely-used technique in user studies in general. Variations are easily-imagined and reasonable for evaluation tasks; for example, an observer may be asked whether a target is in front, between, or behind a set of reference objects distributed in depth. This protocol again offers the advantage of having few encumbrances in AR systems. However, it implies that ordinal depth measures are being collected, which may or may not be the desired protocol. This again is an open-loop task.

3.3 Perceptual Matching

A common method of estimating the metric distance of an object is to have the observer place it at the distance requested. The observer is given some method of controlling the distance to the object, and adjusts the location until it matches that of the reference object. This is a *closed-loop* task, because the user can see the changes being made to the target as adjustments are applied and thus receives feedback about the changing relative size. Since such motions are in themselves a depth cue, this visual feedback clearly informs the user as to the quality of the match. When implemented in an AR system, this implies the need for some interaction with the system. This can be as simple as manipulation with a mouse, or could even task the form of verbal commands such as “closer,” “farther,” and “stop.” This method is quite popular among the AR studies discussed in Sec. 5.

3.4 Reaching

For distance judgments to objects in the near-field (which is often defined as that that the observer can reach), the user may simply be asked to reach out and touch the object. This can be done accurately even when the user is blind-folded. Thus a *blind reaching* protocol may be employed to determine the egocentric distance. This is in a sense a form of perceptual matching if performed while the observer can still see the object whose distance is being perceived, so a blind-fold and object removal are often employed in a reaching task. There remains the possibility of tactile feedback if the user is able to touch a table or other objects continuously or at intervals along the path to the object's location. The user should be informed if the object has been removed while the blind-fold is in place, so that the lack of such feedback is not a surprise and does not induce an increase in the estimated distance. This technique is well within the capabilities of most AR systems, although the physical encumbrances such as various wires and hardware components of many AR systems must be carefully placed in order to avoid unintentional tactile cues. We are aware of only one study that has implemented this in AR up to the current time, however. One should also be very careful with the implementation of video AR systems; if they displace the camera viewpoint outside the eyes, the resulting distance perception will be heavily affected by the offset between the cameras' viewpoints and the eyes' viewpoints [4], assuming a stereo system. A monocular camera system used for bi-ocular (same image to both eyes) display is sure to cause additional problems for the binocular depth cues. This offset is especially important in the near-field, where this estimation protocol can be applied.

3.5 Walking

People are quite good at walking to a location of a previously-seen target. The user can be shown a target and (similarly to reaching), after removal of the object and putting on a blind-fold, walk to the target's location. This task is often used in the perception literature and has been applied to AR. The physical encumbrances of AR systems can come into play, so care must be taken that the user will not trip over wires or other equipment. Further, it has been shown that the perceived distance can depend on the effort to reach the target [50], so if a heavy head-worn AR display or a backpack loaded with devices to support the AR system is being used, this may result in distorted distance perception. Even a hand-held device, if it tires the user's arms, may induce fatigue and thus distort this measure of perceived distance. The caution regarding effort does apply to any of the distance measures discussed here, but seems most applicable to systems where the user knows that walking to the target is imminent. Clearly, this measurement protocol is most applicable to "reasonable" distances, so the medium-field seems to benefit the most from such a measure. However, care must be taken (using AR or not) to make sure that – if the subject returns to the same starting position for each trial – the return path does not give subjects unintended feedback about their accuracy. This is especially true if subjects remove the blindfold between trials.

A variation on this concept is for the subject to conduct *imagined walking* to the location of the target object. This has also been used with success in the perception literature [35]. The subject is asked to start a timer (e.g. a stopwatch or a computer-implemented timer) as he imagines beginning to walk to the target. Then the subject stops the timer after the time has elapsed that he believes he would need to reach the target. This yields (perhaps surprisingly) an accurate estimate of the distance; the only need is to "calibrate" the timing data by having the subject (actually) walk at his or her normal pace for a known distance before the trials begin. This time can then be used to convert the time for each trial into a distance estimate. The factors of fatigue and effort may play a role in this estimated time.

3.6 Throwing

Another visually-directed action that has been successfully applied in the perception literature is throwing a small object that won't bounce (e.g. bean bag) to the target's location. While this clearly requires a minimum level of motor skill on the part of the observer, it has been shown to be as accurate as walking to a target's location. With regard to AR, this again requires a low level of encumbrance from the AR system in order to engage in the action of throwing. Assuming the object to be thrown is light and easily held in the hand, this would appear to be an easy requirement to meet. Mobile AR systems, however, may find some difficulty in applying this, as will those AR systems (mobile or otherwise) that employ hand-held display devices. The combination of walking and throwing has been used with success in the perception literature.

3.7 Triangulation

As a substitute for walking great distances, a user can walk a short distance at an oblique angle to the object (typically at an angle between the line of sight to the object and a direction perpendicular to the line of sight). The subject can be asked to point continuously at the object while walking on this path, known as *triangulation by pointing*. Alternatively, the subject may be asked to walk (without pointing) until told to stop; upon stopping, the subject is asked to turn and face the target or turn and begin walking toward the target. The direction indicated by the pointing, facing, or walking is then used to triangulate the measurement of perceived distance. This can reduce the effort and time associated with a walking protocol. These protocols all work well for AR systems, albeit with the caveats expressed above regarding physical encumbrances. The weight that the subject is carrying would seem to be less of a concern, since the walking distance is presumably not going to be very far;

however, fatigue can still be a factor. One difficulty with this protocol in general is the low precision that results for larger distances. If there is a large range of distances being studied, the subjects may have greater capability for precision with the nearby distances, an argument that can be verified with simple trigonometry.

3.8 Size and Motion

The final methods of measuring perceived distance come directly from the application of depth cues and are thus thought to be less susceptible to interference from the cognitive level processing in the human visual system. The relationship between the size of an object and its distance is simple to express with trigonometric functions. This implies that the perceived size of an object implies an assessment of the perceived distance to that object. Thus asking a subject to assess the size of a distant object provides an indirect measure of the subject's perception of the distance to it. Of the above protocols, verbal estimation, forced-choice, and perceptual matching variations are easy to construct for this measurement.

In a variation of the triangulation protocol, a subject can be asked to view an object while moving. The object of known size serves as a reference for judging the distance that the observer moved; another simple trigonometric relationship converts the distance moved to a perceived distance to the object. The distance judgments through this protocol tend to be lower than through perceiving size.

Both of these protocols are amenable for AR systems. The resolution of the graphics may affect the perception of size and of displacement due to motion, as discussed above for the cues of relative size and motion parallax. For the motion estimate, the caveats regarding moving (expressed above for other protocols in which the observer moves) apply.

3.9 Summary of Protocols

As one can see, there are a number of protocols that have been used in the perception literature to judge the egocentric distance to objects. These protocols generally may be applied to AR systems; the primary limiting factors are the limitations for the depth cues themselves, noted in Sec. 2, and potential limitations on ease of movement for protocols that require the observer to move. Reasonable (if not easy) accommodations have been demonstrated in AR depth perception experiments summarized in Sec. 5. Interested readers should refer to the original papers for details. Loomis and Knapp [31] summarize the application and relative performance of these various measurement protocols in the perception literature. Some of the evaluations in AR described below studied the accuracy of estimates made with multiple protocols.

4 Visualization Metaphors

In response to the perceptual challenges, AR system designers have devised several visualization metaphors to convey ordinal and metric depth information. Some metaphors attempt to emulate the appearance of the real world within the virtual environment in order to allow the user to take advantage of the real-world cues to understand the depth relationships of virtual objects within the real world. Other methods introduce synthetic cues to convey depth relationships with graphical parameters instead of relying on analogies to the real world. Still other applications have simply presented the graphics in superposition and counted on other depth cues to overcome the conflict in occlusion. Some methods were developed in the context of mobile AR systems, which presents some unique design challenges (cf. Chapter 5), and we discuss applicability of the metaphors for a variety of displays and application scenarios. We review these methods in approximate order of their appearance in the literature; however, we allow that some techniques were not published upon first usage. In addition, it should be recognized that graphics in **superposition** over real surfaces have long represented occluded surfaces.

4.1 Opacity

Perhaps the most natural metaphor for X-ray vision is to depict a surface that occludes others as being partially transparent [12]. If the observer can accept the (physically unrealistic) premise that the real surface has turned translucent, then the presence of the graphical objects can convey the correct depth ordering between the real and virtual surfaces. This becomes as direct an implementation of “X-ray vision” as the technology permits, although there is a significant qualitative difference between the appearance of the (real) occluder and (virtual) occluded surface. The technique may be extended to any number of hidden surfaces, and a filtering aspect can be added by enabling a user or an automated control mechanism for various hidden surfaces. The technique may also be inverted by setting transparency of a virtual surface as a function of distance from the observer [29]. This variation does not require modification or explicit representation of the real surface which hides the virtual object’s location, although it does benefit from it. Current 3D graphics hardware – specifically four-byte color and z-buffer depth ordering – simplifies the implementation of this technique considerably. This variation of the technique is illustrated in Fig. 2.

The challenge of this metaphor for human perception is that it can explicitly break the occlusion cue. Both the graphics corresponding to a hidden object and the real surface must to a certain degree be visible in the correct direction. Another natural interpretation is that the graphics, being visible while in competition with a real surface, must therefore be in front of the real surface, which is exactly the opposite of the geometric arrangement that X-ray vision condition attempts to convey. This technique is affected by the display capabilities as well. Most OST AR

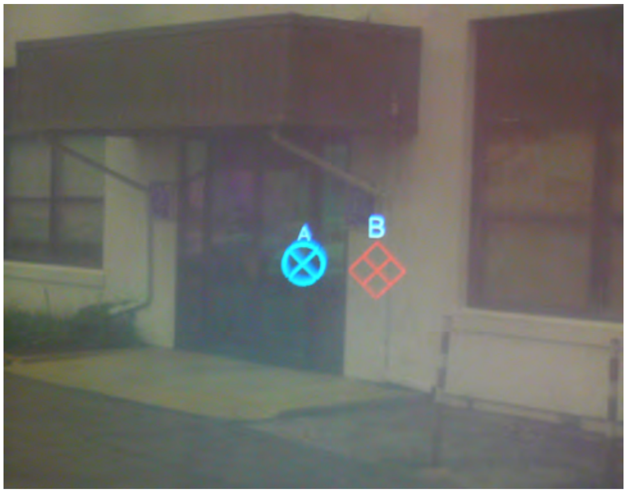


Fig. 2 Setting transparency of a virtual surface as a function of distance simulates the atmospheric perspective cue for the human visual system.

displays are not capable of occluding the real world with graphics on the display. Thus, all graphics are translucent and thus interact with the metaphor on which this technique tries to build. Video overlay displays can completely overwrite the video pixels with graphics, however. The cost of this capability is that the depicted occlusion relationship is thus completely reversed from the desired understanding. This gave rise to the next metaphor.

4.2 "Cutaway" or "virtual hole"

As noted above, despite the dimness of transparent graphics representing occluded objects, they often appeared to be in front of the real surfaces that were logically intended to hide them. This is especially true when the virtual surfaces are represented with wireframe outlines or small filled surfaces, whereas the occluder is a larger, smooth surface. Because the human visual system has a preference for continuous surfaces, the real surface's continuity "pushes" the graphics forward in front of the real surface [11], creating precisely the opposite of the intended perception.

This conflict can be resolved by creating more graphical representations to make it clear that the first set of objects in fact lies behind the real surface. The metaphor of a "cutaway" [12] or a "virtual hole" [2] can be used to create a context for the graphics. The hole breaks the real surface in a way that permits the virtual graphics to be perceived at the intended depth. A cutaway can take the form of the X-ray vision metaphor in the Superman comics which gave rise to the name for the capability, simply making it clear that the occluder is interrupted. (Fig. 3 shows an implementation from the AR system for laparoscopic visualization [14].) Or such a

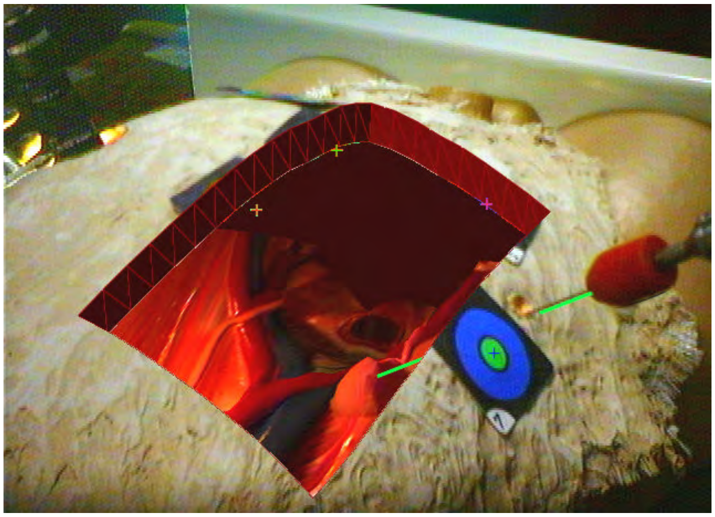


Fig. 3 The virtual hole metaphor in the AR system for laparoscopic surgery visualization [14]; image courtesy Department of Computer Science, University of North Carolina at Chapel Hill.

virtual hole may have a full 3D structure to it, with sides and a virtual bottom that is clearly behind the virtual graphics that fit inside the hole [2, 41]. Either of these metaphors overcomes the perceptual preference for continuity of the surfaces and conveys the ordinal depth relationships, a fact verified by an experiment employing a physical hole [11].

4.3 Stipple

One classic illustration technique to represent hidden surfaces is to depict visible surfaces with solid lines, represent a set of hidden surfaces with dashed lines, and perhaps extending to represent a more distant set of surfaces with dotted lines. This builds upon the depth cue of relative density of texture in the scene to create the impression of depth. Such stipple effects have been a tool for technical illustrators for many decades. They were among the earliest techniques introduced in AR for representing hidden surfaces as well [12]. Because simple stipple patterns have been a technique in hand-drawn illustration for so long, they are among the techniques implemented for drawing lines in computer graphics hardware as well. Extensions of this technique to filled polygonal representations of surfaces are also featured in graphics hardware. Fig. 4 shows an example of this technique using filled polygons.

One difficulty with such techniques, however, is that the hardware implementation of the patterns typically uses screen-space or window-space addressing of the coordinates for the origin of the stipple pattern. This means that if the object moves or the view position or orientation changes even slightly, the stipple pattern

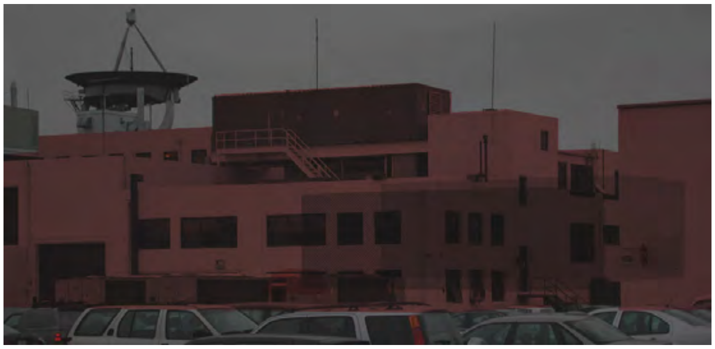


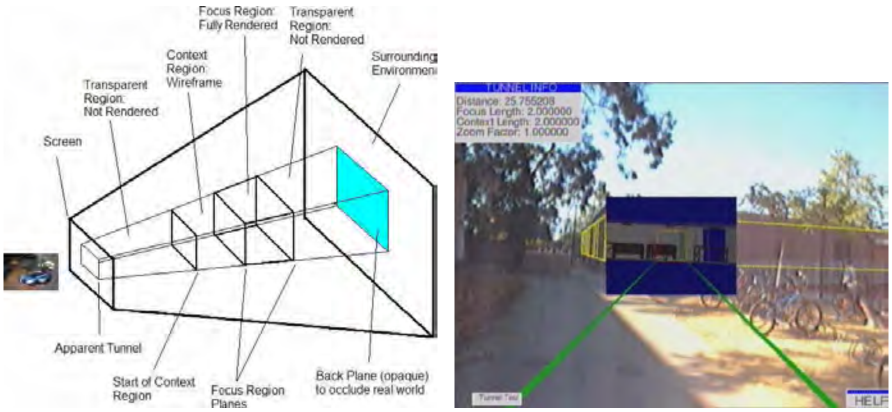
Fig. 4 Using stipple effects available in graphics hardware can provide a depth cue. The decreasing density of the stipple pattern can indicate increasing distance from the viewpoint. This can be reminiscent of the relative density depth cue. One can see that overlapping silhouettes cause interference in the density of stipple patterns, limiting the effectiveness of the cue.

will move with respect to the object. Thus the stipple pattern will appear to "crawl" along the lines or surfaces. While this does not prevent the user from understanding the depth relationships, it is likely to be distracting. Programmable graphics hardware allows for the implementation of an object-space stippling pattern [28].

Stippling patterns would appear to suffer from the converse problem as the opacity-based representation: the stipple is by definition a non-continuous representation of a continuous surface. This visual cue can therefore be reconstructed into something continuous (which is the intention), but then also brought forward in depth (counter to the intention). The familiarity of many people with assembly instructions and other forms of technical illustration mitigates this conflict somewhat; it is simply a convention that dashed lines are behind solid lines, and (for most people) that dotted lines are behind dashed lines. So the technique succeeds in spite of some perceptual ambiguity.

4.4 Shadow projection

Height in the visual field—i.e., distance from the horizon—is a direct cue from the virtual object about its depth within the environment, based on perspective properties of the rendered display. An indirect form of this cue is the position of a shadow cast on the environment [46]. Understanding the shadow may also require understanding the size and shape of the virtual object, as these will also affect the form of the shadow. Once the shadow is understood to be caused by the object, it gives the observer some information about the position of the object above the ground (on the line from the light to the shadow) and thus the distance from the user; this is of course assisted by the existence of shadows cast by real objects, so that the location of the light source(s) may be reliably inferred. Accurately casting a shadow from a


Focus Region: Fully Rendered
Transparent Region: Not Rendered
Context Region: Wireframe
Surrounding Environment
Transparent Region: Not Rendered
Screen
Apparent Tunnel
Start of Context Region
Focus Region Planes
Back Plane (opaque) to occlude real world
HELP

TUNNEL INFO
Distance: 25.755208
Focus Length: 2.000000
Context Length: 2.000000
Zoom Factor: 1.000000

Fig. 5 The virtual tunnel extends the metaphor of the virtual hole to multiple surfaces, creating the effect of a tunnel. *Left*: Diagram showing the four regions of space used in rendering the tunnel. *Right*: An example of the tunnel from the user’s point of view. Images courtesy of Tobias Höllerer and the University of California at Santa Barbara [3].

virtual object onto the real world requires a precise model of the real surface onto which the shadow must be cast, an accurate characterization of the light source(s), and the computational resources to compute the geometric interaction of the light(s) with the object and the surface. While these have been within reach of graphics hardware for a number of years [44], relatively few systems implement such cues.

4.5 Virtual tunnel

Interactive X-ray vision can be achieved through a virtual tunnel [3]. In this tool, a frustum was formed with the user’s position at the apex and rendered rectangles parallel to the view plane. Occluded objects are rendered inside that frustum. This type of rendering provides a feeling of looking down a tunnel, hence it is called a tunnel tool. Inside the tunnel there are four regions created by three planes. The first region, starting from the user’s position to the occluder, remains transparent. The second region is called the Context region, where the occluder is rendered with wireframe outlines, to provide the user a context of what she is looking through without providing too much information. Then the third region, called the Focus region, renders the occluded object with filled shapes. This is the region in which the user is interested. The fourth region, ranging from the end of Focus region to the opaque back plane, is again transparent. The back plane is used to occlude the real world. The user can slide the whole set of planes (regions) to get an interactive X-ray vision tool.

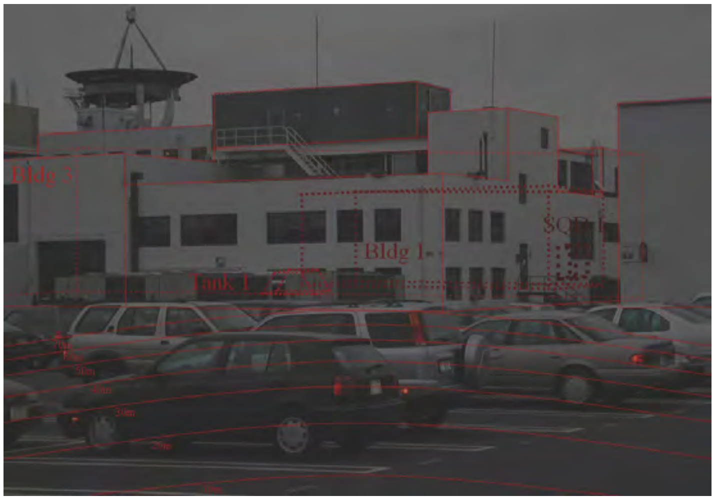


Fig. 6 A ground grid consisting of concentric circles can assist the user with resolving the depths of virtual and (if registration is accurate) real objects in the scene. This image is a concept sketch from 2002.

4.6 Ground grid

Horizontal relationships, i.e. the context, between an occluder and an occluded object can be clarified by a ground grid [49]. Here, a virtual grid on the ground is presented in the occluded region, and the user can determine the size and location of the area hidden by an occluder by interpreting the grid lines. Grid lines may consist of concentric circles (Fig. 6) or a rectilinear grid corresponding to some external coordinate system. While this metaphor can resolve the relationship, it may increase visual clutter and, to some extent, decrease the visibility of the occluded location.

4.7 Image-based techniques

Most of the previous X-ray vision metaphors presented occluded objects as floating on top of the occluder due to the naïve overlay of synthetic data on top of the real world imagery. Inspired by Focus and Context visualization techniques, Kalkofen et al. [21] introduced the notion of context-preserving X-Ray vision. This technique controls the removal of real-world information based on edge maps of the image of the occluder. Avery et al. [1] extended the idea to mobile AR. They detected and overlaid edges of the occluder on the occluded virtual objects to increase the perceptual reference between occluded objects and occluders (Fig. 4.7, left). The idea of edge-map based X-ray vision was extended and a closer understanding of human perception was employed by Sandor et al. [39] to design an X-ray vision

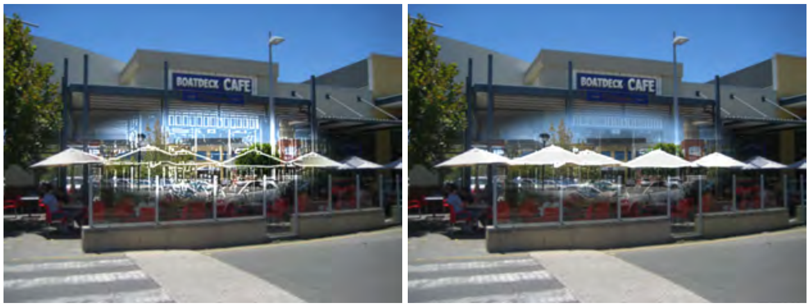


Fig. 7 Two types of image-based techniques: (*left*) only Edge-map based and (*right*) three additional saliency-map based.

based on multiple *saliency-maps* (Fig. 4.7, right). Along with the edges of the occluder, this X-ray metaphor additionally preserved hue, luminosity, and motion as salient features. Image analysis may reveal other salient features that can be used to communicate that a single layer of real surface exists in front of the virtual objects superimposed on them [51]. Textures and highly saturated colors may produce visual saliency in the way that edges do; this can be exploited to determine what pixels should have an extra overlay on them to convey the proper depth ordering.

4.8 Tram lines

Tram lines emerge from a similar concept as the ground grid; they give a direct cue of the distance along the tram lines and/or relative size of the area between the lines. These distance cues and relative size then give cues to the absolute or relative size of virtual objects near the lines. The tram lines were originally designed [26] to match the linear perspective cue provided by an indoor hallway (Fig. 8). The goal in the work was to assist with depth perception, not necessarily not X-ray vision perception. However, as noted above, one belief is that improvements in depth perception of objects – independent of each other – will in turn cause correct perception of relative depth judgments of visible and occluded surfaces, in terms of both ordinal depth and metric depth. Since relative size is considered to be a powerful cue at any distance, this would appear to be a promising method for expressing the distance of virtual objects. However, limited resolution of AR displays may limit the effectiveness of this cue.

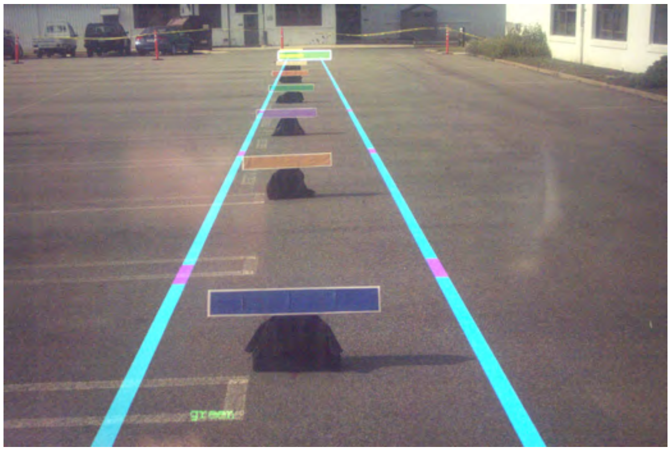


Fig. 8 Virtual tram lines aimed to re-create the linear perspective cue from an indoor hallway in the outdoor AR application context.

4.9 Melting

In the same inspiration of X-ray vision as the image-based techniques, Sandor et al. [40] employed a space-distorting visualization that virtually melts the occluder to show the occluded objects (Fig. 4.9). While Melting provides a clearer view of the occluded location, it does not preserve enough information of the occluder. Hence, the context of occlusion suffers.

4.10 Virtual wall

Despite the success of the edge map technique for expressing the relative order of real and virtual surfaces, it can be of limited applicability due to the expense of computing the edge map or the need for special-purpose hardware. Thus a simpler set of “standard” edges could be applied based on the depth of an object [28]. Edges may be added (or subtracted) with each successive layer of depth behind another real surface. This cue does not build directly on any perceptual cues; it emerges from the occlusion cue, attempting to express a physical surface or set of physical surfaces that intervene between the observer and the virtual object of interest. In the initial implementation, additional “edges” in the virtual wall represent another layer of “occlusion” of the virtual object (Fig. 10).

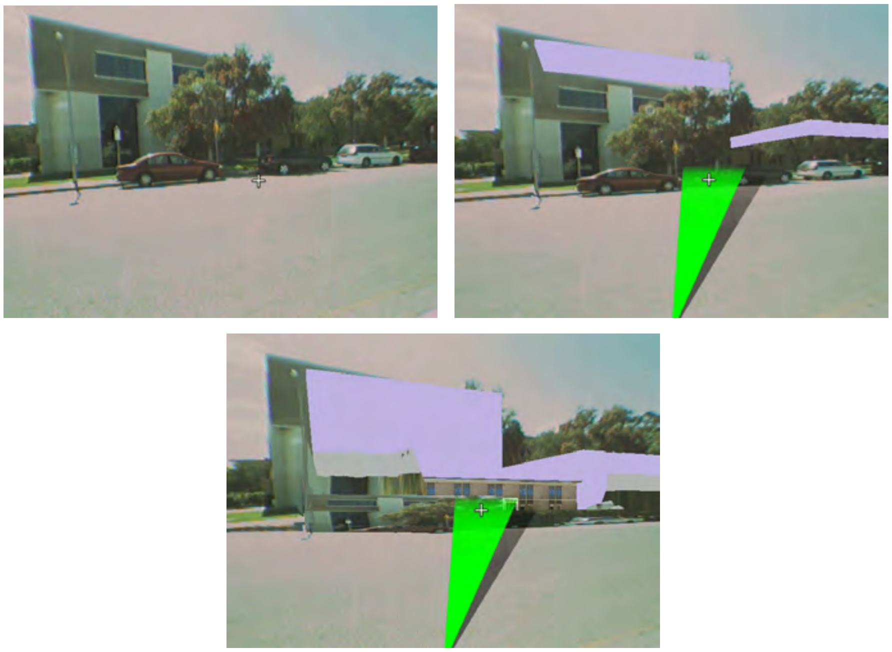


Fig. 9 The Melting metaphor completely removed the occluder to show occluded objects.

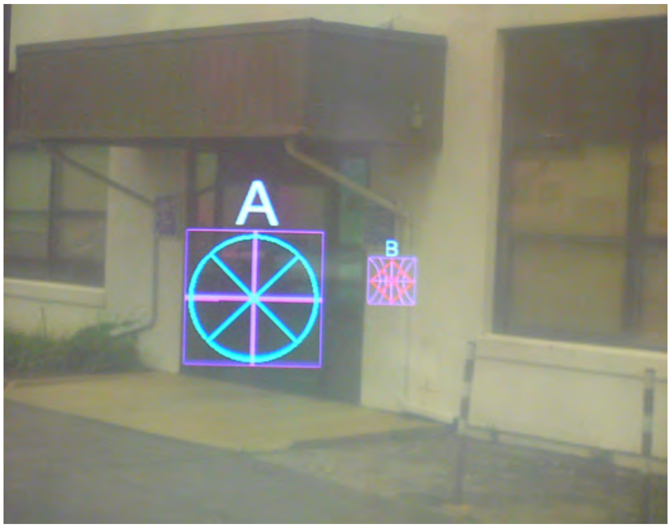


Fig. 10 The Virtual wall aimed to provide the cues of the Edge map, but with a less-cluttered and computationally less-demanding representation.

5 Evaluations

Because the perception of depth occurs solely within the human visual system, it becomes critical to evaluate the perception of depth experienced by users of a particular application. Several studies have been conducted, using both ordinal and metric depth properties. There are several protocols for measuring depth perception; most of these can be adapted to AR scenarios.

We next review results of various studies. Table 1 shows that X-ray vision and metric depth perception have been investigated for a long time; however, most of the experiments were conducted during the last decade using an OST HMD and, until recently, mostly in an indoor location. There are a few key concepts to consider; we classify studies according to these principles in Table 1. Some studies have explicitly studied the metric depth perception of graphical objects that are at locations in the real world that are not occluded from view. Other studies have focused on the case of "X-ray vision," or seeing a graphical object that is represented at a location that is occluded from view. Many of these studies use a baseline case of depth perception in the non-occluded case, which helps establish the relationship between the two cases. Also, the studies have varied in the field of depth they studied; early work focused on the near-field, but work progressed to the medium-field and the far-field in the last decade.

5.1 Work in the Near-Field

Rolland et al. (1995) [36] conducted a pilot study at 0.8 and 1.2 meters, which asked users to identify whether a 40 mm cube or a 13 mm (diameter) cylinder was closer; the objects were both virtual, both real, or one of each. They found large variability in the perception of depth of virtual objects presented in a stereo optical AR display, attributing this to accommodation and convergence errors (since the display used collimated views for the two eyes) as well as differences in the size of shape of the objects. Subjects overestimated at the tested distances.

Ellis et al. (1995) [10] showed that perceived depth of near-field (1.08 meters) virtual objects was linked to changes in binocular convergence. Subjects' perception of the depth of the virtual object was correlated with the change in convergence induced by the presence of a real object either at or closer than the virtual object's distance. The shortening of the estimated distance was less in the case of the real object being closer than the virtual object. One potential cause for this foreshortening of distances was the mismatch between accommodation and convergence. In this study, X-ray vision was expressed by simple superposition of the graphical objects. Ellis and Menges (1998) [11] summarized a series of AR depth judgment experiments, which used a perceptual matching task to examine near-field distances of 0.4 to 1.0 meters and studied an occluder (the X-ray vision condition), convergence, accommodation, observer age, and monocular, bi-ocular, and stereo AR displays. They found that monocular viewing degraded the depth judgment, and that the X-

Table 1 Summary of evaluations (ordered chronologically). Displays used were either optical see-through (OST) of various brands, a head-mounted projective display (HMPD), or hand-held video (HH-Vid) AR displays. The site in which the experimental task implies in what type of environment the target was embedded; we note which experiments were indoor (In), outdoor (Out), or both. The field of distance was within the near-field (N), medium-field (M), or far-field (F), or may have crossed the boundary between two of these fields. Depth measures used could be either metric (Met) or ordinal (Ord), as defined in Sec. 2; some papers used experiments of both types, and one (∗) used subjective evaluation. We note whether X-ray vision was a condition in the experiment; we use the sub-section numbers from Sec. 4 to indicate which X-ray vision metaphors were employed. In addition, a + indicates simple superposition was used, an ellipsis means that some number of depth cues were systematically tested or adapted to the AR system, and a blank indicates that X-ray vision was not a condition. In the case of both X-ray vision and depth cues, “use” may not imply controlled testing; see the text and the original papers for details. Finally, one or more estimation protocols were used to measure subjects’ perceptions: forced-choice (C), perceptual matching (M), verbal report (V), direct walking (W), and/or blind reaching (R).

	Display	Task	Field of	Depth	X-Ray	Estimation
Publication	Used	Site	Distance	Measure	Metaphors	Protocols
Rolland et al. 1995 [36]	OST	In	N	Ord		C
Ellis et al. 1995 [10]	OST	In	N	Ord,Met		C M
Ellis & Menges 1998 [11]	OST	In	N	Met	...	M
McCandless et al. 2000 [33]	OST	In	N	Met		M
Rolland et al. 2002 [38]	OST	In	N	Ord,Met		C M
Livingston et al. 2003 [29]	OST	Out	F	Ord	1,3,...	C
Kirkley 2003 [22]	OST	In	M	Met	+	V
Livingston et al. 2005 [30]	OST	In	M–F	Met		M
Jerome & Witmer 2005 [17]	OST	In	M	Met		M V
Swan et al. 2006 [48]	OST	In	M–F	Met	1	M
Fidopiastis 2006 [13]	HMPD	In	N–M	Met		M
Swan et al. 2007 [47]	OST	In	M–F	Met		M V W
Jones et al. 2008 [19]	OST	In	M	Met		W
Livingston et al. 2009 [26]	OST	In+Out	M–F	Met		M
Dey et al. 2010 [8]	HH-Vid	Out	F	Met	7,9	V
Singh et al. 2010 [43]	OST	In	N	Met	+	M R
Sandor et al. 2010 [39]	HH-Vid	Out	F	∗	7	
Livingston et al. 2011 [28]	OST	Out	M–F	Ord	1,3,5,6,7,10	C
Jones et al. 2011 [18]	OST	In	M	Met		W

ray vision condition caused a change in convergence angle which resulted in depth judgments being biased towards the observer. They also found that cutting a virtual hole in the occluder, which made the depth of the virtual object physically plausible, reduced the depth judgment bias compared to superposition. McCandless et al. (2000) [33] used the same experimental setup and task to additionally study motion parallax and AR system latency in monocular viewing conditions; they found that depth judgment errors increased systematically with increasing distance and latency. They constructed a model of the error showing that lateral projected position of the virtual object and depth judgment error were each linearly related to the latency.

Rolland et al. (2002) [38] compared forced-choice and perceptual matching tasks with a prototype stereo OST display; the display alleviated some conflict between accommodation and convergence noticed in earlier experiments. Four shapes (cube, cylinder, faceted cylinder, and octahedron) and three sizes of objects were used as stimuli; all were at a distance of 0.8 meters. They found a discrimination threshold of 7 mm using constant stimuli and precision of rendered depth of 8-12 mm using adjustments, depth accuracy, and no consistent depth judgment bias.

Fidopiastis (2006) [13] developed protocols to test perceptual effects of head-mounted projective displays (HMPDs), including depth perception. Pilot testing at near-field distances (and edging into the medium-field) were conducted with two prototype displays. With a 52° diagonal FOV and 640 × 480 graphical resolution, five subjects were found to have mean signed errors between -7.7 mm and 9.9 mm at a distance of 800 mm, -5.5 mm to 12.5 mm at 1500 mm, and -11.8 mm to 19.8 mm at 3000 mm. With a 42° diagonal FOV and 800 × 600 graphical resolution, five subjects had mean signed errors between -3.0 mm and 7.7 mm at a distance of 800 mm, between -8.9 mm and 15.6 mm at 1500 mm, and -34.7 mm and 32.2 mm at 3000 mm. The variances in these errors (within subject) grew from a few millimeters at 800 mm to 10-30 mm at 1500 mm to 20-70 mm at 3000 mm. The research indicated that individual differences in the distance at which subjects would focus in dark environments may help predict some differences in depth perception accuracy.

Singh et al. (2010) [43] evaluated the effect of closed-loop (perceptual matching) and open-loop (blind reaching) depth perception tasks for near-field AR (34-50 cm) in the presence and absence of a salient occluder. They reported that a perceptual matching protocol (closed-loop) was significantly better in depth judgment than the blind reaching (open-loop) protocol. However, using both of the protocols, depth was mostly underestimated. The effect of the occluder was complex and interacted with the protocol and the distance. Without an occluder present, error tended to increase with distance. With the occluder present, an interaction with the convergence cue could produce or not produce errors depending on the distance.

5.2 Extension to Medium-field and Far-field

Livingston et al. (2003) [29] studied varied representations of occluded buildings using drawing styles of wireframe, filled, and filled-with-wireframe outlines, varied opacity (constant or based on distance), and varied intensity (constant or based on distance). As discussed in Sec. 4, the last two approximate the aerial perspective cue. Other variables included the use of a consistent ground plane (on and off) and presenting images with or without binocular disparity (in software, with fixed hardware IPD). Subjects were more accurate in identifying ordinal depth among buildings at 60-500 meters with the ground plane consistent, but were found to have no significant difference in performance under conditions of filled-with-wireframe-outline drawing, opacity decreasing with object distance, and intensity decreasing with object distance. Subjects were found to get faster with practice, but not significantly more accurate. This could indicate that the cues were intuitive for users. A follow-up study [30] used a perceptual matching technique and found that similar errors were made when matching the depth of real objects and unoccluded virtual objects against real reference objects.

Kirkley (2003) [22] studied occluders (via superposition), the ground plane, and object type (real, realistic virtual, and abstract virtual) in both monocular (Microvision Nomad) and bi-ocular (Sony Glasstron) optical AR viewing at medium-field distances (3.0-33.5 meters). He found that occluders increased error, placing objects on the ground plane decreased error, and judging the depth of real objects was most accurate. In most cases, users underestimated the distance to virtual objects seen in the head-worn AR displays. Jerome and Witmer (2005) [17] noted issues of registration, monocular viewing, and wide separation of repeated distances (i.e. becoming a memory task rather than a perceptual task) with Kirkley's work. They used an OST display, object distances of 1.5-25 meters with eight objects (four small and four large, four abstract shapes and four familiar objects), and two protocols (perceptual matching and verbal report). They found that female subjects had significantly more error with virtual objects than with real objects, while male subjects showed no significant difference. They also found that error in judging distance to virtual objects was significantly reduced when preceded by a distance judgment of a real object. Finally, their subjects were more accurate with perceptual matching (of a physical robot) to a virtual distance than with verbal reporting of virtual distance. Both of these tests were done with hardware copies of the Battlefield Augmented Reality System [27], used in the two studies mentioned in the preceding paragraph and in the first study discussed in the next paragraph.

Swan et al. (2006) [48] explored medium- and far-field distances (5-45 meters) in optical AR with variables such as position in the visual field, occlusion of the site of the virtual object, and practice on the task using a perceptual matching technique. Subjects underestimated inside of 23 meters but overestimated beyond that distance. However, the subjects may have been using a combination of cues to arrive at estimates of depth of the virtual object with respect to the appropriate real object. These included relative size, disparity, brightness, aerial perspective approximations, or the perhaps-questionable convergence presented in the display (which lacked an IPD adjustment). Thus this apparent "cross-over" point from underestimation to overestimation may have derived from the environment or the equipment used. This cautionary argument is one motivating factor to compare the work against similar experiments. Data in [48] supported an estimate of an 8% increase in the rate at which error increased with distance from the unoccluded to the occluded condition. While this pattern is not surprising, the precise amount again requires verification and qualification as to the conditions under which it occurs in general.

Swan et al. (2007) [47] examined the effect of the experimental protocol and object type (real objects with natural vision, real objects seen through optical AR, virtual objects, and combined real and virtual objects) at medium-field distances (3-7 meters) with optical AR displays. The results indicated underestimation and implicated the restricted FOV and the inability for observers to scan the ground plane as explanations for the bias. Jones et al. (2008) [19] compared optical AR depth perception against VE depth perception and suggested that the virtual background contributes to the underestimation of depth in immersive VE. They found less underestimation in AR than in an analogous VE and no effect of using motion parallax in AR (versus standing still). Another study of metric depth with indoor AR using an OST display was reported by Jones et al. (2011) [18]. A blind walking protocol was used to measure depth perception in medium-field distances (3-7 meters) and, similar to previous studies, found a consistent underestimation of distance.

5.3 Experiments Focused on Outdoor and Mobile AR

Livingston et al. (2009) [26] studied the tram lines they introduced and compared performance indoors (with strong linear perspective cues) and outdoors (without strong cues) using optical AR and perceptual matching at medium-field distances (4.8-38.6 meters). They found a consistent underestimation of depth indoors (in contrast to earlier experiments in the same environment [48]) but overestimation outdoors. The presence of the tram lines decreased the estimated distance both indoors and outdoors, reducing error in the latter but increasing error in the former.

Melting [40] was compared with an edge map in an egocentric depth perception study of X-ray vision by Dey et al. (2010) [8]. The authors employed a verbal reporting protocol in their outdoor study, where participants had to guess the distance of an occluded object placed at far-field distances (69.7-117.0 meters). Contradicting previous findings by Livingston et al. [26], they found that, like indoor AR environments, depth is underestimated in outdoor AR environments as well. Authors reported that the Melt visualization performed more accurately and faster than X-ray vision. However, Melt vision removes the occluder completely and eventually loses the important features of the occluder. Unlike other experiments of X-ray vision, this experiment was performed using a hand-held device.

In another evaluation using a hand-held display at an outdoor location, Sandor et al. (2010) [39] evaluated their *saliency-based* X-ray vision with the previously-presented edge-map metaphor. They found similar results while selecting a target object in the occluded location, though saliency-based X-ray preserved more information of the occluder. Another on-line survey was conducted, where both of these X-ray vision metaphors were applied to three different levels of brightness and edges of the occluder. They found that with higher brightness, edge-overlaid X-ray vision provided more information about the occluded region, whereas under mid-range or lower brightness, saliency-based X-ray vision provided more information about the occluded region. Dey et al. (2011) [9] found that subjects, while navigating in an urban environment, spent more time looking at their hand-held AR display with X-ray vision capabilities than subjects spent looking at traditional maps viewed on their hand-held display (whether oriented with North in the vertical direction or the user's current view direction in the vertical direction on the screen). The AR with X-ray vision condition also had the fewest context switches between the the hand-held display and the environment.

Broad-based comparison of X-ray vision metaphors is rare in the literature. Livingston et al. (2011) [28] compared opacity, stipple, ground grid, edge map, virtual

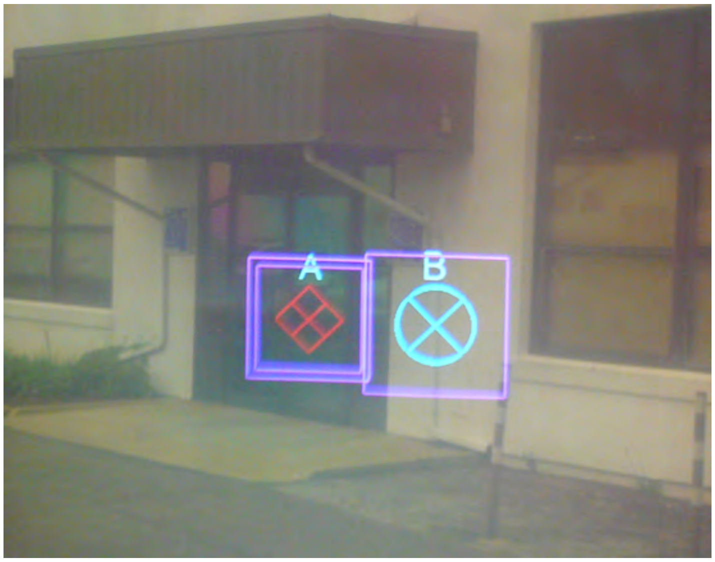


Fig. 11 This customized and simplified version of the virtual tunnel concentrated on the number of planes through which the tunnel was being extended to get to the target object. It was the best method in a broad-based comparison of X-ray vision metaphors [28].

wall, and (a variant of) the virtual tunnel techniques for ordinal depth judgments of virtual icons amongst a set of real buildings given on a map (of which the nearest wall was visible). A customized virtual tunnel (Fig. 11) was found to yield the lowest error, followed by the virtual wall and the ground grid. Additionally, the virtual tunnel, virtual wall, and edge map techniques were found to bias subjects to underestimation, while the opacity technique led to overestimation. Users were fastest with the baseline case of no X-ray vision metaphor and the virtual tunnel; they were slowest with the edge map. It was noted that the edge map technique had the potential to be very slow or require dedicated hardware to maintain in a dynamic environment; this study was conducted in a static environment.

6 Discussion

We conclude our review with discussion of the trends evidenced by the visualization metaphors and studies of them. We present four graphs that summarize some of the experiments described above in the medium-field and far-field. While the medical applications demonstrate the need for X-ray vision to work for applications that are contained within the near field, the trend towards mobile applications of AR as the potential consumer application of the technology pushes the more distant fields to great importance.

6.1 Metric Depth in Medium-field and Far-field

The first graph (Fig. 12) shows data from four experiments that were conducted in the medium-field and the adjacent portion of the far-field using head-worn OST displays. Each experiment compared different conditions; the point marker indicates which experiment is graphed with each line. Most of these conditions were visualizations of virtual objects in locations that were directly visible to the user; an X-ray vision metaphor was applied in only two conditions of two (separate) experiments. The data from Kirkley 2003 is notable in that it diverges so far from the remainder of the data; the final data point (33.5,-13.9) is cut off from the graph (which was zoomed to the majority of the data to enable legibility of the bulk of the data). The X-ray condition in Swan et al. 2006 agrees with Kirkley's data for some medium-field distances. These tests were designed to understand the metric perception of distance to virtual objects. Of the four data sets for direct virtual object viewing indoors (red hues), one stands out as an outlier: the indoor data from Livingston et al. 2009 [26] near the bottom of the visible portion of the graph (although the final point from Kirkley's equivalent condition appears to show the same behavior). That the same physical environment could give rise to such disparate data as the indoor data sets from Livingston et al. 2005 and Swan et al. 2006 (which were acquired in the same hallway) is curious or even suspicious. It underscores the potential difficulty in acquiring good data in these experiments, including the need to control potential confounding factors and to be concerned with the general ability of subjects with depth perception. One potential confound was the larger FOV of the OST display used in these experiments (a Sony Glasstron, Microvision Nomad 1000, and nVis nVisorST were used in various experiments).

The outdoor data from Livingston et al. 2009 is also somewhat different, but the environment offers an obvious hypothesis as to why this may have occurred. The difference in the data may be an effect produced by the loss of the powerful perspective cue from the indoor to the outdoor environment, or it could (again) be an artifact of the different head-worn display. Due to the disruptive nature of collecting indoor data in an office environment, the environment variable was not counterbalanced, but merely separated by seven to fourteen days. With a perceptual task, one would assume that such a design would not produce order effects, but this graph does cause some concern.

Two data sets in this graph show a version of the task with a real target [22, 30]. The data with the real target appears to be quite reasonable; this would appear to validate that the users were in fact making use of the depth cues in the merged environment when the task was moved into AR. One could argue that the perceptual matching was not a depth-based task, but merely a size-matching task, however. Again, this argument underscores the importance of validating the task and experimental design. Taken together, the data in this graph tend to drift from below the veridical line (signed error equal to zero) in the medium-field to above the veridical line in the far-field. This actually contradicts the long-standing observation in virtual environments that users tend to underestimate depth. The majority of the data in this graph in the medium-field indicates underestimation, but it would appear that

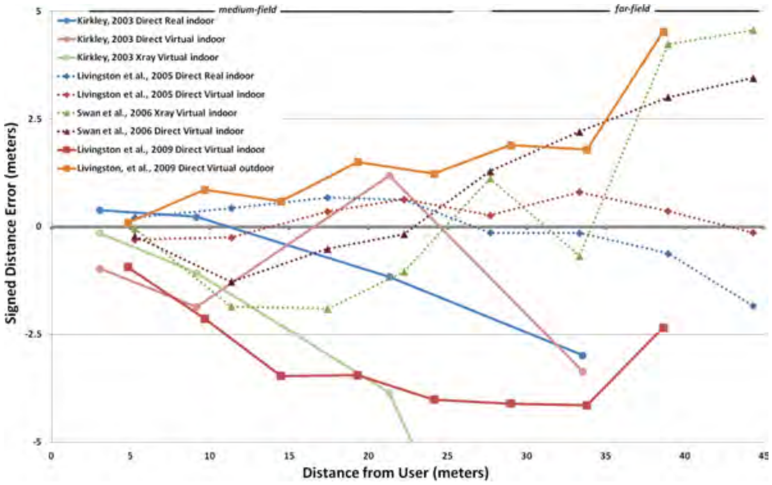


Fig. 12 A graph of data from experiments [22, 26, 30, 48] that were conducted in the medium-field and the neighboring part of the far-field. Each experiment compared varying conditions, but only two directly studied the X-ray condition. The lines on this graph are coded by symbols to differentiate the experiment and by color to identify equivalent conditions. Thus the two green lines for the Xray conditions should be similar; instead, they diverge, with the fourth point of [Kirkley, 2003 Xray Virtual indoor] at (33.5,-13.7), well off the bottom of this graph. The red lines are all measures of depth estimation with virtual objects in spaces that are directly viewed by the subject; these lines also show quite a bit of divergence from each other. The two blue lines from estimation of distance to real objects are somewhat similar, but still do not quite agree on the nature of the errors. In this and the next four graphs, the thick gray line at zero indicates veridical perception.

there is a slight tendency to overestimate depth to virtual objects as they move into the far-field. However, the data are far from in agreement even on this fundamental judgment. This is clearly an area that deserves more detailed exploration.

6.2 Metric Depth Well into the Far-field

The second graph (Fig. 13) shows a metric depth study conducted well into the far-field [8]; this experiment used a hand-held display. As noted above, this graph shows a general trend of underestimation under these experimental conditions. The reasons for the difference between this graph and the previous graph are unexplored, but the variety of conditions could give rise to several hypotheses that could be tested. Clearly, the virtual cue akin to the grid lines and the ground grid was extremely helpful in helping users improve their accuracy with metric depth estimation. Given the difficulty of this task in general and the accuracy achieved, it is likely that users simply relied on the cue and estimated distance by counting the units (tens of meters) from their location to the virtual object. If this is an acceptable mode of operation in

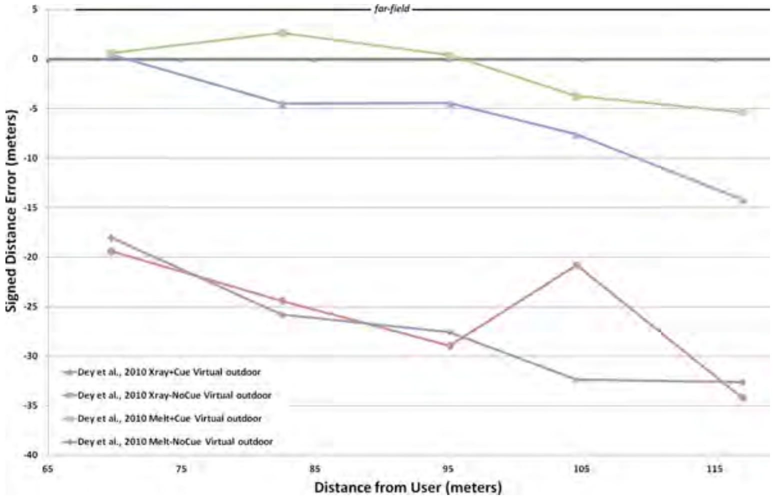


Fig. 13 A graph of data from an experiment [8] that was conducted in the far-field compared two visual metaphors for X-ray vision and a cue akin to a tram line. This cue was extremely helpful. All of these conditions were conducted with hand-held displays.

the target application, then it becomes perhaps the most obviously beneficial technique of all those discussed here.

6.3 Ordinal Depth in the Medium-Field and Far-field

The third graph (Fig. 14) returns to the boundary region between the medium-field and far-field; this data comes from a single experiment conducted on ordinal depth perception using a head-worn display [28]. In contrast to the original paper, this graph separates by X-ray vision metaphor and by the "zone" in which an object was found. A "zone" indicated the number of intervening real surfaces in the known environment that were between the user and the location of the virtual object; zone 1 had no intervening surfaces. The analysis of this graph should begin with the line representing no X-ray vision metaphor being applied, which exhibited the most error in the first three zones, found to be overestimation of the ordinal distance, then crossed over to a balance between overestimation and underestimation, and finally settled into the second-greatest underestimation in the farthest zone. Clearly, we would expect any X-ray vision metaphor to improve upon this performance. One should be careful regarding the first and last zones, since only one direction of error is possible in each, but relative performance differences are meaningful. The Virtual Tunnel stays the closest to correct ordering, with the Virtual Wall coming close to or exceeding its performance until the most distant zone. The Edge Map, by contrast, appeared to suffer from visual clutter. It is also interesting to note the similarity of the graphs for the Opacity and Stipple metaphors. The former simulates the natural

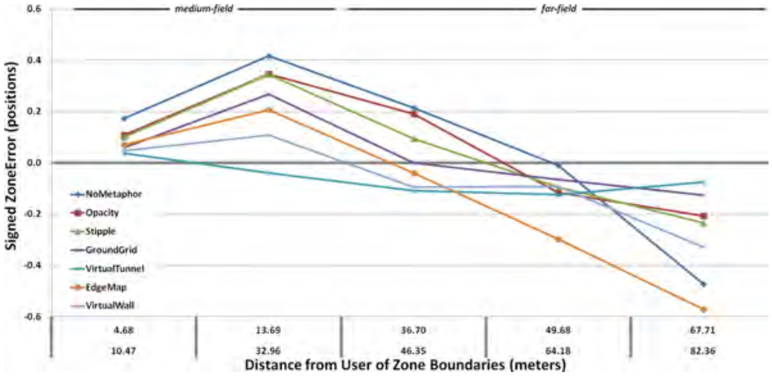


Fig. 14 A graph of data from an experiment [28] that was conducted in the medium-field and far-field compared several metaphors for X-ray vision. This experiment used a head-worn OST display. The occlusion “zones” were the distances (shown on the horizontal axis) for which the number of real surfaces intervening between the user and the virtual object location was constant. Since the users knew the environment, it is safe to assume that they were aware of these surfaces.

depth cue of aerial perspective quite well, whereas the latter is a completely synthetic cue commonly used in technical illustration. This demonstrates that users can condition themselves to think in terms of metaphors quite effectively, and perhaps as well as we intuit natural cues for the human visual system.

6.4 Ordinal Depth well into the Far-field

The fourth graph (Fig. 15) shows the data from an early comparison of various X-ray vision metaphors [29] using a head-worn OST display with far-field distances. The two metaphors that emulated the aerial perspective cue (Opacity and Intensity fading with distance) both improved user performance, and their combination improved it yet further. Users were also aided by the use of a filled shape with a wireframe outline, rather than either representation alone. It is interesting to compare how the users fared with the cue of height in the visual field (represented by the use of a consistent ground plane, denoted “GroundPlaneOn” in the graph legend) in comparison to the X-ray vision metaphors. This early work showed the potential of AR to provide usable X-ray vision and helped to spark renewed interest in this area, including several of the works discussed in this chapter. This data is in unsigned error, so the graph is not as informative as the others.

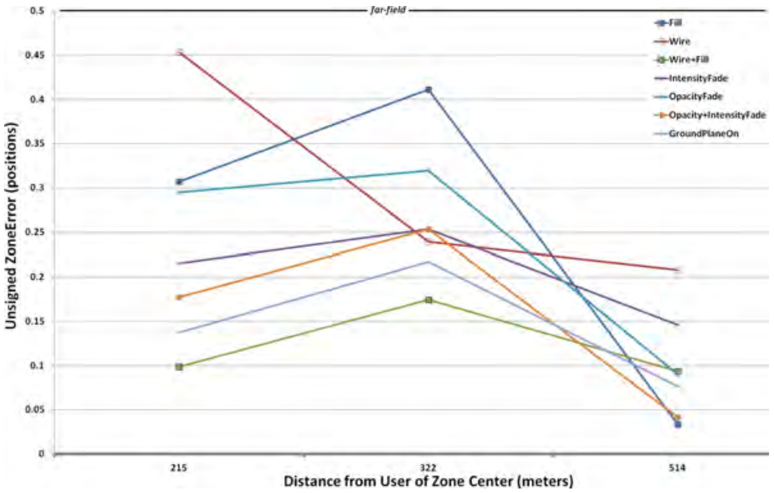


Fig. 15 A graph of data from an experiment [29] that was conducted in the far-field compared three metaphors for X-ray vision. This experiment also used a head-worn OST display. The distances (shown on the horizontal axis) were to the centroid of a building that was the target at the given distance. The targets and the distracting objects were mutually disjoint.

6.5 *Near-field Depth*

The final graph summarizes experiments conducted in the near-field (Fig. 16). While the experimental conditions differ so greatly that any general conclusions are difficult to draw, we can see that much of the data indicates underestimation of distance. The graph depicts data points in X-ray vision conditions with circle glyphs (and connecting lines for the Singh et al. [43] data). All of these points indicate underestimation, except for the Ellis et al. data point for monocular viewing by subjects with advanced age (greatest overestimation of the Ellis et al. data) [11]. This and the other monocular point (most underestimated of the Ellis et al. data) presumably suffered from the poor disparity and convergence cues in the early optical AR displays as well as the decreasing ability of humans to use accommodative cues with increasing age. All three of these cues are quite sensitive in the near-field. This may be similar to the explanation of the overestimation that Rolland et al. [36] found. McCandless et al. [33] also noted that the accommodative was slightly too close for the virtual objects even at the near distance and cited this as a potential source of bias in depth judgments. They also posited that the close proximity of the wall behind the experimental scene may have biased the depth judgments. These two factors may have cause the overestimation of distance for the nearest targets and perhaps some of the underestimation at the farther distances. This gives another specific example of how uncontrolled factors may affect experimental results.

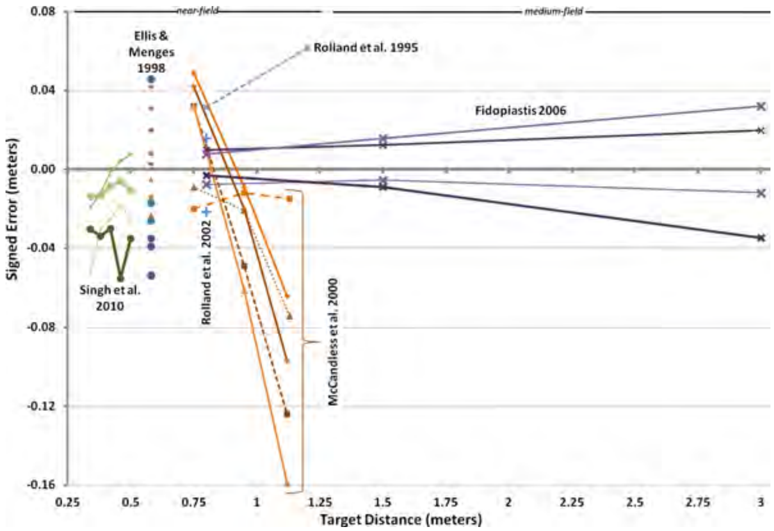


Fig. 16 A graph of data from six papers studying depth perception and/or X-ray vision in the near-field. Labels within the graph indicate the publications. The X-ray vision conditions (two line graphs in Singh et al. and six points in Ellis et al.) are indicated by large circles at the data points. This data largely indicates underestimation, a point discussed in the text.

6.6 *Summary Observations*

The first observation beyond the graphs is how little overlap there is between them. It would have been nice to show all this work on a single graph, but the distances would have been starkly disjoint sets, not to mention the near impossibility of equating the metric and ordinal measurements. A few issues deserve special attention, given the summary of experiments above.

6.6.1 Underestimation vs. overestimation

As noted above, there is a well-observed (though not yet well-understood) trend towards underestimation of distances in VEs. One can easily put forth theories that AR need not behave the same, primarily on the basis of the grounding in the real environment and natural depth cues that users receive from it. Still, we have noted that some experiments with (metric or ordinal) depth perception found underestimation in one or more conditions, a few found overestimation in one or more conditions, and many found a mixture of underestimation and overestimation. Clearly, this is an area that is ripe for further study to understand the phenomenon in AR.

6.6.2 Optical vs. video

We have attempted to note precisely what type of AR display was used in the various experiments above, and the interaction of the two fundamental technologies of OST and video overlay with the natural depth cues used by the human visual system. Despite the relative ease with which we can reason about the expected effects, there is relatively little direct comparison of the benefits of the two types of displays for the perception of depth. (Some comparison for applications has been done [37].) This is another area that would benefit the field to have studied further.

6.6.3 Head-worn vs. hand-held

An issue that may interact with the above question is the type of display used. Head-worn displays were classically the presumed form for AR, but mobile phones appear to have displaced them in favor of hand-held displays. A priori, there is no reason to assume that the perceptual qualities of the two form factors are identical. This may account, for example, for the difference between Fig. 13 and the apparent differences between certain portions of data in Fig. 12 and Fig. 14. (We make this statement despite our acknowledgment of the difficulty of comparing the graphs.) Since each type of display has a large potential user base, it would again seem to be a rich area for research. One twist, however, is that when studying the effects of these types of displays, it may be more important than for the other empirical questions to embed the task within an expected or prototypical application context. For example, if mobile phones are to be used as AR navigation aids, then a consumer-level registration accuracy and task should be the basis for an evaluation. If a military application plans to integrate a head-worn AR capability into (for example) a pilot's helmet, then the depth perception or X-ray vision task should come in the context of his role in the battlespace.

6.6.4 Binocular vision

One underexplored area in recent research is near-field localization; while this was heavily featured in early work, advances in displays – notably adjustments in binocular disparity and focal distance, have the potential to significantly improve the depth cues that are critical in the near field. This distance is important for a number of proposed AR applications, including medicine and manufacturing. Studies that determine the utility of adjustments in convergence, binocular disparity, and accommodation in display devices might provide some guidance for hardware designers.

6.6.5 Outdoor environments

While a few recent experiments [8, 26, 28] studied outdoor X-ray vision, this is still an under-explored direction. The use of X-ray vision is not restricted to indoor environments. Recent advances in mobile phones and portable devices have led to their rise as AR platforms and opened new possibilities for outdoor mobile AR applications, such as navigation tools. It is required to conduct experiments on AR X-ray vision in outdoor locations. Unlike indoor environments, outdoor environments are practically impossible to fully control. The random variation of the environment is expected to have different effects on human perception, which are necessary to understand for the extension and improvement of X-ray vision. The contradictory results need to be clarified and quantified.

6.6.6 Comparison of visual metaphors

We note that a few tests directly compared two or more metaphors; however, the set of all possible combinations of comparison is far from complete. In part, this is due to the recent introduction of methods, but it also stems from the volume of data needed to compare more than a small number of X-ray vision metaphors. An ambitious study could fill this gap in the knowledge of the field, although an experimenter should clearly be prepared to exhibit patience in such an endeavor. Also, there are parameters associated with various techniques; ideally, each technique to be compared against another technique should be optimized for the task in order to achieve a fair comparison.

6.6.7 Intuitiveness and interaction

Sutherland’s original vision for the virtual world – which was in fact an augmented reality, not the immersive virtual worlds we conceive today with the term – was that the virtual objects would be indistinguishable from reality. We argue that to fulfill this vision, we must achieve visualization metaphors in AR that are as intuitive for the human visual system as natural depth cues. This may seem impossible, but we noted above that stipple effects appear to have achieved this status. How to test the intuition of the responses is not an easy question, but one obvious first choice is the reaction time with which a user can (accurately) determine the (metric or ordinal) depth. Many of the experiments described herein measured reaction time, but rarely with anything close to the speed of human vision in assessing depth. Many experiments used only a few repetitions of the task, and even fewer reported results with respect to repetition. This could give some insight into how easily and how well users can learn to accept the X-ray visualization metaphors. Similarly, only one tool ([3]) used interaction as a means to understand depth; while perhaps not appropriate for all applications, this is another area that is under-explored in the field.

7 Conclusion

X-ray vision is among the most flashy of the capabilities offered by AR technologies. As mobile applications begin to enter the consumer market and industrial and military applications continue to mature and find new uses, making AR systems work at the perceptual level embodied in the X-ray vision perception is critical. Even for (conceptually) simple applications such as navigation, which can be implemented without X-ray vision, accurate distance perception of graphical cues can be an obvious benefit to the user. It should be clear from the significant research efforts described here that the best methods are far from clear for even narrow applications, let alone across the field of AR. The analysis across experiments identified several directions for future research that are likely to bring great benefit to the field and scientific grounding to the techniques. Furthering the latter, we connected the techniques through the metaphors to the perceptual cues applied by the human visual system to the understanding of distance; it is important to understand the fundamental relationship between the AR cues and the visual system in order to determine the reasons for the success or failure of a technique and to suggest improvements in the case of the latter. It will also be increasingly important to push the tests of distance perception and X-ray vision closer to the applications envisioned, so that the metaphor of X-ray vision can truly be simulated in augmented reality.

Acknowledgments

The authors would like to thank Steve Ellis, Cali Fidopiastis, Henry Fuchs, LT Gregory O. Gibson, Tobias Höllerer, Christian Jerome, Sonny Kirkley, Jannick Rolland, Andrei State, and Ed Swan for their assistance in the preparation of this chapter.

References

1. Avery, B., Sandor, C., Thomas, B.H.: Improving spatial perception for augmented reality X-ray vision. In: IEEE Virtual Reality, pp. 79–82 (2009)
2. Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. Computer Graphics (Proceedings of SIGGRAPH'92) **26**(2), 203–210 (1992)
3. Bane, R., Höllerer, T.: Interactive tools for virtual X-ray vision in mobile augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 231–239 (2004)
4. Biocca, F.A., Rolland, J.P.: Virtual eyes can rearrange your body: Adaptation to visual displacement in see-through, head-mounted displays. Presence: Teleoperators and Virtual Environments **7**(3), 262–277 (1998)
5. Bruce, V., Green, P.R., Georgeson, M.A.: Visual Perception: Physiology, Psychology, and Ecology, 3rd edn. Psychology Press (1996)
6. Cutting, J.E.: Reconceiving perceptual space. In: H. Hecht, R. Schwartz, M. Atherton (eds.) Looking into Pictures: An Interdisciplinary Approach to Pictorial Space, chap. 11, pp. 215–238. MIT Press (2003)

7. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: Handbook of perception and cognition, Vol. 5: Perception of space of motion, pp. 69–117. Academic Press (1995)
8. Dey, A., Cunningham, A., Sandor, C.: Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In: 17th ACM Symposium on Virtual Reality Software and Technology, pp. 211–218 (2010)
9. Dey, A., Jarvis, G., Sandor, C., Wibowo, A.K., Mattila, V.V.: An evaluation of augmented reality X-ray vision for outdoor navigation. In: The 21^{st} International Conference on Artificial Reality and Telexistence, pp. 28–32 (2011)
10. Ellis, S.R., Bucher, U.J., Menges, B.M.: The relationship of binocular convergence and errors in judged distance to virtual objects. In: Proceedings of the International Federaion of Automatic Control, pp. 297–301 (1995)
11. Ellis, S.R., Menges, B.M.: Localization of virtual objects in the near visual field. Human Factors **40**(3), 415–431 (1998)
12. Feiner, S.K., Seligmann, D.D.: Cutaways and ghosting: Satisfying visibility constraints in dynamic 3D illustrations. The Visual Computer **8**(5–6), 292–302 (1993)
13. Fidopiastis, C.M.: User-centered virtual environment assessment and design for cognitive rehabilitation applications. Ph.D. thesis, Department of Modeling and Simulation, University of Central Florida (2006)
14. Fuchs, H., Livingston, M.A., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J.R., Rademacher, P., Drake, S.H., Meyer, A.A.: Augmented reality visualization for laparoscopic surgery. In: Proceedings of First International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 934–943 (1998)
15. Furness, LT T.A.: The application of head-mounted displays to airborne reconnaissance and weapon delivery. Tech. Rep. TR-69-241, U.S. Air Force Avionics Laboratory, Wright-Patterson AFB (1969)
16. Furness, T.A.: The super cockpit and its human factors challenges. In: Proceedings of the Human Factors Society 30th Annual Meeting, pp. 48–52. Human Factors and Ergonomics Society (1986)
17. Jerome, C.J., Witmer, B.G.: The perception and estimation of egocentric distance in real and augmented reality environments. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 49, pp. 2249–2252 (2005)
18. Jones, J.A., Swan II, J.E., Singh, G., Ellis, S.R.: Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, pp. 29–35 (2011)
19. Jones, J.A., Swan II, J.E., Singh, G., Kolstad, E., Ellis, S.R.: The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, pp. 9–14 (2008)
20. Julier, S., Lanzagorta, M., Baillot, Y., Rosenblum, L., Feiner, S., Hollerer, T., Sestito, S.: Information filtering for mobile augmented reality. In: Proceedings of IEEE and ACM International Symposium on Augmented Reality, pp. 3 –11 (2000)
21. Kalkofen, D., Mendez, E., Schmalstieg, D.: Interactive focus and context visualization for augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 191–201 (2007)
22. Kirkley, Jr., S.E.H.: Augmented reality performance assessment battery (ARPAB: Object recognition, distance estimation, and size estimation using optical see-through head-worn displays. Ph.D. thesis, Instructional Systems Technology, Indiana University (2003)
23. Kiyokawa, K., Kurata, Y., Ohno, H.: An optical see-through display for mutual occlusion of real and virtual environments. Computers and Graphics **25**(5), 765–779 (2001)
24. Klein, G., Murray, D.: Simulating low-cost cameras for augmented reality compositing. IEEE Transactions on Visualization and Computer Graphics **16**(3), 369–380 (2010)
25. Liu, S., Hua, H., Cheng, D.: A novel prototype for an optical see-through head-mounted display with addressable focus cues. IEEE Transactions on Visualization and Computer Graphics **16**(3), 381–393 (2010)

26. Livingston, M.A., Ai, Z., Swan II, J.E., Smallman, H.S.: Indoor vs. outdoor depth perception for mobile augmented reality. In: IEEE Virtual Reality, pp. 55–62 (2009)
27. Livingston, M.A., Brown, D., Julier, S.J., Schmidt, G.S.: Military applications of augmented reality. In: NATO Human Factors and Medicine Panel Workshop on Virtual Media for Military Applications (2006)
28. Livingston, M.A., Karsch, K., Ai, Z., Gibson, LT G.O.: User interface design for military ar applications. Virtual Reality **15**(2–3), 175–184 (2011)
29. Livingston, M.A., Swan II, J.E., Gabbard, J.L., Höllerer, T.H., Hix, D., Julier, S.J., Baillot, Y., Brown, D.: Resolving multiple occluded layers in augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 56–65 (2003)
30. Livingston, M.A., Zanbaka, C., Swan II, J.E., Smallman, H.S.: Objective measures for the effectiveness of augmented reality. In: IEEE Virtual Reality, pp. 287–288 (2005)
31. Loomis, J.M., Knapp, J.M.: Visual perception of egocentric distance in real and virtual environments. In: L.J. Hettinger, M.W. Haas (eds.) Virtual and Adaptive Environments, pp. 21–46. Lawrence Erlbaum Associates (2003)
32. McCandless, J.W., Ellis, S.R.: The effect of eye position on the projected stimulus distance in a binocular head-mounted display. In: Stereoscopic Displays and Virtual Reality Systems VII, Proceedings of SPIE Vol. 3957, pp. 41–48 (2000)
33. McCandless, J.W., Ellis, S.R., Adelstein, B.D.: Localization of a time-delayed, monocular virtual object superimposed on a real environment. Presence: Teleoperators and Virtual Environments **9**(1), 15–24 (2000)
34. Nagata, S.: How to reinforce perception of depth in single two-dimensional pictures. In: Pictorial Communication in Virtual and Real Environments, chap. 35, pp. 527–545. Taylor & Francis, Inc. (1991)
35. Plumert, J.M., Kearney, J.K., Cremer, J.F., Recker, K.: Distance perception in real and virtual environments. ACM Transactions on Applied Perception **2**(3), 216–233 (2005)
36. Rolland, J.P., Ariely, D., Gibson, W.: Towards quantifying depth and size perception in virtual environments. Presence: Teleoperators and Virtual Environments **4**(1), 24–49 (1995)
37. Rolland, J.P., Fuchs, H.: Optical versus video see-through head-mounted displays. Presence: Teleoperators and Virtual Environments **9**(3), 287–309 (2000)
38. Rolland, J.P., Meyer, C., Arthur, K., Rinalducci, E.J.: Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. Presence: Teleoperators and Virtual Environments **11**(6), 610–625 (2002)
39. Sandor, C., Cunningham, A., Dey, A., Mattila, V.V.: An augmented reality X-ray system based on visual saliency. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 27–36 (2010)
40. Sandor, C., Cunningham, A., Eck, U., Urquhart, D., Jarvis, G., Dey, A., Barbier, S., Marner, M., Rhee, S.: Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. In: IEEE Virtual Reality (2010)
41. Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B., Schmalstieg, D.: Handheld augmented reality for underground infrastructure visualization. Personal and Ubiquitous Computing **13**(4), 281–291 (2009)
42. Schiffman, H.R.: Sensation and Perception, 5th edn. John Wiley & Sons (2001)
43. Singh, G., Swan II, J.E., Jones, J.A., Ellis, S.R.: Depth judgment measures and occluding surfaces in near-field augmented reality. In: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, pp. 149–156 (2010)
44. State, A., Hirota, G., Garrett, W.F., Chen, D.T., Livingston, M.A.: Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In: Computer Graphics Proceedings, Annual Conference Series (Proceedings of SIGGRAPH'96), pp. 429–438 (1996)
45. Stix, G.: See-through view: Virtual reality may guide physician's hands. Scientific American **267**(3), 166 (1992)
46. Sugano, N., Kato, H., Tachibana, K.: The effects of shadow representation of virtual objects in augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 76–83 (2003)

47. Swan II, J.E., Jones, A., Kolstad, E., Livingston, M.A., Smallman, H.S.: Egocentric depth judgments in optical, see-through augmented reality. IEEE Transactions on Visualization and Computer Graphics **13**(3), 429–442 (2007)
48. Swan II, J.E., Livingston, M.A., Smallman, H.S., Brown, D., Baillot, Y., Gabbard, J.L., Hix, D.: A perceptual matching technique for depth judgments in optical, see-through augmented reality. In: IEEE Virtual Reality, pp. 19–26 (2006)
49. Tsuda, T., Yamamoto, H., Kameda, Y., Ohta, Y.: Visualization methods for outdoor see-through vision. IEICE Transactions on Information and Systems **E89–D**(6), 1781–1789 (2006)
50. Witt, J.K., Proffitt, D.R., Epstein, W.: Perceiving distance: A role of effort and intent. Perception **33**(5), 577–590 (2004)
51. Zollmann, S., Kalkofen, D., Mendez, E., Reitmayr, G.: Image-based ghostings for single layer occlusions in augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 19–26 (2010)