Improving Spatial Perception for Augmented Reality X-Ray Vision

Benjamin Avery Christian Sandor Bruce H. Thomas

Wearable Computer Laboratory
University of South Australia

(a) (b) (c)

Figure 1 – Our Augmented Reality x-ray vision system creates an image-based reconstruction of a remote scene (b) based on captured images and models. The reconstruction is overlaid on the user's view of the environment (a) to provide x-ray vision (c).

ABSTRACT

Augmented reality x-ray vision allows users to see through walls and view real occluded objects and locations. We present an augmented reality x-ray vision system that employs multiple view modes to support new visualizations that provide depth cues and spatial awareness to users. The edge overlay visualization provides depth cues to make hidden objects appear to be behind walls, rather than floating in front of them. Utilizing this edge overlay, the tunnel cut-out visualization provides details about occluding layers between the user and remote location. Inherent limitations of these visualizations are addressed by our addition of view modes allowing the user to obtain additional detail by zooming in, or an overview of the environment via an overhead exocentric view.

Keywords: Outdoor Augmented Reality, Wearable Computers, Image-Based Rendering, Visualization, Depth Perception.

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Virtual Reality. J.9.e [Mobile Applications]: Wearable computers and body area networks.

1 INTRODUCTION

Augmented Reality (AR) enables the overlay of computer graphics registered to the user's view of the physical world. One exciting application of AR is to provide users with x-ray vision capabilities. To simulate this 'superman'-like ability of x-ray vision, computer-generated views of occluded objects and locations are rendered on the user's view of the environment. As an example, a user observes a physical wall (Figure 1(a)), a camera behind the wall provides images (b) and the simulated x-ray vision of this occlude scene is shown in (c). AR x-ray vision has been used in medical scenarios [3] and maintenance tasks [4].

Our previous AR x-ray vision system employed wire-frame models, textured with video images to render x-ray images (see Figure 2) [1]. While that system made the shapes easy to understand, the rendered images appeared to be floating on top of the occluding surfaces. These were conflicting depth cues, as the hidden outdoor areas are located behind the occluding surfaces. The problem of occluded objects appearing on top of surfaces rather than behind them is a well known problem in AR [3, 7, 10].

We have designed new visualizations that extend the concepts presented by Kalkofen et al. [8]. Our visualization can be used to perform x-ray vision depicting real remote objects and locations as opposed to purely virtual occluded objects. We made additions to the concept to be suitable for use with real scenes. Our *Edge Overlay* visualization provides depth cues for x-ray vision not available in our previous system. One limitation of this new visualization is the lack of information about occluding layers between the user and the remote location. There could exist any number of occluding layers between the user and the remote scene (see Figure 3(d)). We have designed a visualization that displays a representation of these occluding layers when observing a remote scene. Our visualization, *Tunnel Cut-out,* is of similar appearance as the cut-away views by Coffin and Höllerer [6]; however our system operates automatically. An egocentric viewpoint provides no ability to see details in objects from a large distance or any overview of the surroundings. This is particularly limiting for x-ray vision as occluded objects can be at a large distance from the user, and spatial relationships between occluding and occluded

Figure 2 – Our old AR x-ray vision system did little to convey depth. The x-ray images appeared to be floating in front of the walls, rather than behind.

e-mail: ben@benavery.net
e-mail: chris.sandor@gmail.com
e-mail: bruce.thomas@unisa.edu.au

IEEE Virtual Reality 2009
14-18 March, Lafayette, Louisiana, USA
978-1-4244-3812-9/09/$25.00 ©2009 IEEE

objects are difficult to determine. We implemented a set of view modes (*detail, walking* and *overview)* that the user can freely transition between to address these issues.

Through an iterative design process, showing the system to user interface researchers along the way, we have developed new features in our x-ray vision system that work together to complement the limitations of each other. The edge overlay is utilized at all times to provide depth cues, and if geometry information is available, we can use the tunnel cut-out to depict occluding layers. The view modes can be used at all times, together with the visualizations to see details or an overview of the environment.

2 BACKGROUND

Our previous x-ray vision system [1] provides a purely egocentric viewpoint. The occluded locations are rendered as textured 3D models on the display, but tend to appear to be floating over the surrounding environment, rather than behind it (see Figure 2). The user receives very little sense of having the powers of x-ray vision.

The system operates on a belt-mounted wearable computer and a video see-through head-mounted display (HMD). This allows complete freedom to move around outdoors. The pose is tracked with GPS and an orientation sensor. Image based rendering (IBR) techniques are used to create photorealistic views of occluded locations using 3D models and video images streamed from a remote camera. The research in this paper uses manually created models, and pre-captured video. The reconstructions are displayed on the HMD over the user's view of the world. This differs from much of the previous research on x-ray interfaces that showed purely virtual objects [7, 8].

The problem of presenting occluded information in AR was first observed by Bajura et. al when displaying ultrasound information captured from a patient [3]. Bajura noted that "*the ultrasound images did not appear to be inside the subject, so much as pasted on top of her.*" This problem was investigated by Furmanski et. al in a user study [7] to determine how well users could judge the depth of a virtual object when it was rendered in a location in behind or at the same distance as a real wall. Participants tended to assume the box was always in front of the wall, regardless of actual virtual distance, even when motion parallax suggested otherwise.

The work by Kalkofen et. al [8] proposed that when showing occluded objects, to preserve the context of occluding structures by rendering just the edges. Very little of the occluded object is obscured by the thin edges, but there are enough visual cues to give a compelling sense of depth. Kalkofen et. al primarily used pre-defined 3D models, and distinct virtual occluded objects. Their investigations differ from our research, in that we are interested in real occluded scenes.

Bane and Höllerer developed some display and interaction techniques for x-ray vision in augmented reality [4]. The Tunnel-Tool technique was based around a section of a view frustum which could be moved forwards and backwards through the scene. Everything inside of this area would be visible, regardless of obstructions. This technique is useful for exploring 3D volumetric data such as heat distributions. The Cut-away Views technique by Coffin and Höllerer [6] uses constructive solid geometry operations to remove geometry. The user is able to manually cut away walls by marking out points on the surface and cutting through them. Our tunnel visualization has a similar appearance to the cut-away views, but operates automatically. It is tailored to provide details about occluding layers, without obscuring the occluded scenes.

The first x-ray vision system for mobile outdoor AR using real images was created by Kameda et. al [9]. Their see-through vision system made use of a handheld display to show the user images from fixed surveillance cameras overlaid on a camera image of the current environment. They made use of wire-frame models and semi-transparency to aid the user in depth perception, and spatial relationships [12]. A study performed by Livingston et. al [10] investigated users' ability to determine depth of multiple occluded virtual objects at a variety of distances. Filled wire-frame models were found to be optimal as the wire-frame shows the shape, while the filled geometries convey correct occlusions. Both of these systems use semi-transparency to provide depth cues to the user. Since our design goal is to maximize visibility of the remote locations, obscuring them with semi-transparent overlays would be inappropriate. In a study previously conducted, users found it difficult to differentiate between the x-ray and real regions on the display [2]. The overlays were placed directly over the video with no transparency. The difficulty arose due to colors and textures of buildings and trees in both occluded and occluding sections of the display being quite similar. Blending the similar scenes with transparency would make this problem worse.

Transitioning between different viewpoints was employed by Bell et. al [5]. They developed an AR interface that used the pitch of the user's head to smoothly transition from an egocentric view to an exocentric view of a world-in-miniature. Our system uses similar smooth transitions between viewpoints, but with the addition of a zoom view and manual control.

3 VISUALIZATION DESIGN

In this section we present the design of visualizations that provide spatial awareness and depth cues for AR x-ray vision.

3.1 Edge overlay

The *Edge Overlay* visualization aims to provide depth cues when viewing occluded objects. In urban environments the occluding objects are primarily buildings. To show structure of the buildings while looking through them, we need to place detail back on top of the x-ray images. In our visualization, we apply an edge detection filter to the original AR video image to detect sharp changes in luminance. Each of these detected edges is represented by a thin white line that is drawn over the x-ray image. By only drawing the lines describing the major shape and structure of the building, very little of the x-ray image is occluded, allowing the user to see most of the detail. The edge overlay of the bricks on the wall can be seen in Figure 1(c) and Figure 3(a-c). The edges are only shown over the section of the display containing the x-ray vision.

In our initial observations of this technique, we found that any solidly colored lines in the x-ray images could cause confusion as they could easily be mistaken for foreground edges that do not match the real surroundings. We removed wire-frames from occluded geometry, as the white edges were difficult to distinguish from the edge overlay. This may make the remote shapes harder to understand, but results in a more realistic effect.

Hard edges between the x-ray overlay and the background image break depth perception. Any sudden transitions between occluding and occluded sections of the display would cause the appearance of the occluded objects to be floating in front of the occluding objects. Such a hard edge can be seen on the top-center of the x-ray vision in Figure 2. Making use of faded edges significantly alleviates this problem. We have implemented soft edges; the edges surrounding the x-ray images fade gradually from opaque to transparent providing a smooth transition between the occluding and occluded imagery. The combination of the edge overlay with the design constraints regarding lines and edges produces a sense of realism that the user really does have x-ray vision capabilities.

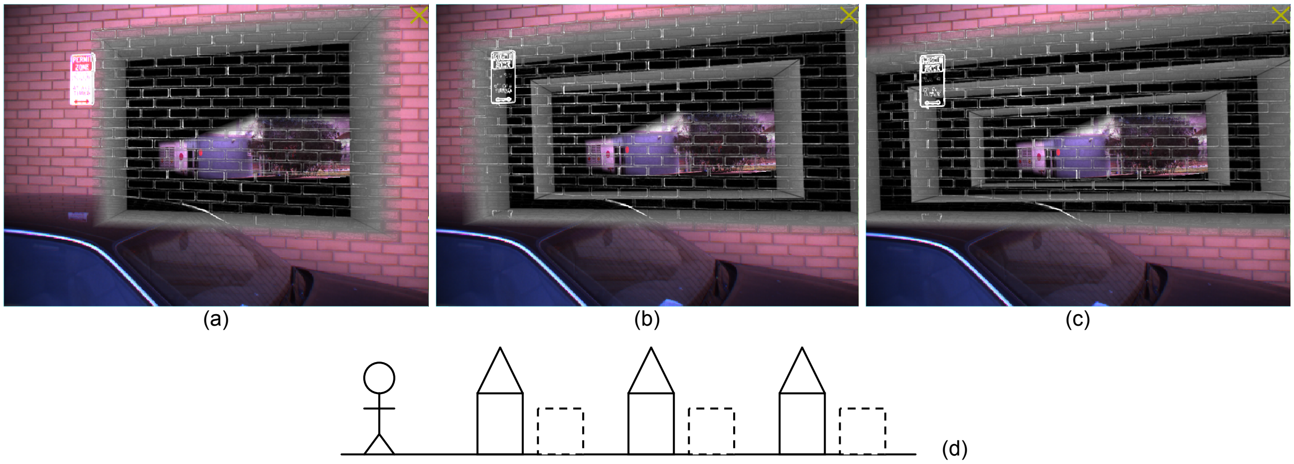

Figure 3 – Our Tunnel Cut-out visualization displays a representation of occluding layers that may exist between the user and the remote scene. The tunnel shows the difference between looking through one (a), two (b), or three (c) buildings. An overview of these scenarios is shown in (d).

3.2 Tunnel Cut-out

If the x-ray vision system has knowledge about objects between the user and the remote location they are observing, providing information about these objects to the user is an important visual cue for situation awareness. The edge overlay technique is only able to convey a single layer of occlusion. In Figure 3(d) the dashed lines represent reconstructed areas while the solid objects represent real buildings. The three areas would appear identically to the user when rendered using only the edge overlay. The computer system does not always contain a complete model of the environment. Depending on how the system is being used, there could be anywhere from no environmental knowledge, to complete knowledge of all geometry and images of every surrounding object. Our previous iterations of x-ray vision worked with the assumption that the system has limited knowledge of the environment; therefore images and models were captured and created in real-time as required. As services such as Google Earth and Google Street View become easy to integrate into mobile AR systems, this type of environmental knowledge will become available.

If we now consider that the system has geometry information available, we can begin to provide additional context to the user. We have created a visualization, *Tunnel Cut-Out*, to provide context about objects between the user and the remote location. The tunnel cut-out visualization creates a dynamic cut-out through each known object between the user and the occluded area of interest. These cut-outs are shown in Figure 3. The dimensions of the tunnel are determined such that it is large enough to view the entire remote area. The tunnel cuts through all geometry along this path and draws a representation (a grey box) of each object around the outside of the tunnel. As seen in Figure 3 it is easy to determine how many buildings the user is looking through.

As the tunnel cut-out itself represents occluded geometry, we have applied the edge overlay here. The front edges of the cut-out were created with a transparent falloff to avoid hard edges, and the edge overlay is extended to cover the entire tunnel. The outlines in the tunnel were drawn in dark colors to avoid being confused with the edge overlay.

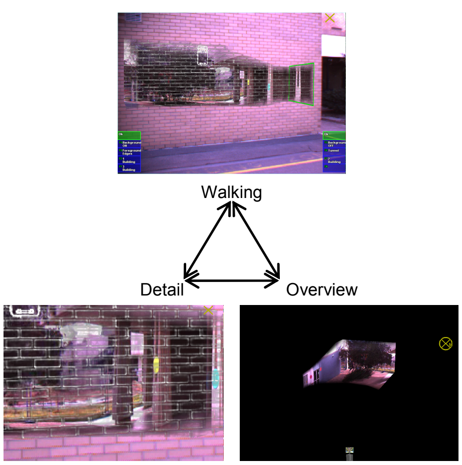

Figure 4 – We employ three view modes in our x-ray vision systems. The user can freely change between modes for normal use, viewing additional detail, or an overview of the surroundings.

3.3 Detail and Overview

We have implemented additional viewpoints to supplement the limitations of a purely egocentric interface. Occluded objects viewed outdoors at a distance from the user appear very small on the display, making details difficult to see. In a similar manner, the user is also unable to get an overview of the environment, as they view the world from close to ground level. Our x-ray vision systems implements three view modes: *walking*, *detail*, and *overview*. The user is able to freely transition between each of these three view modes as depicted in Figure 4.

The normal state is the *walking* mode, during which the image on the display matches the normal field-of-view of the user. In order to see additional *detail* in x-ray visualizations we have implemented a digital zoom function that modifies the view of the virtual world and real-world view, maintaining correct AR registration. To provide an *overview* of the environment, the viewpoint is animated from the users head location to an overhead position 30m behind and 5m above their head. The orientation is slaved to the user's view direction allowing them to look around the environment from anywhere on the 30m radius sphere surrounding the user. These modes are activated by holding a button on a wireless remote unit. This begins a smooth animation from the regular view to the detail/overview view.

4 IMPLEMENTATION

Our x-ray vision implementation runs on a mobile wearable AR system similar to the Tinmith system [11]. In this section, we summarize implementation details for the edge overlay, tunnel cut-out and detail and overview modes.

The edge overlay visualization is implemented using a GPU fragment shader and stencil buffering. The video image captured from the head-mounted camera is rendered as a 2D background on the display before any other rendering occurs. The occluded objects are rendered to the display and stencil buffer. Occluding layers are not rendered. After this, the original video image is re-rendered to the display with a GLSL fragment shader program enabled. The shader operates on each pixel of the video image and performs a 3x3 Sobel edge operator on surrounding pixels, and outputs white pixels for any edges. Stencil tests ensure that edges are only drawn over the occluded objects.

The soft edges are achieved by an algorithm that alters the alpha values at the edges of the textures to achieve a smooth gradient before they are applied to the 3D models. This effect could also be achieved by rendering to an offscreen buffer, then drawing the buffer back to the screen while applying a blur filter to the alpha channel.

The tunnel cut-out is generated dynamically by determining the minimum cut-out that can be made to the furthest wall from the user such that all of the occluded objects are visible through it. This is performed using ray-firing. The edges of the cut-out are fired back through all geometry in a direction parallel to the vector from the occluded object to the user. Each of the intersection points define a tunnel of constant width and height that runs from the user's location to the occluded objects. The tunnel is rendered using these intersection points to render grey polygons for anything inside a building, and black for between buildings.

5 Conclusions and Future Work

We have presented an x-ray vision system with new visualizations that provides additional depth cues and spatial awareness to users. A standard egocentric view alone is not sufficient to allow the user to adequately understand a remote location. Thus, we developed zoom and exocentric views.

While this research currently lacks a formal user evaluation, we have made numerous iterations in our development and believe the screen captures shown in this paper, in addition to the supplemental video, demonstrate that these visualizations are easy to understand and provide a clear improvement over previous systems. A formal evaluation of the new system will be conducted as future research.

We found that while the edge overlay performs well at providing a sense of depth, careful visualization design is required when displaying the occluded objects. If edges in the remote scene can be confused with the foreground overlay, or if the transition from occluded to occluding images is too harsh, the depth effect is reduced. Therefore, we avoided the use of wireframes. Future research will investigate image processing techniques to remove edges appearing in the remote scene. We have shown that if these factors are considered, the edge overlay technique is effective for real occluded scenes, not only virtual occluded objects.

A current limitation of the edge overlay is that the edge detection operates with a fixed sensitivity. If the background is too cluttered then too many edges are drawn and the remote scene can be difficult to see. If the background is too plain or lacks contrast, then few edges are drawn and little sense of depth is achieved. In future we will investigate an adaptive threshold to ensure a consistent number of edges.

The edge overlay requires a head-mounted video camera to capture the edges so is more suited to systems with video see-through displays. Although we have not tried, optical see-though displays might also be used; but, this would require an extremely accurate calibration between the user's eyes, HMD, and camera.

The Tunnel Cut-out performs well at representing occluded layers. It is a simple visualization that provides context to the user, without obscuring the occluded scenes. As visible in Figure 3, we use black to represent the 'nothingness' between the different layers. This makes the intent of the visualization clear, but is not an efficient use of screen space. If images of occluding buildings were available in the system, it would be interesting to experiment with using IBR techniques to show parts of the occluding layers as the tunnel passes through them.

We have presented new visualizations for x-ray vision of real occluded objects. We believe that these visualizations could be applied in a variety of areas where remote, photorealistic information is rendered, such as surveillance, video conferencing, or augmented virtuality.

Acknowledgements

We would like to thank Wayne Piekarski for developing the Tinmith platform. Christian Sandor had worked on an unpublished prototype with Denis Kalkofen, Erick Mendez, and Dieter Schmalstieg, that has inspired parts of this publication.

6 References

[1] Avery, B., Piekarski, W., and Thomas, B. H. Visualizing Occluded Physical Objects in Unfamiliar Outdoor Augmented Reality Environments. In *6th Int'l Symposium on Mixed and Augmented Reality*. p 285-286. Nara, Japan 2007.

[2] Avery, B., Thomas, B. H., and Piekarski, W. User Evaluation of See-Through Vision for Mobile Outdoor Augmented Reality. In *7th Int'l Symposium on Mixed and Augmented Reality*. p 69-72. Cambridge, UK 2008.

[3] Bajura, M., Fuchs, H., and Ohbuchi, R. Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery within the Patient. In *19th Int'l Conference on Computer Graphics and Interactive Techniques*. p 203-210. 1992.

[4] Bane, R. and Höllerer, T. Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality. In *3rd Int'l Symposium on Mixed and Augmented Reality*. p 231-239. Arlington, VA, USA 2004.

[5] Bell, B., Höllerer, T., and Feiner, S. An Annotated Situation-Awareness Aid for Augmented Reality. In *ACM Symposium on User Interface Software and Technology*. p 213-216. Paris, France 2002.

[6] Coffin, C. and Höllerer, T. Interactive Perspective Cut-away Views for General 3D Scenes. In *IEEE Symposium on 3D User Interfaces*. p 25-28. Virginia, USA 2006.

[7] Furmanski, C., Azuma, R., and Daily, M. Augmented-reality Visualization Guided by Cognition: Perceptual Heuristics for Combining Visible and Obscured Information. In *1st Int'l Symposium on Mixed and Augmented Reality*. p 215-224. 2002.

[8] Kalkofen, D., Mendez, E., and Schmalstieg, D. Interactive Focus and Context Visualization for Augmented Reality. In *6th Int'l Symposum on Mixed and Augmented Reality*. p Nara, Japan 2007.

[9] Kameda, Y., Takemesa, T., and Ohta, Y. Outdoor See-Through Vision Utilizing Surveillance Cameras. In *3rd Int'l Symposium on Mixed and Augmented Reality*. p 151-160. DC, USA 2004.

[10] Livingston, M. A., Swan, J. E., Gabbard, J. L., Höllerer, T., Hix, D., Julier, S., Baillot, Y., and Brown, D. Resolving Multiple Occluded Layers in Augmented Reality. In *2nd Int'l Symposium on Mixed an Augmented Reality*. p 56-65. Tokyo, Japan 2003.

[11] Piekarski, W. and Thomas, B. Interactive Augmented Reality Techniques for Construction at a Distance of 3D Geometry. In *Immersive Projection Technology / Eurographics Virtual Environments*. p 19-28. Zurich, Switzerland 2003.

[12] Tsuda, T., Yamamoto, H., Kameda, Y., and Ohta, Y. Visualization Methods for Outdoor See-Through Vision. In *15th Int'l Conference on Artificial Reality and Telexistence*. p 1-8. Christchurch, New Zealand 2005.