Spatially Augmented Reality

Ramesh Raskar, Greg Welch, Henry Fuchs
{raskar, welch, fuchs}@cs.unc.edu , (919)-962-1761
Department of Computer Science, University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599, U.S.A.

Abstract

To create an effective illusion of virtual objects coexisting with the real world, see-through HMD-based Augmented Reality techniques supplement the user's *view* with images of virtual objects. We introduce here a new paradigm, *Spatially Augmented Reality* (SAR), where virtual objects are rendered directly within or on the user's *physical space.*

A key benefit of SAR is that the user does not need to wear a head-mounted display. Instead, with the use of spatial displays, wide field of view and possibly high-resolution images of virtual objects can be integrated directly into the environment. For example, the virtual objects can be realized by using digital light projectors to "paint" 2D/3D imagery onto real surfaces, or by using built-in flat panel displays.

In this paper we present the rendering method used in our implementation and discuss the fundamentally different visible artifacts that arise as a result of errors in tracker measurements. Finally, we speculate about how SAR techniques might be combined with see-through AR to provide an even more compelling AR experience.

1. Introduction

In *Spatially Augmented Reality (SAR)*, the user's physical environment is augmented with images that are integrated directly in the user's environment, not simply in their visual field. For example, the images could be projected onto real objects using digital light projectors, or embedded directly in the environment with flat panel displays. For the purpose of this paper we will concentrate on the former. While the approach has certain restrictions, it offers an interesting new method to realizing compelling illusions of virtual objects coexisting with the real world. The images could appear in 2D, aligned on a flat display surface, or they could be 3D and floating above a planar surface, or even 3D and floating above an irregular surface.

For 2D non-head-tracked SAR, the images representing virtual objects do not continuously change with user motion. For example, in the Luminous Room [UnderKoffler97], the user's environment is enhanced with synthetic images projected on flat surfaces. However the user can be head-tracked and the images update dynamically to create an illusion that virtual objects are registered to real objects. Shuttered glasses can be used to facilitate stereo imagery, further enhancing the 3D effect of the virtual imagery. For the purpose of this paper we will focus on technologies for head-tracked SAR where virtual objects are rendered on irregularly shaped real objects. While not appropriate for every application, in this method the user does not need to wear a head-mounted display. Instead, with the use of spatial displays, wide field of view and possibly high-resolution images of virtual objects can be integrated directly into the environment.

It is our work on the Office of the Future [Raskar98a] that led to the realization and implementation of the SAR paradigm as described and analyzed in this paper. Specifically, when exploring the use of irregular (non-planar) surfaces for spatially immersive displays, we realized that the registration problem becomes somewhat unusual one of having to register 2D imagery with 3D physical geometry. This is similar to conventional AR applications, except for one crucial point: the 2D imagery exists on a virtual image plane that is attached to the (fixed) physical displays surface instead of

being attached to the user's (moving) head. However, as with conventional AR techniques, we realized that what is most important in terms of results is what the *viewer* sees. In retrospect it makes sense, but it was only after some further thinking that we realized the fundamentally different manifestation of visual registration error in this paradigm. Similar to the situation in a CAVE™ [Cruz-Neira93] (spatially immersive display), such error is virtually unaffected by viewpoint orientation error, and viewpoint position error results in a form of image shear rather than a simple mis-registration.

The occlusion relationships in SAR are also different than in see-through AR systems. In SAR, a real object can occlude the virtual object. Thus, for example, bringing your hand in front of the face will occlude the virtual object behind maintaining the illusion that the virtual object exists in the real world. On the other hand, a virtual object cannot obstruct the view of a real object even if it is intended to float in front of that object.

User interaction with virtual objects in SAR which involves tracking multiple parts of user body in 3D, such as head and hands, is dependent on the accuracy of tracking devices. In head-mounted display virtual reality (HMD-VR), where a virtual interface device manipulates the virtual world, small tracking measurement errors are not noticeable. In SAR, however, an accurate spatial relationship between user's head, interacting body part and the virtual object needs to be maintained. Thus, user interaction issue in SAR is a subset of registration issues in AR.

SAR techniques could be used for many different applications. For example, architectural designing applications could benefit from the perceived ability to visually modify portions of a real physical environment of tabletop architectural models. Similarly, engineers could use the approach to "paint" alternate appearances on or inside a life-sized automobile mockup. The approach could also be used for product training or repair: one could set the product in the SAR environment and have the system render instructions directly on the product. Doctors could use the SAR to jointly visualize and discuss virtual information that is projected onto a patient or mannequin, while simultaneously visualizing remote collaborators whose imagery and voices are spatially integrated into the surrounding environment.

2.   Previous work

Various levels of integration of virtual objects in user's physical environment are seen in current augmented reality (AR) systems [Milgram94a]. HMD-VR has been widely used to generate synthetic images inside head-tracked head mounted displays that occlude the view of the real world but give the illusion of spatial and temporal context in the user's physical world. Optical and Video See-through Augmented Reality (OST-AR and VST-AR) systems combine real scene viewed by the user and a virtual scene generated by the computer to augment the view with additional information [Milgram94b] [State96].

Some systems have integrated synthetic images with real scenarios for a static user. Dorsey et al provide a useful framework in the context of theater set design [Dorsey91] where a pre-distorted image appears correct when projected onto a curved backdrop of the theater. Luminous room [UnderKoffler97] is a partially immersive spatially integrated environment. The system projects and then generates 2D images on flat surfaces in a room to enhance the user's environment. The HoloGlobe exhibit uses High Definition Volumetric Display to display huge amounts of data concerning global change on a four-foot, 3-D floating image of the Earth with precision optical components such as parabolic mirrors and beam splitters [Hologlobe]. Viewers can walk around the 3D image and see it from different angles. The users do not need to wear any gear. The Office of the Future (OOTF) [Raskar98a] is the other extreme where the user is surrounded by synthetic images such as in spatially immersive displays (SID). CAVE [Cruz-Neira93}

and ARC's dome shaped displays [Benette98] are other examples of SID. However, in OOTF the display surfaces are not limited to the designated flat walls (or parameterized surfaces) and could be everyday surfaces.

3. Methods

To create an illusion that virtual objects are registered to real objects for a moving user, we need to know the position of the user, projection parameters of the display devices, the shape of the surfaces of real objects in the physical environment and be able to render virtual objects on those surfaces. Here, we will describe a method for each of these in a unified projector-based SAR system.

3.1. Display surface shape extraction

The 3D surface shape extraction can be achieved using a calibrated [Tsai86] projector-camera pair where structured light patterns are projected and observed by the camera. [Raskar98c] describes a near-real time method to capture the 3D shape of the display surface and [Raskar98a] describes a unified approach to capture and display on irregular (non-planar) surfaces.

3.2. Rendering and viewing method

Projecting images on irregular surfaces so that they appear correct to a static user have been described in [Dorsey91] [Max91] [Jarvis97] [Raskar98d]. In [Raskar98a], a real time technique to generate such images for a moving head-tracked user was introduced. Here, we will describe how it can be used for SAR even when not all the viewing parameters are known. Let $V = V_i * V_e$ represent the user's eye perspective projection matrix, where, V_i is the projection matrix and V_e is the transformation matrix (subscripts i and e are used for intrinsic and extrinsic parameters, respectively). $E = E_i * E_e$ is an intermediate projection matrix, which shares the center of projection (COP), with the user's eye and P represents the projector's perspective projection matrix. Display surface model is D and G is the graphics model we wish to render. We will use the notation $I(r,g,b,z) = M*[G]$ to indicate that image I is generated from 3D colored model G using perspective projection matrix M (using any rendering method). Note that $M^{-1}*I(r,g,b)$ represents a set of 'colored' rays and $M^{-1}*I(r,g,b,z)$ represents colored surface interpolating 3D points due to termination of the rays at the corresponding depth. We want to present an image $I = V*[G]$ to the user with *only* the COP for V known. (Note that in general it is difficult to correctly estimate the COP for V, due to the obvious difficulty in co-locating the tracking sensor at the human eye COP. However, as we point out below, erroneously estimating the COP results in a fundamentally different visual artifact than head-mounted AR, an artifact that would be less noticeable for typical applications.)

Rendering

Step I :
(a)   Compute the desired image color by rendering G:

$$I1(r,g,b) = E*[G]$$

(b) Update only the depth buffer (without overwriting color) by rendering the display surface model D.
$$I1(z) = E*[D]$$

Step II :
Apply a 3D warp to image **I1** to effect view transformation from projection system defined by **E** to projection system defined by **P**. This is similar to image based rendering of images with depth. (This can also be achieved by projective texture mapping: project texture **I1**(r,g,b) onto **D** and render using **P** as described in [Raskar98b]).

$$\textbf{I2}(r,g,b,z) = \textbf{P}*\textbf{E}^{-1}*\textbf{I1}(r,g,b,z)$$

Viewing

While viewing in the real world, if **I2** is projected by the projector on the display surface and viewed by the viewer, we get image **I3**,

$$\textbf{I3}(r,g,b) = \textbf{V}*[\textbf{D}], \text{ where } \textbf{D} \text{ is now display surface colored by projected light}$$
$$= \textbf{V}*\textbf{P}^{-1}*\textbf{I2}\ (r,g,b,z)$$
$$= \textbf{V}*\ \textbf{E}^{-1}*\textbf{I1}(r,g,b,z)$$

Note that, although $\textbf{E}^{-1}*\textbf{I1}(r,g,b,z)$ is not the same as [**G**] because the depth values in **I1** are due to display surface model [**D**], $\textbf{E}^{-1}*\textbf{I1}(r,g,b)$ represents 'colored' rays due to [**G**] meeting at COP of the projection system of **E**. Thus, if COP for the systems represented by **V** and **E** are the same, **I3** is the intersection of the colored rays with image plane for **V**. This is the perspective projection of [**G**] such that $\textbf{I3}(r,g,b) = \textbf{V}*[\textbf{G}]$.

Repeating this for the second eye with a different **E** (i.e. different COP) allows one to create stereo images. In practice, **E** is chosen to approximately match the parameters (especially, field of view and orientation) of the user's eye to minimize the size of required framebuffer for rendering.

3.3.   Registration artifacts

For purely virtual environments, it suffices to know the approximate position and orientation of the user's head in a fixed world coordinate system. Small errors are not easily discernible because the user's visual sense tends to override the conflicting signals from his or her vestibular and proprioceptive systems. However, in see-through AR or SAR, preserving the illusion that virtual and real objects coexist requires proper alignment and registration of virtual objects to real objects [Azuma94]. Traditional AR methods use body-centric coordinate system to render synthetic objects and SAR methods use a fixed world coordinate system to render them. However, in both, the registration errors are caused by a number of factors such as system delay, optical distortion and tracker measurement error, and are difficult to correct with existing technology [Holloway95].

The tracking requirements for registration in SAR are similar to SID-VR systems because real and virtual objects lie in the same fixed world-coordinate system. Thus, static calibration errors can play an important role in registration. They include correct estimate of transformations between display devices, tracker and world coordinate system. In video-see through AR, optical distortion in camera lenses introduces error when computer generated imagery is combined with video images [State96]. On the other hand, registration error in SAR is introduced by optical distortion in the projector lenses.

Let's compare how errors in the actual and estimated perspective projection parameters of the system result in visible artifacts. In see-through AR, such errors result in virtual object 'swimming' with respect to the real objects [Holloway95]. In SAR, these errors lead to fundamentally different types of visible artifacts.

As noted in the previous section, the system needs only the location of user's eye and not the orientation to render perspectively correct images of virtual objects on real surfaces. Hence, change in only the orientation of the user does not change the way real objects are 'painted'. Similarly, an error in the measurement of orientation in the tracker will not result in misregistration error. An error in measurement of position of the user will result in shearing of virtual objects that are projected on the real surfaces with which they are expected to be registered. The shear visible at any point on the virtual object is proportional to the distance between the virtual point from real display surface on which the object is projected. Therefore, a virtual object not in 'contact' with a real surface but fixed in world will exhibit shear as well as swimming when there is position measurement error.

4.   Advantages of SAR

A key benefit of SAR is that the user does not need to wear a head-mounted display. In [Bryson97] various advantages of spatially immersive displays over head-mounted displays have been noted. SAR shares similar benefits. In SAR, large field-of-view images can be generated with greater amount of integration of virtual objects with real world and also to improve sense of immersion if necessary. Projector-based SAR allows possibly higher resolution and bright images of virtual objects, text or fine details. Since virtual objects are typically rendered near their real-world location, eye accommodation is easier.

5.   Problems with SAR

The most crucial problem with projector-based SAR is its dependence on display surface properties. A light colored diffuse object with smooth geometry is ideal. It is practically impossible to render vivid images on highly specular, low reflectance or dark surfaces. The ambient lighting can also affect the contrast of the images. This limits application of SAR to controlled lighting environments with restrictions on type of objects with which virtual objects will be registered. For front-projector-based SAR, shadows of the user can create problems. This can be partially overcome using multiple projectors. SAR also allows only one active head-tracked user at any instant in the environment because the images are created in the physical environment rather than in individual user space. Time multiplexed shuttered glasses can be used to add more users that are active and head-tracked.

6.   Future Work

We have built a proof-of-concept system and demonstrated the SAR ideas in our laboratory, however we are anxious to test the approach with a real application. For instance, something like one of the many applications listed in the introduction, or an application that involves a more complicated physical display surface e.g. room-sized terrain visualization.

While the majority of our efforts to date have been focused solely on projector-based SAR, a hybrid environment could be built with, for example, digital light projectors and a see-through HMD. While this would require the user to wear an HMD, we believe that such a hybrid has the potential to offer the best of both worlds, combining all of the advantages of conventional see-through AR and this new SAR paradigm. As we have experience building both conventional AR systems and projection-based SAR systems, we look forward to building and experimenting with such a hybrid environment.

7.   Conclusion

Others have certainly used light projectors and even LCD panels to add virtual imagery to real environments. However, this is essentially an augmented reality problem, albeit an unusual one, and we are excited about the opportunities that accompany this realization. On the one hand, some of the conventional wisdom about AR can be applied to address certain SAR problems, and on the other hand, SAR can be used to address some difficulties with conventional AR for some applications. We look forward to refining our ideas and the related algorithms, further analyzing the relevant error sources, and to pursuing some of the many applications we have in mind.

References

[Azuma94] Azuma, R., Bishop, G. Improving Static and Dynamic Registration in an Optical See-through HMD. Proceedings of SIGGRAPH 94 (Orlando, Florida, July 24-29, 1994). In *Computer Graphics* Proceedings, Annual Conference Series, 1994, ACM SIGGRAPH, pp. 197-204.

[Bennett98] David T. Bennett. Chairman and Co-Founder of Alternate Realities Corporation, 215 Southport Drive Suite 1300, Morrisville, NC 27560, USA. Cited 29 March 1998, available at http://www.virtual-reality.com.

[Bryson97] Bryson, Steve, David Zeltzer, Mark T. Bolas, Bertrand de La Chapelle, and David Bennett. "The Future of Virtual Reality: Head Mounted Displays Versus Spatially Immersive Displays," SIGGRAPH 97 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison-Wesley, pp. 485-486, August 1997.

[Cruz-Neira93] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. 1993. "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE," SIGGRAPH 93 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison Wesley.

[Dorsey91] Julie O'B. Dorsey, François X. Sillion, Donald P. Greenberg. 1991. "Design and Simulation of Opera Lighting and Projection Effects," SIGGRAPH 91 Conference Proceedings, Annual Conference Series, Addison-Wesley, pp 41-50.

 [Holloway95] Holloway, R. "Registration Errors in Augmented Reality Systems," PhD Thesis. University of North Carolina at Chapel Hill, 1995.

[Hologlobe] (Cited July 10, 1998) http://www.3dmedia.com

[Hornbeck95] Hornbeck, Larry J., "Digital Light Processing for High-Brightness High-Resolution Applications," [cited 21 April 1998]. Available from http://www.ti.com/dlp/docs/business/resources/white/hornbeck.pdf, 1995

[Jarvis97] Kevin Jarvis, "Real Time 60Hz Distortion Correction on a Silicon Graphics IG," in Real Time Graphics, Vol. 5, No. 7, pp. 6-7, February 1997.

[Max91] Nelson Max. 1991. "Computer animation of photosynthesis," Proceedings of the Second Eurographics Workshop on Animation and Simulation, Vienna, pp. 25-39.

[Milgram94a] P Milgram and F Kishino. "A taxonomy of mixed reality visual displays", *IEICE (Institute of Electronics, Information and Communication Engineers) Transactions on Information and Systems*, Special issue on Networked Reality, Dec. 1994.

[Milgram94b] P Milgram, H Takemura, A Utsumi and F Kishino. "Augmented Reality: A class of displays on the reality-virtuality continuum". *SPIE Vol. 2351-34, Telemanipulator and Telepresence Technologies,* 1994.

[Raskar98a] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. 1998. "The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays," to appear in SIGGRAPH 98 Conference Proceedings, Annual Conference Series, Addison-Wesley, July 1998.

[Raskar98b] Ramesh Raskar, Matt Cutts, Greg Welch, and Wolfgang Stürzlinger. 1998. "Efficient Image Generation for Multiprojector and Multisurface Displays," UNC Computer Science Technical Report TR98-016, University of North Carolina at Chapel Hill, March 1998.

[Raskar98b2] Raskar, Ramesh, Matt Cutts, Greg Welch, Wolfgang Stüerzlinger. "Efficient Image Generation for Multiprojector and Multisurface Displays," Rendering Techniques '98, Drettakis, G., Max, N. (eds.), Proceedings of the Eurographics Workshop in Vienna, Austria, June 29-July1, 1998

[Raskar98c] Raskar, Ramesh, Henry Fuchs, Greg Welch, Adam Lake, Matt Cutts. "3D Talking Heads: Image Based Modeling at Interactive Rate using Structured Light Projection," University of North Carolina at Chapel Hill, Dept of Computer Science, Technical Report TR98-017, 1998

[Raskar98d] Ramesh Raskar, Greg Welch, Henry Fuchs. 1998."Seamless Projection Overlaps Using Image Warping and Intensity Blending," to appear in Fourth International Conference on Virtual Systems and Multimedia, Gifu, Japan. November 1998.

[State96] State, A., Hirota, G., Chen, D.T., Garrett, W.F., Livingston, M.A. "Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking". Proceedings of SIGGRAPH `96 (New Orleans, LA, August 4-9, 1996). In *Computer Graphics* Proceedings, Annual Conference Series, 1996, ACM SIGGRAPH.

[Segal92] Mark Segal, Carl Korobkin, Rolf van Widenfelt, Jim Foran, and Paul E. Haeberli. 1992. "Fast Shadows and Lighting Effects using Texture Mapping," SIGGRAPH 92 Conference Proceedings, Annual Conference Series, Addison Wesley, volume 26, pp. 249-252, July 1992.

[Tsai86] Tsai , Roger Y. "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 364-374, 1986.

[UnderKoffler97] John Underkoffler. "A View From the *Luminous Room*," Springer-Verlag London Ltd., Personal Technologies (1997) 1:49-59.