

ACHIEVING INTERACTIVE-TIME REALISTIC ILLUMINATION IN MIXED REALITY

Charles E. Hughes, Erik Reinhard, Jaakko Konttinen, Sumanta Pattanaik
School of Computer Science
University of Central Florida
Orlando, FL 32816-2362

ABSTRACT

Properly lighting virtual objects and having them cast shadows on each other based on dynamic lighting conditions is a computationally hard challenging. Even more complex are the issues raised when this must be done in a Mixed Reality environment where virtual objects and virtual light interact with real objects and real light. This paper presents several interactive-time algorithms to enable interaction between real and virtual objects, including color matching, lighting and shadowing techniques.

1. INTRODUCTION

Proper lighting of synthetic objects is critical to the effectiveness of any immersive experience, and even more so when these objects are seen in the same context as real objects. The work we present is applicable across the entire range of Mixed Reality, from augmented reality, where synthetic objects augment the real world, to augmented virtuality, where real objects augment a virtual world. In particular, our techniques can be used to react rapidly to changes in real world ambient lighting, and to dynamic changes in virtual lights, including the addition, repositioning and reorientation of these lights.

This research is applied to military training and future warrior concept validation in the context of the MR MOUT (Mixed Reality for Military Operations in Urban Terrain) project, an augmented reality application supported by the U.S. Army RD&E Command (RDECOM) and hosted at the RDECOM STTC facility in Orlando. This application uses video see-through HMDs to create an anytime, anywhere MOUT training facility. Our techniques are also being applied to a collaborative augmented virtuality system, the DEMO DOME, and to the augmentation of a science exhibit, MR Sea Creatures (Hughes et al., 2004). This latter application does not involve an HMD, but rather uses an MR Dome Projection system (a Dome Projection system with attached camera(s) to capture the real world), an approach that helps address issues of cost, hygiene and maintenance in a high-use setting such as a museum.

2. CONTRIBUTIONS

We present two distinct contributions to combining live video with rendered content. These techniques were developed by the Graphics Group at UCF's School of Computer Science and then integrated with the MR System at the Media Convergence Laboratory.

The first method uses changes to the colors of objects in the real world, as captured by video cameras (typically those in the HMD), to shift and scale the pixel data in each virtual object so that the mean and variance of the pixel colors in its rendering move in the same way as those of the real scene (Figure 1). In its simplest form, this means that a shift in the colors of the real world, e.g., their darkening in reaction to the lights being dimmed, results in a similar shifting in the colors of all virtual objects. The key here is to choose the color space correctly and adjust for differences in the composition of each virtual object relative to that of the real scene. We have found that RGB is a poor choice because of the correlation of the three dimensions, whereas $L\alpha\beta$ color space (Ruderman, 1998), with its essential independence across dimensions, appears to be quite effective. Of course, this paradigm can be reversed, in the sense that a change to virtual ambient light can be used to adjust the colors of real world objects in the same manner. Although our approach of having the virtual react to change in the real world works in any see-through technology (video or optical), the paradigm reversal only works in the context of a video see-through HMD where we can alter reality before the user gets to see it.

Our second contribution is biased toward video see-through technologies, including video see-through HMDs, MR Dome Projection systems and MR Windows (screens that can pivot to change orientation and that have attached real world capture camera(s)). Here we perform lighting computations, using dynamic virtual lights to illuminate real and virtual objects, based on the location, direction, beam width and radius (distance the light travels). We also use the silhouettes of virtual objects to create shadows on real objects, as well as other virtual objects. In a similar manner, the silhouettes of real objects block the propagation of virtual light to objects that they

(partially) occlude. The key in this work is the creation of algorithms that are implemented as shaders found in modern graphics cards, e.g., those from ATI and nVidia. These GPU implementations achieve the interactive rate required for a successful MR experience.

3. COLOR TRANSFER

In mixed reality settings, rendered imagery is superimposed onto captured video data. There is generally a mismatch between the visual quality of these data streams. The lighting of the rendered scene may be different from the filmed scene; the shadows may be in the wrong places; and the lighting of either scene does not interact with the objects in the other scene. Compositing of rendered and captured data may therefore not yield the visual quality necessary for effective simulation and training.

In many cases, the rendered images “stand out” from the filmed data. There are a number of different possible causes for this – the materials and colors chosen for the rendered geometry may be unnatural; the light sources may be brighter or dimmer than the lighting in the filmed scene; and the light sources used in the rendered scene may be the wrong color.

Without directly addressing the causes for visual mismatches, we favor a simple yet effective image-based correction step which brings the color content of the rendered imagery within the same approximate range as the filmed data. We match each rendered frame to its filmed companion using a simple statistical approach. Our method may be combined with the final compositing step. The simplicity of the approach makes the method suitable for interactive applications.

Our real-time solution is based on color transfer between images, a technique which relies on a carefully chosen color space and analysis of two images in terms of means and variances (Reinhard et al., 2001). Here, a decorrelated color space is used to analyze both example and target images – the statistical measures being mean and variance. The pixels in the target image are shifted and scaled so that this image has the same mean and variance as the example image. This method is straightforward to implement and computationally efficient.

In RGB color space, the values of one channel are almost completely correlated with the values in the other two channels. This means that, if a large value for the red channel is found, the probability of also finding large values in the green and blue channels is high. Color transfer between images in RGB color space would therefore constitute a complex 3-dimensional problem with no obvious solution.

On the other hand, it is possible to run Principle Components Analysis (PCA) on image data. The effect of applying PCA is that the data is rotated such that the three color axes become decorrelated (Ruderman et al., 1998). In practice, the axes are close to independent. The resulting color space is termed $L\alpha\beta$.

The implication for color transfer is that our complex 3-dimensional problem may be broken into three 1-dimensional problems with simple solutions (Reinhard et al., 2001). The color transfer algorithm thus converts example and target images to $L\alpha\beta$ color opponent space. Here, the mean and variance of all pixels in both images are computed. This results in three means and three variances for each of the images. The target image is then shifted and scaled so that in color opponent space the data along each of the three axes has the same mean and variance as the example image. The result is then converted back to RGB space.

In our mixed reality application, we apply this algorithm to modify each rendered frame to match its companion captured frame, before compositing the result. In terms of visual quality, we have found that if we shift and scale the pixel data such that the means and variances are precisely the same as the mean of the video capture, the results are too dramatic. For instance, if the video depicts a red brick wall, adjusting the rendered data to be displayed in front of the wall to have the same means and variances will result in the data becoming fully brick-colored. We have found that by adjusting the variance of the rendered data to be in-between its original variance and the variance of the filmed frame, usually improved results are obtained.

We combine this adjustment with the actual compositing step so that a pixel is either copied from the video capture, or it is a rendered pixel which is adjusted by the above formula. Once each pixel has either been copied or adjusted, the resulting image is converted back to RGB space for display. We demonstrate the behavior of our algorithm in Figure 1. Simply compositing the results would cause the color content of the sphereflake to stand out and not mesh very well with the background. By adjusting the values to alter the variances of the sphereflake’s pixels to shift 40% in the direction of the background’s pixel variances we obtain the right-most frame in Figure 1 (Reinhard et al, 2004).

Figure 2 shows a Pegasus, the University of Central Florida’s logo, as it passes by a vending machine and then the window looking out from the second floor of the Computer Science Building. Although we are not running an illumination algorithm, the background appears to reflect off the logo in both cases.



Figure 1. Video frame, rendered frame, result of compositing, and result of color adjustment algorithm.



Figure 2. Apparent reflection from vending machine and washing out of colors from direct sunlight.

4. LIGHTING AND SHADOWS IN MR

Accurate lighting is one of the key elements to realism in rendered images of virtual objects. The previous section describes a post-processing approach to manipulate the color of the synthetic objects to the color of the real world into which the virtual object is integrated. This method is intended as a reasonable first approximation to light interaction between real and virtual worlds. However, careful observation will tell us that the light distribution on the virtual objects is not representative of the distribution that will result if that object were actually present in the real world, and more importantly it does not cast any shadows on the real-world objects. In addition to accurate lighting, shadows also provide a key cue to visual integration of real world and virtual objects in MR. We describe here our approach to address these issues.



Figure 3. The Ladybug camera for real-time capture of environment light. Details about the camera and its capabilities can be found from Pointgrey Research (<http://www.ptgrey.com/products/ladybug>).

To correctly handle the lighting and shadow problems in our MR system we must capture the distribution of the real world light source around the location of the virtual object. We also require geometric models of the visible objects in the real world. It is currently possible to instrument the capture of dynamic changes of light at fast

rates. For this purpose, we make use of a camera specifically designed to capture the lighting in the hemisphere surrounding a point (LadyBug, Figure 3). The camera captures environment maps at sub-second intervals. If we assume that the major light source direction does not change very fast, then this captured illumination can be used for lighting all the virtual objects in the environment.

We do not yet have an automated solution to capture the geometry of real world objects. Thus, at present, we pre-design this geometry for known static real world objects. In Section 5 we describe an interface to interactively capture the geometry of the planar faces of the objects.

In Figure 4 we show insertion of a virtual object into a 3D scene with and without shadow. The room model (floor, wall and cabinets) was created offline in 3D Studio MAX from photographs of the scene taken from various viewpoints. The geometry of the room model is used to access and modify the pixel values belonging to the real world to introduce shadow and any other lighting effects. Note that the lighting on the virtual object and the shadow cast by the virtual object on the floor are not appropriate to the actual illumination in the scene.



Figure 4. A real scene with and without shadow cast from the virtual object.

The captured real world light source is predominantly omni directional in nature and hence rendering of virtual objects with such a light source is not straightforward. We pre-compute and store accurate lighting effects of a number of basis light sources on the vertices of the virtual objects. At the time of rendering for MR, we approximate the captured environment light into a linear combination of the basis lights. We make use of the GPU vertex engine to compute the lighting at each vertex by modulating the

stored lighting effects corresponding to each basis light with the corresponding approximation coefficient and summing the modulated values.

The images in Figure 5 show a virtual object (bunny) accurately lit using the captured light of the scene. For shadow computation on the floor, we associate a well-tessellated phantom quadrilateral to the bottom of the virtual bunny. For each vertex of the phantom mesh, we pre-compute the lighting effect of the basis lights with and without the virtual object. We store the ratio of these coefficients at the mesh vertices. During rendering, we carry out the same computation at the phantom mesh vertices as we do at the vertices of the virtual object. The computation result at the mesh vertices gives the attenuation factor. The phantom is rasterized and the color of the pixels of the captured real-world overlapping with the rasterized phantom are attenuated by the interpolated factor. For the images shown here we used Spherical Harmonics basis lights (Ramamoorthi, 2001). This results in a smooth shadow appropriate to the lighting in the scene. Notice the realistic shadow appearance on the floor around the bunny in the left image. The image on the right is without shadow.

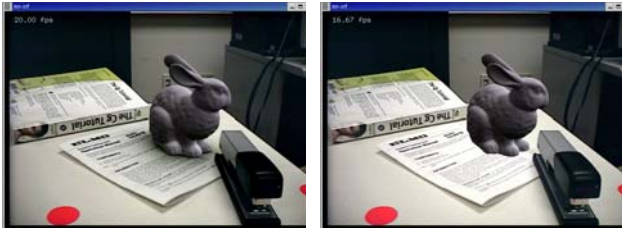


Figure 5. Accurately illuminated virtual bunny inserted into the real scene as viewed from a video see through HMD. The bunny on the left casts a shadow on the tabletop.

Spherical Harmonics basis is an appropriate choice for lighting with environmental light with low frequency light distribution. Point basis sources are more appropriate for high frequency light source distribution. In Figures 6 and 7 we show examples using points as basis lights (Hughes et al., 2004b). Using points as bases makes the computation much simpler. Instead of any pre-computation we carry out our computation on the fly.

Figure 6 shows an example of the lighting of the real world using a dynamic virtual fire. Using ARToolkit marker tracking utility (Kato et al, 2000) we tracked special markers to locate the positions for phantoms of real world objects. We modeled a virtual fire using a particle system consisting of approximately 300-500 particles. We used a highly simplified motion model consisting of largely constant upward velocity with some random variation to make the flames appear more chaotic. The flames were rendered by drawing a bill-boarded “flame texture” at each particle’s location and using

additive alpha blending to create a continuous shape. For the lighting calculation, we chose a point approximation of the light source. Each point source approximated the illumination of a particle group. The total light contribution is then computed as the sum of the light contributions from each point light. We computed this light per-pixel for pixels corresponding to the phantom of the real objects (floor and box sides) in a programmable hardware fragment shader, and performed the shading operation by using alpha blending between the lighting contribution and the original pixel intensity.

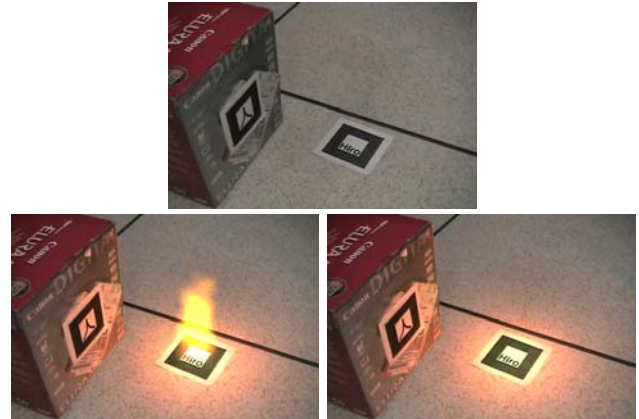


Figure 6. Sequence of virtual fire brightening a real scene, in clockwise direction from top: (a) Original intensity, (b) Illumination of real scene without virtual flames, (c) final scene with composited virtual flames.



Figure 7. Sequence of virtual flashlight illuminating a real scene, clockwise from top: (a) Original intensity, (b) virtual-to-real shadowing, (d) lit virtual and real objects with shadow in a darkened surrounding.

In Figure 7, we demonstrate the use of a virtual light to illuminate and cast shadow of virtual objects on a darkened real scene. Again, we used markers to locate the phantoms of the real objects and the virtual flashlight in the scene. The user can “shine” the flashlight at real objects which should then be lit correctly. The flashlight illuminates the virtual objects and real objects it is shined

towards. The shadow in the scene is computed using a stencil shadow volume technique run on a GPU. Shadowing on the real object is carried out by alpha blending dark shadow with the actual intensity of the visible pixels corresponding to the phantoms of the real objects.

5. INTERFACE FOR SIMPLE GEOMETRY CAPTURE

Most of the static objects in our environment have simple planar surfaces. We therefore developed an interactive interface for planar geometry capture (Hughes et al., 2004b). The interface uses ARToolkit for tracking markers in the captured image and extracting camera pose from them. It allows the user to quickly trace the surface of the planar object in the image as a polygon. Each vertex of the traced polygon is initially defined in screen space. It is then transformed to world space using the transformation matrix obtained from the camera pose. The polygon is then converted to a triangle mesh and is available for lighting and shadow computation. Figure 8 shows an example of the process.

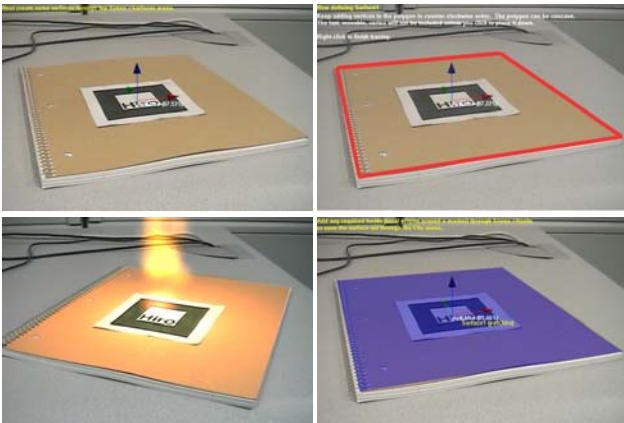


Figure 8. Sequence of interactive phantom generation for a notebook, clockwise from top-left: (a) notebook with a marker added to scene, (b) surface polygon being traced, (c) 3D phantom surface shown in blue, (d) notebook lit by virtual fire.

6. CAPTURING SILHOUETTES

Earlier we referred to the concept of an MR Demo Dome. This novel augmented virtuality environment uses a combination of a unidirectional retro-reflective screen and camera-mounted cold-cathode lights. Since unidirectional retro-reflective screen reflects light only to its source, the light emitted from the HMD is reflected back with negligible chromatic distortion or degradation from the surrounding environment. As a result, consistent lighting keys are achieved. The essential idea is that users are surrounded by a retro-reflective curtain, almost like being inside a cabana. When a user looks at this

surrounding material through the video see-thru HMD, the image received by the HMD's cameras is a blue (or green or whatever) screen. This chroma key can be replaced by virtual assets, so that users are surrounded in a virtual setting. Of course, each user must have a unique color associated with the lights on his or her HMD to avoid the other user's light appearing as the chroma key. With this approach, real objects in the environment that are not cloaked in the retro-reflective material stand out. Thus, each person can see any others who are in the same MR Demo Dome. Thus, the cameras capture the people and the compositing process treats them as real objects that should not be overlaid with virtual content. It is then trivial to compute the silhouettes of real objects; that's just the stuff that doesn't look like the chroma key.

To evaluate the MR Demo Dome, we seated two users inside a uni-directional retro-reflective cylinder wearing video see-thru HMD's. The experience begins with a virtual guide describing the scene. The users see a virtual 3D 360 degree view of a science center. The science center then floods with water and Cretaceous sea life begins swimming around the venue. The users feel as if they are underwater in a vehicle and are encouraged to move around the science center and observe and investigate the various creatures, plant life and other ancient artifacts in the scene. Navigation is done through a table that users can manipulate to move their vehicle. Each user also has a wand that casts virtual light to allow the users to select items in the scene and gain information on these chosen objects. Overall the sea creatures act and react according to their ancient behaviors (or at least what we presume are those behaviors). Virtual lights affect both real and virtual objects, casting shadows and changing luminance as appropriate. The ease with which the silhouettes of humans can be captured – they are the tracked objects that don't look like the chroma key – makes it possible to dynamically create imposters that are then used for lighting and shadows. The consequence is an engaging multi-modal immersive learning experience.

Figure 9 is an artist's rendition of the MR Demo Dome. The door area is shown open here to depict a user sitting at a table. Typically another user sits on the other side of the table. The table can be used as a navigation mechanism (tilting it navigates your point-of-view in the chosen direction). Other navigation paradigms are possible, and the table can be removed if that is appropriate. Also, the opening can be covered by draping the retro-reflective material. This creates a total immersion of the participants in a shared virtual setting. Surprisingly, creases and wrinkles in the material have no adverse effects; it still reflects light directly back to the source.

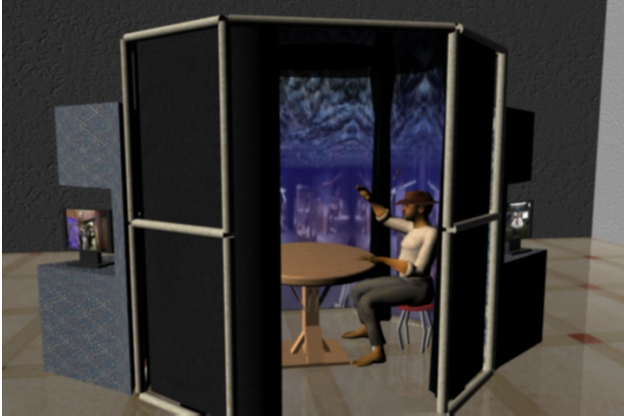


Figure 9. 3D Model of MR Demo Dome

CONCLUSIONS

In this paper we demonstrate that visual realism in Mixed Reality experiences can be achieved without sacrificing interactivity. Our specific contributions are the creation of two primary means of blending virtual objects into real scenes, and vice versa, in the context of dynamic lighting.

The first method focuses on having objects quickly respond to their changing environment without the need for any complex illumination algorithms. This approach is based on observing the changing statistical properties of the pixels of real objects and imposing those shifts on the pixels of all virtual assets. While we have not done so, this paradigm can be reversed to impose changes in the virtual light on real objects, provided the world is captured as video and then altered prior to the viewer seeing it. The second method deals explicitly with illumination models, providing shadows and the effects of reflection. This scheme is discussed in a variety of Mixed Reality settings from Augmented Reality to Augmented Virtuality. The latter involves a novel use of unidirectional retro-reflective material to create a low-cost alternative to a CAVE (Cruz-Neira et al., 1992); one that allows multiple people with multiple points of view to interact naturally in a shared virtual setting.

ACKNOWLEDGEMENTS

The research reported here is in participation with the Research in Augmented and Virtual Environments (RAVES) supported by the Naval Research Laboratory

(NRL) VR LAB. The MR MOUT effort is supported by the U.S. Army's Science and Technology Objective (STO) Embedded Training for Dismounted Soldier (ETDS) at the Research, Development and Engineering Command (RDECOM). We are also indebted to our undergraduate students, Jaakko Konttinen and Mark Colbert, who implemented these algorithms, to the members of UCF's Media Convergence Laboratory for their creation of context for this work and to its director, Christopher Stapleton, who first conceived of the MR Demo Dome. Special thanks are due to the Mixed Reality Laboratory, Canon Inc., whose support of our efforts through the generous provision of their Coastar Head Mounted Display has made this research possible.

REFERENCES

- Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R. and Hart, J.C., 1992: The CAVE, Audio Visual Experience Automatic Virtual Environment, *Communications of the ACM*, **35**(6), 64-72
- Hughes, C. E., Smith, E., Stapleton, C. B. and Hughes, D. E., 2004a: Augmenting Museum Experiences with Mixed Reality, *Proceedings of KSCE 2004*, St. Thomas, V.I., November 22-24, 2004, in press.
- Hughes, C. E., Konttinen, J. and Pattanaik, S. N., 2004b: The Future of Mixed Reality: Issues in Illumination and Shadows, *Proceedings of I/ITSEC 2004*, Orlando, December 6-9, 2004, in press. Also available from <http://www.cs.ucf.edu/graphics/MAR-Sumant/>.
- Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., & Tachibana, K. 2000: Virtual Object Manipulation on a Table-Top AR Environment. *Proceedings of ISAR 2000*, Oct 5-6, 2000, 111-119.
- Ramamoorthi, R. and P. Hanrahan, 2001: An Efficient Representation for Irradiance Environment Maps, *Proceedings of SIGGRAPH 2001*, Los Angeles, August 12-17, 497-500.
- Reinhard, E., Ashikhmin, M., Gooch, B. and Shirley P., 2001: Color Transfer between Images. *IEEE Computer Graphics and Applications*, **21**, 34-41.
- Reinhard, E., Akuyz, A. O., Colbert, M., Hughes, C. E. and O'Connor, M., 2004: Real-time Color Blending of Rendered and Captured Video. *Proceedings of I/ITSEC 2004*, Orlando, Dec 6-9, 2004, in press.
- Ruderman, D. L., Cronin, T. W. and Chiao C-C., 1998: Statistics of Cone Responses to Natural Images: Implications for Visual Coding. *J. Opt. Soc. Am. A*, **15**(8), 2036-2045.