Mixed Fantasy: An Integrated System for Delivering MR Experiences

Charles E. Hughes
School of Computer Science &
Media Convergence Laboratory
University of Central Florida
ceh@cs.ucf.edu

Christopher B. Stapleton
Darin E. Hughes
Scott Malo
Matthew O'Connor
Media Convergence Laboratory
University of Central Florida
{chris,darin,scott,matt}@mcl.ucf.edu

Paulius Micikevicius
Computer Science Department
Armstrong Atlantic State Univ.
paulius@drake.armstrong.edu

1. Introduction

Mixed Reality (MR), the landscape between the real, the virtual and the imagined, creates challenges in science, technology, systems integration, user interface design, experimental design and the application of artistic convention.

The primary problem in mixing realities (and not just augmenting reality with overlays) is that most media techniques and systems are designed to provide totally synthesized disembodied experiences. In contrast, MR demands direct engagement and seamless interfacing with the user's sensory perceptions of the physical. To achieve this, we have developed an integrated system for MR that produces synthesized reality using specialized content engines (story, graphics, audio and SFX) and delivers content using hybrid multi-sensory systems that do not hinder the human perception of the real world.

Figure 1. Blend of virtual explosion on real window

The primary technical contributions we will discuss include a unique hybrid application of real-time composting involving techniques of chroma-key mattes, image recognition, occlusion models, rendering and tracking systems. The primary integration contribution is a component-based approach to scenario delivery, the centerpiece of which is a story engine that integrates multimodal components provided by compelling 3-D audio, engaging special effects and the realistic visuals required of a multimodal MR entertainment experience. The primary user interface contributions are a set of novel interaction devices, including plush toys and a 3-DOF tabletop used for collaborative haptic navigation. Our work on human performance evaluation in multi-sensory MR environments is just now starting, with plans for extensive testing by cognitive psychologists, usability engineers and team performance researchers using data available from our delivery system. Our contributions to artistic convention include techniques derived from a convergence of media and improvisational theater.

2. Rendering and Registration

While most of our work transcends the specific technology used to capture the real scene and deliver the mixed reality, some of our work in rendering and registration is highly focused upon our use of a video see-through head mounted display (HMD). The specific display/capture device used in our work is the Canon Coastar video see-through HMD [1]. Thus, this section should be read with such a context in mind.

In our MR Graphics Engine, the images displayed in the HMD are rendered in three stages. Registration is performed during the first and the third stages and includes properly sorting the virtual and real objects, as well as detecting when a user selects a real/virtual object.

2.1. Rendering

Rendering is done in three stages. In the first, images are captured by the HMD cameras, and processed to correct for lens distortion and scaling. Since the cameras provide only RGB components for each pixel, an alpha channel is added by chroma-keying, as described in the subsection below. Alpha-blending and multi-pass rendering are used to properly composite the virtual objects with the captured images. For example, a virtual character passing by a portal is properly clipped. Additionally, where possible, the depth of real objects is determined by pre-measurements, visual markings and/or embedded tracking.

In the second stage the virtual objects are rendered and combined with the output of the first stage. This often involves the rendering of invisible objects that serve as occluders, e.g., a model of a building can be aligned with the actual building, rendered using only the alpha channel, thus serving to totally or partially occlude virtual objects that are at greater depths.

In the third stage, the output of the second stage is used to determine what objects the user has selected (if selections have been made). This phase can detect the intersection of either real beams (produced by lasers) or virtual beams (produced by projecting rays from tracked selection devices into the rendered space).

2.2. Object selection

A number of Mixed Reality scenarios require that users have the ability to select virtual objects. A variation of the selection process is registering a hit of an object (real or virtual) by a virtual projectile emanating from a real device, e.g., a futuristic-looking laser blaster.

A number of systems use tracking devices on the selection mechanisms for this purpose. The limitations of this approach are that it is susceptible to tracking errors and is constrained by the maximum number of tracking elements in the system. As an alternative, a laser emitter can be combined with laser-dot detection in the images captured by the HMD. Laser-detection is performed in real-time by our system. In fact, for efficiency reasons, it is part of the chroma-keying process in our system. The problem with this scheme is that it requires that the targeted object be seen by the HMD.

The limitations noted above have led us to implement both schemes (tracking and vision-based). We have found that the story is the primary driver for emphasizing one technique over the other.

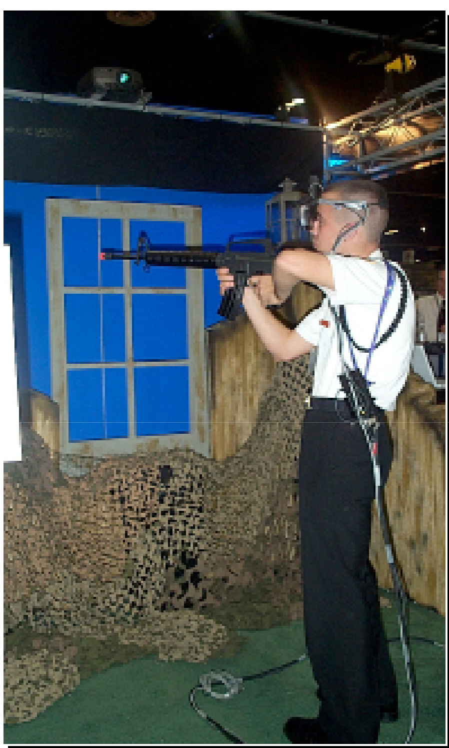

Figure 2. Tracked user and weapon

2.3. Retro-reflection chroma-key

As noted above, we often employ chroma-keying techniques to properly sort virtual objects with respect to real-world objects, especially when live-capture video is inserted into portals such as windows or doors in a real scene. One approach is to key off of a monochromatic screen. A blue or green screen has become a standard part of television and film studios. Unfortunately, lighting conditions greatly affect the color values recorded by digital cameras and thus must be tightly controlled. Uniform lighting of the screen is necessary to avoid "hot-spots." Such control is problematic for MR systems as they often operate in environments with dynamic lighting. Furthermore, the lighting of a reflective or backlit surface spills light on physical objects degrading a precise matte.

Figure 3. Blue chroma key and SFX lights

As an alternative, we use a combination of a unidirectional retro-reflective screen and camera-mounted cold-cathode lights. Since unidirectional retro-reflective screen reflects the 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, dramatically reducing the computational complexity and at the same time allowing for dynamic venue and environmental lighting..

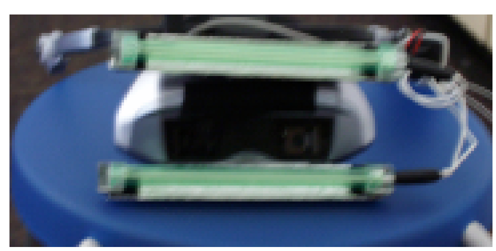

Figure 4. HMD rigged with cathode lights

3. Audio

In designing the MR Sound Engine, we required a system that could provide real-time spatialization of sound and support acoustical simulation in real physical spaces. This required a hybrid approach using techniques from simulation, theme parks, video games and motion pictures. This led to the integration of novel techniques in both audio engine and environmental display. Real-time synthesis of sound was not a primary concern.

After studying many alternatives from APIs to DSP libraries, we settled on EAX 2.0 in combination with OpenAL. We also chose to incorporate the ZoomFX feature from Sensaura to simulate emitter size of sound sources.

There are three main properties that affect 3-D sound: directionality, emitter size and movement [2]. The first can be divided into omni-directional versus directional sounds, the second into point sources with emitter size zero versus voluminous sources, and the third into stationary versus moving sounds.

The MR Sound Engine combines these attributes to provide the ability for real-time spatialization of both moving and stationary sound sources. For stationary 3-D sounds, the source can be placed anywhere along the x, y, and z axes. Moving sounds must have a path and a timeline for that movement to take place. In our case, the path is typically associated with a visible object, e.g, a person walking, a seagull flying or a fish swimming.

Simulation of emitter size, sound orientation, and rotating sound sources are supported. Emitter size can only be altered for sounds that are spatialized. Sounds that do not have spatialization are referred to as ambient sounds and play at a non-directional level in surround sound speakers.

Standard features such as simultaneous playback of multiple sounds (potentially hundreds using software buffers), ramping, looping, and sound scaling are supported.

4. Special Effects

Special effects devices in a mixed reality allow for rich enhancement of the real space in response to actions by either real or virtual characters. When coupled with our story engine, this system creates the next level of mixed reality – one in which the real world knows and is affected by items in the virtual, and vice versa.

4.1. Effects

We apply traditional special effects, the same that might enrich a theme park attraction, theater show, or museum showcase. The following is a breakdown of some of the types of effects we use.

We lace our experiences with lighting – to illuminate the area of the experience and also as localized effects. Area lighting can signify the start of a scenario, change the mood as it is dimmed, and reveal new parts of the set on demand. Effect lighting allows us to illuminate smaller areas, or simply act as an interactive element. Heat lamps allow the user to feel the temperature of an explosion and flickering LED lighting can simulate a fire.

In one of our simulations, users are instructed to shoot out lights. A motion tracked shot allows us to flicker out the light (over a door or window). In another scenario, users are navigating along the dirt paths of an island and through its underwater regions. Their speed of motion directly affects wind in their faces or at their backs if they are moving in reverse. This effect is achieved simply by turning on rheostat controlled fans. We thought about blowing mist in their faces when they were underwater, but we quickly realized that this would adversely affect the cameras of the HMD.

We also use compressed air for various special effects. Pneumatic actuators allow us to move pieces of sets, such as flinging a door open to reveal a virtual character or toppling a pile of rocks. Compressed air can be blown in a short burst to simulate an explosion or to scatter water, sawdust, or even smells. Smoke can be used to simulate an explosion, a rocket taking off, or can be combined with lighting to simulate an underwater scene. We use low-cost commonly available smoke machines to great effect.

Our system allows anything that can be controlled electrically (or mechanically, with pneumatic actuators) to be connected. We have connected fans and vibrating motors to the system, and even driven shakers with the bass output of our audio system.

The use of all this equipment creates a rich real world environment that transcends the virtual and assaults the senses with the type of effects popularized by theme park attractions. The difference is that these effects are truly interactive, able to be controlled both by elements in the virtual environment as well as the real participant.

5. User interfaces

How one intuitively interacts with an MR experience can have a profound effect on how compelling the experience is. Using a gun device to disable evil alien creatures is reasonable, but using the same device to control a dolphin's action is obscene and inappropriate, especially for children.

Inspired by the work of MIT's tangible media group, we dissected a plush doll (a cuddly dolphin), placed wireless mouse parts and a 6-DOF sensor inside it (interchangeably using a Polhemus magnetic or an Intersense acoustical/inertial tracker), and stitched it back up. This allows players (children preferred) to aim the dolphin doll at a virtual dolphin and control the selected dolphin's actions. In our case, you press the right (left) flipper down to get the dolphin to move right (left); you press the doll's nose to cause your dolphin to flip a ball off its nose (provided you already guided it to a ball); and press its rear fin to get your virtual dolphin to send sonar signals in the direction of others. That latter action (sending sonar) can be used to scare off sharks or to communicate with other dolphins. The communication aspect of this might be used in informal science education, as well as in pure entertainment.

Recently we created a novel MR environment that performs like a CAVE but looks like a beach dressing cabana and greatly reduces square footage, costs and the infrastructure of projection displays. We call this the Mixed Simulation Demo Dome (MS DD), and it consists of a unidirectional retro-reflective curtain that surrounds you when you enter this very portable setting. Inside the dome is a round table with two chairs and two HMDs. Users sit on each side of the table, which is itself covered with retro-reflective material. .

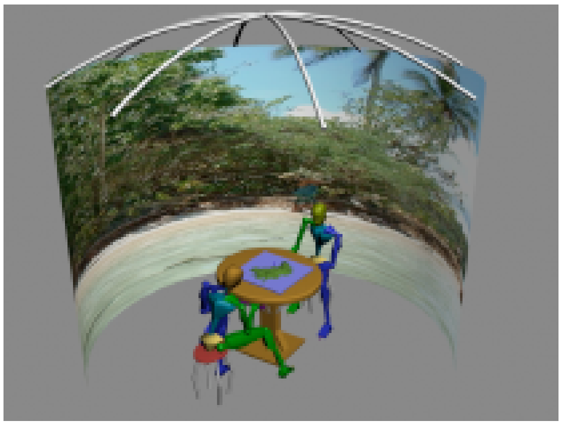

Figure 5, MS DD concept drawing

When the users start the MR experience, each wearing an HMD, they see a 3d map embedded in the table. The map is of an island with an underwater coral reef surrounding it. An avatar, representing both users, is on the island. The users have a God's-eye view of their position and movement, and must cooperate to move their character around the island and through the coral reef from opposite points of view. Tilting the 3-DOF tabletop left or right changes the avatar's orientation; titling it forward or backward causes the avatar to move around the world, basically in the direction of its current orientation. An underlying ant-based behavior model for the avatar prevents it from going up sharp slopes or through objects, such as coral reefs.

In providing both a "God's eye" point-of-view on the table and an immersive first person perspective around the users, the level of interactive play dramatically increases, yet remains intuitive due to the tangible haptic feedback. The desired effect is achieved by creating a dual chroma mask on the retro-reflective material that is produced by opposing green and blue lights on each of the user's HMDs. This first person scene is a shared one, although the users are typically looking in different directions. Since your fellow user sees you as part of the MR experience, your pointing and gazing are obvious. To encourage such collaboration, we have added many enticing virtual objects: trees, seagulls, a shark, coral fish, butterflies, a sting ray and a giant tortoise.

We have also added a little humor by changing the head of your fellow user to a fish head whenever you two are underwater. Since the fish head is made up of one-sided polygons, you don't realize that you look as silly as your colleague.

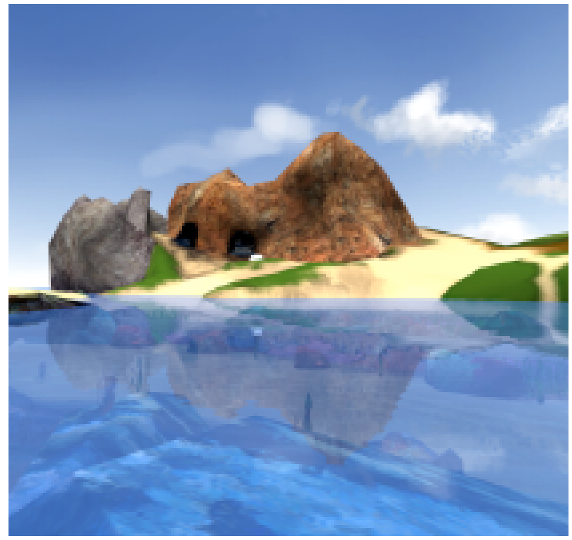

Figure 6. MS Isle augmented virtuality experience

6. The system

The Mixed Fantasy system encompasses both a scenario authoring and a scenario delivery system. These are integrated in the sense that the authoring system produces an XML document that is imported into the delivery system, resulting in the instantiation of all objects needed to run the scenario.

Figure 7. Integration of MR system

6.1. Authoring

At present, users author scenarios by selecting actors that reflect the synthetic and active real objects of your world. For instance, a scenario might contain a synthetic helicopter, several synthetic people, a number of real lights, a real soda can, a tracked moveable table on which a map appears, an abstract object that is acting like a story director and, of course, the real person who is experiencing the MR world.

An actor component, when brought into the authoring system, will have a set of default behaviors, some complex enough that they will have been pre-programmed. However, a behavior object, e.g., a finite state machine, a physical model or an AI-based software agent is the primary means of expressing an actor's behavior.

In the case of an FSM behavior object, each transition emanates from a state, has a set of trigger mechanisms (events) that enable the transition, a set of actions that are started when the transition is selected, and a new state that is entered as a consequence of effecting the transition. A state can have many transitions, some of which have overlapping conditions. States and transitions may have associated listeners (attached via hookups in the authoring environment) causing transitions to be enabled or conditions to be set for other actors.

6.2. The communication protocol

Other behavior models use other schemes, but all must eventually lead to external notifications of observable state changes. This notification is in the form of a message sent to all interested parties (graphics, audio and special effects engines). In our protocol the same message is delivered to all and each interprets (or ignores) it as appropriate. For example, the messages

```
SHARK_PATH MAKE PATH PATH1
SHARK_PATH LOOP
SHARK ASSOCIATE SHARK_PATH
SHARK SHOW
SHARK LOOP SWIM 200
USER ASSOCIATE SHARK
USER MOVETO 0 0 –1000 3000
```

cause the Graphics Engine to place the selected path in a "gallery" of objects that are updated at 30 fps. The shark object associates itself to the path, meaning that its position and orientation will move with the path. The shark will then become visible and start swimming, taking 200 milliseconds to blend from its stationary stance to its swimming animation. Finally, the user's viewpoint will be moved to 1000 millimeters behind the shark (time for a theme park ride), taking 3 seconds to transition to this position. Note that the user is free to move away from or towards the shark, but will still be dragged along as the shark moves.

These same messages, when seen by a 3D audio engine, lead to the sound of the shark's swimming being placed along the path, moving as the shark moves. Similarly, these messages may cause a special effects engine to change the lighting in the room to a sinister blood red color.

One critical aspect of the above is that only the story engine understands the semantics of events. All other parts of the system are either purely reactive (audio and special effects) or can also perform some specialized action, such as the graphics engine analyzing a scene, but do not determine the consequences of this analysis.

7. Future directions

Our team is made up of technical and artistic creators, each group facing its own challenges, but each supported by the other. The development of our system and the content it delivers has as much to do with integrating the implicit conventions and semantics of disciplines as it does in providing explicit standards and methods for execution. Our work is human centered and requires collaborations with human performance scientists and specialists from usability, cognition and entertainment.

7.1. Technical directions

The technical team is continuing its efforts to make the real and virtual as indistinguishable as possible. This work, in collaboration with other faculty at our institution, is currently focused on issues of real-time rendering and registration. The rendering tasks include illumination (real lighting affecting virtual objects and virtual lights affecting real objects), shadowing (also concerned with real-virtual interaction), visibility-based level-of-detail (LOD) management at run-time, and the evolution of plant models from grammatical descriptions (L-systems). [3]

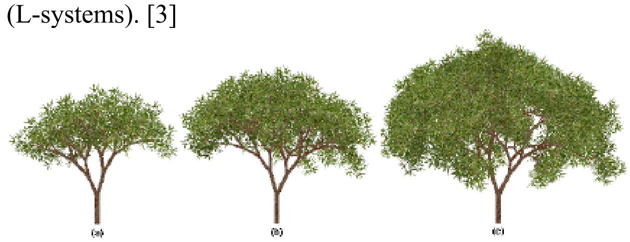

Figure 8. L-system evolved trees

Recent publications on rendering have demonstrated offline integration of real and virtual objects. However, extending such work for real-time computation is not straightforward [4]. Our approach creates accurate, interactive rendering of synthetic scenes with proper illumination and shadowing (at present shadows from virtual objects on both real and virtual, with the goal of adding real shadowing on virtual objects, as well). Our LOD work has focused on visibility-based complexity control. As with the illumination work, we are designing and implementing fast algorithms using the parallel capabilities of graphics processors and the inexpensive computational power of clusters.

7.2. Content directions

The content team is interested in bringing the dimension of imagination into MR [5], [6]. It is also interested in expanding the applications of MR into cross-industry application and interoperability such as informal education, marketing, business training, design and manufacturing, while continuing to expand its use in training and entertainment.

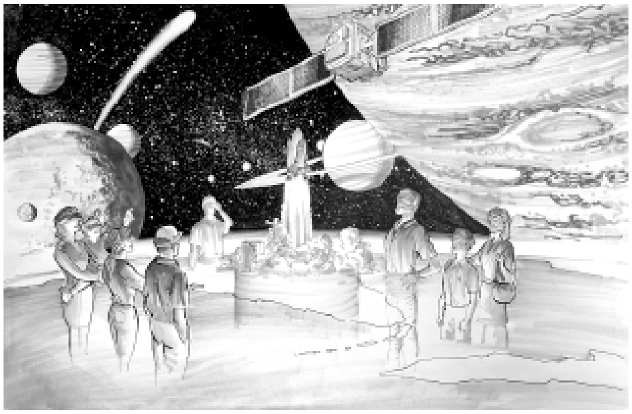

Figure 9. Concept drawing of MR planetarium

We view the MR continuum as having three dimensions: the real, the virtual and the imaginary. Moreover, we believe that a single modality (visual only) approach is disabling, and that the emotions evoked by the audioscape, the intuitiveness provided by special effects, and the tactile impact of appropriately chosen interaction devices are critical to MR's progress and acceptance. Thus, our content team is pushing the frontiers in each of these areas by emphasizing the rich layering of all of them over the fidelity of any one.

Figure 10. Timeportal experiential trailer

Figure 11. Timeportal at Siggraph 2003

8. Acknowledgements

The work reported here has been supported by the Army Research Development and Engineering Command, the Office of Naval Research, the Naval Research Laboratory, the National Science Foundation (under grant EIA 9986051) and the Canon Mixed Reality Systems Laboratory in Tokyo. The authors are also grateful for the support provided by UCF;s Institute for Simulation and Training, School of Computer Science and School of Film and Digital Media

9. References

[1] S. Uchiyama et al., "MR Platform: A Basic Body on Which Mixed Reality Applications Are Built," *ISMAR* 2002, Darmstadt, Germany.

[2] M. P. Hollier, et al., "Spatial Audio Technology for Telepresence," *British Telecommunications Technology Journal* 15(4), pp. 33-41, 1997.

[3] O. Deussen, P. Hanrahan, B. Lintermann, R. M_ch, M. Pharr, P. Prusinkiewicz. Realistic modeling and rendering of plant ecosystems, *Proc. of SIGGRAPH 98*, pp. 275-286, 1998.

[4] S. Gibson and A. Murta, "Interactive Rendering with Real-World Illumination," *Rendering Techniques 2000*, pp. 365-375, June 2000, Brno, Czech Republic.

[5] C. B. Stapleton, C.E. Hughes, J.M. Moshell, P. Micikevicius and M. Altman., "Applying Mixed Reality to Entertainment," *IEEE Computer* 35(12), pp. 122-124, December 2002.

[6] C. B. Stapleton and C. E. Hughes., "The Interactive Imagination: Tapping the Emotions through Interactive Story for Compelling Simulations," *IEEE Computer Graphics and Applications* 23(5), pp. 11-15, September-October 2003.