Sharing Space in Mixed and Virtual Reality Environments Using a Low-Cost Depth Sensor

Evan A. Suma* David M. Krum* Mark Bolas*†

*USC Institute for Creative Technologies ‡USC School of Cinematic Arts

ABSTRACT

We describe an approach for enabling people to share virtual space with a user that is fully immersed in a head-mounted display. By mounting a recently developed low-cost depth sensor to the user's head, depth maps can be generated in real-time based on the user's gaze direction, allowing us to create mixed reality experiences by merging real people and objects into the virtual environment. This enables verbal and nonverbal communication between users that would normally be isolated from one another. We present the implementation of the technique, then discuss the advantages and limitations of using commercially available depth sensing technology in immersive virtual reality applications.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality;

Keywords: mixed reality, depth-sensing cameras, HMDs

1 INTRODUCTION

Head-mounted displays (HMDs) can provide a highly immersive experience, but at the cost of isolating the user from the real world. While these displays can be very effective for single-user applications such as immersive architectural walkthroughs (e.g. [1]) and treatment of psychological disorders (e.g. [2]), their usefulness is limited for applications that require communication or collaboration between multiple users. Subsequently, applications such as military training often use immersion projection systems when multiple users are required (e.g. [3]). Mixed reality environments, which merge stimuli from both the real world and virtual environment, can overcome this limitation by using immersive head-mounted displays (HMDs) with attached cameras or optical-see through augmented reality displays. In either case, this introduces two significant challenges: (1) real world people and objects must be segmented from the background, and (2) they must rendered in the scene at the correct depth in the virtual environment so that occlusions between real and virtual stimuli are handled properly. While this is difficult with standard RGB cameras, recent advances in commercial depth-sensing technology by PrimeSense, the company that designed the sensing technology in the Microsoft Kinect, have made it possible to generate real-time depth maps using a single, low-cost sensor with a compact footprint. In this paper, we describe an approach that leverages this technology to enable sharing of virtual space with a user that is fully immersed in an HMD, which would normally be an isolated experience. This approach allows communication through body language and gesture, which will better support collaborative tasks in the virtual environment, and could be used to allow multiple users, each wearing their own HMD, to see each other.

*e-mail: {suma, krum, bolas}@ict.usc.edu

IEEE International Symposium on Virtual Reality Innovation 2011
19-20 March, Singapore
978-1-4577-0037-8/11/$26.00 ©2011 IEEE

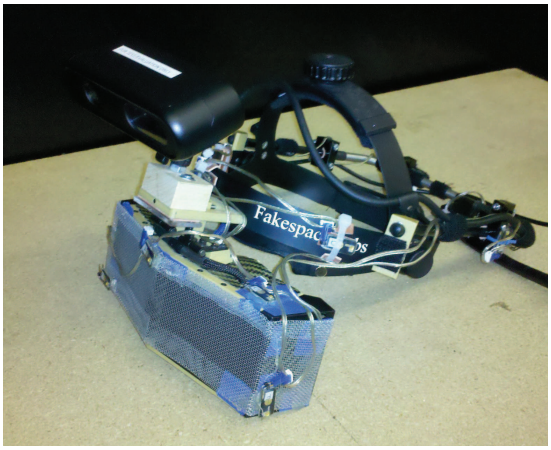

Figure 1: The PrimeSensor™ was mounted to the front of the Fakespace Wide 5 head-mounted display. This setup provided real-time depth maps of the real scene as the user's head moved.

2 SHARING SPACE IN THE VIRTUAL WORLD

Our implementation uses the PrimeSensor™ Reference Design, a USB plug-and-play device featuring an IR projector along with RGB and IR CMOS image sensors. In order to generate the depth map, the sensor uses coded IR light and a sophisticated black-box algorithm to resolve the light pattern. The system has a field-of-view of 58 degrees horizontal and 45 degrees vertical, and is capable of generating a depth map with a resolution of either 640x480 at 30fps or 320x240 at 60fps. In our initial tests, we found that the higher resolution depth map at the slower framerate was preferable, and so we used those parameters for all the techniques described in this paper. The software component of the sensor package, known as OpenNI, was used to retrieve the depth and RGB images in our VR applications and solve for the camera's intrinsic properties.

Due to the sensor's compact footprint, we began to explore the potential applications of the PrimeSensor™ for head-mounted immersive virtual environments. To this end, we built a custom mount for the sensor on a Fakespace Wide 5 HMD (see Figure 1). This HMD has a total FOV of 150 degrees horizontal and 88 degrees vertical, and a variable resolution with higher pixel density in the central region and lower resolution in the periphery. To track the HMD, we used a PhaseSpace Impulse Motion Capture System with an array of 52 high-resolution detectors that track seven LED markers mounted around the display. As the user moves around, the head-mounted sensor generates a depth map for the real world scene within the central region of their field-of-view, which can then be resolved into world coordinates based on the rigid offset between the sensor and the HMD's origin. Since the sensor is placed near the user's eyes, the resulting depth map will roughly correspond what would normally visible or occluded from the user's viewpoint. Given that the display has a much wider vertical field-of-view than the sensor, in our initial tests we found that the setup that best cap-

(a) (b) (c)

Figure 2: (a) A person strikes a pose in front of the immersed user wearing the HMD. (b) The person is segmented from the background and rendered in 3D in the virtual world in front of the immersed user. (c) A comparison between the unaltered pixels from the sensor's RGB camera and the "hologram" effect intended to give the impression that the person is projected into virtual world.

tured the visible scene required tilting the sensor slightly off-axis, approximately 15 degrees downward from the user's look vector. Using this apparatus, we began to experiment with merging real stimuli into the virtual environment.

We initially intended to experiment with the sensor's built-in segmentation and feature detection capabilities in a head-mounted setting, with the goal of rendering virtual avatars that represent real people that were interacting with the user immersed with the HMD. Unfortunately, these black-box algorithms assume a fixed sensor, and fail whenever the sensor is moved around. While it may be possible to replicate the sensor's existing feature detection capabilities and include a perspective correction to cancel out the user's head motion (using the head-tracking data), we chose to adopt a different approach. Instead, we merge the depth map with the image map returned from the sensor's RGB camera, and render real people directly into the virtual world as textured 3D point clouds (see Figure 2). This approach allows people to interact with the immersed user using gesture, body language, and facial expressions.

Due to the fact that our motion tracking area consists of open space within the tracking volume, segmenting the background from the depth map is a simple process. After resolving each depth map point from projective to real world coordinates, we place each point in the virtual environment, adjusting for the position of the HMD and the relative offset of the sensor. During this process, any point that falls outside the tracking volume or lies near the floor is masked out. In our tests, we found that a threshold of approximately six inches from the floor was sufficient to reliably mask out the background even at far distances without cutting off too much of a person's body. Thus, any points that remain correspond to people or other objects in the foreground. Of course, due to the resolution of the depth map, there will always be some noise present in the segmentation, particularly for complicated shapes such as human figures (see Figure 2.b). While this is very obvious in static example images, as the user and the observed figure moves around, the border around the figure is transient, resulting in a shimmering effect around the figure. Additionally, segmentation becomes noticeably less accurate for objects at distances greater than 12 feet, which is not surprising since the sensor was designed for video game applications in spaces the size of a typical living room.

Rendering a real person in a virtual scene does introduce an interesting conflict, since the appearance of the human figure does not match the virtual environment. Thus, we experimented with visual effects to make this discrepancy less objectionable, for example, by imitating the appearance of holographic projections from popular science fiction. Figure 2.c shows a "hologram" effect that we implemented by adjusting the RGB and alpha values of each point as they were read from the sensor's image map. In this example, we adjusted the red value down by 70%, green down by 25%, and the alpha down by 25%. This results in a partially transparent, blue-green tinted virtual "projection" that has been also been commented by observers as "ghost-like." Additionally, the reduced opacity of the figure also appears to make the noise from the depth map segmentation less noticeable.

3 Discussion and Conclusion

Our initial hope was to also use the sensor for displaying the immersed user's hands in the virtual world, since this is a feature that is frequently requested by users of our head-mounted display setup. Unfortunately, it appears that the sensor cannot resolve the depth for objects at a distance less than approximately three feet. We believe that this may be a limitation of the proprietary black-box coded light algorithm used to generate the depth map. It would still be possible to segment out these pixels and render the hands as a 2D overlay, as is normally done with standard RGB cameras, but without depth information they would not appear in 3D, nor would they properly handle occlusions with virtual objects.

While the technique described in this paper was possible using previously available, but comparatively expensive technology, the novelty is the use of a single, low-cost sensor that is compact enough to be easily mounted on top of the user's head. Though we have not yet performed any formal user studies, based on our qualitative observations, we suggest that low cost depth sensors such as the one from PrimeSense have great potential for enabling people to share virtual space in a variety of setups such as virtual environments that use immersive HMDs. When debugging the calbration of the sensor while wearing the HMD, one of the authors of this paper, engaged in an extensive conversation that included body language and gesture with another co-located person that was merged into the virtual world. We believe that this sharing of virtual space is an important component for delivering compelling and effective multi-user virtual reality experiences.

References

[1] G. Bruder, F. Steinicke, and K. H. Hinrichs. Arch-explore: A natural user interface for immersive architectural walkthroughs. In *IEEE Symposium on 3D User Interfaces*, pages 75–82, 2009.

[2] M. Gerardi, B. O. Rothbaum, K. Ressler, M. Heekin, and A. Rizzo. Virtual reality exposure therapy using a virtual Iraq: Case report. *Journal of Traumatic Stress*, 21(2):209–213, 2008.

[3] S. Koepnick, R. Hoang, M. Sgambati, D. Coming, E. Suma, and W. Sherman. Rist: Radiological immersive survey training for two simultaneous users. *Computers & Graphics*, 34(6):665–676, 2010.