MetaSpace: Full-body tracking for immersive multiperson virtual reality

Misha Sra
MIT Media Lab
Cambridge, MA, 02142 USA
sra@media.mit.edu

Chris Schmandt
MIT Media Lab
Cambridge, MA, 02142 USA
geek@media.mit.edu

ABSTRACT
Most current virtual reality (VR) interactions are mediated by hand-held input devices or hand gestures and they usually display only a partial representation of the user in the synthetic environment. We believe, representing the user as a full avatar that is controlled by natural movements of the person in the real world will lead to a greater sense of presence in VR. Possible applications exist in various domains such as entertainment, therapy, travel, real estate, education, social interaction and professional assistance. In this demo, we present MetaSpace, a virtual reality system that allows co-located users to explore a VR world together by walking around in physical space. Each user's body is represented by an avatar that is dynamically controlled by their body movements. We achieve this by tracking each user's body with a Kinect device such that their physical movements are mirrored in the virtual world. Users can see their own avatar and the other person's avatar allowing them to perceive and act intuitively in the virtual environment.

Author Keywords
Multi-person Virtual Reality; Full-body immersion; First person view; Locomotion; Physical to virtual mapping.

ACM Classification Keywords
H.5.1. Information Interfaces and Presentation (e.g. HCI): Multimedia Information Systems

INTRODUCTION
New VR head-mounted devices (HMDs) are light, have dual high resolution screens, provide spatial audio, allow a wide field of view, offer head tracking with low latency, and use custom hand-held haptic input/output devices for navigation and interaction in the virtual world. Despite these characteristics, the lack of proprioceptive cues related to locomotion could limit a user's sense of presence in the virtual environment. Our focus in this demonstration is related to active situations where two users are moving in physical and virtual space while interacting with each other in the real world (see Figure 1). Their real world interactions are mirrored by their avatars in the virtual world.

UIST'15 Adjunct, November 8–11, 2015, Charlotte, NC, USA.
ACM 978-1-4503-3780-9/15/11.
http://dx.doi.org/10.1145/2815585.2817802

Figure 1. Two people wearing Oculus Rifts with their full body being tracked by two Kinect devices in a MetaSpace multiuser VR experience.

First person experiences of the real world represent a standard to which all mediated experiences are compared, either mindfully or otherwise [3]. Our awareness of our body in space i.e. our body image is a *gestalt* through which all external space appears meaningful [1, 2]. Third person video games allow us to control a marionette through the physical extension of the game controller, making the avatar part of that through which the game world comes into existence. Bodily immersion in VR is rooted in the way the body is able to redirect a perception of itself as an object into virtual space (e.g. proprioception) and through this mirror image, the familiar body is also made the embodied subject during interaction [2].

Presence in VR is based on the perception of input through visual, auditory, or kinesthetic senses. For an enhanced sense of presence, VR needs to incorporate the participant as a part of the environment such that, head movements result in motion parallax, locomotion results in translation in space, proprioceptive cues are incorporated, vestibular responses are stimulated, and the user has agency. In most existing VR experiences, a seated user is presented with two contradictory motion cues; the visual stimulus signals movement to the brain, while the vestibular system indicates a lack of movement. These conflicting inputs can produce motion sickness and postural instability resulting in a sub-optimal immersive experience. For exploring VR environments, walking in place is not sufficient as it lacks the proprioceptive cues of actual walking [4]. In MetaSpace users can move in physical space

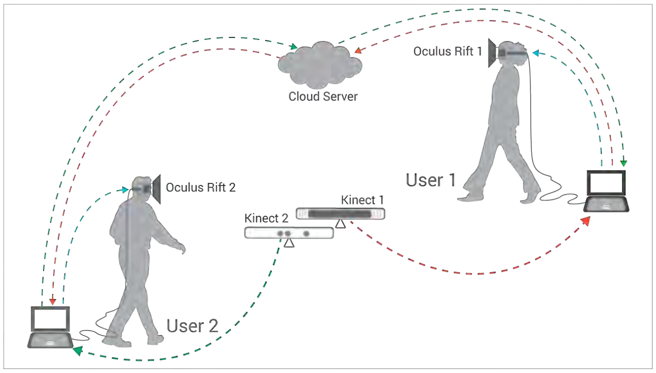

Figure 2. MetaSpace system configuration. Each user wears an Oculus Rift virtual reality head mounted display and can see themselves (first person view) and the other user's avatar on their screen. Each user is fully tracked with a Kinect and body movements like walking, waving etc of each user are transmitted through the cloud server such that everyone can see everyone and feel their presence in the virtual world similar to that in the real world.

(Figure 4), thereby maintaining coherence between the visual and vestibular inputs and experiencing VR in a more natural way. We use two Microsoft Kinect devices to track the full-body of each user and map their movements to an avatar in the virtual world. This allows users to see (first person perspective) their hands and legs move as they walk as well as see other users move around in a shared multi-person virtual world. In this demo we focus on the full body tracking and movement in real and virtual space while leaving the physical to virtual mapping for a separate full writeup.

METASPACE SYSTEM

The MetaSpace system consists of two parts (Figure 2). The first is two Oculus Rift head mounted display devices, one for each person, to view the virtual world. The second is skeleton tracking of each user using Kinect devices (Figure 4). In skeletal tracking, a human body is represented by a number of joints that represent body parts such as head, neck, shoulders, and arms. Each joint is represented by its 3D coordinates and mapped to the joints in a rigged 3D model. For a synchronous multiuser virtual experience, it is essential that all persons receive the same exact system state at all times. Therefore, MetaSpace features a client-server architecture that is made of four main entities: multiuser middleware and service (Photon Server), local clients (laptops), Microsoft Kinect devices, and HMDs (Oculus Rift). Data between the server and laptops is transmitted over wireless technology. The HMDs are wired to the laptops (for now) and present virtual world state such that each user can see the other move in real-time. Communication between the users and the virtual world happens through physical movement tracked using Microsoft Kinect devices. These mechanisms help create a natural hands-free immersive multiuser experience.

The demo VR was created by doing a 3D reconstruction of the hallway outside our offices using depth data, cleaning up and texturing the resulting 3D model of the space and objects, and adding other 3D elements like mountains and a valley to build a scene that is visually different from the hallway (Figure 3). Physical world barriers like walls and storage cabinets

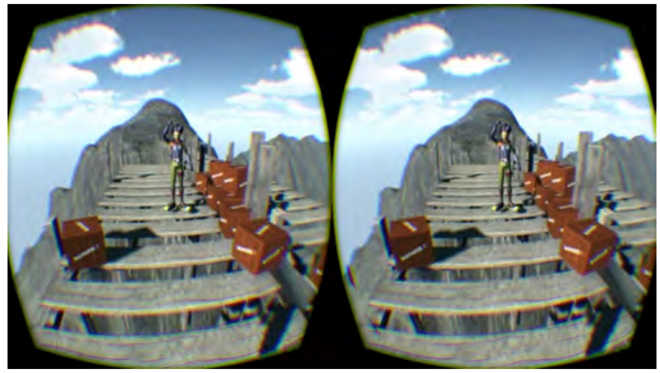

Figure 3. Virtual scene showing a bridge with one drop off corresponding to a physical wall and the crates corresponding to storage cabinets in the real world. The view through the first users' HMD shows the second user waving to them.

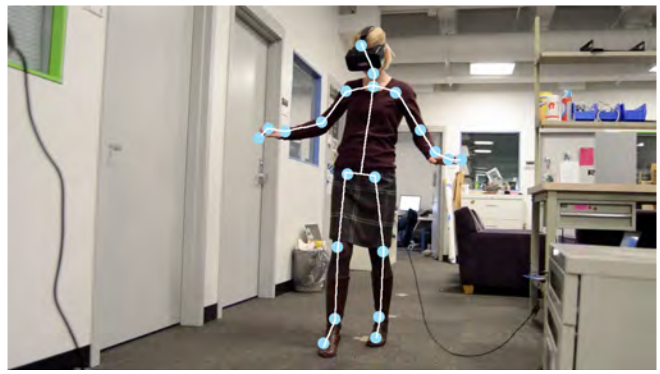

Figure 4. User 1 walking down the hallway (real) and bridge (virtual) toward User 2 (Figure 3) waving at them. User 1's skeleton is tracked by the Kinect and their physical movements are mapped to their avatar's movements.

appear in the virtual world as bridge edges and crates at the right locations and users avoid running into them intuitively as they do in real life. In the demo, the audience will have the opportunity to embody an avatar and walk in physical space to navigate a virtual bridge while wearing the Oculus Rift. They will also be able see another person's avatar and interact with them in the real world as well as the virtual world.

REFERENCES

1. Merleau-Ponty, M., and Smith, C. *Phenomenology of perception*. Motilal Banarsidass Publishers, 1996.
2. Sageng, J. R., Fossheim, H. J., and Larsen, T. M. *The Philosophy of Computer Games*, vol. 7. Springer Science & Business Media, 2012.
3. Steuer, J., Biocca, F., and Levy, M. R. Defining virtual reality: Dimensions determining telepresence. *Comm. in the age of virtual reality* (1995), 33–56.
4. Williams, B., Narasimham, G., Rump, B., McNamara, T. P., Carr, T. H., Rieser, J., and Bodenheimer, B. Exploring large virtual environments with an hmd when physical space is limited. In *Proc. of the 4th symp on Applied perception in graphics and visualization*, ACM (2007), 41–48.