# Enhanced Interaction in Immersive Virtual Environments

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Loren Frank Puchalla Fiore



IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Victoria Interrante

May, 2016

# Acknowledgements

First, I would like to thank the 3M Corporation for their graduate fellowship that allowed me to attend graduate school. I would also like to thank the Linda and Ted Johnson Digital Design Consortium and the National Science Foundation for lab funding.

I would like to thank my colleagues in the DDC lab, Lane Phillips, Peng Liu, and Koorosh Vaziri who have served as coauthors and friends during my time at the University of Minnesota. They have on many occasions helped with running user studies, debugging software, and providing useful feedback during the PhD process. I also thank Ying He, who assisted with the user studies described in Chapter 3.

To my adviser, Victoria Interrante, I am deeply grateful. Vicki has been very helpful throughout my years as a graduate student in matters of writing, research, and teaching.

Finally, I would like to thank my family. My parents, Leonard and Joy, for their support during my collegiate studies. And most importantly my wife, Heather, for her encouragement all through my years as a graduate student and before. You endured many late nights and uneventful weekends, helped proofread my papers, and gave me the motivation to work through the difficult projects. I could not have done this without you.

**Abstract**

Virtual reality (VR) has many uses across diverse areas such as scientific visualization, flight training, and architecture design. The spatial awareness and feeling of presence created by immersive virtual environments (IVE) assists with learning and reasoning tasks within VR and is one reason for its adoption. However, there are still problems of perception within IVEs that can limit the effectiveness of their use. One example is the compression of egocentric distances when using a head-mounted display to view the IVE. The goal for this dissertation is to investigate whether IVEs can be made more effective through the development of enhanced locomotion and interaction methods that provide more accurate visual and vestibular feedback to the user. We investigate the use of color and depth (RGB+D) cameras to generate real-time video-based self-avatars that perfectly match the user's own body without the need for markers or per-user calibration. We perform a user study to determine if the video avatars reduce the errors in egocentric distance perception experienced in virtual environments. Next, we discuss the geometric and perceptual errors present in multi-viewer single-view virtual reality displays, such as CAVEs. Since the calibration and setup of these displays is crucial to limit these errors, we have created open source software for calibration of these displays. Finally, we look at walking in IVE and discuss the benefits of natural movements over *indirect* interfaces such as keyboards and introduce the novel technique of redirected driving. We show that redirected driving has the potential to offer the benefits of walking, such as better spatial understanding and mapping recall, while still allowing movement of large distances.

# Contents

# List of Figures

# 1

# Introduction

*He inferred that persons desiring to train this faculty must select localities and form mental images of the facts they wish to remember and store those images in the localities, with the result that the arrangement of the localities will preserve the order of the facts, and the images of the facts will designate the facts themselves, and we shall employ the localities and images respectively as a wax writing tablet and the letters written on it.*

– Cicero, *de Oratore 2.86.354*



The method of loci, described in the quote above and occasionally referred to as the memory palace technique, is a mnemonic device that uses visualization to organize and remember information. It was used in antiquity by Cicero, Quintilian, and others to remember speeches, as paper and ink were still too valuable for speaking notes. It is still used by many in the present day to for memory contests, such as the record $2^{16}$ digits of $\pi$ [1], and everyday items such as shopping lists.

The method at its core is straightforward. Simply visualize a place that you have been and know well, such as your house, and within each room of your house place one piece of the information to be remembered. Then at a later time the visualization of the well known place will bring back the ordering and contents of the memory. The method is effective because it is well-suited to how the human brain has evolved to have a keen sense of spatial understanding. Functional magnetic resonance imaging (fMRI) has shown that the brains of individuals in the process of remembering using the method have activation in the areas of the brain thought to control spatial awareness such as the medial parietal cortex and the right posterior hippocampus [2].

The method of loci shows the strength of the human mind at visualizing spatial information and how one can leverage that strength to help with other unrelated tasks. Leveraging the mind's ability of spatial understanding is also one of the motivations behind the field of virtual reality. Virtual reality (VR) is defined as a computer-generated simulation of a three-dimensional image or environment that can be interacted with in a seemingly real manner. VR can take the form of large-screen immersive displays (LSID), such as those found in motion simulators for planes and automobiles, and head-mounted displays (HMD) that allow the wearer to view only the virtual environment. Having the ability to view, navigate, explore, and interact with a virtual environment has been shown to increase task performance on a variety of tasks versus attempting the same task with a traditional monitor and keyboard computer setup [3–7].

## 1.1 Immersive Virtual Reality

When a virtual environment simulates the user being in the environment and in such a way that they themselves feel physically present, then it is said to be an immersive virtual environment (IVE). The immersion can be as straightforward as head-tracking, so that the user's head motion causes the VR view to update accordingly, to something as involved as physics-based interaction with the environment. These IVEs have a diverse range of uses across areas such as scientific visualization, flight training, design review, and architectural design, to name a few [8, 9]. One reason for this wide-spread adoption is that virtual reality provides the ability to immerse the user in the virtual environment in a way that is fundamentally different from traditional keyboard and monitor setups. Immersive virtual environments can take advantage of the human mind's keen sense of spatial awareness, and systems using head-tracking can provide a metaphor for movement and exploration that is in many ways the same as real life. For instance, users can change their point of view by moving themselves and by turning their head instead of pressing buttons or moving a mouse. Additionally, many immersive virtual reality systems employ stereoscopic displays that allow each eye to see slightly different images allowing for full three-dimensional viewing of the virtual environment.

However, there are limitations to the current implementations of these systems. For one, tracking of any kind is only possible when the user is within the tracked space. Larger tracked spaces require larger lab spaces to move within and additional tracking equipment. Also, stereoscopic displays require calibration and when miscalibrated can create a false sense of space that can fool users into drawing incorrect inferences about the virtual environment. For example, when designing a building in an IVE an architect must be able to trust that the space they and the clients are viewing looks the same as it does in the real world. If there is a miscalibration that causes distances to appear shorter when viewed in the IVE the clients may insist on a requiring a space larger than needed, resulting in wasted resources and additional cost. In the following sections we will discuss the promises and limitations of the specific areas of immersive virtual reality addressed in this dissertation.

### 1.1.1 Video-Based Self-Avatars

Immersive virtual environments experienced through head-mounted displays (HMDs) confer several benefits. If head-tracking is used, the display can update in real-time in response to the motion of the wearer who can move around the virtual environment and examine it from different angles in a natural manner. Another benefit of the HMD system is that, most of the time, the real-world is completely obscured from view, forcing the user to focus on the virtual environment and enhancing the feeling of being physically present. However, this is also a major disadvantage because the obscuring of the real-world also obscures the user's view of his or her own body. This can be quite disorienting when navigating a virtual environment and has a detrimental effect on the immersion. A solution to this problem is the give the user a VR avatar.

An avatar is a graphical representation of the user in the virtual environment. Avatars are often created from generic models, sometimes with minor customization options, and rendered using the tracking data of the user's body. Depending on the application and the tracking system used, the avatar's arms and legs may move in the virtual environment to match the motions of the user in the real world, or the limbs may be static and only the head position of the avatar follows the user movement. In [10], Slater *et al.* show that immersive virtual environments employing a generic avatar increase the feeling of presence and immersion in the environment compared to the same environment without any avatar. However, a generic avatar is not as realistic as the real-world. The physical appearance, clothing, and even the gender of the avatar may not match that of the user. Also, the tracking systems available often require per-user calibration resulting in long setup times before the virtual reality system can be used.

We propose, in Chapter 3, to use a video see-thru HMD, that is one with attached video cameras, in order to create a video-based self-avatar for the user. This will allow the user to see an avatar matching his or her own likeness within the virtual environment. Our hypothesis is that this method of avatar creation will show an increase in presence greater than that of a generic avatar. The use of a camera-based system instead of one involving tracking markers, such as optical motion capture, makes the system easier to use because no per-user calibration is needed prior to use.

### 1.1.2 Immersive Projector Calibration

A popular display modality in the virtual reality community is the Large Screen Immersive Display (LSID). These displays are usually on the scale of two meters or more in both width and height and are sometimes referred to as *display walls* or *virtual windows*. A LSID may have a single screen surface that subtends a large portion of the viewer's field of view or multiple surrounding screen surfaces such as in a CAVE system. An advantage of these large immersive displays is that they are large enough for multiple users to view the scene and, at the same time, interact with each other in a natural way. This is often useful in design review or data visualization tasks to facilitate discussion between the designers and researchers viewing the virtual environment.

A critical step in the creation of a LSID is calibration of the display surface. This is especially true with stereoscopic LSIDs using head-tracking, where the center of project of the rendered view must match the vantage point of the user. Normally, in stereo vision, feature points between images seen by both eyes can be matched and triangulated by our eyes to obtain depth information about a scene. When the rendered views do not match, or are rendered incorrectly, the geometric solution may not exist and the human brain will approximate a solution [11]. This approximate solution, if believed, can lead users to experience the virtual environment differently than if it were a real physical space. For example, when viewing a CAD design of a building a user may believe a virtual wall to be 2.5 meters distant, but in the design the wall is only 2 meters distant. In the use case of architectural design this can, and will, cause incorrect designs because the user's senses are tricked into perceiving the environment incorrectly, and this incorrect understanding will affect any decisions made to alter the design.

In Chapter 4, we describe the development of an open-source software package for the calibration of arbitrary shaped rear-projection large-screen immersive displays. This software was developed at the University of Minnesota in order to calibrate the LSID at Digital Design Consortium lab. This particular screen is saddle-shaped (i.e. curving in opposite directions along two axes) and as such was not capable of calibration using the traditional parametric approaches. The software makes it easy, using only a webcam, several printed fiducial markers, and time to calibrate any shape of screen by a piece-wise planar approximation. Techniques for rendering using the calibrated surface are also discussed in this chapter.

### 1.1.3 Locomotion in Large Environments

Creating the experience of natural self-motion in virtual environments is a fundamental problem facing the field of virtual reality [12]. While tracking technologies allow users of virtual reality to perform motions in the virtual world that match their real-world motions, many such setups provide this natural interaction over the range of little more than a few square meters. This can be because of technical limitations of the tracking method, the cost of covering a larger area, or simply the limiting size of the available physical space. To overcome these limitations, users of virtual environments often have to revert to indirect forms of traveling involving keyboard, joystick, or wand input. These systems allow the users to travel in a large virtual environment but at the expense of losing the ability for natural motion. This is detrimental because several groups have shown that users of indirect locomotion systems have substantially greater difficulty in remembering their motions and the layout of the virtual environment than users who are able to move about naturally [3, 13, 14].

Natural walking redirection techniques have shown great potential for enabling users to travel in large-scale virtual environments while their physical movements are limited to a much smaller laboratory space [15]. In the real world, walking is primarily used to cover short distances, but a variety of methods of transportation are employed when traversing longer distances. In Chapter 5 we introduce a novel approach to the concept of redirection by using physical traveling devices. We show that using a modified electric wheelchair it is possible to redirect users using visual rotations as well as direct re-steering of the wheelchair itself. To evaluate this novel locomotion technique we designed and conducted two psychovisual experiments comparing the technique to existing redirection and locomotion strategies in terms of effectiveness at environment spatial understanding, redirection sensing thresholds, and rate of cybersickness. We have designated this technique *redirected driving* and believe it fills an important niche of natural-seeming locomotion covering large distances as it allows users to draw on their everyday experiences of driving or riding in moving vehicles.

## 1.2 Our Approach

### 1.2.1 Thesis Statement

The goal of this dissertation is to increase the effectiveness of immersive virtual environments by investigating the following thesis statement:

> *Immersive virtual environments can be made more effective through the development of enhanced locomotion and interaction methods that provide more accurate visual and vestibular feedback to the user.*

Our research accomplishes this investigation through three focused research tasks:

- Develop and evaluate video-based methods for virtual self-avatars that operate without the need of body tracking or per-user calibration. (Chapter 3)
- Create open-source projector display calibration software that allows for arbitrary screen shape and flexible projector configuration using an off-the-shelf web camera for calibration input. (Chapter 4)
- Propose and evaluate the novel locomotion technique of *redirected driving* that allows for natural metaphors for the exploration of large virtual environments via direct and indirect redirection of user movements. (Chapter 5)

In the pursuit of these research avenues, we have developed new software and algorithms for virtual reality that are applicable to a large variety of immersive virtual environments. Throughout this dissertation we will focus on the driving use case of architectural design, however, we expect the contributions to be useful in other domains. For the self-avatar and redirected driving research, we have performed multiple user-studies of the effectiveness and will present and discuss the results.

## 1.3 Dissertation Overview

The dissertation begins in Chapter 2 with an overview of the research creating virtual reality, discovering its limitations, and several of the key experiments and efforts by others to overcome these limitations. Then, in Chapter 3, we investigate the use of color (RGB) and color plus depth (RGB+D) cameras for the purpose of generating real-time video-based self-avatars to give users an avatar that perfectly matches their own body yet requires no markers or per-user calibration. Next, in Chapter 4, we discuss the types of geometric and perceptual errors present in multi-viewer single-view virtual reality display setups, such as CAVEs or other large stereoscopic displays. We show how these errors can result from corresponding errors in calibration and describe the development of an open source software package for arbitrary shaped, multiple projector, screen calibration. Finally, in Chapter 5, we look at walking in immersive virtual environments and discuss the benefits of natural movements over more *indirect* interfaces such as keyboards and joysticks. Furthermore, we introduce the novel technique of *redirected driving*. We demonstrate the potential of redirected driving to confer the benefits of walking while maintaining its usefulness for covering great distances similar to the methods involving indirect interfaces, such as flying using a wand or joystick interface.

# 2

# Background & Related Work



Immersive virtual environments (IVEs) are made possible by a specialized set of computer hardware and software. In this dissertation we focus on two types of immersive virtual environments, those using head-mounted displays (HMDs) and those using large-screen immersive displays (LSIDs).

An HMD contains two small computer screens positioned one in front of each eye. By rendering slightly different images to each screen a three-dimensional stereoscopic view is presented to the wearer. An example HMD, and the one used in this dissertation, is the SX60 manufactured by NVIS [16]. Recently, consumer grade HMDs have begun to appear on the market using smartphone technology in order to drastically reduce prices such as the Oculus Rift [17], Samsung Gear VR [18], and Google Cardboard [19].

The CAVE Automatic Virtual Environment (CAVE), first proposed and built by Cruz-Neira *et al.* in 1992 [20, 21] is one type of large-screen immersive display, and the second display modality discussed in this dissertation. A CAVE is a cubic room with full-extent stereoscopic screens on several of its surfaces. The original CAVE had four of its six surfaces as screen (three walls and the floor) but several modern CAVEs, such as the ones built by Mechdyne Corporation at Iowa State University and KAUST [22], have all six surfaces as displays. Another type of LSID is the large projector wall. Typically several meters in both width and height this can be flat or curved. An extreme example of a curved LSID wall is the CAVE2 at the University of Illinois at Chicago [23] that is a cylindrical room with a tiled 37-megapixel stereoscopic display wrapping 320-degrees around the user.

Regardless of the display modality being used any IVE application will also require some form of user tracking. Tracking the user's head position and orientation is critical for rendering their correct first-person viewpoint of the virtual world. The current state-of-the-art tracking system is the optical tracking system, such as that manufactured by Vicon [24], that uses reflective markers attached to the HMD and watched by multiple cameras placed around the perimeter of the tracked physical space. The video from the cameras is sent to a central computer server that localizes the marker position in each image and uses that to triangulate its real-world position. By attaching multiple markers to the HMD a full six degree of freedom (6-DOF) tracking can be achieved. Markers may also be placed on the arms, legs, and body of the user for full-body motion capture. An interesting variation on the optical tracking system is the Hiball [25], that switches the roles of the markers and the camera. In the Hiball system the cameras are placed on the HMD and the markers, each a flashing infrared light, are place on the ceiling of the room. This enables greater stability in the localization at the expense of a more complex head-worn display and sensor unit.

Another commonality between display modalities is the software. Depending on the application an immersive virtual environment may require physically-based rendering, physics simulation, artificial intelligence for virtual agents and a host of other complicated software. To ease the burden of this integration reusable software is often used to create IVEs. A notable example of a commercial software is Vizard by WorldViz [26]. Another advantage for virtual reality is the recent shift in the video games industry towards independent, or indie, developers. This has created many affordable video game engines, several of which are open source. In this dissertation we make use of the G3DEngine [27], Unity3D [28], and, most recently, the Unreal 4 Engine [29] to create immersive virtual environments.

Existing research related to this dissertation falls into three broad categories: Avatars & Embodiment, Large-Screen Immersive Display (LSID) Design & Calibration, and Natural Locomotion in IVEs. We will now briefly discuss each of these categories including the established techniques and technologies in each as well as any remaining unsolved problems. More detailed and in-depth review of prior work is left for the start of each related chapter.

## 2.1 Avatars & Embodiment

An avatar is a graphical representation of the user. In immersive virtual environments this typically appears as a three-dimensional model of person. The avatar can be designed to match the user's physical appearance in the real-world or be purposefully different depending on the application. In multi-user environments it is possible to show avatars for the other users as well as virtual agents for AI characters. In this dissertation we focus specifically on first-person avatars of the user.

Giving the user a virtual avatar creates a greater sense of realism in the IVE as well as an awareness of the effects of his or her actions. The avatar also provides several spatial cues to help the user understand the extent of the virtual space and his or her location within. It is strongly believed that virtual avatars increase a user's sense of presence in the virtual environment [10]. Lok [30] *et al.* has shown that the use of virtual avatars can lead to significant improvement in performing tasks involving reaching distances.

In many virtual reality systems the only tracked object is the head-mounted display. A common occurrence in this instance is to not render any avatar for the user, creating an abrupt visual disconnect from the laws of physics as the user experiences the immersive virtual environment from a disembodied view. An alternative is the give the user a rigid, non-articulated, avatar that is always aligned with the position of the tracked head-mounted display. There will still be some disconnect as the motion of the user's arms and legs will not be reflected in the virtual environment.

Adding additional tracked objects to the virtual reality system is usually the next step for more robust avatars. Wands and hand markers can be used to track the location of the user's hands and feet in order to articulate the arms using inverse kinematics. More elaborate systems involving full-body motion capture can allow accurate tracking of the entire skeleton and avoid the jumps in motion often present in inverse kinematics as the solver fluctuates between multiple correct solutions.

An open area of research in this field is reducing the number of tracking markers that must be used in order to create an avatar. In the extreme case the goal is marker-less avatar creation. Marker-less systems promise to reduce the costs of providing articulated avatars in virtual reality and also simplify the process of using the virtual reality system as there will be no markers or special clothing to wear in order to use the system.

## 2.2 LSID Design & Calibration

A popular display modality in virtual reality is the Large Screen Immersive Display (LSID). These displays are usually on the scale of 2m or more in height, and 2m or more in width and are sometimes referred to as *display walls* or *virtual windows*. A LSID may have a single screen surface that subtends a large portion of the viewer's field of view, or multiple surrounding screen surfaces such as in a CAVE system.

A critical step in the construction of any LSID is the determination of the mapping from display pixels to world space. This mapping is needed in order to display the correct stereo image given a specific vantage point. The process of determining this mapping is known as screen calibration.

Calibration is often done manually, using specially designed screen hardware and mounting. However, this is time consuming and does not scale well. Several researchers have therefore investigated automatic calibration of LSIDs using feedback from one for more video cameras, [31–39]. However, even with all of this knowledge relating to automated calibration of LSIDs, the the authors were unable to find any readily available software package for performing automated calibration of arbitrary LSID configurations. Therefore, in Chapter 4, we present the results of our work to create such a software package.

## 2.3 Natural Locomotion in IVEs

One of the primary means of interaction within immersive virtual environments is exploration. Users need to be able to move about within the environment and examine it from different perspectives. The most straight-forward way of exploring a virtual environment when viewed through an HMD is to use head-tracking and walk about in a natural way with motions in the real-world mapping one-to-one with motions in the IVE. However, because of the physical limitations of the available physical space and the tracking equipment involved it is very often the case that the virtual environment displayed is much larger than the tracked space. A notable exception is the HIVE, Huge Immersive Virtual Environment, where the tracking system along with all the required VR hardware is wearable and can be used outdoors in ample space [40]. To overcome the size limits of the tracked space many researchers have studied additional locomotion techniques.

The most common form of locomotion in immersive virtual environments is natural walking, where real-world movements map one-to-one to virtual movements, limited to the tracked workspace with button-controlled flying or jumping to new locations within the virtual environment [41]. When the tracked space is too small to allow actual walking, walking-in-place is an alternative. In [42], Slater *et al.* develop a walking-in-place algorithm that analyzes tracked head motion to count steps. In [43], Usoh *et al.* discuss the results of a user study showing that walking interfaces have much higher levels of presence and immersion that those of walking-in-place or flying.

In [44], Interrante *et al.* propose the "seven-league boots" metaphor for exploring virtual environments. In this mode the user holds a tracked wand device with a single button. Without using the wand, motion within the virtual environment is matched exactly to the head-tracked motion. When the button is pressed and the user walks forward the program determines the direction of travel using the head tracking information and amplifies the motion along this direction. This allows for rapid movement between different regions of the virtual environment when the button is pressed, and for natural movement when the user has reached the destination. A disadvantage of this system is its unnatural modal nature requiring the user to press a button to switch locomotion modes.



An interesting fusion between walking-in-place and true walking is given by methods that move the floor under the user. These systems allow the user to take physical steps while remaining in the same place. In [45], Souman *et al.* describe the CyberWalk omni-directional treadmill that is used for immersive virtual environments. Iwata *et al.*, in [46], develop the CirculaFloor alternative to a omni-directional treadmills. The CirculaFloor consists of several flat, rectangular robotic pads that rearrange themselves under the user's feet to produce the same effect as an omni-directional treadmill but without needing a special building or room to house the treadmill. A low-cost alternative to these setups is the VirtuSphere system [47]. The VirtuSphere uses a 10-foot hollow sphere that the user walks within. The sphere rests atop a set of roller wheels, so that the user's steps only cause the sphere to rotate in-place instead of moving about the physical room.

A promising new form of locomotion in immersive virtual environments is *redirected*