Integrated Space: Authoring in an Immersive Environment with 3D Body Tracking

Paulo F.U. Gotardo, Alan Price*
The Advanced Computing Center for the Arts and Design
The Ohio State University, Columbus (OH)

1 Introduction

Our research explores the use of real-time computer vision techniques and a pair of standard computer cameras to provide 3D human body awareness in an inexpensive, immersive environment system, Fig. 1. The goal is to enhance the user experience of immersion in a virtual scene that is displayed by a 3D screen. We combine stereo vision and stereo projection to allow for both the user and the virtual scene to become aware of each others 3D presence as part of a single, integrated 3D space.

We focus on enabling authoring applications based on the direct manipulation of virtual objects, with users interacting from a first-person perspective. This emphasis contrasts with the avatar-based, mostly reactive focus often employed in the design of computer game interfaces. We note that our work is part of an effort to develop low-cost solutions that provide interaction designers with means to prototype ideas easily and in anticipation of future releases of similar technology in the mainstream of HCI applications. Here, we present a prototype system that demonstrates the interaction paradigm above.

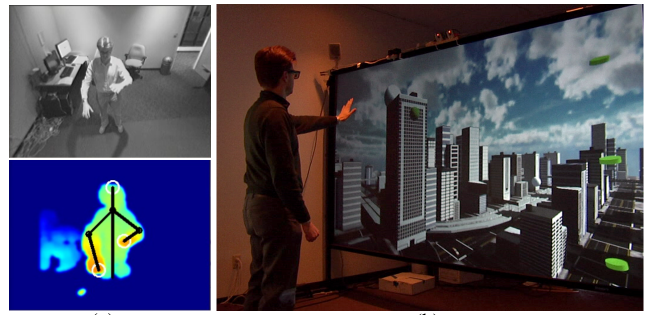
No markers or other special user-born equipment is required. The user's presence in front of the projection screen is automatically detected, Fig. 1(a), and head tracking dynamically updates the camera frustum, adjusting the 3D output. This capability significantly enhances the experience of immersion as it allows for 3D objects to be seen from different points of view as the user moves sideways or closer/farther from the screen. Hand detection enables the user to directly interact with virtual objects by touching and moving them in the scene. Rapid hand gestures allow captured objects to be released.

In the metro scene in Fig. 1(b), the user is presented with a scale model of a city and can reorganize its layout by grabbing and relocating buildings. The 3D location of the user's hand is indicated by a small green sphere that serves as an optional hand avatar. The user can trigger different actions by touching a simple heads-up-display (HUD) controller that virtually floats on the right of the 3D space. These actions are: grabbing buildings with a white, spherical object that extends the user's reach farther into the scene; drawing or sculpting 3D curves that turn into 3D scene objects; and "flying" (moving) to a different location of the city. Simulation of rigid body dynamics enhances the user interaction experience in this new 3D space.

While we do not present a new stereo vision algorithm or a novel 3D projection device, we propose a simple and effective two-way visual interface unencumbered by controllers, tracking markers, or other user-borne devices. Only the use of polarized glasses is required for the perception of the 3D output. Our system does not recognize fine details such as individual fingers due to limitations in the current technology of stereo imaging. Even state-of-the-art stereo algorithms are still incapable of providing depth estimates in real-time, for a small 0.08 megapixel (320×240 pixels) image pair. However, our interactive system is an engineering solution and creative interaction design that provides a quality real-time user experience balanced against computation limitations.

*e-mail: {pgotardo,aprice}@accad.osu.edu

SIGGRAPH 2010, Los Angeles, California, July 25 – 29, 2010.
ISBN 978-1-4503-0210-4/10/0007



(a) (b)

Figure 1: *Integrated space: (a) one of the two input images and the computed depthmap showing all detected body parts; (b) the 3D projection screen, stereo cameras, and infrared illuminators.*

2 Technical Approach

Our 3D projection system, Fig. 1(b), was built in-house following a known procedure. The large-scale rear screen projection of a virtual scene provides depth perception for users wearing 3D glasses. Centered atop the projection screen is a pair of cameras with wide-angle lenses. Two infrared illuminators, on the top left and right corners of the screen, allow for stereo imaging to be done with low visible light, as required for optimal visualization of the 3D output.

We use the fast stereo imaging routines of the computer vision library OpenCV (*http://opencv.willowgarage.com*) and compute sparse depthmaps at a coarse image resolution. This is done at 30 fps on a quad-core desktop processor. We also developed our own software that smoothly extrapolates depth estimates to nearby pixels with unknown depth, originated by low-contrast image regions. Our computer vision routines also perform body and head detection with a random sampling algorithm that generates and validates hypothetical, whole-body cylindrical models. A similar subsequent procedure looks for arms and hands using an articulated arm model. Our software can tolerate pixels with poor or unknown depth estimates and successfully analyze rough depthmaps computed with fast but less accurate stereo algorithms.

A possible extension enabled by our current system is to use gesture recognition by matching the 3D curves, originated by the time-trajectory of a users hand, against predefined curves. This functionality can be used to complement or replace our HUD virtual controller used to trigger different actions.

Our system provides a visual experience of immersion that is similar to expensive GeoWall™ systems. However, this capability is achieved without the need for the user to learn how to handle complex trackers or control devices, allowing for passers-by to immediately begin interaction with applications. Our video-based tracking system allows for multiple users with simultaneous points of interaction. Our system comprises a combination of inexpensive, off-the-shelf hardware and software that is reproducible, reconfigurable, and expandable to accommodate new technological advances as they become readily available.